International Conference of Science, Engineering & Environmental Technology (ICONSEET), 3(15): 100-110, 2018
ISSN 0794-9650
www.repcomseet.com

Towards detecting anomalies in diabetes mellitus patients' data: An association rule based approach

Olusola F. Babafemi[a]*, Olatubosun Olabode[b], Busayo C. Olaniyi[a], Funsho A. Olatunji[a], John C. Nwaiwu[b].

[a] *Computer Science Department, Federal Polytechnic Ede, Nigeria.*
[b] *Computer Science Department, Federal University of Technology Akure, Nigeria*

Abstract- Diabetes Mellitus is an accumulation of metabolic infections in which a human being has elevated blood sugar due to a number of associated symptoms. These symptoms include regular urination, excessive eating, weight loss, joint pain, increased need for liquids amongst others. Discovering interesting patterns within symptoms of diabetes mellitus data remains a major data mining application. This paper presents an association rule based approach towards detecting anomalies in diabetes mellitus patients' data. This approach extracts interesting frequent symptoms pattern, mines association rules and detects anomalies in the dataset using the mined rules. Dataset of diabetes mellitus patients containing 150 records and 48 symptoms is sourced from Obafemi Awolowo University Teaching Hospital, Ile-Ife Nigeria. The method is implemented in Python Integrated development environment. The performance is evaluated based on number of frequent symptoms, mined rules and anomalies detected. The strongest rule recorded from the rule mining has a confidence threshold of 96%. This means that the occurrence of Polydipsia will result to a 96% probability of a presence of Polyuria in a diabetes patient. The result from the anomaly detection shows that an average percentage of 27% anomalies are detected in the diabetes data. The paper shows that anomalies in diabetes mellitus diagnosis can be detected using our approach.

Keywords: Anomaly detection, Association Rule Mining, Diabetes Mellitus, Frequent Patterns, Symptoms.

1.0 Introduction

Diabetes, one of the growing scourges of non-communicable disease affects greater proportion of the world population. Diabetes is a disease in which the ability of the body to produce hormone insulin is impaired, resulting in abnormal metabolism of carbohydrates and elevated levels of glucose in the blood. The deficiency of insulin results from inadequate insulin secretion otherwise diminished tissue responses to insulin at the complex pathways of hormonal reaction in the body (Dhivya and Merlin, 2015; Sangeetha and Moorthy, 2015). This symptom leads to conventional signs of polyuria (regular urination), polyphagia (increased starvation) and polydipsia (increased need for liquids) (Karthikeyan and Vembandasamy, 2015).

In current medical studies, the identification of risk factors for diabetes and designing a diagnostic or prediction model are generally based on multivariate statistical analysis utilizing logistic regression (Abassi *et al*, 2012). Huge amounts of data generated by health care systems and epidemiological studies contain hidden knowledge, which is unfeasible to uncover using traditional methods, thus, application of data mining is more adapted for medical studies in identification of risk factors for diabetes (Francisci, Brisson, and Collard, 2003). Data mining is the process of extracting hidden patterns or knowledge from large data sets. This is achieved by a combination of statistical and artificial intelligence methods with big data.

Association Rule Mining (ARM) is an important data mining technique that exhaustively mines hidden patterns between variables in a dataset. This characteristic makes it ideal for the discovery of predictive rules from medical databases (Han, Kamber, and Pei, 2006; Ordonez, Ezquerra, and Santana, 2006). An association rule (AR) is a pair (X, Y) of sets of attributes, denoted by X → Y, where X is the antecedent and Y is the consequent of the rule X → Y. The rule states that if X happens, then Y does happen. In general, a set of items, such as X or Y, which are disjoint, is called an item set (Han, Kamber, and Pei, 2006). In medical diagnosis (for example diabetes mellitus), the use of association rules is valuable because apart from quantifying the diabetes risk, they also readily provide the physician with a justification namely the associated set of conditions (Sangeetha and Moorthy, 2015). These conditions provide accurate guidelines towards treatment of a more personalized and targeted preventive care or diabetes management. Several applications of ARM in the medical domain include discovering disease co-occurrences (Cao, Markatou, Melton, Chiang, and Hripcsak, 2005), identifying adverse effects of drugs (Wang, Guo, Xu, Wu, Sun, and Ye, 2012), public health surveillance (Mullins, Siadaty, Lyman, Scully, Garrett and Miller, 2006), detecting risk factors for heart disease and diabetes (Nahar, Imam, Tickle , Y-P, 2013; Simon, Schrom, Castro, Li and Caraballo, 2013) and determining relations among complications or the various diseases that accompany type 2 diabetes (Kim, Shin, Kim and Kim, 2012).

Anomaly detection deals with the process of identifying unexpected trends or patterns in datasets, which differ from the normal behaviour. These non-conforming patterns are mostly referred to as anomalies, outliers or discordant observations. A suitable definition of an abnormality is given as "an observation that deviates so much from other observations as to arouse suspicions that it was caused by a different mechanism" (Dhivya and Merlin, 2015).

Anomaly detection in the medical and public health domains typically work with patient records; the data can have anomalies due to several reasons such as abnormal patient condition, instrumentation errors or recording errors. Thus, anomaly detection is a sensitive problem and requires high degree of accuracy. In contrast to standard classification tasks, anomaly detection is often applied on unlabeled data, taking only the internal structure of the dataset into account (Goldstein and Uchida, 2016).

This paper gears towards detecting anomalies in diabetes mellitus patients' data using rule based approach. This paper adopts association rule mining technique. First, frequent symptoms itemsets or patterns are generated from the dataset. Then association rules are mined from the frequent patterns. Lastly, anomalies are detected using a rule based approach whereby an anomaly count is encountered when a patient symptoms itemset record violates a given association rule. The dataset contains 48 features of 150 diabetes patients' diagnosis. The work is implemented in Python IDLE and performance evaluation is analyzed based on number of frequent patterns generated, number of rules mined and anomalies detected.

This paper gears towards detecting anomalies in diabetes mellitus patients' data using rule based approach. This paper adopts association rule mining technique. First, frequent symptoms itemsets or patterns are generated from the dataset. Then association rules are mined from the frequent patterns. Lastly, anomalies are detected using a rule based approach whereby an anomaly count is encountered when a patient symptoms itemset record violates a given association rule. The dataset contains 48 features of 150 diabetes patients' diagnosis. The work is implemented in Python IDLE and performance evaluation is analyzed based on number of frequent patterns generated, number of rules mined and anomalies detected.

1.1. Related works

Anomalies are patterns in data that do not conform to a well-defined notion of normal behaviour. An outlying observation, or outlier, is one that appears to deviate markedly from other members of the sample in which it occurs (Goldstein and Uchida, 2016). Some literature geared towards anomaly detection in diabetes data using association rule mining and related techniques are presented.

Data mining methods were explored to study the diabetes complication patterns and to unveil the latent association mechanism between treatments and symptoms from large volume of electronic medical records (Liu, Tang, Cheng, Agrawal, Liao, and Choudhary, 2013). The aim of the paper was to utilize the heterogeneous medical records to aid the clinical treatments of Diabetes Mellitus. The paper presented a probabilistic graphical model to link the patient symptoms and physicians' diagnosis, treatment together and to mine diabetes complication and treatment patterns simultaneously. It further studied the demographic statistics of patient population with respect to complication patterns in real data and discovered some interesting insights.

In (Ramezankhani *et al*., 2015), an application of association rule mining to extract risk pattern for type 2 diabetes using Tehran lipid and glucose study database was presented. The aim of this study was to identify risk patterns for type 2 diabetes incidence using association rule mining (ARM). The study used Apriori algorithm in the implementation of the ARM. The study showed that ARM is a useful approach in determining which combinations of variables or predictors occur together frequently, in people who will develop diabetes. There is need for extension of this work in detecting anomalies in the diabetes mellitus instances and use of patient symptoms as variables.

The authors (Vani and Priyadharshni, 2016) presented the use of Association Rule Mining algorithm in decision support systems for diagnosis of diabetes mellitus. The method utilized hierarchical clustering of data to automate the process of association rule mining. The paper presented a clinical decision support system based on data mining to identify whether a patient can be diagnosed with diabetes with probability high, low or medium. This could be used as an effective method because it discovered hidden patterns from the collected facts, it enhanced real-time indicator and discovered bottlenecks and improved information visualization. The approach didn't take into account the detection of anomaly in the diabetes instance cases.

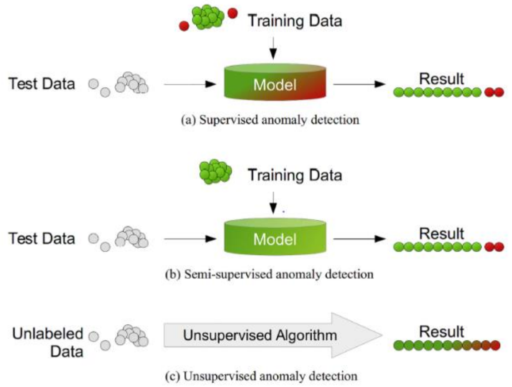
A novel fuzzy clustering (FCM) method for detecting outliers in diabetes data was presented (Vembandasamy and Karthikeyan, 2016). The paper provided the solution for classifying the diabetes found in the data through analyzing the classification by using fuzzy clustering with outlier detection method. The approach used fuzzy clustering with outlier detection method in the data mining with admittance information from patient medical records. Computation of small clusters was done which resulted to outlier detection. The outlier detected data were classified using MPSO-LS-SVM technique for efficient classification diabetes patients. The evaluation was carried on PIMA dataset. A comparative analysis of accuracy and execution time of Modified Particle Swarm Optimization-Least Square Support Vector Machine (MPSO-LS-SVM) and the proposed FCM with outlier method is carried out. The proposed method achieved better accuracy with less execution time than the MPSO-LS-SVM method. This method tends to perform poorer in small datasets due to the density estimation involved in clustering.

2.0 Methodology

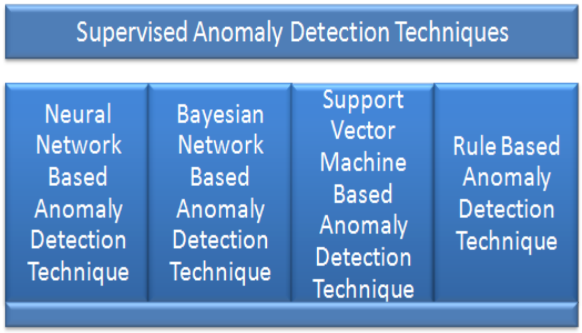
The methodology used in this work is rule based approach which falls under supervised anomaly detection technique. This work further stressed that the choice of anomaly technique is based on the type of anomaly under investigation. Supervised anomaly detection is a technique whereby the data under investigation comprises of fully labelled training and test data sets. Semi-supervised anomaly detection uses training and test datasets but the training data only consists of normal data without any anomalies. Unsupervised anomaly detection does not require any labels. Classification of anomaly detection techniques based on test data is described in figure 1a, while figure 1b shows the supervised anomaly detection techniques.

2.1 Dataset description

The raw dataset of Diabetes patients' symptoms were sourced from Obafemi Awolowo University Teaching Hospital Ile-Ife, Nigeria. The raw data comprise of eight fields and 150 tuples (records). Each attribute describes serial number (S/N), initials, hospital number, gender, age, diagnosis, status and symptom. The initials field was removed for reasons of confidentiality. The last attribute in the raw dataset (symptoms) is the main attribute to be investigated.



(a)



(b)

Figure 1: (a) Different anomaly detection techniques (b) Supervised anomaly detection techniques

2.2 Dataset pre-processing

Data preparation is carried out by selecting and converting the target field (symptoms) in dataset into useable format for the data mining process. The target symptoms field contains string values so the normalization technique used is encoding of each field value (symptom) with numbers from 1 to 48. The number of symptoms present in each diabetic patient instance record as represented with code numbers. The patient record instance has at least one symptom and at most nineteen symptoms.

2.3 Proposed rule based approach

This paper proposes an association rule based approach for detecting anomalies in diabetes mellitus data as shown in figure 2. First, frequent symptoms itemsets or patterns are fetched from the dataset. Then association rules are mined from the frequent patterns. Lastly, anomalies are detected using a rule based violation approach whereby an anomaly count is encountered when an individual patient symptoms set record violates a given association rule.

2.4 Finding frequent symptoms pattern

Finding frequent pattern is based on the Apriori principle (Agrawal and Srikant, 1994) which states that if an itemset is frequent, then all of its subsets are frequent. This principle helps reduce the number of possible interesting itemsets.

Apriori algorithm is used to implement the frequent item set mining over the patient database.

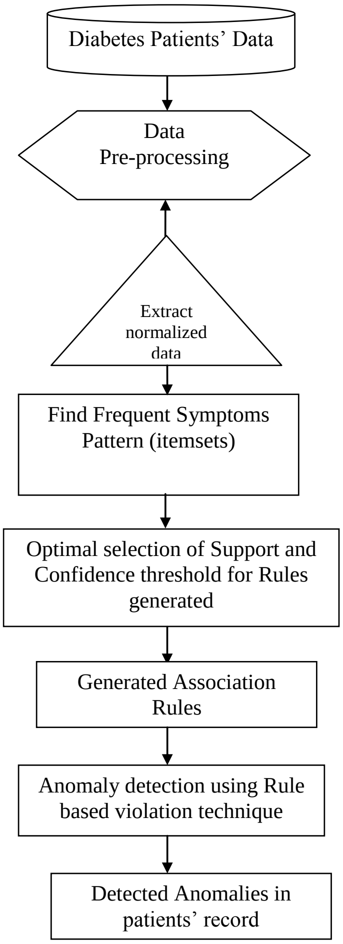


Figure 2: Proposed Rule based Approach for anomaly detection

The steps of the Apriori algorithm are given as:

Listing.1: (Apriori Algorithm)
Step 1: While the number of symptom items in the set is greater than 0
Step 2: Create a list of symptom itemsets of length k
Step 3: Scan the dataset to see if each symptom itemset is frequent
Step4: Keep frequent symptom itemsets to create symptom itemsets of length k+1

Generation of the patient symptom itemsets is carried out by setting a function to create an initial set, and scanning the dataset in search of items that are subsets of transactions. Steps for scanning the dataset are described as;

Listing 2: (Scan Dataset)
Step 1: For each transaction in tran the dataset:
Step 2: For each patient symptom itemset, pat:
Step 3: Check to see if pat is a subset of tran
Step 4: If true, increment the count of pat
Step 5: For each patient symptom itemset:

Step 6: If the support meets the minimum, keep this item
Step 7: Return list of frequent symptom itemsets

2.5 Association rules generation

This loop in listing 2 continues until an empty symptom itemset is reached, and then the list of k-1 itemsets is returned. The resultant rules are generated based on Support and Confidence. The support (S) of the rule (X → Y) is defined as the percentage of records (transactions) in the dataset that contain both X and Y (XUY). Confidence (C) of the rule (X → Y) is the percentage of records in a dataset containing X symptom(s) that also contain Y symptom(s) and is defined by;

$$conf(X \rightarrow Y) = \frac{(\sup(X \cup Y))}{(\sup(X))} \tag{1}$$

A symptom itemset is quantified as frequent if it meets minimum support level. There is a unique measurement for association rules. This metric is shown in equation 1.Applying same apriori principle, if a rule doesn't meet the minimum confidence requirement, then subsets of that rule also won't meet the minimum.

2.6 Association rules base syntax

This describes the integrity constraints imposed on a dataset for the generation of frequent itemsets and association rules. Schemas are defined for each frequent itemset and association rule generated. Let,

L[ ]: [fronzenset([ ])], [support] (2)

Equation 2 is the conforming frequent itemset schema, where L[ ] represents list number, fronzenset( ) is the list of frozen frequent itemsets and [support] is the support value of each list of frozen frequent itemset.
Also let,

R[ ]: fronzensetX([ ]) → fronzensetY([ ]), conf: ( ) (3)

Equation 3 is the conforming association rules schema, where R[ ] represents the rule number, fronzensetX( ) is the list of frozen itemsets on the antecedent part, fronzensetY( ) is the list of frozen itemsets on the consequent part and conf: (value) is the confidence threshold of each rule.

2.7 Rule based anomaly detection

This paper seeks to detect point anomalies in the patients' dataset. A rule based anomaly detection technique is used to detect anomalies in the patient data instance. In this rule based anomaly detection approach, an anomaly occurs for each violation of the rules for any given record. For each patient test instance (Pat_x) in the dataset that violates a rule (R_m) is considered as an anomaly (point anomaly) as shown in expression;

$$Anomaly = \forall Pat_{(x)}, \begin{cases} 1, if\_R_m[itemsets] \not\subset Pat_x(itemsets) \\ 0, if\_R_m[itemsets] \subset Pat_x(itemsets) \end{cases} \tag{4}$$

Anomalies are detected using a rule based violation approach whereby an anomaly count is encountered when a patient symptoms itemset record violates a given association rule.

3.0 Results and performance evaluation

The approach is implemented in Python and evaluation metrics is based on number of frequent patterns generated, number of rules mined and anomalies detected. Table 1 shows total number of frequent itemsets generated for different minimum support levels (from 0.1 to 0.5). Table 2 shows the number of generated association rules at various support levels with minimum confidence thresholds. Thus, optimal mining results occurred at minimum support of 0.4 and minimum confidence threshold of 50% to 70%

beyond which only one strong rule was encountered. Table 3 shows the number of anomalies detected at minimum support of 0.2 and varying minimum confidence level from 0.5 to 1.0 (i.e. 50% to 100% rule confidence levels. The results show that beyond minimum support of 0.4, no anomaly was detected.

Table 1: Total number of frequent itemsets generated from different minimum support

Minimum support	Number of frequent itemsets
0.10	98
0.20	19
0.30	7
0.40	4
0.45	2
0.50	1

Table 2: Total number of generated association rules at various minimum support and minimum confidence thresholds

		Minimum confidence thresholds				
		0.5	0.6	0.7	0.8	0.9
Min. Support	0.10	50	29	13	6	3
	0.20	10	7	4	1	1
	0.30	4	3	3	1	1
	0.40	2	2	2	1	1
	0.45	0	0	0	0	0
	0.50	0	0	0	0	0

Table 3: Total number of anomalies detected and percentage at minimum support 0.2 for different minimum confidence levels

Minimum confidence	Number of Anomalies detected	Percentage in dataset
1	0	0
0.9	3	2%
0.8	3	2%
0.7	36	24%
0.6	69	46%
0.5	88	58.7%

Table 4 shows the number of anomalies detected and percentage of anomalies in dataset at minimum support of 0.3 and varying minimum confidence level from 0.5 to 1.0 (i.e 50% to 100% rule confidence levels). Table 5 shows the number of anomalies detected and percentage of anomalies in dataset at minimum support of 0.4 and varying minimum confidence level from 0.5 to 1.0 (i.e 50% to 100% rule confidence levels). It can be seen that from Table 3 to5, no anomaly was detected when minimum confidence is set 1. This means no anomaly was at 100% chance of occurrence. This is because the model didn't generate any rules for all minimum confidence level at both minimum confidence of 1.

3.1 Performance evaluation

Three evaluation criteria are used due to their widespread relevance in most related literature. They include number of frequent itemsets and rules generated, and anomalies detected (percentage anomaly).
Anomalies detection is reported based on number of patient instances in the diabetes dataset that violates the mined rules using at various minimum confidence levels. Figure 3 shows a chart of anomaly detection of the model using support thresholds of 0.2, 0.3 and 0.4 respectively.

Table 4: Total number of anomalies detected and percentage at minimum support 0.3 for different minimum confidence levels

Minimum confidence	Number of Anomalies detected	Percentage in dataset
1	0	0
0.9	3	2%
0.8	3	2%
0.7	35	23.3%
0.6	35	23.3%
0.5	61	40.7%

Table 5: Total number of anomalies detected and percentage at minimum support 0.4 for different minimum confidence levels

Minimum confidence	Number of Anomalies detected	Percentage in dataset
1	0	0
0.9	3	2%
0.8	3	2%
0.7	25	16.7%
0.6	25	16.7%
0.5	25	16.7%

Table 6: Top ten anomalous diagnostic data points in the diabetes dataset

S/N	Sex	Age	Status	Diagnosis	Anomaly Score
001	Female	59	Alive	diabetic foot ulcer	100%
145	Male	63	Alive	TYPE 11 DM	100%
139	Female	45	Dead	TYPE 11 DM/ Foot ulcer	93%
012	Male	71	Alive	Type 11 DM	60%
016	Female	65	Dead	Type 11 DM	60%
022	Male	48	Alive	DM Nephropathy/CKD	60%
027	Female	60	Dead	Type 11 DM	60%
032	Female	67	Alive	-	60%
033	Female	55	Dead	Diabetic foot gangrene	60%
034	Female	65	Alive	Diabetic foot	60%

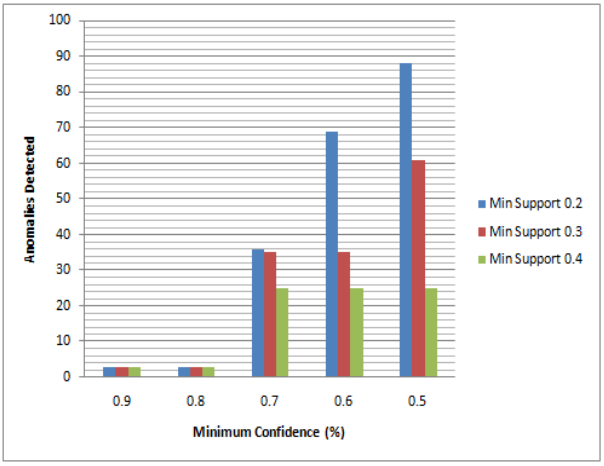


Figure 3: Chart showing Anomalies detected in dataset

The top ten anomalous diagnoses in the diabetes data is shown in table 6. The highest percentage of anomaly detection as shown in figure 3 is approximately 26.54%.

3.2 Discussion on Results

This presents a rule based approach for identifying frequent itemsets patterns, mining association rules and detecting anomalies in the diabetes mellitus patients' dataset.

Drawing inference from the results, it is discovered that of all notable symptoms of diabetes, Polyuria, Polydipsia, Weakness or General body weakness are the top three major symptoms responsible for diabetes disease. The results from the anomaly detection show that an average percentage of 27% anomalies are present in the diabetes dataset which means 27% of the dataset do not conform to expected pattern. It was observed that high number of anomalies resulted as minimum support threshold was increased. Thus, at a minimum support of 0.45 and above, no anomaly was detected. However, top ten anomalies in the diabetes data were discovered with their respective anomaly score (in percentage).

Table 6 depicts the diabetic patients' instances showing the patient S/N, gender, age, status, diagnosis, and anomaly score. It further shows a percentage probability of anomalous diagnosis in the diabetes patients' dataset using the proposed approach. Table 6 shows that Patient with serial number 001 has 100% anomaly score which implies that none of the rules generated from frequent itemset applied to him/her that is all the symptoms from his/her records do not conform to any of the rules generated. Also Patient with serial number 012 in Table 6 has 60% anomaly score means that when the rules generated were compared with his/her symptoms, 60% which is the percentage in dataset do not conform to expected pattern from the rules. From the anomalies, it can be inferred that most patient already have complications before being admitted to the hospital which is the major reason they have unusual pattern in their symptoms. Seven out of top ten top patients' diagnosis with anomalies are females of which four have death status. It can be seen that anomalous diagnosis of diabetes mellitus disease occurred more in women than in men given the dataset under study.

4.0 CONCLUSION

Diabetes Mellitus (DM) is an accumulation of metabolic infections in which a human being has elevated blood sugar, either for the reason that the pancreas does not generate sufficient insulin, or because cells don't react to the insulin that is generated. This paper applies a rule based approach in detecting point anomalies in diabetes patient's data.

The strongest rule recorded from the rule mining has a confidence of 96%. This means that the occurrence of Polydipsia will result to a 96% probability of a presence of Polyuria in a diabetic patient. The results from the anomaly detection show that an average percentage of 27% anomalies are detected in the diabetes dataset which shows those concerned are already having complications. Early detection of diabetes Mellitus in Patients can help to prevent them from complications. The strong rules generated from the association rule can help individual to examine themselves if there is need to go for diagnosis as soon as possible. The paper further shows a percentage probability of anomalous diagnosis in the diabetes patients' dataset as discovered from the rule base mining. Further studies can be done to investigate why anomaly diagnosis occur more in women than in men.

Future research will be geared towards using unsupervised techniques to detect anomalies in patients' data.

References

Abbasi, A., Peelen, L. M., Corpeleijn, E., Van-der, S. Y. T., Stolk, R. P., Spijkerman, A. M., Van-der, A. D. L., .G., Navis, G., Bakker, S. J., and Beulens, J. W., (2012). Prediction Models for Risk of Developing Type 2 Diabetes: Systematic Literature Search and Independent External Validation Study. doi: 10.1136/bmj.e5900.

Agrawal, R., and Srikant, R., (1994). Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Data Bases, VLDB*, pages 487-499, Santiago, Chile, September 1994.

Cao, H., Markatou, M., Melton, G. B., Chiang, M. F., and Hripcsak, G., (2005). Mining A Clinical Data Warehouse to Discover Disease-Finding Associations Using Co-Occurrence Statistics. AMIA Annual Symposium Proceedings, American Medical Informatics Association; pp. 106.

Dhivya, S., and Merlin, M., (2015). Distributed Association Rule Mining and Summarization for Diabetes Mellitus and Its Co-Morbid Risk Prediction Strategy Using Fuzzy Classifier. *International Journal of Engineering and Applied Sciences (IJEAS),* Vol. 2, Issue 11, pp. 58-62, ISSN: 2394-3661.

Dongaonkar, P. P., and Gaikwad, A., (2015). Diabetes Mellitus Estimation Risk Using Association Rule Mining. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 3(5): 470-472, ISSN: 2321-9653.

Francisci, D., Brisson, L., and Collard, M., (2003). A Scalar Evolutionary Approach to Rule Extraction. Rapport de Recherche;

Goldstein, M., and Uchida, S., (2016). A Comparative Evaluation of Unsupervised Anomaly Detection Algorithms for Multivariate Data. *PLoS ONE*, 11(4):1-31, doi:10.1371/journal.pone.0152173

Han, J., Kamber, M., and Pei, J., (2006). Data Mining: Concepts and Techniques. Morgan Kaufmann pub. 2006. Karthikeyan, T., and Vembandasamy, K., (2015). A Novel Algorithm to Diagnosis Type II Diabetes Mellitus Based on Association Rule Mining Using MPSO-LSSVM with Outlier Detection Method. *Indian Journal of Science and Technology,*Vol. 8(S8), pp. 310–320, DOI: 10.17485/ijst/2015/v8iS8/53631.

Kim, H. S., Shin, A. M., Kim, M. K., Kim, Y. N., (2012). Comorbidity Study on Type 2 Diabetes Mellitus Using Data Mining. *Korean J Intern Med*., Vol. 27, No. 2, pp. 197–202.

Koh, H. C., and Tan, G., (2011). Data Mining Applications in Healthcare. *Journal Health Inf. Manag*. Vol. 19, No. 2, pp. 65.

Liu, L., Tang, J., Cheng, Y., Agrawal, A., Liao, W-K., and Choudhary, A., (2013). Mining Diabetes Complication and Treatment Patterns for Clinical Decision Support. CIKM'13, Oct. 27–Nov. 1, 2013, San Francisco, CA, USA. ACM, http://dx.doi.org/10.1145/2505515.2505549.

Mao, R., Yin, Y., and Pei, P., (2004). Data Mining and Knowledge Discovery. Kluwer Academic Publishers.

Mullins, I. M., Siadaty, M. S., Lyman, J., Scully, K., Garrett, C. T., Miller, W. G., (2006). Data Mining and Clinical Data Repositories: Insights From A 667,000 Patient Data Set. *Comput. Biol. Med*, Vol. 36, No.12, pp. 351–377.

Nahar, J., Imam, T., Tickle, K. S., Chen, Y-P, P., (2013). Association Rule Mining to Detect Factors Which Contribute to Heart Disease in Males and Females. *Expert Systems with Applications*, Volume 40, Issue 4, pp. 1086–1093, https://doi.org/10.1016/j.eswa.2012.08.028.

Ordonez, C., Ezquerra, N., Santana, C. A., (2006). Constraining and Summarizing Association Rules in Medical Data. *Knowledge Information System*, 2006, Vol. 9, No. 3, pp.1–2.

Ramezankhani, A., Pournik, O., Shahrabi, J., Azizi, F., and Hadaegh, F., (2015). An Application of Association Rule Mining to Extract Risk Pattern for Type 2 Diabetes Using Tehran Lipid and Glucose Study Database. [15]*International Journal of Endocrinol Metab*, Vol. 13, No. 2, DOI: 10.5812/ijem.25389.

Sangeetha, R., and Moorthy, M. V., (2015). Predicting Relative Risk for Diabetes Mellitus Using Association Rule Summarization in EMR.*International Journal of Science and Research (IJSR)*, Vol. 4, Issue 3, pp. 1118- 1122, ISSN (Online): 2319-7064.

Simon, G. J., Schrom, J., Castro, M. R., Li, P. W., Caraballo, P. J., (2013). Survival association Rule Mining Towards Type 2 Diabetes Risk Assessment. *AMIA Annu Symp Proc*.2013, pp.1293–1302.

Vani, B., and Priyadharshni, J., (2016). Discovering the Diagnosis of Diabetes Mellitus by Using Association Rule Mining, *International Journal Research Instinct*, Vol. 3, Issue 2, pp. 87-95, ISSN: 2348-2095.

Vembandasamy, K., and Karthikeyan, T., (2016). Novel Outlier Detection in Diabetics Classification Using Data Mining Techniques, *International Journal of Applied Engineering Research*, 11(2):1400-1403, ISSN 0973-4562.

Wang, C., Guo, X. J., Xu, J. F., Wu, C., Sun, Y. L., Ye, X. F., (2012). Exploration of The Association Rules Mining Technique For The Signal Detection of Adverse Drug Events in Spontaneous Reporting Systems. *PLoS One*, 7(7).