International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2019): 7.583

Prevention and Detection of Diabetes (Type-I & Type-II) using Data Warehousing and Data Mining Techniques in Andaman & Nicobar Islands

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
Email: *s.deepaonline[at]gmail.com*

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

Abstract: ***One of the most significant health issue faced by all the human being these days is diabetes. Diabetes is one of the leading causes of mortality and morbidity worldwide. The common sites of Diabetes have varied distribution in different geographical locations. The present study is conducted to detect and prevent Diabetes of two major types i.e., Type – I and Type - II using data mining and warehousing techniques in Andaman and Nicobar Islands. The study uses data mining techniques such as classification, clustering and prediction to identify potential diabetes patients.For that a multidimensional architectural diabetes data warehouse will be built specifically to store and process diabetes-related database which include patient's general and medical records and also adata mining model is proposed to be build and will be implemented within the diabetes data warehouse which can predict a person's predisposition towards diabetes and generate the risk level for a particular type of diabetes and the exact method of clinical diagnosis. The k-means clustering algorithm is used for partitioning the data into diabetes and non-diabetes clusters, where the initial cluster center is represented by the mean value of the weightage of significant patterns.***

Keywords: Andaman & Nicobar Islands, Diabetic patients, Data Mining and Warehousing Techniques, Multidimensional Star Schema, k- means Algorithm, OLAP Operations, Types of Diabetes- Type-I & Type-II, WAM

1. Introduction

According to the World Health Organization (WHO), India had 69.2 million people living with diabetes in 2015. Nearly 98 million people in India may have type 2 diabetes by 2030, according to a study published in the 'Lancet Diabetes & Endocrinology' journal, found that the amount of insulin needed to effectively treat type 2 diabetes will rise by more than 20% worldwide over the next 12 years.

"The number of adults with Type-II diabetes is expected to rise over the next 12 years due to aging, urbanization, and associated changes in diet and physical activity," said Sanjay Basu from Stanford University, who led the research.

As an alternative to the tedious physical storage of resources it is important to develop a data warehouse specific to diabetes disease and a data mining model to predict diabetes earlier. If a machine learning technique is developed to store a person's medical and general record and predict his predisposition towards diabetes, its type and exact diagnostic method, physicians can directly start treatment immediately without wasting the precious time in different methods of diagnosis. There have been multiple data mining techniques in health care and allied industries and specifically with respect to type-I & type-II of diabetes.

The present study is carried out for 50% of the total population of diabetic patients of Andaman & Nicobar Islands in which the datas will be collected from Andaman Nicobar Islands Institute of Medical Science (ANIIMS), Port Blair, the only government tertiary care hospital in Andaman and Nicobar Islands.

All the patients diagnosed or suspected with diabetes, registered in G.B. Pant Hospital. This research focuses on the building of multidimensional diabetes data warehouse and development of data mining model for the early detection of two major types of diabetes namely, Type I and Type II, hence prevention is also possible.

2. Aims and Objectives of the Research

The aim of the study is to develop a multidimensional architectural diabetes data warehouse built specifically to store and process diabetes-related database which include patient's general and medical records.

The diabetes data warehouse is proposed to be built on OLTP and OLAP technologies simultaneously, thereby retrieving necessary information using query engines. A data mining model is also proposed to be built and will be implemented within the diabetes data warehouse which can predict a person's predisposition towards diabetes and generate the risk level for a particular type of diabetes and the exact method of clinical diagnosis.

3. Review of Related Literature

From the literature review it is learned that use of evolving IT systems in medical sciences to eradicate, diagnose, prevent diseases like diabetes and ameliorate the standard of living of patients with such life-threatening diseases has

garnered the attention of IT researchers worldwide. Review of the literature on diabetes related databases helped to realize the fact that suitable datawarehouse architecture should be implemented for the efficient functioning of the objective data analysis.

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

4. Methodology of Research

Data warehousing is a collection of decision support technologies, aimed at enabling the knowledge worker (executive, manager, and analyst) to make better and faster decisions. A data warehouse (or scale data mart) is a specially prepared repository of data will be designed to support decision making.

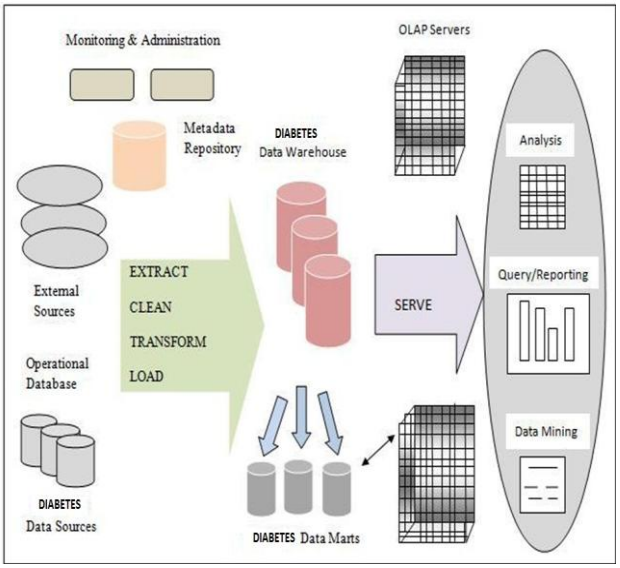


Figure 1: DIABETES Data Warehousing Architecture using ECTL, OLTP, and OLAP Servers

The data required for the research will be collected from Andaman Nicobar Islands Institute of Medical Science (ANIIMS), Port Blair, Andaman & Nicobar Islands. The study uses data mining techniques such as classification, clustering and prediction to identify potential diabetes patients. A multidimensional data warehouse specific to diabetes disease will be built and implemented and further it is to be used for a data mining work to detect a person's predisposition towards diabetes. Finally, a detection and prevention system will be developed to analyze the risk levels which help in prognosis.

The data will be comes from operational systems and external sources. To create the data warehouse, diabetes data will be extracted from source systems like questionnaire, diabetes institute database, etc. which will be cleaned (e.g., to detect and correct errors), transformed (e.g., put into subject groups or summarized), and loaded into a data store (i.e., placed into a data warehouse).

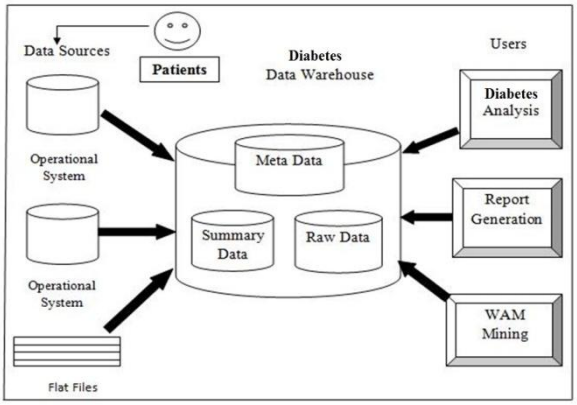


Figure 2: Diabetes Data Warehouse Architecture

A. Multidimensional Star Schema

The basic building block going to be used in dimensional modeling is the star schema. A star schema will consists of one large central table called the fact table, and a number of smaller tables called dimension tables. The fact table forms the "center" of the star, while the dimension tables form the "points" of the star. A star schema may have any number of dimensions. The fact table will contain measurements (e.g. Patient History, Risk Factor, Diabetes, Symptoms, Treatment, and Diagnosis) which may be aggregated in various ways.

The dimension table will provide the basis for aggregating the measurements in the fact table.

The fact table will be 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.

The advantage of using star schemas to represent data is that it reduces the number of tables in the database, the number of relationships between them and therefore the number of joins required in user queries.

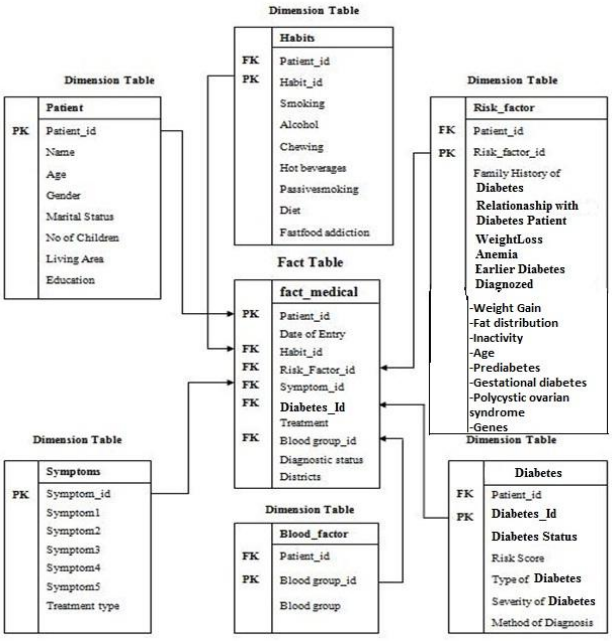


Figure 3: Star Schema representing Diabetes Data Warehouse

B. OLAP Operations of Medical Diabetes Data Warehouse

OLAP is performed on diabetes data warehouse or diabetes disease data marts. The primary goal of OLAP is to support ad hoc query needed to support decision support system. The multidimensional view of diabetes data is fundamental to OLAP function. OLAP is a practical view, not a data structure or schema. The complex nature of OLAP process requires a multidimensional review of the diabetes data. OLAP Operations in Multidimensional Diabetes Data Warehouse (MDDW),

1) Roll-Up
2) Drill Down
3) Slice and Dice
4) Pivot

Roll Up (Drill-Up)

- It is performed by climbing up hierarchy of a dimension or by dimension reduction (reduces the cube by one or more dimensions).
- The roll up operation is based on location (roll up on location) and is equivalent to grouping the data by districts.
- Roll-up operations do not remove any events but change the level of granularity of a particular dimension.

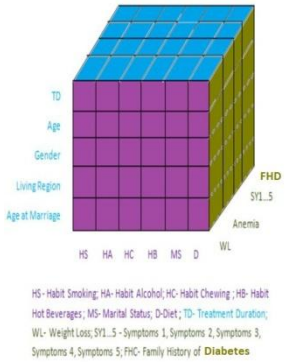


Figure 4(a): Original view of Diabetes Data

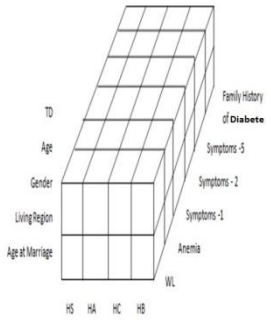


Fig4(b): Roll-Up View Warehouse

Drill down (Roll Down)

- It is the reverse of roll-up.
- Navigates from less detailed data to more detailed data by -
- Stepping down a concept hierarchy for a dimension to Introducing additional dimensions
- Drill down operations does not remove any events but change the level of granularity of a particular dimension.

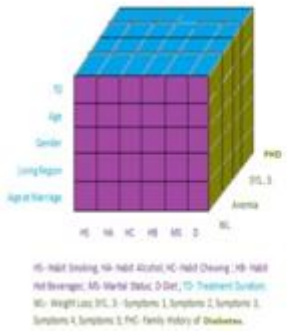


Figure 5 (a): Original view of Diabetes

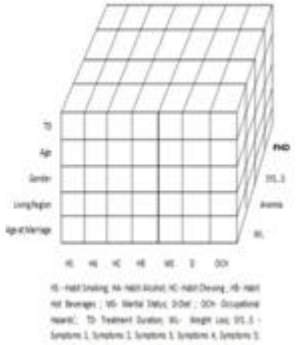


Fig5(b): Drill-Down View Data Warehouse

Slice and Dice

- The slice operation performs a selection on one dimension of the given cube, resulting in a sub-cube.
- The slice operation produces a sliced OLAP cube by allowing the analyst to pick specific value for one of the dimensions.

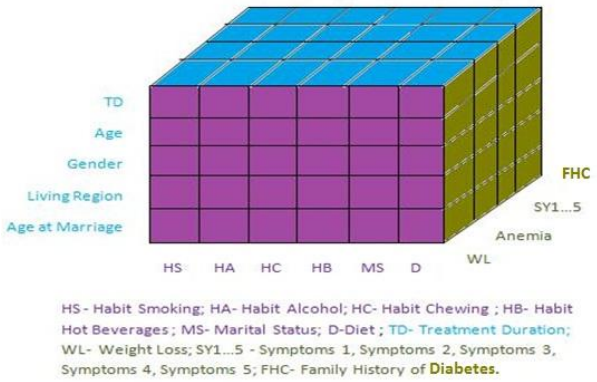


Figure 6(a): Original view of Diabetes Data Warehouse

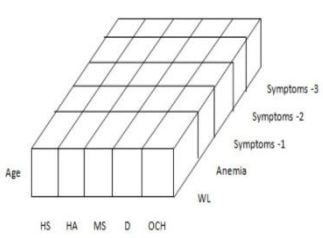


Figure 6(b): Slice View

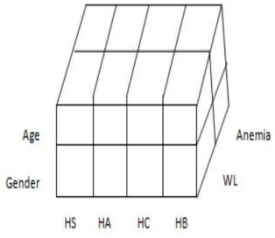


Figure 6(c): Dice View

Pivot

Visualization operation that rotates the data axes in view in order to provide an alternative presentation of the data. It removes a measure.

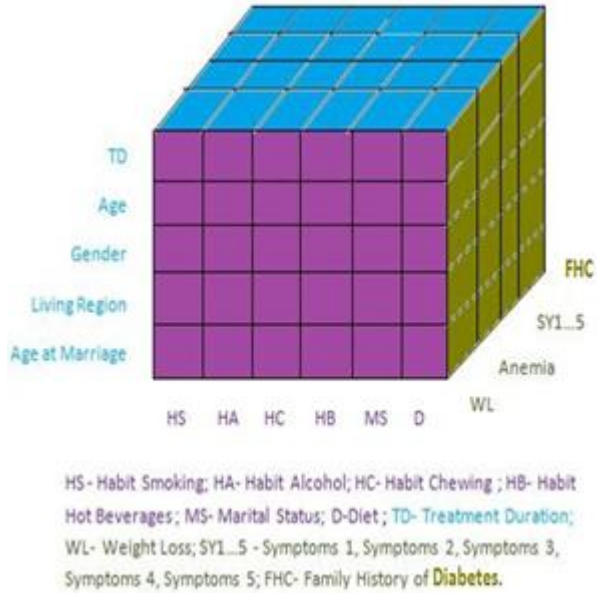


Figure 7(a) Original view of Diabetes Data Warehouse

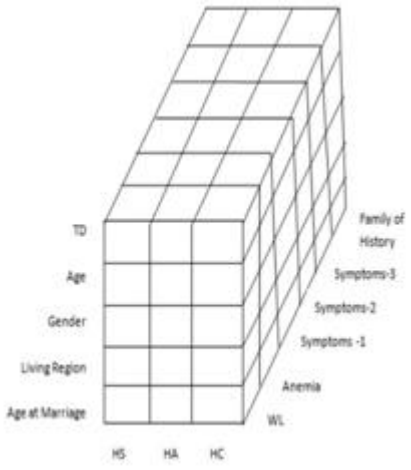


Figure 7(b): Pivot view

Table 1: Site wise distribution of Diabetes

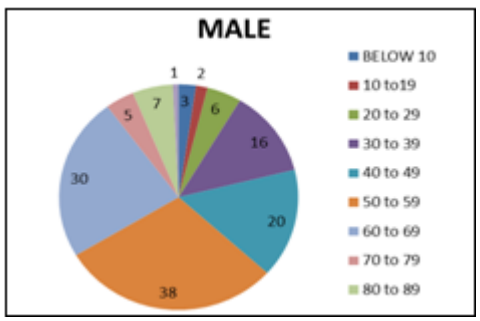
S. No.	Type of Diabetes	Male	Female	Total
1.	Type - I	(50% of selected population)	(50% of selected population)	(50 % of Total Population)
2.	Type - II			

Table 2: Correlation of Age and Sex of Diabetes patient

Age Group	Male	Female	Total
<10	(50% of selected population)	(50% of selected population)	(50 % of Total Population)
10-19			
20-29			
30-39			
40-49			
50-59			
60-69			
70-79			
80-89			
>90			

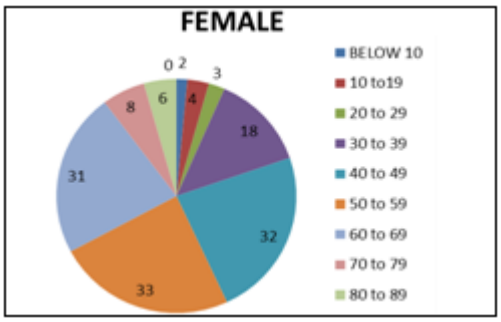
Data Cube dimension of 'Male' Factor of Diabetes data

Eg:-



Data Cube dimension of 'female' factor of Diabetes data

Eg:-



Correlation of Age and Sex of Diabetic patients

Through this study we can easily correlate the various age groups of diabetic patients with respect to their sex.

Eg:-

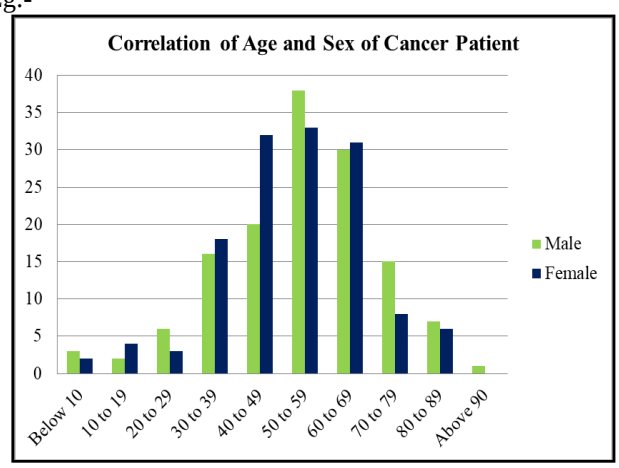
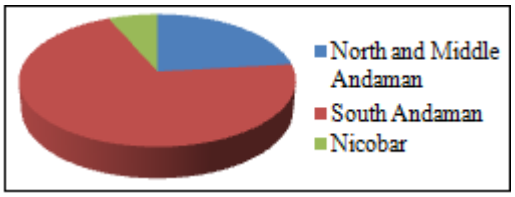


Table 3: District-wise data of Diabetes Patients

District	Male	Female	Total
North and Middle Andaman District	50 % of the selected population	50% of the selected population	50% of the total population
South Andaman District			
Nicobar District			

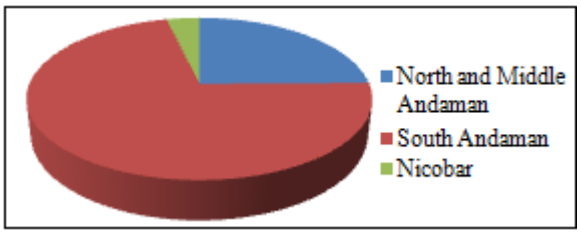
District - Wise Diabetic Patients distribution– Male

Eg:-



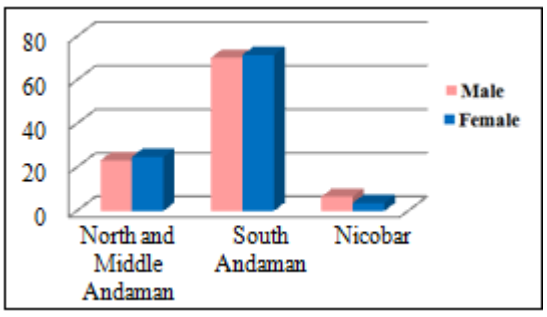
District - Wise Diabetic Patients distribution - Female

Eg:-



District – Wise Patients distribution – Male Vs Female

Eg:-



Like correlation between age and sex of diabetic patients, we can easily segregate the patients according to three major

districts namely North and Middle Andaman District, South Andaman District and Nicobar District.

5. Developing a Data Mining Model to Diagnose Diabetes Disease

The disease diagnosis is a major process to treat the patients who are affected by diabetes disease. 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 disease once the diabetes detection is accomplished.
In the present study, a proposed data mining model will usetwo different techniques whichperform consecutively. The techniques are classification and clustering method of conceptual modeling. So that thediabetes data would be converted into a knowledge base which is called as training data.

k-means Clustering for Classified Significant Pattern
The instances will be 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. The aim of clustering is that the data object is assigned to unknown classes that has a unique feature and hence maximize the intra-class similarity and minimize the interclass similarity. The weightage scores of the significant patterns mined will be fed into k-means clustering algorithm to cluster and divide it into diabetes and non - diabetes groups. The diabetes group is further subdivided into two groups with each cluster representing a type of diabetes.

The data in the cluster is again fed into *k-means* clustering algorithm to further subdivide it. The resulting two clusters are separated based on particular symptoms associated with any one type of diabetes i.e. Type-I and Type-II. Finally all the data is partitioned into two types of clusters and sub-clusters of the diabetes cluster. The k-means clustering algorithm is used for partitioning the data into diabetes and non-diabetes clusters, where the initial cluster center is represented by the mean value of the weightage of significant patterns.

Weighted Average Method k-means Clustering Based Diabetes Detection
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. WAM considers the patient population distribution at a system to reflect the impact of behavior/genetic factors (family history).

The WAM *k*-means algorithm follows an iterative optimization similar to *k*-means, and by consequence it is affected by some of its strengths, such as itsconvergence in a finite number of iterations to improve the centroid of clusters.

Weighted Average cluster k-means could cluster the patient data of medical records to different groups, and divided into two groups mapped as diabetes types based on Db1 and Db2.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.

6. Summary of Findings

Through this research, a novel multilayered method combines Data warehouse and Data Mining techniques to build the diabetes risk detection and prevention system will be develop. The most effective way to reduce the diabetes deaths is to detect and prevent diabetes disease. The developing of detection and prevention system may provide an easy and a cost-effective way for screening diabetes and may play a pivotal role in disease diagnosis process for different types of diabetes and provide useful, preventive strategy.

The multidimensional data has processed from 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.

7. Conclusions

There's no cure for type 1 diabetes. It requires lifelong disease management. But with consistent monitoring and adherence to 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.

Diabetes control aims to reduce the incidence or instance, morbidity and mortality of diabetic and to improve the quality of life of diabetic patients in a defined population, through the systematic implementation of evidence.

An implementation of our new system to expose the diabetic risk factors and to ensure that people are provided with the information and support they need to adopt in a healthy lifestyles.

Four basic components of diabetes control–prevention, early detection, diagnosis, treatment and painkilling care–thus avoid and cure many diabetes, as well as palliative the suffered patients. Diabetes control aims to reduce the incidence or instance, morbidity and mortality of diabetes and to improve the quality of life of diabetes patients in a defined population, through the systematic implementation of evidence. An implementation of our new system to expose the diabetes risk factors and to ensure that people are provided with the information and support they need to adopt in a healthy lifestyles.

Diabetes detection and prevention is still a challenging for the upgraded and modern medical technology. After researching a lot of statistical analyses which is based on those people who are affected in various diabetes types are based on some general risk factors and symptoms have been discovered.

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] SeyedAtaaldinMahmoudinejadDezfuli 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 TarunGangil (2019), "Analysis of diabetes mellitus for early prediction using optimal features selection", Article No. 13.
[4] DeeptiSisodia and Dilip Singh Sisodia (2018), "Prediction of Diabetes using Classification Algorithms",Procedia Computer Science, Volume 132, Pages 1578-1585.
[5] DeerajShettyet. al. (2017), "Diabetes disease prediction using data mining", International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), **INSPEC Accession Number:** 17558894,DOI: 10.1109/ICIIECS.2017.8276012 (IEEE).
[6] NimnaJeewandaret. 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] LoannisKavakiotiset. 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 BhambriPankaj (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] RavneetJyot Singh and Williamjeet Singh(2012), "Data Mining in Healthcare for Diabetes Mellitus", International Journal of Science and Research (IJSR),ISSN (Online): 2319-7064
[11] MiroslavMarinovet. 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