International Journal on Emerging Technologies **11**(2): 110-115(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Data Driven Diagnosis of Cervical Cancer using Association Rule Mining with Trivial Rule Expulsion Approach

K. Logeswaran[1], P. Suresh[2], S. Savitha[3], K. R. Prasanna Kumar[4] A. P. Ponselvakumar[5] and A. Rajiv Kannan[6]
[1]Assistant Professor, Department of Information Technology, Kongu Engineering College (Tamilnadu), India.
[2]Associate Professor, Department of Information Technology, Kongu Engineering College (Tamilnadu), India.
[3]Assistant Professor, Department of CSE, K.S.R. College of Engineering (Tamilnadu), India.
[4]Assistant Professor, Department of Information Technology, Kongu Engineering College (Tamilnadu), India.
[5]Assistant Professor, Department of Information Technology, Kongu Engineering College (Tamilnadu), India.
[6]Professor, Department of CSE, K.S.R. College of Engineering (Tamilnadu), India.

(Corresponding author: K. Logeswaran)
(Received 04 December 2019, Revised 30 January 2020, Accepted 04 February 2020)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Cervical cancer is the fourth most prevalent cancer in women according to World Health Organization estimates. In low- and middle- countries such as India, 90% of women diagnosed with cervical cancer may lead to death due to insufficient detection, early diagnosis, successful screening and treatment services. However, if detected correctly at an early stage, cervical cancer is an a vertible and curative form of cancer. Now an immense amount of clinical data is created from numerous sources such as electronic health reports, data from clinical trials, health assessments, and registries of patients and diseases. These data can be used to forecast cancer disease correctly in early stages in an optimal way. As one of the machine learning methods, Association Rule Mining (ARM) is used to extract the fascinating and accurate rule from cervical dataset. A data driven approach called Validation of Association Rule Mining using Test Train Approach (VARMTTA) is proposed. By using a train-test validation method, VARMTTA reduces the volume of rules produced from the dataset. It is important to prevent ambiguity as to how the adequacy and utility of the proposed solution is represented, for which in this method traditional metrics called Sensitivity, Precision and Total Accuracy are used.

Keywords: Association Rule Mining, Cervical Cancer, Cancer, Rule Mining, Trivial rule expulsion, train-test validation, support, confidence, Frequent Itemset Mining, Sensitivity, Specificity and Total Accuracy.

I. INTRODUCTION

More than 16,000 billion cells make up the human body, they are the unit that shapes the tissues that make up the organs including liver, heart, and lungs. As the body had to divide our cells into two and remove those cells which are damaged or end their lifetime. It allows the tissue to maintain its shapes and their respective functions over time. Thus, each cell was programmed to multiply and die. This order and complex system is regulated by the cell centre, the nucleus that contains chromosomes that hold several DNA genes. Often some of these genes underwent a transition, then the nucleus detects anomalous orders and the cells go wrong. It uncontrollably multiplies, and takes on its own name. Each generated cell contains the same defect. The cells regularly proliferate and form a tumour. That may be as low but often long operation. Ten to thirty years will distinguish the birth of first abnormal cells from around one cubic centimeter of the tumor's appearance. The tumor forms several in order blood vessels to survive, which will provide oxygen and nutrients to the tumor to allow it to live and develop. It is dubbed the phenomenon of of angiogenesis. The cancerous cells begin to assault through the blood vessels and affects the surrounding organs of the body. The tumour will really become dangerous when it spreads to other organs of the body. These cells are carried to other parts of the body, multiplied and produce new tumours. Metastasis is the term used for representing this spreading process of cancer.

There are many factors that exists to different degrees which act as reasons to cause cancer. Factors such as here ditary genetic anomalies have been identified. Exposure to some viruses such as HIV, Hepatitis B, C and D, Epstein-Barr and papilloma virus. Exposure to poisonous agents, toxins, radiation and sunlight. Unhealthy habits, including alcohol and tobacco use. Diet too rich in fat, and low in vegetables and fruits. Today, more than twenty-five million people in the world are living with cancer, and seven million of them die every year. That ease is the leading rate of mortality for those in the 65. Lung cancer is the widest spread followed by breast cancer, colorectal cancer, stomach cancer and prostate cancer. Based on the type of cancer and the degree of cancer severity, different therapies are either used individually or in combinations. The aim of the treatment is to facilitate the removal of tumor and cancer cure in the early stages of a patient or like an electronic disease to monitor its growth. There are three main treatment forms. Surgery requires the removal of a portion of the tumor. Radiotherapy, exposes the tumor to gamma radiation, removes the cancer cells and kills them. Chemotherapy consists of taking a systemic medical drug to suppress or prevent the spread of cancer cells. Chemotherapy and radiotherapy work on cancer cells, but with some side effects on healthy cells too.

But today, a greater understanding of the characteristics of cancer cells makes it possible to develop more precise therapies that address the mechanism for cancer development. Every year there are more and more popular treatments that go perfect. But, in many instances these therapies are effective only for short interval and after a time period, cancer cells evolves and comes back in more aggressive and resistant form. This results in disease which is very heterogeneous. It is also identified that each patient with cervical cancer may have different causes and heterogeneous behaviour of cells. So, it is anticipated that all common treatment should not be used for patient with cervical cancer. Most challengeable thing is that how to personalize medical drugs and medical treatment for each patient. More specialized and customized treatment is possible only through data driven approach of machine learning. As a part of machine learning, Association Rule Mining is used here to mine the more interesting rules from the cervical cancer dataset and those rules can be further used to predict the customized treatment required for each patient. Many Machine Learning (ML) algorithms are used in current methods to forecast the diseases. However, it has some common pitfalls that include ML taking enough time to allow the algorithms to learn and improve sufficiently to accomplish their function with sufficient accuracy and relevance. It takes huge funding for it to function as well. That can mean additional machine resource requirements for you. Another major obstacle is the ability to analyze the ML algorithm-generated results correctly. Even you have to pick the algorithms carefully for your function. Machine learning is autonomous but is particularly vulnerable to mistake. Suppose you train an algorithm that includes data sets small enough not to be inclusive. Through identifying usefulness trends from data generated by clinical services, the proposed VARMTTA method would eliminate all of these obstacles.

II. RELATED WORK

Existing studies exposes that association rule mining is mainly used for market basket analysis, to find the frequent pattern. By using mined rules, ecommerce analyst can improve the profit of their business by incorporating appropriate changes in their business using the mined rules. Due to growing technology in medical field, later studies have applied the association rule mining for medical and disease dataset. From the rules mined from the medical and disease dataset, various risk factors precipitate to disease and side effects in the future days of patient after taking treatment for the disease. Yang and Chen (2015) interrelated the clinical knowledge and pathology description, used to estimate a patient's prognosis, and can help physicians to design suitable plan for treatment [1]. Alwidian *et al.*, (2018) used ARM for Association Classification (AC) techniques to enrich the classification process [2]. Most of the AC algorithm generates rules with moderate and minimum accuracy due to prioritization at attribute level and evaluation of rules based on some estimated measure. Alwidian *et al.,* (2018) attempted to solve this problem by introducing statistical measure based prediction and pruning techniques to improve the accuracy of association rule and association classifier [2]. To enhance the trait of oral cancer care data mining techniques were tested on patients pharmaceutical records [3, 4]. Chen and Xu (2014) studied cancer comorbidities that includes common disorders such as Hypertension, cardiovascular disease, diabetes mellitus, Depression and anxiety in patients with colorectal cancer [5]. The accuracy of any cancer predictive model largely depends on the quantity of information about comorbid conditions of patients stored in dataset and by which better diagnostic and treatment decisions can be made by doctors [6]. Although these studies experiment the temporal association between toxicities and cancer treatment it has two drawbacks. First, it generates huge number of rules, as it does not epitomize relevant diagnosis code. Second, it is futile in admitting the chronological evolution of toxicities of cancer treatments. To dazed this drawback, Temporal Association Rule (TAR) is used where rule is framed by mapping diagnosis code in right hand side with diagnosis code on left hand side [7, 8, 9]. HUIM is designed as multi objective optimization problem, which includes more than one support functions, and an appropriate support functions for particular instance can be selected using Reinforcement Learning (RL) [10]. A data driven approach with SVM based approach is used diagnosis the malignant sample of cervical cancer and performs classification based on four target variables such as Hinselmann, Schiller, Cytology and Biopsy out of 30 risk factors [13]. When Association Rule Mining (ARM) is used to mine the alluring rules from cervical cancer dataset, it generates huge amount of irrelevant which leads to more memory consumption and mislead the decision making. To overcome this drawback, a partition based approach is followed which validate the mined rules on each partition and reduce the size of final rules mined [14]. Apriori algorithm is used to mine the interesting association rule from breast cancer dataset [15]. The given heart disease dataset is divided into three partitions named as train_partitions, test_partition and validate_partition and constraint based is search is used to mine rules and finally validated against the validate_partition [16]. To mine the utility trend from the provided transaction dataset, list structure and evolutionary appraoch are used [17].

III. PROPOSED METHODOLOGY

A. Validation of Association Rule Mining using Test Train Approach (VARMTTA)

The proposed VARMTTA needs the following information as an input parameter: the Dataset (D) which can be extracted from the given real-time transaction database, Minimum Support Threshold (δ), Minimum Confidence Threshold (ε) and Minimum Lift Threshold (ϕ). The parameters δ, $\mathcal{E}$, and φ are used to curb the frequent itemset and rules mined using ARM. The size of partition is managed by Train Sample Fraction (λ) and Test Sample Fraction (μ). The output of the VARMTTA is the validated rules (R). The algorithm generates reliable and interesting rules from transaction dataset.

At initial stage, the algorithm follows the Fisher-Yates shuffle algorithm [11] to randomly picks the transaction id from the transaction set, sequentially until there is no

more transaction id in the transaction. The time complexity involved in shambling the transaction is O(n). It works in neutral manner so that it produce as most firm and interesting rules from the shambled transaction set. Logically non-imbricating section are generated from the shambled transaction set. These segments are referred as partitions such as Training Partition (P^{tr}), Testing Partition (P^{te}) and Validation Partition (P^{va}). The parameter λ controls the dealing set to be additional into the partition Ptr. The parameter controls the dealing set to be additional into the partition P^{te} with the constraint $\mu \leq \lambda$. The partition P^{va} size is not considered here.

Now the frequent itemset of size 1 is found and denoted as L1. By using F1, frequent itemset of size 2 is found and denoted as L2. Similarly, frequent itemset of size upto k is found and denoted as Lk respectively. This process of generating Lk continues until there is no more itemset in a transaction set. These k-itemsets are called as Candidate itemset (Ck). Ck is going to be the initial set of input itemset for the algorithm. Now compute the train_support sup p(X, P^{tr}) from equation(1) for each itemset. If the train_support is less than the δ, then that itemset is eliminated from Ck. Similarly all the itemset having train_support less than the δ is removed from the Ck. Finally, only the remaining itemset from the Ck is used to involve in the future process. Association rule is derived from these frequent itemset which are available in P^{tr}.

Once the initial set of rules are mined, Eqns. (2) and (3) are used to calculate the train_confidence and train_lift respectively. Now, the rules which are not satisfying the constraint train_support< δ or train_confidence< Ɛ or train_lift< φ are eliminated and remaining rules are designated into R^{tr}

$$\sup p(x \Rightarrow y) = \frac{|P_{xy}^{tr}|}{|P^{tr}|} \quad (1)$$

$$\mathrm{conf}(x \Rightarrow y) = \frac{P_{xy}^{tr}}{P_{x}^{tr}} \quad (2)$$

$$\mathrm{lift}(x \Rightarrow y) = \frac{|P^{tr}| \times |P_{xy}^{tr}|}{|P_{x}^{tr}| \times |P_{y}^{tr}|} \quad (3)$$

Validate the rules in R^{tr} using test partition P^{te} and designate the valid rules into R^{te}. Using Eqn. (4) to measure the test_support sup(X, P^{te}) for each regular itemset X. For each rule $X \Rightarrow Y \in P^{te}$, using Eqn. (5) and (6) to measure thetest_confidence and test_lift on the P^{te}. Now validate the rule using the constraint $\mathrm{supp}(x \Rightarrow y) < \delta$ or $\mathrm{conf}(x \Rightarrow y) < Ɛ$ or $\mathrm{lift}(x \Rightarrow y) < \phi$ and eliminate the rules from R^{te}

$$\sup p(x \Rightarrow y) = \frac{|P_{xy}^{te}|}{|P^{te}|} \quad (4)$$

$$\mathrm{conf}(x \Rightarrow y) = \frac{P_{xy}^{te}}{P_{x}^{te}} \quad (5)$$

$$\mathrm{lift}(x \Rightarrow y) = \frac{|P^{te}| \times |P_{xy}^{te}|}{|P_{x}^{te}| \times |P_{y}^{te}|} \quad (6)$$

Validate the rules in R^{te} using validation partition P^{te} and designate the valid rules into R^{va}. For each frequent itemset X compute the test_support sup(X, P^{va}) using Eqn. (7). Forany statute $X \Rightarrow Y \in P^{va}$, compute the test_confidence and test_lift on the P^{va} using Eqns. (8) and (9). Now validate the rules using the onstraint $\mathrm{supp}(x \Rightarrow y) < \delta$ or $\mathrm{conf}(x \Rightarrow y) <$ Ɛ or $lift(x \Rightarrow y) < \phi$ and designate all valid rules in to R^{I}.

$$\sup p(x \Rightarrow y) = \frac{|P_{xy}^{va}|}{|P^{va}|} \quad (7)$$

$$conf(x \Rightarrow y) = \frac{P_{xy}^{va}}{P_{x}^{va}} \quad (8)$$

$$lift(x \Rightarrow y) = \frac{|P^{va}| \times |P_{xy}^{va}|}{|P_{x}^{va}| \times |P_{y}^{va}|} \quad (9)$$

The above task is repeated for n times and at the end, the algorithm will produce n set of rules. Average metric of rules are computed using the Eqns. (10), (11) and (12).

$$\sup p(x \Rightarrow y) = \frac{1}{t}\sum_{i=1}^{t} \sup p[(x \Rightarrow y), D_i] \quad (10)$$

$$conf(x \Rightarrow y) = \frac{1}{t}\sum_{i=1}^{t} \mathrm{conf}[(x \Rightarrow y), D_i] \quad (11)$$

$$lift(x \Rightarrow y) = \frac{1}{t}\sum_{i=1}^{t} \mathrm{lift}[(x \Rightarrow y), D_i] \quad (12)$$

From final evaluated rule, the cervical cancer can be classified across four target variable Hinselmann Colposcopy, Schiller, Cytology and Biopsy.

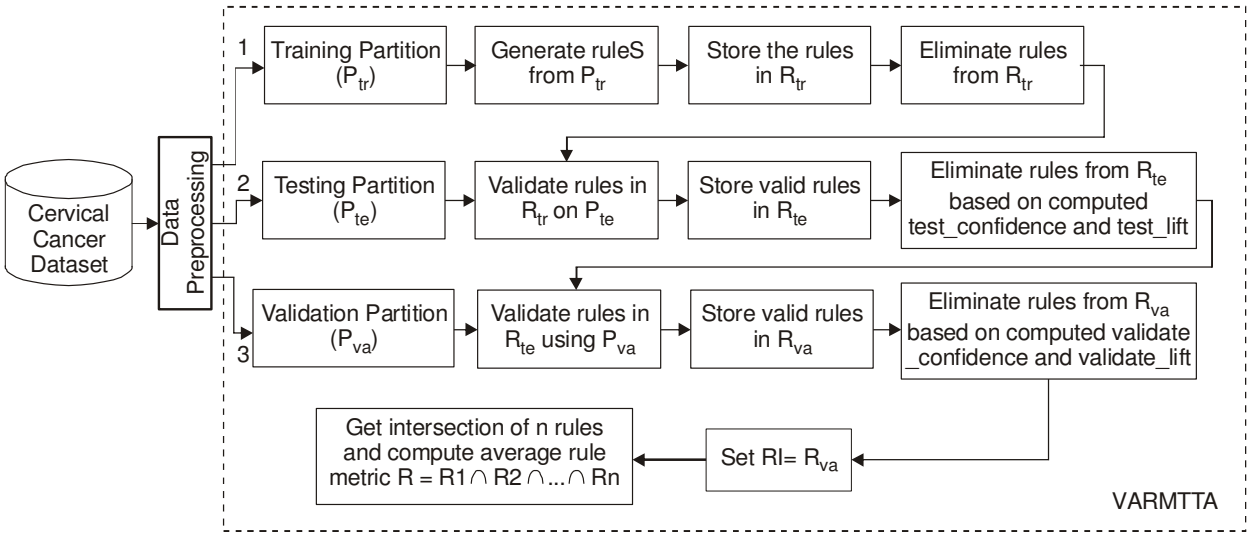

Fig. 1. VARMTTA.

B. Cervical Cancer Data

The Cervical cancer dataset is collected [12]. This dataset was actually collected at 'Hospital Universitario de Caracas' in Caracas, Venezuela. This dataset includes the multivariate attributes such as significant medical records from patient history, numeric information and behavioural habits of patient. 658 patient records are included in dataset and each record is described by 32 attributes. Due to privacy concern, few patients are not interested to disclose some information and it creates missing value in the records. 32 attributes will correspond to the risk factors involved in causing cervical cancer. Meanwhile the attribute such as Hinselmann Colposcopy, Schiller, Cytology and Biopsy are considered as target variable [13]. German physician Hans Hinselmann invented this Colposcopy test and it is a pharmaceutical examination procedure used to view an flashed, deepen sample of the tissue around cervix and vagina. The main aim of colposcopy is to isolate the precancerous lesions in earlier stage and prevent cervical cancer. Schiller test is preamble procedure to identify the cervical cancer. In schiller test, iodine solution is painted on the cervix under direct vision. After painting, normal tissue in cervix will appear as brown where as cancerous tissue will appear as white or yellow due to inability to observe the stain as the these cells are deficient in glycone. In Cytology, tissues are collected from various area of the body which are prone to cause cancer and these tissues are examined under microscopic view. At the end of examination, pathologist will prepare the report. In biopsy representative sample of tissue is removed from a suspicious lesion and kept under microscopic examination. Out of 32 risk factors, risk factor 27 and 28 are removed from the dataset due to imbalanced caused by the missing values and one 30 risk factors are considered for actual process.

C. Simulation Experiments and Analysis

For biomedical data the evaluation metric for an algorithm should be strong enough to measure the accuracy of an approach being used in that algorithm and for correct diagnosis [12]. In some cases, the anticipation provided by the algorithm will have higher total accuracy even though it is not pursued in actual case. For example, consider an algorithm working on cervical cancer dataset containing only few data records of benign condition and an algorithm may conceive all the cancer samples as malignant condition. In such a case, sometimes the algorithm will relish higher total accuracy when comparing with correctly predicting the benign cancer condition. This makes the concept of Negative Predictive Rate (NPR), Positive Predictive Rate (PPR), specificity and sensitivity to come into the picture which makes the diagnosis condition can acquire plenty clarifications.

$$\text{NPR} = \frac{\text{Number of TN}}{\text{Number of TN} + \text{Number of FN}} \quad (13)$$

$$\text{PPR} = \frac{\text{Number of TP}}{\text{Number of TP} + \text{Number of FP}} \quad (14)$$

$$\text{Sensitivity} = \frac{\text{Number of TP}}{\text{Number of TP} + \text{Number of FN}} \quad (15)$$

$$\text{Specificity} = \frac{\text{Number of TN}}{\text{Number of TN} + \text{Number of FP}} \quad (16)$$

$$\text{Total Accuracy} = \frac{\text{Number of TP} + \text{Number of TN}}{\text{No. of (TP} + \text{TN} + \text{FP} + \text{FN)}} \quad (17)$$

Table 1: Cervical cancer dataset attribute description.

S. No	Attribute Name	Attribute Type
1.	Age	int
2.	Number of sexual partners	Int
3.	First sexual intercourse	Int
4.	Num of pregnancies	Int
5.	Smokes	Bool
6.	Smokes (years)	Bool
7.	Smokes (packs/year)	Bool
8.	Hormonal Contraceptives	Bool
9.	Hormonal Contraceptives (years)	Int
10.	IUD	Bool
11.	IUD (years)	Int
12.	STDs	Bool
13.	STDs (number)	Int
14.	STDs: condylomatosis	Bool
15.	STDs: cervicalcondylomatosis	Bool
16.	STDs: vaginalcondylomatosis	Bool
17.	STDs: vulvo-perinealcondylomatosis	Bool
18.	STDs: syphilis	Bool
19.	STDs: pelvic inflammatory disease	Bool
20.	STDs: genital herpes	Bool
21.	STDs: molluscumcontagiosum	Bool
22.	STDs: AIDS	Bool
23.	STDs: HIV	Bool
24.	STDs: Hepatitis B	Bool
25.	STDs: HPV	Bool
26.	STDs: Number of diagnosis	Int
27.	STDs: Time since first diagnosis	Int
28.	STDs: Time since last diagnosis	Int
29.	Dx: Cancer	Bool
30.	Dx: CIN	Bool
31.	Dx: HPV	Bool
32.	Dx	Bool

True Positive (TP) indicates the sample properly diagnosed as having malignant cervical cancer. False Positive (FP) indicates the sample improperly classified as having cervical cancer. True Negative (TN) corresponds to the sample without cervical cancer and correctly predicted as negative. False Negative (FN) corresponds to the undiagnosed sample. Sensitivity measure the proportion of benign cancer samples, which have been exactly identified as benign in all samples. Specificity corresponds to the measure of proportion of malignant cancer samples, which have been exactly identified as malignant in all those samples. Negative Predictive Rate (NPR) is the probability that subjects to precisely classified benign samples and all the benign samples.

Positive Predictive Rate (PPR) is the probability that subjects to precisely classified malignant samples and samples which are malignant.

IV. RESULT COMPARISON AND ANALYSIS

The proposed approach is performed in system with 2.40 GHz Intel (R) Core™ i5-4285U CPU, 8 GB DDR3 RAM, 1 TB HDD and running Windows 7 operating system and implemented using Java 1.8.

A. Accuracy for Target Variable

Under each target variable 638 benign samples and 30 malignant samples were considered. Total Accuracy, Sensitivity, Specificity, PPR and NPR is calculated for each target variable and its result is shown in below Table 2.

Table 2: Total Accuracy, Sensitivity, Specificity, PPR and NPR values.

	Hinselmann Colposcopy	Schiller	Cytology	Biopsy
Total Accuracy	92.89	91.91	93.43	92.34
Sensitivity	99.27	89.88	99.21	94.14
Specificity	90.12	81.97	88.34	90.12
PPR	85.66	79.12	85.00	89.19
NPR	100	96.81	88.12	97.55

B. Most Interesting Rule

The proposed VARMTTA was tested on Cervical Cancer Dataset. The VARMTTA was compared with traditional Apriori algorithm. Table 3 reveals no correlation rule mined from the dataset on cervical cancer. The minimum support is set as 60% and minimum confidence set as 60%. The summary finding in table 3 reveals that as compared with the Apriori algorithm, VARMTTA reduces rules by 24.04 percent.

Table 3: Number of Rules generated in VARMTTA and Apriori.

Dataset	Apriori	VARMTTA	% Reduced
Cervical Cancer	2317	1760	24.04%

The Fig. 2 show the comparative analysis of VARMTTA and Apriori algorithm. In Fig. 1, the x axis comprises of support % range from 10 to 100 and y axis comprises of execution time in seconds. Confidence and lift factors were not included in the graphs instead those values are fixed as 60% and 30% respectively. From the empirical analysis, it is evidence that VARMTTA performs better than Apriori.

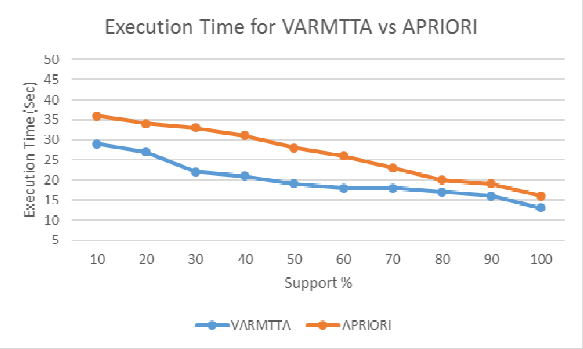

Fig. 2. Execution Time for VARMTTA vs APRIORI.

The Fig. 3 shows the comparative analysis of VARMTTA and Apriori algorithm. In Fig. 2, the x axis comprises of support % range from 10 to 100 and y axis consist of memory consumed in MB. From the empirical analysis, it is evidence that VARMTTA consumes less memory when comparing with Apriori.

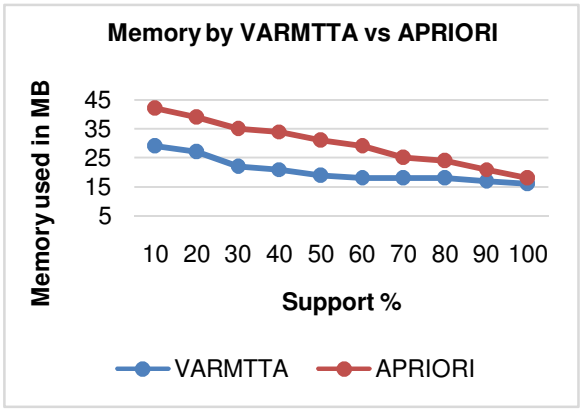

Fig. 3. Memory consumed for VARMTTA vs APRIORI.

V. CONCLUSION

VARMTTA is implemented to mine the rules from cervical cancer dataset. The entire dataset is divided in to three disjoint partitions. All the generated rules are validated against the train-test-validation approach and only the reliable most interesting rules are considered for final evaluation by eliminating irrelevant rules. In order to measure the degree of closeness of evaluated values from the actual value, total accuracy, specificity, sensitivity, PPR and NPR is used. Cervical Cancer Dataset is studied and experiment performed on that dataset shows that the elimination of deceptive rules have greater impact on the execution time and memory of the algorithm. Also, the number of rules generated are reduced to considerable quantity by maintaining the accuracy of prediction. In future evolutionary algorithm can be applied on the disease dataset to mine the association rule.

VI. FUTURE SCOPE

As a potential area, utility-based itemset mining can be implemented on clinical dataset to improve prediction accuracy. Evolutionary algorithms inspired by evolution may also be used for the mining utility itemset.

ACKNOWLEDGEMENTS

The writers are indebted to the experienced referees for bringing this report to our notice. Authors are also thankful to the referees for their insightful recommendations which improved the paper's content and presentation.

Conflict of Interest. K. Logeswaran, Dr. P. Suresh, S. Savitha, K. R. Prasanna Kumar A. P. Ponselvakumar and Dr. A. Rajiv Kannan declare that they have no conflict of interest.

REFERENCES

[1]. Yang, H., & Chen, Y. P. P. (2015). Data mining in lung cancer pathologic staging diagnosis: Correlation between clinical and pathology information. *Expert Systems with Applications*, *42*(15-16), 6168-6176.

[2]. Alwidian, J., Hammo, B. H., & Obeid, N. (2018). WCBA: Weighted classification based on association rules algorithm for breast cancer disease. *Applied Soft Computing*, *62*, 536-549.
[3]. Shanholtz, C. (2001). Acute life-threatening toxicity of cancer treatment. *Critical care clinics*, *17*(3), 483-502.
[4]. Cox, J. D., Stetz, J., & Pajak, T. F. (1995). Toxicity criteria of the radiation therapy oncology group (RTOG) and the European organization for research and treatment of cancer (EORTC). *International Journal of Radiation Oncology• Biology• Physics*, *31*(5), 1341-1346.
[5]. Chen, Y., & Xu, R. (2014). Mining cancer-specific disease comorbidities from a large observational health database. *Cancer informatics*, *13*, 37-44.
[6]. Zolbanin, H. M., Delen, D., & Zadeh, A. H. (2015). Predicting overall survivability in comorbidity of cancers: A data mining approach. *Decision Support Systems*, *74*, 150-161.
[7]. Hanauer, D. A., & Ramakrishnan, N. (2013). Modeling temporal relationships in large scale clinical associations. *Journal of the American Medical Informatics Association*, *20*(2), 332-341.
[8]. Munson, M. E., Wrobel, J. S., Holmes, C. M., & Hanauer, D. A. (2014). Data mining for identifying novelassociations and temporal relationships with Charcot foot. *Journal of Diabetes Research*, 214353.
[9]. Nguyen, D., Luo, W., Phung, D., & Venkatesh, S. (2018). LTARM: A novel temporal association rule mining method to understand toxicities in a routine cancer treatment. *Knowledge-Based Systems*, *161*, 313-328.
[10]. Logeswaran, K., Suresh, P., Savitha, S., & KR, P. K. (2020). Optimization of Evolutionary Algorithm Using Machine Learning Techniques for Pattern Mining in Transactional Database. In *Handbook of Research on Applications and Implementations of Machine Learning Techniques*, 173-200. IGI Global.
[11]. Fisher, R. A. & Yates, F. (1938). Statistical tables for biological, agricultural and medical research (3rd ed.). London: Oliver & Boyd., 26–27.
[12]. https://archive.ics.uci.edu/ml/datasets/Cervical+cancer+%28Risk+Factors%29
[13]. Wu, W., & Zhou, H. (2017). Data-driven diagnosis of cervical cancer with support vector machine-based approaches. *IEEE Access*, *5*, 25189-25195.
[14]. Rameshkumar, K., Sambath, M., & Ravi, S. (2013). Relevant association rule mining from medical dataset using new irrelevant rule elimination technique. In *2013 International Conference on Information Communication and Embedded Systems (ICICES)*, 300-304.
[15]. Pala, T., YÜCEDAĞ, İ., & Biberoğlu, H. (2017). Association rule for classification of breast cancer patients. *Sigma J. Eng. & Nat. Sci.*, *8*(2), 155-160.
[16]. Ordonez, C. (2006). Association rule discovery with the train and test approach for heart disease prediction. *IEEE Transactions on Information Technology in Biomedicine*, *10*(2), 334-343.
[17]. Logeswaran, K., Suresh, P., Savitha, S., & Kannan, A. R. (2018). Contemplate Study of Contemporary Techniques for HUIM. In *2018 International Conference on Intelligent Computing and Communication for Smart World (I2C2SW)* (pp. 282-286). IEEE.

How to cite this article: Logeswaran, K., Suresh, P., Savitha, S., Kumar, P. K.R., Ponselvakumar, A. P. and Kannan, A. R. (2020). Data Driven Diagnosis of Cervical Cancer using Association Rule Mining with Trivial Rule Expulsion Approach. *International Journal on Emerging Technologies*, *11*(2): 110–115.