

Clinical Decision Support Systems Using Sequential Pattern Mining Algorithms for Cardio Vascular Diseases

Cherukuri Harini¹; Dr.V. Maria Anu²

¹Assistant Professor, Department of CSE, Adikavi Nannaya University, Rajahmundry, AP, India.

²Associate Professor, School of Computing, Sathyabama Institute of Science and Technology, India.

²maniwebinar@gmail.com

Abstract

In the medical field, Cardio Vascular disease (CVD) considered to be the treacherous diseases in all aspects also it leads to the stroke, heart attack, angina (chest torment) etc Generally the classifier used to analyze and foretell the disease are Support Vector Machine Classifier (SVMC), Logistic Regression Classifier (LRC), Random Forest Classifier (RFC), Decision Tree Classifier (DTC) and K-nearest neighbours Classifier (KNNC). While diagnose in this way, classifier may misclassify due to the vast amount of data being generated in all the fields and also its time consuming in order to detect some hard disease in early stage. So we propose an improved sequential pattern mining algorithm (Two phase) combined with association pattern mining (APM) method. Here we grouped data's based on their similarity of symptoms also examined by use of discriminant analysis (DA) to check the grouping significance. Then we build a clinical decision support system by using sequential pattern mining along with association pattern mining. Then the results are compared by using evaluation metrics.

Key-words: Cardio Vascular Disease, Machine Learning, Sequential Pattern Mining, Association Pattern Mining.

1. Introduction

Globally the cardio vascular disease is one of the remarkable precarious diseases. When compared to few another ailments, the inhabitant's substantial section chunks the dust from the cardio vascular disease (CVD) per annum. As per the report [1], in 2016, 17.9 million persons are died on due to CVD. It is nearly 31% per individual demise globally. Among these deaths, 85% are died due to the heart failure and stroke. In the downcast countries, greater than three fourth of cardio vascular deaths are turned out. In the year 2015, because of non-infectious ailment,82% are in the hopeless countries

and also 37 % are conveyed through the cardio vascular disease out of 17 million fewer than ideal closures [2]. Most of the cardio vascular disease are might be destroyed to naturally disposed towards distinguished dangerous aspects. Using tobacco, uncertain eating habit and bigness, devastating exertion of liquor, and the physical dormancies are some of the examples of dangerous aspects. Either the person with CVD or who are at greater cardiovascular symptoms (due to proximity of one hazard aspect at least once. For instance, ie. Diabetes, hyperlipidaemia also hypertension which are very adverse and not resolved easily and have require more than a step. At the initial step, as the prescriptions provided by the directorate was not susceptible easily, also as a set apart. As a evolution of greasy stores [3], Cardio vascular disease is puffed out also within the conduits (atherosclories) and the blood clusters in all disciplines had an extreme growth. In such a way, have a chance to create very harmful to the courses in some organs which is heart, kidney, mind and some other minacious diseases. Also Cardio vascular diseases paves the way to extreme dangerous function and major chance of getting lots of disabilities and leads to death also. Moreover it can frequently to a higher boundary must be eliminated with the push of standard of lifestyle. Earnest moments preferred the coronary episodes and strokes. This is one the most treacherous obstruction and also prevent bloodstream in such a way most of the heart and mind stress get resolved. This broadly analysed the disease and the evaluation of fawning stores to splitters of veins. So the failure of cardiovascular disease and stroke which usually produce proximity of the stir and leads to the dangerous symptoms in some aspects. For instance, because of eating habit is not in a sequenced way, consume large amount of tobacco and fattiness.

Data mining technology has the distinctive features especially in the clinical field like as its multi-formality, moral deliberation, economical criteria and medicine's particular level. so due to this, data mining technology in the clinical area has grown into a remarkable study concept, in particular extensive study of the health pharmaceutical analyses record and foretelling ailments from them.[5-7]. In spite of the finest understanding, some of the records have attempted to the forecast the patterns of regular hospital visits for clinical analysis by regarding association patter depends on clinical information .It forecast the inmate's health record whether the person is sick or not. Machine learning methodology had used to foretell the sick person's delicacy (a patients supervision require) in the before research [8]. Sequential pattern matching (SPM) was utilized for auguring few varieties of the infections [6] and association rule mining (ARM) for doctor report depends on the chronicle of nursling's report was proposed by [9]. SPM is generally utilized to find the either most used patterns or characteristics which present in the datasets in the clinical aspects. Machine learning classifiers are

decided to be a perfect techniques as per the before study due to its easily classify and forecast either the sufferer have infection or not.

Even it implies for the prediction process only, it does not exactly forecast the disease in sequenced manner. Hence the foretelling is not completed at the beginning step earlier. But there is a chance of obtaining the results in the final step. Due to this demerit, we introduced a sequential pattern mining along with association rules.

2. Related Study

2.1. Sequential Pattern Mining (SPM)

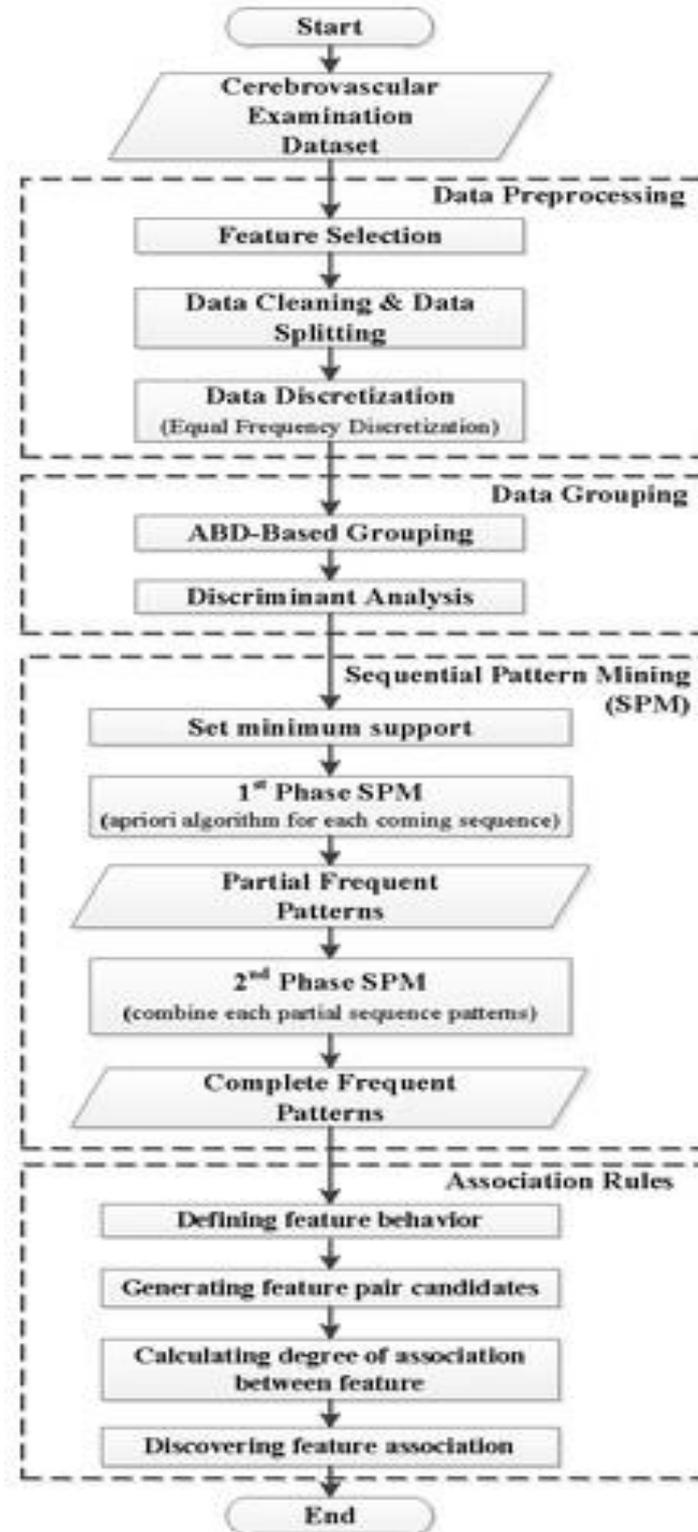
Totally 12 association mining in various reports, examining the co-prescribing pattern of antacid medicines for a occasion [13]. Discovering the comorbidity of attention deficit/hyperactivity disorder (ADHD)[17] and analysing the sets of medicines integrated with the Stevens johnson syndrome reception many researchers have presented the significance of forecasting the needs in clinical filed usually because of utilizing the tactic plan. Foretelling the needs in an extremity section(MC Carthy t.al.,2008) and health protection functions [18], and forecasting the seek person's calls at an immediate wellness centre[19] are proposed. Also, minor reports have examined features of recall patterns of sufferer with our distinctive understanding particularly utilizing a cerebro vascular health analysing dataset. Hence this research integrated few methods utilized by before studies to escape into the problems. The implemented technique includes three important sections. They are data discretization, SPM and association mining (AM).

In our proposed method [20], the information discretization portion injects same equal discretization (|EFD). The core section of this process is priori dependent sequential pattern mining implemented by [21]. The next method is utilized to identify important data of regular (fully or half fully ordered) continuation .It is mostly used in promoting to discovering clients characters. This research adapts the technique for discovering out sufferer's recall character leads to frequent health check-up process regarding a risk of stroke. The final section of the implemented technique utilized a priori dependent integration mining that is used to examine the sets of medicines combined with SJS remedy [22]. The another techniques used to examine sets of medical aspects entities integrated with the out sufferer's revisit characters. A minor accommodation requires to be opinion to integrate the AM method with pattern mining in before phase.

3. Work Flow of the Proposed Method

3.1. Selecting Normalised Features

Figure 1 - Proposed Flow Chart



While a selecting normalised feature from the dataset, the instance reference of sufferer send back reports to the existing selected instance and also depends on detecting the stroke's determinant condition of age, History regarding to blood pressure and also history regarding to diabetes. The forwarding examination of this study will be interlinked to ABD. Future more another seven factors regarding health also the instance references correspondingly. Frequently it is presented in the before analysis on the clinical check-up like blood pressure, body mass index, neck vascular ultrasound, magnetic resonance imaging, blood sugar cholesterol and triglycerides. To guarantee these variable were selected appropriately, researchers were negotiated beyond the selection of literature dependent variables.

3.2. Association Pattern Mining

According to the full regular sequence patterns of every c-visit groups, the association mining technique might be injected to get integrations of regular sequence patterns that have greater association calculation. The AM technique might be operated by make use of following procedures [9]

Step 1. Defining Health Factor Variable Behaviours

The existence of health aspect variables in every visit sequence for whole c-visit group is explained by makes use of particular notation The current value of health factor variable in FSP of three visit group is represented by

$$\text{Sequence} = F_{ij}$$

Where,

- f- Factors regarding health instance reference.
- i- Existence abnormal instance reference of value F at the initial visit which means i='1' also that denotes the value of abnormal had been present, else it is i='0'.
- j- Existence of an abnormal value at the 2nd visit.
- Remember that the subscript alphabet letter come after the frequency value of visit sequences from the c-visit groups, in which (c-1).
- For example, there are four visit groups, then, c-1=4-1=3 visit sequences. Then this is indicated by f_{ijk} and so on.

EXAMPLE: In case the blood pressure abnormal (BPA) and neck vascular ultrasound abnormal(NVA) are included in the FSP variable set in the three visit group. the abnormal blood

pressure is identified in the first visit itself, if the previous variable behaviour is presenting BPA₁₀. Blood pressure might be normal during the second visit. Therefore the variable behaviour (BP₁₀) is indicated by the general behaviour notation of f₁₀. Additionally, the latter variable behaviour is indicated by NVA₁₁. Next for the first and second visits, NVA continued. The general behaviour notations for NVA₁₁ is f₁₁

Step 2. Generating Variable Pair Candidates

Sets of variable behaviours in accomplished Frequency Sequence Pattern (FSP) are created to show the association among the couple variables in pairs. For two visit group there is only possible pair. Similarly six possible pairs for three visit groups

- $\{f_1\} \wedge \{f_1\}$.- only possible pair in the two visit group
- $\{f_{11}\} \wedge \{f_{11}\}$, $\{f_{11}\} \wedge \{f_{10}\}$, $\{f_{11}\} \wedge \{f_{01}\}$, $\{f_{10}\} \wedge \{f_{10}\}$, $\{f_{10}\} \wedge \{f_{01}\}$, and $\{f_{01}\} \wedge \{f_{01}\}$. – six possible pairs in the three visit group.

According to all possible pairs of usual characters, the variable pair applicants can be created from the perfect accomplished frequent pattern of c visits.

Step 3: Calculating the Degree of Association Between Variables

The degree of association for every pair is measure after all whole variable pair were created. The two symmetric measures such as Cosine and Jaccard were picked out due to the association of every variable present in the pair is same. Hence, there is no if- then association in every pair. Azevedo and Jorge indicated that cosine and Jaccard are utilized to calculate the degree of overlap in the middle of cases balanced by every variables. Jaccard assess the distance allying the two variables as the fraction of cases offset by variables accordance with fraction of instance make up by one among them. Greater values of Jaccard shows in eq (2) that is two variables which are responsible to makeup the homogenous cases. Contrastingly, Cosine is utilized to calculate the distance which unites the variable while they are observed as two binary vectors eq(3), if the cosine value is 1, then that two variables are reviewed to be match. If the cosine value is 0, then those vectors does not have overlap

$$\text{JACCARD}(B) = \frac{P(A \cap B)}{P(A) + P(B) + P(A \cap B)}$$

$$\text{COSINE}(A) = \frac{P(A \cap B)}{\sqrt{P(A) + P(B) + P(A \cap B)}}$$

- P(A) & P(B) – Probabilities that variable A and B present in the data.

- $P(A,B)$ - probability that variable A and B present at the sometime in the dataset.

Step 4: Discovering Variable Associations

Regarding to cosine and jaccard values brought from the before step, whole pairs of variables were listed. A greater value represented that a set of variables had a powerful combination. Hence those particular pairs may be examined as the cause why our sufferers supervised for a health check-up.

3.3. Sequential Pattern Mining (SPM)

Since the issues of mining focus on subsequences in a pair of sequences, Agrawal and Srikant proposed a SPM method[4]. Naturally, it was applied after the sketching. After the conversion of time series into sequences, it can also be applied to time series using discretization techniques.

Sequence $S_p = (p_1, p_2, \dots, p_n) \rightarrow$ said to be length of k or k -sequence if it has k items

In other words if $k = |p_1| + |p_2| + \dots + |p_n|$.

For example, the sequence $(\{a, b\}, \{r\}, \{f, g\}, \{v\}, \{e\})$ is a 7-sequence. A sequence database SDB is a list of sequences $SDB = (S_1, S_2, \dots, S_P)$ containing sequence identifiers (SIDs).

Table 1 - Database in the Sequential Manner

SID	SEQUENCE
1	$(\{p, q\}, \{u\}, \{u, v\}, \{v\}, \{t\})$
2	$(\{p, s\}, \{r\}, \{q\}, \{p, q, t, u\})$
3	$(\{p\}, \{q\}, \{u\}, \{t\})$
4	$(\{q\}, \{t, v\})$

A sequence $S_P = (P_1, P_2, \dots, P_n)$ is said to be contained in another sequence $S_Q = (Q_1, Q_2, \dots, Q_m)$ if and only if there exist integers $1 \leq i_1 < i_2 < \dots < i_n \leq m$ such that $P_1 \subseteq Q_{i_1}, p_2 \subseteq Q_{i_2}, \dots, P_n \subseteq Q_{i_n}$ (denoted as $S_A \subseteq S_B$). For example, the sequence $(\{q\}, \{u, v\})$ is contained in sequence $(\{p, q\}, \{r\}, \{u, v\}, \{v\}, \{t\})$, while the sequence $(\{q\}, \{v\}, \{u\})$ is not. If sequence S_a is contained in a sequence S_q , S_a is said to be a subsequence of S_b .

The main objective of SPM is to identify focusing subsequence's in a sequence database i.e., sequential associations alluring items are intended for the client. Several measures might be used to evaluate how intended a subsequence is. The support measure is used in the true issue of SPM. The

support (or absolute support) of a sequence S_p in a sequence database (SDB) is described as the no. of sequences that contain S_p and is denoted by $\text{sup}(S_p)$. In other words,

$$\text{sup}(S_p) = |\{S \mid S \vee S_p \wedge S \in \text{SSQ}\}|.$$

Example: The support of the sequence $(\{q\}, \{f, g\})$ in the database of table 1 is 3 due to the sequence present in the three sequences (1, 2 and 4). Remember that the few articles explain the support of sequence x as the ratio. This explanation called relative support is

$$\text{relSup}(S_p) = \text{sup}(S_p) / |\text{SSQ}|$$

i.e. the no. of sequences including s_a divided by the no. of sequence in the database.

The relative support of the item set $(\{q\}, \{f, g\})$ is 0.75 for a instance

In the below table 0 indicates 'No' and 1 indicates 'Yes'.

Table 2 - Sequential Pattern Found in the Table 1

PATTERN	SUP.	CLOSED?	MAXIMAL	GENERATOR
({p})	3	1	0	0
({p}, {v})	2	1	0	1
({p}, {v}, {e})	2	1	1	0
({p}, {u})	3	1	1	0
({p}, {u}, {e})	2	1	1	0
({p}, {r})	2	1	0	1
({p}, {r}, {u})	2	1	0	0
({p}, {r}, {t})	2	0	0	0
({p}, {q})	2	0	0	1
({p}, {q}, {u})	2	0	0	0
({p}, {q}, {t})	2	1	0	0
({p}, {t})	3	1	1	1
({a, b})	2	1	1	0
({q})	4	0	1	1
({q}, {v})	3	0	1	1
({q}, {v}, {t})	2	0	0	1
({q}, {u})	4	1	1	0
({q}, {f, g})	3	1	1	1
({q}, {u}, {t})	2	0	1	0
({q}, {t})	3	1	0	1
({r})	2	0	0	1
({r}, {u})	2	0	0	1
({r}, {t})	2	0	0	1
({t})	3	0	0	1
({u})	4	0	0	1
({f, g})	3	0	0	1
({u}, {t})	2	0	0	1
({v})	3	0	0	1
({v}, {t})	2	0	0	1

4. Experimentation and Result Analysis

Now we are going to examine the results of foretelling the cardiovascular infections. The jupyter web application environment utilizes the quad core i5 system with the RAM size of 6GB pandas, Ipython, Scipy, StatsModels and the matplotlib. We did the experimental analysis in two levels. In the first level, three process are done. i) Dataset is cleaned by make use of pandas tool ii) Extracting normalised features iii) sequential patterns are found with respect to the small data integrated with association rules. The final normalized dataset is procured by the end of feature selection process as illustrated in the table 3.

Table 3 - Resultant Dataset after Feature Selection

S.no	Perceived inputs	Instance reference
1	0.0362	(normalized) Age
2	0.12	(normalized) Sex=Female
3	3 - 1.67491	(normalized) Cp=1: typical angina
4	1.4256	(normalized) Cp=2: atypical angina
5	1.5893	(normalized) Cp=3: non-anginal
6	1.3885	(normalized) Cp=4: asymptomatic
7	10.8347	(normalized) Trestbps
8	0.1542	(normalized) Chol
9	0.002	(normalized) Fbs=no
10	4.379	(normalized) Restecg=0: normal
11	3.9892	(normalized) Restecg=1: ST-T wave abnormality
13	0.1921	(normalized) Thalach
14	141.28	(normalized) Exang=no
15	150.0348	(normalized) Oldpeak
16	0.08	(normalized) Smoke=1
17	0.03	(normalized) Cigs
18	0.1568	(normalized) Years
19	0.5	(normalized) Famhist=1
20	0.90557	(normalized) Activity=Seldom
21	0.69975	(normalized) Activity=Daily
22	0.63876	(normalized) Activity=Yearly
23	0.64956	(normalized) Activity=Once
24	0.68987	(normalized) Activity=Monthly
25	0.87865	(normalized) Activity=Often
26	0.83875	(normalized) Activity=Weekly
27	0.83924	(normalized) Activity=Never

Then the normalized dataset is passed into a two phase of SPM. In that SPM has been set some minimum support value to the dataset of each attributes. Then phase I of SPM algorithm applies apriori

algorithm. By means that partial frequent pattern is obtained. Then phase II takes the result of phase I, and applies the apriori algorithm again. Then the result as complete frequent patterns. After that resultant is passed to the association rules, then the resultant showed final extracting pattern of cardiovascular dataset. Hence decision support system predicts the diseases in a much advanced way. Then the resultant accuracy is compared with the existing classifier accuracy.

Here the result are compared and evaluated through the accuracy. The graph shows that our proposed SPM along with association rules produces better accuracy compared with all other existing classifiers.

The performance of algorithms is typically developed by a confusion matrix as illustrated in Table 4 (for a 2-class problem). The rows are actual class and the columns are predicted class.

Table 4 - Confusion Matrix

	Predicted as Negative	Predicted as Positive
Actually it is Positive	FN	TP
Actually it is Negative	TN	FP

In the confusion matrix,

- TN –Usually it is negative but classifier predicted as positives (True Negatives).
- FP – Usually it is negative but classifier predicted as positive(False Positives).
- FN-Usually it is positive but classifier predicted as positive (False Negatives).
- TP –Usually it is positive but classifier predicted as negative (True Positives).

Consequently, Predictive accuracy is the performance measures generally associated with pattern mining algorithms.

Table 5 - Accuracy of Existing Classifier

Classifier	Accuracy	Inaccuracy
RFC	85.72	14.28
DTC	78.29	21.71
LRC	75.23	24.77
SVMC	77.29	22.71
KNNC	69.63	30.77

Table 5 shows the accuracy of existing classifiers. Out of all existing classifiers random forest produces good accuracy while dealing with cardiovascular dataset.

Table 6 - Accuracy of Pattern Mining Algorithms

Pattern mining Algorithms	Accuracy	Inaccuracy
Sequential pattern mining	83.41	16.59
Association pattern mining	82.06	17.94
sequential pattern mining along with association rules	84.02	15.98

Table 6 shows the accuracy of various pattern mining algorithms. In that SPM along with association rules produces the better accuracy (84.02).

Table 7 - Accuracy of Improved Sequential Pattern Mining Algorithms

Pattern mining algorithms	Accuracy	Inaccuracy
Feature selection dataset	69.82	30.18
Improved sequential pattern mining(Two Phase)	89.65	10.35
Improved sequential pattern mining along with association rules	92.01	7.99

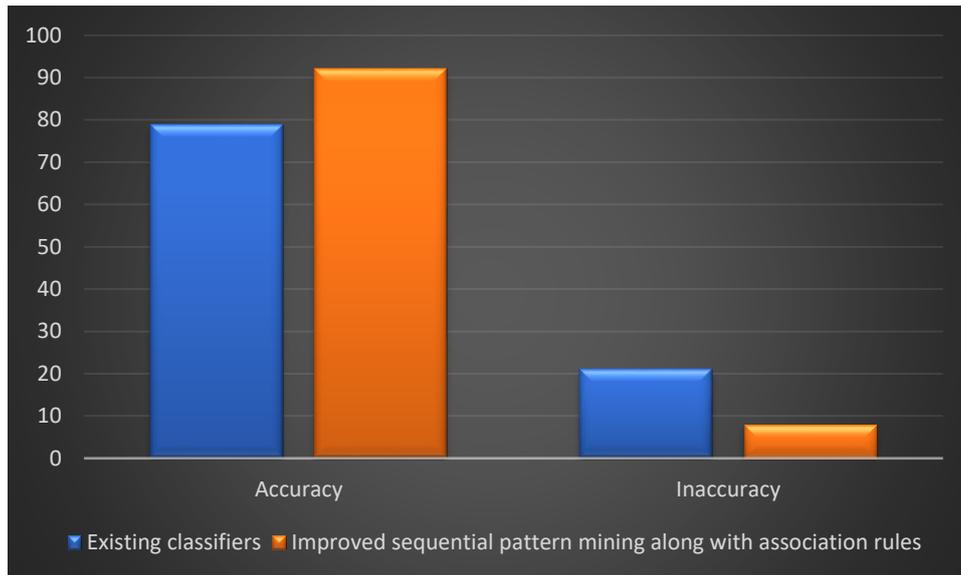
Table 7 shows the accuracy of our proposed methodology. Our proposed improved SPM along with association rules produces best accuracy (92.01).

Table 8 - Comparison of overall Classifier Accuracy with Proposed Improved SPM Algorithms

Existing Classifiers and pattern mining algorithms	Accuracy	Inaccuracy
Existing classifiers	78.9	21.1
Improved sequential pattern mining along with association rules	92.01	7.99

Note: Existing classifiers means the average of all existing classifiers [Random Forest Classifier(RFC), Decision Tree Classifier(DTC), Logistic Regression Classifier(LRC), Support Vector Machine Classifier(SVMC), K-Nearest Neighbor Classifier(KNNC)]

Table 8 demonstrates the overall comparison of accuracy of the existing classifiers with our proposed improved SPM algorithms. The comparison shows our proposed improved SPM along with association rules produces best accuracy (92.01).



5. Conclusion

In this research work, we build a clinical decision support system by use of improved sequential pattern mining along with association rules to predict cardiovascular diseases. In the existing literature survey, they used classifier to predict the disease, but it does not helpful to find the diseases at the earlier stage. Hence we applied the advancement of sequential pattern mining(SPM) ie. Improved SPM(Two phase) along with association rules has been applied to the dataset also we achieved a best accuracy when compared with existing classifiers. Our proposed SPM with ARM algorithm produces the best accuracy **92.01%** while predicting the disease. Hence our clinical decision support system diagnose and predict the disease at the earlier stage. In future work, this approach can be applied in the big data environment along with map reduce technique.

References

[https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvd\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvd)).

Kelly, B.B., & Fuster, V. (Eds.). (2010). *Promoting cardiovascular health in the developing world: a critical challenge to achieve global health*. National Academies Press.

Poirier, Paul, et al. "Obesity and cardiovascular disease: pathophysiology, evaluation, and effect of weight loss: an update of the 1997 American Heart Association Scientific Statement on Obesity and Heart Disease from the Obesity Committee of the Council on Nutrition, Physical Activity, and Metabolism." *Circulation* 113.6 (2006): 898-918.

Bhatnagar, Prachi, et al. "Trends in the epidemiology of cardiovascular disease in the UK." *Heart* 102.24 (2016): 1945-1952.

- Koh HC, Tan G. 2005. Data mining applications in healthcare. *Journal of Healthcare Information Management* 19(2):64-72
- Ohara C, Inoue K, Kashima S, Inoue M, Akimoto K. 2013. Undiagnosed diabetes has poorer profiles for cardiovascular and metabolic markers than known diabetes: the Yuport medical checkup center study. *Diabetes Research and Clinical Practice* 101(2): e7–e10
- Suka M, Odajima T, Okamoto M, Sumitani M, Nakayama T, Sugimori H. 2014. Reading comprehension of health checkup reports and health literacy in Japanese people. *Environmental Health and Preventive Medicine* 19(4):295-306
- Kontio E, Airola A, Pahikkala T, Lundgren-Laine H, Junttila K, Korvenranta H, Salakoski T, Salanterä S. 2014. Predicting patient acuity from electronic patient records. *Journal of Biomedical Informatics* 51:35-40
- Ou-Yang C, Agustianty S, Wang HC. 2013. Developing a data mining approach to investigate association between physician prescription and patient outcome—a study on re-hospitalization in Stevens-Johnson syndrome. *Computer Methods and Programs in Biomedicine* 112(1):84-91
- Agrawal R, Srikant R. 1995. Mining sequential patterns. In: Proceedings of the eleventh international conference on data engineering. *Piscataway. IEEE.* 3-14
- Berlingerio M, Bonchi F, Giannotti F, Turini F. 2007. Time-annotated sequences for medical data mining. In: *Seventh IEEE international conference on data mining workshops (ICDMW 2007)*. 133-138
- Ali R, ElHelw M, Atallah L, Lo B, Yang G-Z. 2008. Pattern mining for routine behaviour discovery in pervasive healthcare environments. In: 2008 ITAB 2008 *international conference on information technology and applications in biomedicine*. *Piscataway. IEEE.* 241-244
- Chen TJ, Liu JY, Hwang SJ. 2002. Non-steroidal anti-inflammatory drug and ant acid co-prescription in Tai wan: analysis of national insurance claims. *Chinese Medical Journal* 65:588-593
- Srikant R, Agrawal R. 1996. Mining sequential patterns: generalizations and performance improvements. In: EDBT'96 proceedings of the 5th international conference on extending database technology: advances in database technology. Berlin. Springer. 1-17
- Soni J, Ansari U, Sharma D, Soni S. 2011. Predictive data mining for medical diagnosis: an overview of heart disease prediction. *International Journal of Computer Applications* 17(8):43-48
- Ji Y, Ying H, Dews P, Mansour A, Tran J, Miller RE, Massanari RM. 2011. A potential causal association mining algorithm for screening adverse drug reactions in post marketing surveillance. *IEEE Transactions on Information Technology in Biomedicine* 15(3):428-437
- Tai YM, Chiu HW. 2009. Comorbidity study of ADHD: applying (ARM) to national health insurance database of Taiwan. *International Journal of Medical Informatics* 78(12): e75–e83
- Al Nuaimi N. 2014. Data mining approaches for predicting demand for healthcare services in Abu Dhabi. In: Innovations in information technology (Innovations), 2014 10th international conference on. *Piscataway. IEEE.* 42-47.
- Batal H, Tench J, McMillan S, Adams J, Mehler PS. 2001. Predicting patient visits to an urgent care clinic using calendar variables. *Academic Emergency Medicine* 8(1):48-53.
- Wong AKC, Chiu DKY. 1987. Synthesizing statistical knowledge from incomplete mixed-mode data. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 9(6):796-805.

- Garcia S, Luengo J, Sáez JA, Lopez V, Herrera F. 2013. A survey of discretization techniques: taxonomy and empirical analysis in supervised learning. *IEEE Transactions on Knowledge and Data Engineering* 25(4):734-750
- Agrawal R, Srikant R. 1995. Mining sequential patterns. In: Proceedings of the eleventh international conference on data engineering. Piscataway. IEEE. 3-14
- Dinesh, Kumar G., et al. "Prediction of Cardiovascular Disease Using Machine Learning Algorithms." 2018 *International Conference on Current Trends towards Converging Technologies (ICCTCT)*. IEEE, 2018.
- Mezzatesta, Sabrina, et al. "A machine learning-based approach for predicting the outbreak of cardiovascular diseases in patients on dialysis." *Computer Methods and Programs in Biomedicine* 177 (2019): 9-15.
- Terrada, Oumaima, et al. "Classification and Prediction of atherosclerosis diseases using machine learning algorithms." 2019 5th *International Conference on Optimization and Applications (ICOA)*. IEEE, 2019.
- Alić, Berina, Lejla Gurbeta, and Almir Badnjević. "Machine learning techniques for classification of diabetes and cardiovascular diseases." 2017 6th *Mediterranean Conference on Embedded Computing (MECO)*. IEEE, 2017.
- Awan, Shahid Mehmood, Muhammad Usama Riaz, and Abdul Ghaffar Khan. "Prediction of heart disease using artificial neural network." *VFAST Transactions on Software Engineering* 13.3 (2018): 102-112.
- Fathalla, Karma M., et al. "Cardiovascular risk prediction based on Retinal Vessel Analysis using machine learning." 2016 *IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2016.
- Gjoreski, Martin, et al. "Chronic Heart Failure Detection from Heart Sounds Using a Stack of Machine-Learning Classifiers." 2017 *International Conference on Intelligent Environments (IE)*. IEEE, 2017.
- Metsker, Oleg, et al. "Dynamic mortality prediction using machine learning techniques for acute cardiovascular cases." *Procedia Computer Science* 136(2018): 351-358.
- Balasubramanian, Vineeth Nallure, et al. "Support vector machine based conformal predictors for risk of complications following a coronary drug eluting stent procedure." 2009 36th *Annual Computers in Cardiology Conference (CinC)*. IEEE, 2009.
- Groselj, C., et al. "Machine learning improves the accuracy of coronary artery disease diagnostic methods." *Computers in Cardiology* 1997. IEEE, 1997.
- Zhou, Yijiang, et al. "Machine Learning-Based Cardiovascular Event Prediction For Percutaneous Coronary Intervention." *Journal of the American College of Cardiology* 73.9 Supplement 1 (2019): 127.
- Singh, Manpreet, et al. "Building a cardiovascular disease predictive model using structural equation model & fuzzy cognitive map." 2016 *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*. IEEE, 2016.
- Alaa, Ahmed M., et al. "Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK Biobank participants." *PloS one* 14.5 (2019): e0213653.

Exarchos TP, Papaloukas C, Lampros C, Fotiadis DI, editors. Protein classification using sequential pattern mining. Engineering in medicine and biology society. In: 2006 EMBS '06 28th annual international conference of the IEEE; 2006 [August 30, 2006–September 3, 2006].

Aseervatham S, Osmani A. Mining short sequential patterns for hepatitis type detection. *Proc ECML/PKDD 2005 Discov Challenge*; 2005.

Barrientos LA. *Insider trading sequential pattern mining (INTRASPAM)*; 2012.

Julea A, Méger N, Trouvé E, Bolon P, editors. On extracting evolutions from satellite image time series. In: Geoscience and remote sensing symposium, 2008 *IGARSS 2008 IEEE international. IEEE*; 2008.

Buchta C, Hahsler M, Buchta MC. Package 'arules Sequences'; 2012.

Parikh R, editor. *Cesarean section rate variation across hospital referral regions in Texas: a claims analysis of privately insured population from 2008–2011*. In: 141st APHA annual meeting (November 2–November 6, 2013). APHA; 2013.