Modified Apriori Algorithm for the Diagnosis of Tuberculosis

T. Baskar[1], Dr. M. Kannan[2]

[1]Ph.D Research Scholar, Department of Computer Science and Applications, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Kanchipuram, India, Email: basindya@gmail.com
[2]Assistant Professor, Department of Computer Science and Applications, Sri Chandrasekharendra Saraswathi Viswa Mahavidyalaya, Kanchipuram, India, Email: saikannan1999@kanchiuniv.ac.in

Received: 10.07.2024 Revised: 16.08.2024 Accepted: 24.09.2024

ABSTRACT
Data mining is a technique that involves the knowledge discovery and analysis of patterns within extensive databases to extract new information that may be difficult to identify otherwise. It is an interdisciplinary topic of computer science and statistics that seeks to extract information from a dataset and convert it for further use. Data mining has widespread applications in healthcare sector in the analysis of potential outcomes and relationships among the variables in the healthcare dataset, enabling professionals to predict patterns in patients' medical conditions and behaviours.In the recent decades, data mining has been extensively utilized in the prediction and diagnosis of diseases. Datamining algorithms has the potential to offer distinct approach to aid in the diagnosis of several critical illness including tuberculosis (TB). Association Rule Mining (ARM) is the commonly used data mining methodology for uncovering intriguing and unforeseen rules from large data sets. This method generates a substantial number of rules, some of which are intriguing while others are redundant. It restricts the evaluation of rules to just two metrics: support and confidence. Association rule mining (ARM) is an effective method for identifying relationships in datasets, with the Apriori algorithm being one of the most used and impactful algorithms in this domain. This research aimed to create a predictive model using the modified Apriori algorithm for diagnosing pulmonary tuberculosis. A preliminary diagnosis was established exclusively on patient demographic data, medical history, and physical examination findings. Experiments were conductedto assess the performance of individual classifiers such as Linear Discriminant Analysis (LDA), Support Vector Machine (SVM), C4.5 Decision Tree Classifier, K-Nearest Neighbour (KNN) Algorithm, Binary Logistic Regression (BLR), k-means, Apriori and the proposed modified Apriori based on parameters like accuracy, precision, sensitivity, specificity, recall, and F-measure. The data for the experiments were obtained from the medical records of TB patients across several hospitals in the Chennai area, Tamil Nadu, India. The results demonstrated that the modified Apriori approach has outperformed other individual classifiers across all assessment metrics.

Keywords: Tuberculosis diagnosis, Classification, Association Rule Minin, Apriori Algorithm

1. INTRODUCTION

In recent decades, there has been a significant increase in the collection, distribution and analysis of healthcare data, including information from pharmaceutical studies, pharmacological research, cancer treatment investigations, genomic studies, and proteomics research (Asha et al. 2011a). Progress in data mining methodologies has led to the development and execution of efficient and scalable systems for extracting knowledge and valuable information from extensive databases. Medical data mining is a significant study domain within data mining, concentrating on the processing of extensive information housed in medical databases pertinent to intricate clinical issues and related disorders in patients (Yang et al. 2020). The prevalence of relationships and repeating patterns within this data might provide unique insights pertinent to the medical sector, as shown by several data mining applications in this field (Subrahmanya et al. 2022). The data classification process, using insights from historical data, is an extensively studied subject in statistics, decision science, and computer science.

Data mining methods have been used in several medical applications, including forecasting the effectiveness of surgical interventions, medical diagnostics, pharmacotherapy, and uncovering relationships among clinical and diagnostic data (Kolling et al. 2021). The data mining technique is very beneficial for medical practitioners in uncovering concealed medical information. Traditional pattern matching and mapping tactics would be ineffective and imprecise in prognosis or diagnosis without the introduction of data mining tools (Garg&Rupal, 2014). Computerized data mining and decision support techniques are used to aid doctors in the diagnosis of various diseases. These technologies assist physicians in analyzing extensive data from previous instances to provide a probable diagnosis based on essential characteristics (Javaid et al. 2022). Numerous comparisons have been conducted between different classification and prediction methodologies, and this issue remains an area of ongoing study. No one approach has been universally shown to be optimal for all data kinds.

2. Tuberculosis (TB)

Globally, the bacterial illness known as tuberculosis (TB) has caused a significant number of deaths relative to other infectious diseases (Sánchez et al. 2009). Tuberculosis is a lethal infectious disease caused by Mycobacterium tuberculosis in humans. It often disseminates by airborne transmission and affects several regions of the body, including the lungs, bones, and brain. Many impoverished countries have a considerable problem owing to restricted access to diagnosis and treatment. Tuberculosis has the greatest mortality rate among illnesses caused by a single species of bacteria. Tuberculosis is a major worldwide health concern, particularly in India.

Diverse methodologies, such as "clinical symptoms," "tuberculin test," "sputum-smear microscopy," and "chest radiography," have been used for the diagnosis of tuberculosis (Radzi et al. 2011). These methods exhibit several limitations, including time consumption, suboptimal performance, challenges in obtaining sputum samples from paediatric patients, the necessity for live Mycobacterium tuberculosis, the requirement for advanced measurement instruments operated by highly trained medical personnel, and, as a result, high costs (Osman et al. 2010).

The symptoms of tuberculosis include fever, cough, expectoration, haemoptysis, weight loss, and anorexia. The symptoms are common not just to lung cancer but also to other conditions (Bhatt et al. 2012; WHO, 2006). This leads to delayed proper diagnosis, exposure to inappropriate medicine, misdiagnosis, and possibly mortality (Kusiak et al. 2000). Misdiagnosis often arises from inadequate information supplied by the patient or their relatives (Uzoka et al. 2011). A lengthy delay in detecting pulmonary TB impedes timely treatment and leads to the person being uninsulated. Furthermore, those who do not have adequate treatment are at an increased risk of acquiring multidrug-resistant TB (Sánchez et al. 2009).

3. BACKGROUND

Asha et al. (2011b)have used s Classification based on Predictive Association Rules (CPAR) and Predictive Rule Mining (PRM)and First Order Inductive Learner (FOIL) in conjunction with statistical tests and Laplace accuracy as rule assessment metrics for diagnosis of Pulmonary Tuberculosis (PTB) and Retroviral Pulmonary Tuberculosis (RPTB). The evaluation of these techniqueson 700 real records of patients suffering from TB obtained from a state hospital, Tamilnadu showed that CPAR and PRM exhibited comparable accuracy and a superior number of rules in comparison to FOIL.

Jahantigh and Ostovare (2019) investigated hiddedn trends within tuberculosis patient datasets. The Entropy-Shannon approach discerned the paramount attributes, whilst the APRIORI method formulated the data association rule. The R language was used to implement the recommended techniques using data from 548 tuberculosis patients. The Entropy-Shannon technique produced 18 components. The APRIORI algorithm proposed nine correlation criteria linking maximum lift values with minimal support and confidence values. The use of data mining and rule extraction using the APIRIORI algorithm shown its capacity to reduce several rules. Nine permitted association rules for the patient dataset were identified after the removal of these criteria.

Dehghani and Yazdanparast (2023) have implemented Apriori algorithm for association rule mining to identify symptom patterns in COVID-19 patients. The researchanalyzed2,875 patient records, found the predominant signs and symptoms as apnea (72%), cough (64%), fever (59%), weakness (18%), myalgia (14.5%), and sore throat (12%). The suggested strategy offers doctors critical insights into illness management and treatment efficacy.

Hernando and Samonte(2024)have applied Apriori algorithm and classification-based association (CBA) model to analyze demographic data—specifically Age, Sex, Marital Status, Educational Attainment, Wealth Index, and Region—to effectively predict TB-positive patients. Meticulous data preparation, including cleansing and feature engineering, improved the model's predicted accuracy. The CBA model attained a support value of 0.04 and a confidence value of 0.73, demonstrating a True Positive Rate (TPR) of 73.32% in the training set and 77.14% in the test set. False positives were reduced, achieving low False Positive Rates (FPR) of 30.78% and 30.43% in the training and test datasets, respectively. The research utilized post-processing association rules via clustering and objective metrics (PAR-COM) to identify tuberculosis patterns, condensing 244 rules to 30 within 10 clusters.

Research investigations on tuberculosis diagnosis have included auditory, visual, and changeable inputs. Researchers in studies using sound used "coughing sound detection algorithms" and "lung auscultation software" that harness lung sound waves to enhance the tuberculosis diagnostic process with elevated accuracy and specificity (Tracey et al. 2011; Lestari et al. 2012). Numerous studies used pictures of tuberculosis in tissue to assist pathologists. The employed methodologies comprise feed-forward Neural Network (Osman et al. 2009), Zernike Hybrid Moments and multilayered Perceptron Network (Osman et al. 2010), Genetic algorithm - neural network (Osman et al. 2010), compact single hidden layer feed-forward neural network (Osman et al. 2011), and hybridization signal amplification method (Wang et al. 2011). Data mining methods have been used in many studies to detect TB based on clinical signs. The current work intended to develop a classification model for the preliminary diagnosis of pulmonary tuberculosis. The input comprises patient personal information, medical history, and physical examination data (WHO, 2006). The study results may provide a basis for other studies in the same domain.

4. METHODOLOGY

4.1 Apriori Algorithm

Definitions

Here is the traditional definition of association rules.Let { T1, T2, ..., Tn } be a collection of transactions, and let I be a set of items, I = { I1, I2, ..., Im }. An association rule is an implication expressed as A→B, where A and B are disjoint subsets of I such that A∩B=∅.A is designated as the antecedent, whereas B is designated as the consequent of the rule. An itemset refers to any aggregation of items, including antecedents or consequents. Each itemset has a corresponding measure of statistical significance known as support. The support of an itemset A, represented as support(A) = s, is the proportion of transactions in the database that include A. The rule has a strength metric known as confidence, defined as the ratio of support(A∪B) to support(A) (Han &Kamber, 2006).

Apriori signifies a substantial development in the domain of association rule mining. Apriori is effective during the candidate generation stage. The Apriori technique posits that if a collection of objects is common, then all of its subsets must also be frequent (Ingle &Suryavanshi, 2015). If item set X is not substantial, then the item set "X" including item sets X will never be substantial (Ingle &Suryavanshi, 2015). Aprioriis designed to operate on transaction-based databases. The Apriori principle is beneficial as it minimizes the number of items analyzed by considering only item sets with a support count above the minimum support threshold.

Association rule mining aims to identify all rules from a collection of transactions T that meet the criteria of support being greater than or equal to the minimum support threshold and confidence being greater than or equal to the minimum confidence threshold.

The mining of association rules involves a two-step methodology.

- **Generation of Frequent Item sets**
  Generate all itemsets with support more than or equal to the minimum support threshold.
- **Generation of Rules**
  Generate rules with high confidence from each frequent itemset, where each rule is aBinary splitting of a prevalent itemset.

Table 1: Pseudocode of Apriori Algorithm

```
Input:
  - Transaction database D = { T_1, T_2, ..., T_n }
  - Minimum support threshold: min_support
  - Minimum confidence threshold: min_confidence
Output:
Set of association rules that satisfy min_support and min_confidence
Step 1: Generate all frequent itemsets
1. L_1 = { frequent 1-itemsets in D that satisfy min_support }
2. k = 2
3. While L_(k-1) is not empty:
   a. C_k = Candidate itemsets of size k generated from L_(k-1) (Apriori join step)
b. For each transaction T in D:
i. Increment the count of candidates in C_k that are subsets of T
   c. L_k = { itemsets in C_k that satisfy min_support }
d. k = k + 1
Step 2: Generate association rules from the frequent itemsets
4. For each frequent itemsetF inL_k (for all k ≥ 2):
a. For each non-empty subset A of F:
```

i. B = F \ A
ii. Calculate confidence of rule A → B:
confidence(A → B) = support(A ∪ B) / support(A)
iii. If confidence(A → B) ≥ min_confidence, then:
- Output the rule A → B
End.

4.2 Drawbacks in Traditional Apriori Algorithm

The conventional Apriori approach must examine the transaction database to identify the candidates of the item sets. It must scan the full database at each stage, which requires considerable space. The procedure requires extra time. If n represents the average length of item sets and m is the number of transactions in the database, the conventional Apriori approach requires O(m*n) scans to identify the 1-frequent item sets. Furthermore, the time O(Lk-1*Lk-1) is required to identify the candidates Ck. The computation of support requires O(m*Ck) time.

4.3 Modified Apriori Algorithm

The enhanced Apriori method presented in this research first filters the data and removes transactions that do not satisfy the specified criteria. Considering that the data satisfying the requirements have no effect, establish a sparse representation of the data and store it as a transaction database using a "1 - 0" format. Based on this, the development of frequent item sets is executed, while excluding non-useful item sets. Non-frequent item sets may be eliminated from the matrix. It is unnecessary to scan the original database; it just has to perform operations on the matrix using the vector operation "AND" and the random access capabilities of the array to immediately construct the k-frequent item sets. Ultimately, eliminate the items until they fail to meet the minimal support count. Consequently, it may significantly enhance the algorithm's efficiency.The pseudo code for the modified Apriori algorithm can be explained as follows:

Table 2: Pseudocode of Modified Apriori Algorithm

Generate initial itemsets:
L1={itemsets with support greater than the minimum support threshold from the transactionset T}
Construct C_2
$C_2=L_1\times L_1$
Prune candidate itemsets:
L_2={itemsets in C_2 that meet the minimum support}
Iterate for larger itemsets:
For k=3,…,while $L_{k-1}\neq\emptyset$
Prune L_{k-1}based on non-frequent subsets: Prune(L_{k-1})
Generate candidates by self-joining: $C_k=\{L_x\cap L_y:L_x\in L_{k-1},L_y\in L_{k-1}\}$
Check the support for each itemset in the candidate set: If $L_x^{(1)}=L_y^{(1)}\wedge L_x^{(2)}=L_y^{(2)}\wedge\ldots$
Prune and refine: $c=L_x\cap L_y$
If the support of c exceeds the minimum threshold, add to L_k
Check the termination condition:
If all itemsets in C_k have been processed and no further candidates can be generated, terminate.
Intersect transaction IDs to calculate the support:
L_k=New itemset ∩ transaction ID set of T
Repeat until no more frequent itemsets can be generated:
After generating all possible frequent itemsets from transaction data, output the result.

By employing these steps, the improved Apriori algorithm generates and refines itemsets, continuously pruning non-frequent sets based on minimum support criteria. It iterates until all possible frequent itemsetsare found.

5. Data Collection

The authors of the study conducted personal visits to several hospitals in the Chennai Region, Tamil Nadu, and gathered data from patient records of individuals diagnosed with tuberculosis. A total of 350 genuine patient records were used into the final research. All data were aggregated into a single file including many records. Each record has the most relevant information pertaining to symptoms and particular test data for each patient. The research examined 13 symptoms (attributes), with the last attribute identified as the class characteristic (outcome variable) in associative categorization. The 13 attributes encompassed age, duration of chronic cough (weeks), symptoms of weight loss, duration of intermittent fever (days), presence of night sweats, presence of sputum, presence of hemoptysis, presence of chest pain, HIV status, Diabetes Mellitus status, radiographic abnormalities, presence of wheezing, and type of tuberculosis. Table 3 presents the names of 13 characteristics together with their respective data types.

Table 3: Dataset (N=350)

Attribute No.	Name	Datatype
1	Age	Numeric
2	Chronic cough (Weeks)	Numeric
3	Weight loss	Categorical
4	Intermittent fever (Days)	Numeric
5	Night sweats	Categorical
6	Blood cough	Categorical
7	Diabetes Mellitus (DM)	Categorical
8	Chest pain	Categorical
9	HIV	Categorical
10	Radiographic findings	Categorical
11	Sputum	Categorical
12	Wheezing	Categorical
13	TB Type	Categorical

6. Experimental Setup

The experiment used the open-source program Weka at many phases. Weka is a collection of advanced machine learning algorithms intended for diverse data mining activities such as data preprocessing, feature selection, clustering, and classification. Weka has been used in prior research in clinical data mining and bioinformatics.

Weka functionalities use two graphical user interfaces: Explorer and Experimenter. We may traverse acquired findings, evaluate models constructed on diverse datasets, and graphically depict models and datasets, including categorization mistakes. The Experimenter feature facilitates the automation of executing classifiers and filters with diverse parameter configurations on a dataset collection, collecting performance metrics, and conducting significance tests. Proficient users may use the Experimenter to distribute the computational load over many computers using Java Remote Method Invocation.

7. RESULTS AND DISCUSSION

The performance of different classifiers on the TB dataset was evaluated using different parameters like sensitivity, specificity, accuracy, precision, recall and F-measure. Figure 1 presents the comparison of accuracy values between different classifiers (like LDA, SVM, C4.5, k-NN, BLR, k:mean, Apriori and Modified Apriori) for the TB dataset.

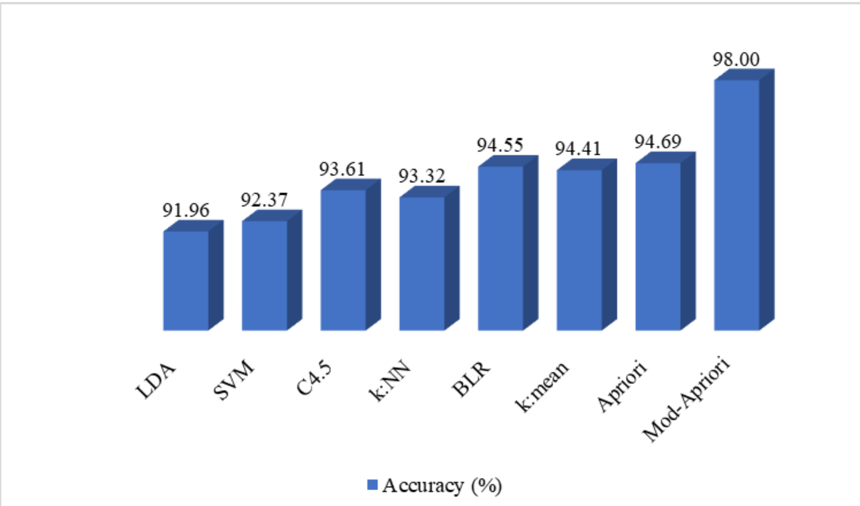

Figure 1: Accuracy of Different Classifiers on TB Dataset

The above figure shows that the Apriori algorithm has produced superior performance in terms of accuracy value (98.00%) when compared with other classification approaches. Figure 2 presents the comparison of specificity values between different classifiers for the TB dataset.

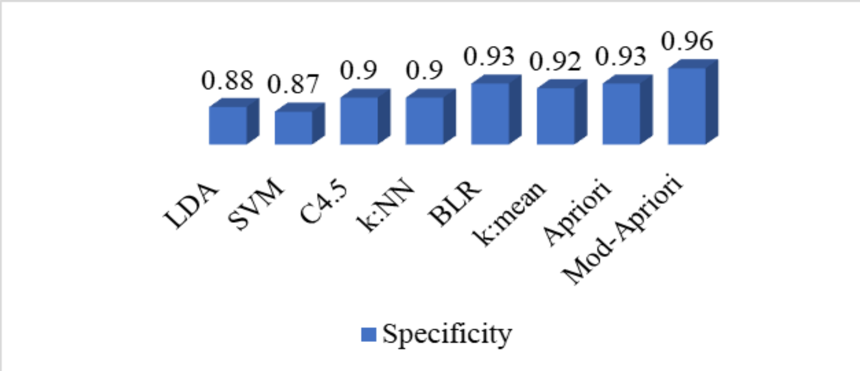

Figure 2: Specificity of Different Classifiers on TB Dataset

The above figure shows that the Modified Apriori algorithm has produced superior performance in terms of specificity value (0.96) when compared with other classification approaches. Figure 3 presents the comparison of sensitivity values between different classifiers and Modified Apriori for the TB dataset.

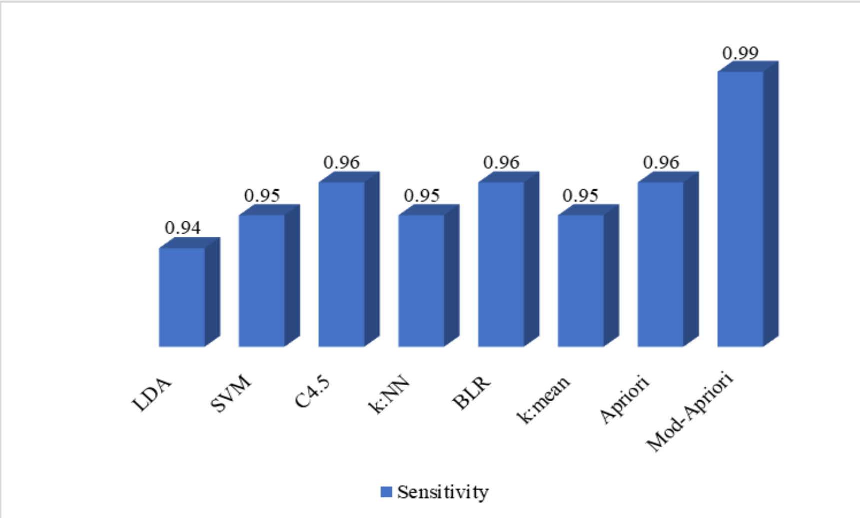

Figure 3: Sensitivity of Different Classifiers on TB Dataset

The above figure shows that the Modified Apriori algorithm has produced superior performance in terms of sensitivity value (0.99) when compared with other classification approaches. Figure 4 presents the comparison of precision values between different classifiers and Modified Apriori for the TB dataset.

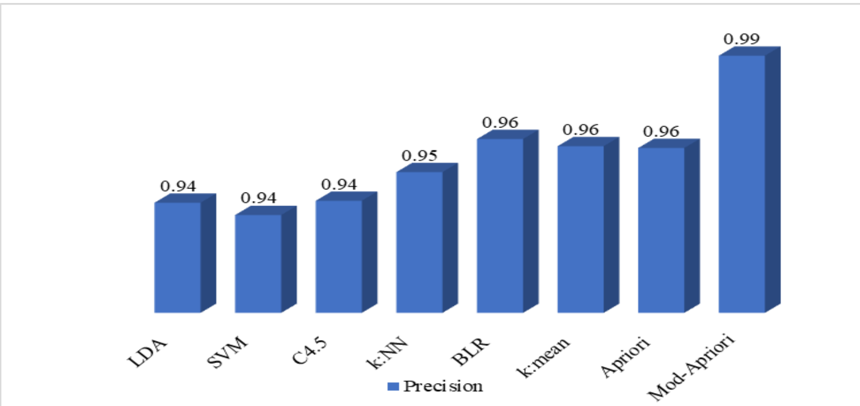

Figure 4: Precision of Different Classifiers on TB Dataset

The above figure shows that the Modified Apriori algorithm has produced superior performance in terms of precision value (0.99) when compared with other classification approaches. Figure 5 presents the comparison of recall values between different classifiers and Modified Apriori for the TB dataset.

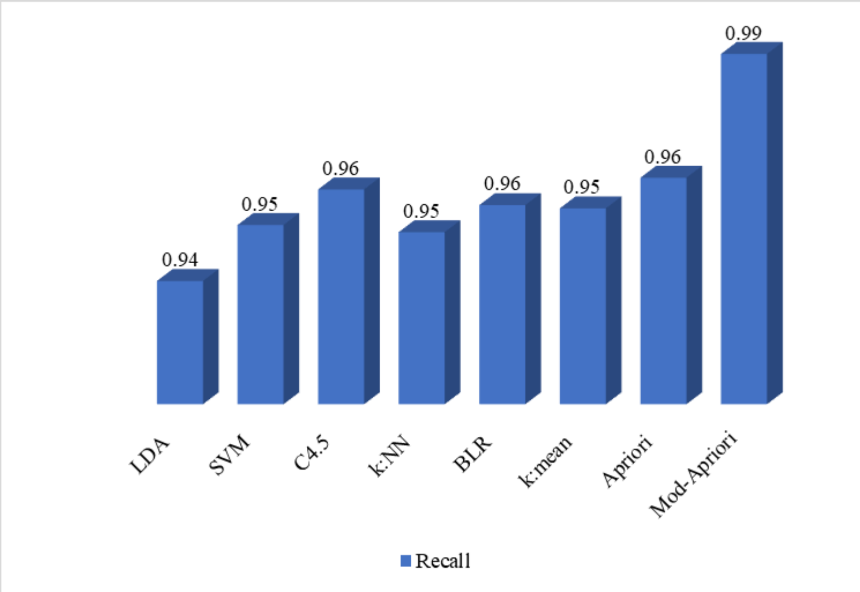

Figure 5: Recall of Different Classifiers on TB Dataset

The above figure shows that the Modified Apriori algorithm has produced superior performance in terms of recall value (0.99) when compared with other classification approaches. Figure 6 presents the comparison of F-measure values between different classifiers and Modified Apriori for the TB dataset.

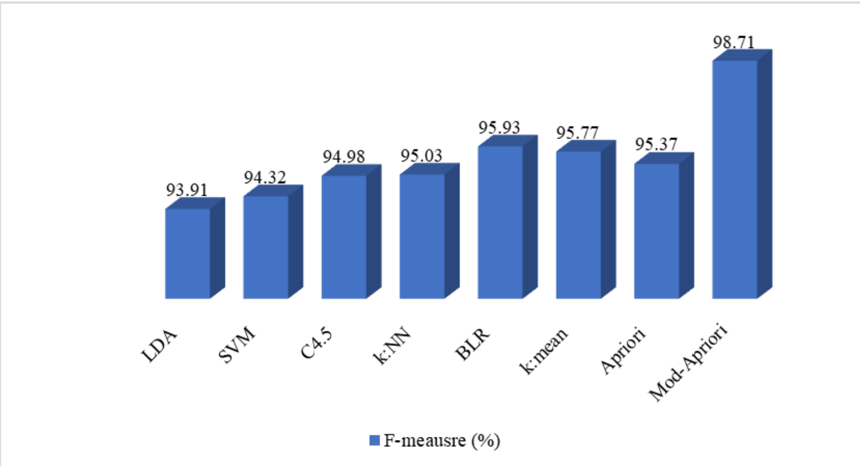

Figure 6: F-measure of Different Classifiers on TB Dataset

The above figure shows that the Modified Apriori algorithm has produced superior performance in terms of F-measure value (98.71%) when compared with other classification approaches. In addition, the computation time of Modified Apriori algorithm was comparable and better than majority of the classifiers.

Table 4: Comparison between Different Algorithms on TB Dataset (N=350)

Algorithm	CT (ms)	Accuracy (%)	Specificity	Sensitivity	Precision	Recall	F-measure (%)
LDA	1282	91.96	0.88	0.94	0.94	0.94	93.91
SVM	1185	92.37	0.87	0.95	0.94	0.95	94.32
C4.5	1050	93.61	0.9	0.96	0.94	0.96	94.98

k:NN	985	93.32	0.9	0.95	0.95	0.95	95.03
BLR	1025	94.55	0.93	0.96	0.96	0.96	95.93
k:mean	1095	94.41	0.92	0.95	0.96	0.95	95.77
Apriori	1137	94.69	0.93	0.96	0.96	0.96	95.37
Mod-Apriori	1176.80	98.00	0.96	0.99	0.99	0.99	98.71

Overall, Modified Apriori algorithm has reported superior performance on the evaluation parameters like sensitivity, specificity, accuracy, precision, recall and F-measure when compared with other classifiers like LDA, SVM, C4.5, k-NN, BLR, k:mean and Apriori algorithm.

8. DISCUSSION

Considering the persistently high global burdenof tuberculosis (TB), the incorporation of machne learning and data mining techniques into diagnostic processes might improve patient outcomes and aid in achieving the ultimate objective of TB eradication, as outlined by the WHO EndTB initiative (WHO, 2023). Machine learning techniques are widely used in the prediction of TB

9. CONCLUSIONS

Tuberculosis is a communicable disease that can affect anyone who contracts it. Tuberculosis has been recognized as the primary cause of death in most underdeveloped countries, and ememrging economies like India. This study compared the efficacy of several data mining algorithms in detecting tuberculosis using data obtained from many hospitals in the Chennai Region of Tamil Nadu, India. The dataset has twelve primary symptoms (attributes) and one categorical attribute. As a contribution in research, this work developed a modified Apriorialgoritm for the prediction of TB. The proposed modified Apriori algoritm was evaluated by comparing the performance with standalone classifiers like LDA, SVM, C4.5, k:NN, BLR, k:mean and Apriori.

The findings indicated that the Modified Apriori algorithm exhibited superior performance, achieving an accuracy of 94.69% and a precision of 0.96. The findings demonstrate that most classifier rules substantially enhance the correct prediction of TB, assisting professionals in their diagnostic choices. The findings unequivocally indicated that, among the single classifiers examined in the research, the apriori algorithm had superior performance across the majority of the assessment metrics. The efficiency of Apriorimight be improved by optimization by using ensemble approaches like as bagging and boosting.

REFERENCES

[1] Asha, T., Natarajan, S., & Murthy, K. N. B. (2011a). Effective classification algorithms to predict the accuracy of tuberculosis-A machine learning approach. International Journal of Computer Science and Information Security, 9(7), 89.

[2] Asha, T., Natarajan, S., & Murthy, K. N. B. (2011b). A study of associative classifiers with different rule evaluation measures for tuberculosis prediction. International Journal of Computer Applications, 3, 15-20.

[3] Bhatt, M. L. B., Kant, S., &Bhaskar, R. (2012). Pulmonary tuberculosis as differential diagnosis of lung cancer. South Asian journal of cancer, 1(01), 36-42.

[4] Garg, S., &Rupal, N. A. (2014). Data Mining Approach to Detect Tuberculosis Using Clustering and GANN Techniques. International Journal of Science and Research (IJSR)IJSR, 4(10), 1841 – 1844.

[5] Hernando, K. B., &Samonte, M. J. C. (2024, January). Using Data Mining to Uncover Association of Philippines' Demographic Data to Tuberculosis. In Proceedings of the 2024 7th International Conference on Software Engineering and Information Management (pp. 49-57).

[6] Wiriyapongsanon, Atit, Warunun Inthakon, and Narawadee Phudolsitthiphat. "Common Attractive Point Theorems for a Finite Family of Multivalued Nonexpansive Mappings in Banach Spaces." Results in Nonlinear Analysis 5.3 (2022): 372-386.

[7] Hernando, K. B., &Samonte, M. J. C. (2024, January). Using Data Mining to Uncover Association of Philippines' Demographic Data to Tuberculosis. In Proceedings of the 2024 7th International Conference on Software Engineering and Information Management (pp. 49-57).

[8] Jahantigh, F. F., &Ostovare, M. (2019). Application of Data Mining for Exploring Hidden Patterns in Tuberculosis Patients. Proceedings of the International Conference on Industrial Engineering and Operations Management Bangkok, Thailand, March 5-7, 2019.

[9] Javaid, M., Haleem, A., Singh, R. P., Suman, R., &Rab, S. (2022). Significance of machine learning in healthcare: Features, pillars and applications. International Journal of Intelligent Networks, 3, 58-73.

[10] Kolling, M. L., Furstenau, L. B., Sott, M. K., Rabaioli, B., Ulmi, P. H., Bragazzi, N. L., & Tedesco, L. P. C. (2021). Data mining in healthcare: Applying strategic intelligence techniques to depict 25 years of research development. International journal of environmental research and public health, 18(6), 3099.
[11] Kusiak, A., Kernstine, K. H., Kern, J. A., McLaughlin, K. A., & Tseng, T. L. (2000, May). Data mining: medical and engineering case studies. In Industrial Engineering Research Conference (pp. 1-7). Cleveland.
[12] Osman, M. K., Ahmad, F., Saad, Z., Mashor, M. Y., &Jaafar, H. (2010, November). A genetic algorithm-neural network approach for Mycobacterium tuberculosis detection in Ziehl-Neelsen stained tissue slide images. In 2010 10th international conference on intelligent systems design and applications (pp. 1229-1234). IEEE.
[13] Radzi, R. U. K. R. M., Mansor, W., & Johari, J. (2011, June). Review of mycobacterium tuberculosis detection. In 2011 IEEE Control and System Graduate Research Colloquium (pp. 189-192). IEEE.
[14] Sánchez, M. A., Uremovich, S., &Acrogliano, P. (2009). Mining tuberculosis data. Data mining and medical knowledge management: Cases and applications, 332-349.'
[15] Kavitha, M. "Advances in Wireless Sensor Networks: From Theory to Practical Applications." Progress in Electronics and Communication Engineering 1.1 (2024): 32-37.
[16] Subrahmanya, S. V. G., Shetty, D. K., Patil, V., Hameed, B. Z., Paul, R., Smriti, K., ... &Somani, B. K. (2022). The role of data science in healthcare advancements: applications, benefits, and future prospects. Irish Journal of Medical Science(1971),191(4), 1473-1483.
[17] Uzoka, F. M. E., Osuji, J., Aladi, F. O., &Obot, O. U. (2011, May). A framework for cell phone based diagnosis and management of priority tropical diseases. In 2011 IST-Africa Conference Proceedings (pp. 1-13). IEEE.
[18] World Health Organization (2023). Global Tuberculosis Report 2023. Geneva: World Health Organization, 2023.
[19] World Health Organization. (2006). Tuberculosis Coalition for Technical Assistance. International Standards for Tuberculosis Care (ISTC). The Hague: Tuberculosis Technical Assistance.'
[20] Yang, J., Li, Y., Liu, Q., Li, L., Feng, A., Wang, T., ...&Lyu, J. (2020). Brief introduction of medical database and data mining technology in big data era. Journal of Evidence-Based Medicine, 13(1), 57-69.
[21] Kumar, P., Dangwal, D., &Puri, N. (2012). Diagnosis of tuberculosis using association rule method. Journal of Information and Operations Management, 3(1), 133.
[22] Dehghani, M., &Yazdanparast, Z. (2023). Discovering the symptom patterns of COVID-19 from recovered and deceased patients using Apriori association rule mining. Informatics in Medicine Unlocked, 42, 101351.