An Efficient Framework for Cryotherapy with Apriori based Probability Tree Classifier (APTC)

[1]A.Kowshika, [2] Dr.M.Gobi,

[1]Research Scholar, [2]Assistant Professor,
Department of Computer Science,
Chikkanna Government Arts College, Tirupur , India

***Abstract*: Machine-learning (ML) methods are applied in many different medical applications such as understanding the disease developments, diagnosing and choosing a treatment method. Machine-learning (ML) methods have the great importance when it applied interdisciplinary. Besides several areas, ML methods save cost factor and time in medical applications. In this study, we experimented several ML methods with different approaches on Cryotherapy dataset, which are applied on wart treatment. Apriori based Probability Tree Classifier (APTC) and naive bayes used to predicts if warts can be healed by the cryotherapy treatment.**

***IndexTerms*: Cryotherapy, warts, apriori, healing, location, patients, treatment.**

I. INTRODUCTION

Today, in the medical industry machine learning has found many applications using machine learning models to predict the potential outcome of a medical treatment. A cryotherapy data set provided by UC Irvine machine learning is used to predict if warts can be healed by the treatment method known as cryotherapy. The data set includes information about 90 patients. The information of each of these patients that is includes patient's sex ,age, time elapsed before treatment, number of warts, the type of warts, area of warts, and result of treatment whether the treatment was successful in healing the warts. The data set includes information of 42 patients for whom the treatment was not successful in healing the warts and 48 patients for whom the treatment was successful in healing the warts.

The data are not concentrated predict the wart healing treatment. It is considered only cryotherapy patients details data for prediction. Thus, proposal of a new classifier named APTC which predicts wart healing pattern based on area and result of treament on patients data. The proposed classifier is used with naïve bayes classifiers to obtain better prediction results. Backdrop consists of wart locations, patient's age, area, result of treatment and each individual's trajectory patient's data collected from reports. The area are framed by analyzing the most frequently appeared wart locations in the entire record. The staying time threshold at each location is calculated as the sum of the average of probability across all locations and standard deviation of that set of location probabilities.

In this study, aptc and naive bayes will be comparatively explored as an approach for predicting whether the cryotherapy treatment will be successful in healing the warts and it will be compared to the most optimal and the least optimal approaches that were used in the previous study. The reason for using this method is that in the data set whether the treatment will or will not be successful in treating wart is denoted as a 1 for saying that the treatment will be successful in treating the warts and a 0 for 48 saying that the treatment will not be successful in treating the warts.

II. BACKGROUND AND MOTIVATION

The aptc and naïve bayes classifier will be explored as an approach for predicting the result of the cryotherapy treatment. The model will be comprised of building and training the model. The data set will be split into training and testing data. The first 69 samples in this data set will be used as training data and the last 21 samples will be used to evaluate the performance of the model. In the training data set, the data was made up of 69 samples that corresponded to 34 patients whose cryotherapy treatment was successful in healing warts and 35 patients whose cryotherapy treatment was not successful in treating warts. This leaves the testing data to be comprised of 13 patients whose cryotherapy was successful and 8 patients whose cryotherapy treatment was not successful.

III. APPROACH USED

Apriori based Probability Tree Classifier (APTC) which predicts if warts can be successfully healed by the cryotherapy treatment. The prediction results gives the new patient with warts will get healed by the cryotherapy treatment or not. The data set includes information about 90 patients. The information of each of these patients that is includes patient's sex, age, time elapsed before treatment, number of warts, the type of warts, area of warts, and result of treatment whether the treatment was successful in healing the warts.

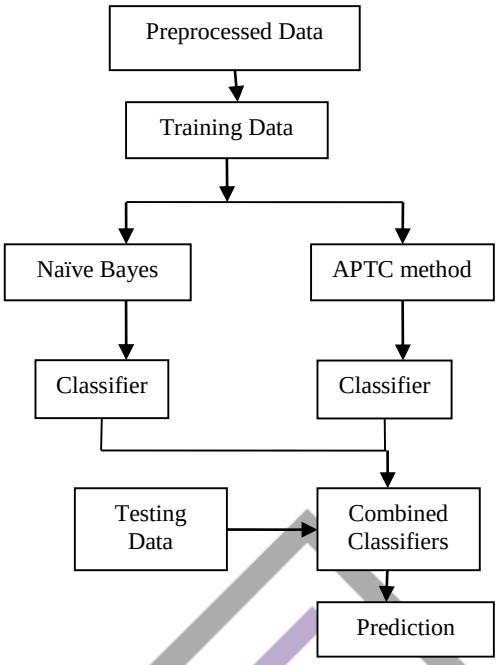

Fig.1 Ensemble model with naïve and aptc

There may probability of certain situation such as,
1.A person is infected with common warts in hands.
2.There is possible to warts occur among regions like fingers.
3.He may also infected with plantar warts around feet.
The Cryotherapy data set from the UCI database contains records of ninety patients. The following table provides more information about the patients whose information was recorded on table1:

Table1: Cryotherapy dataset details

Features	**Values**
Age	15-67
Gender(Male to Female ratio)	46-44
Time elapsed before treatment	0-12
Number of warts	1-12
Area of warts(squared millimeters)	4-750
Type of wart	1-3(common/plantar/both)
Result of Treatment	Healed or Not Healed

IV. NAIVE BAYES METHOD

In this Naive Bayesmethod is a collection of classification algorithms based on **Bayes' Theorem**.It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features equal weight and independent of each other.

Consider a cryotherapy dataset that describes the patient details of wart treatment. Given the patients report, each tuple classifies the whether the warts healed("1") or not healed("0").The dataset is divided into two parts, namely, **feature matrix** and the **response vector**.

1. Feature matrix which contains all the vectors(rows) of the dataset in which each vector consists of the value of **dependent features**. features are 'sex', 'age', 'time','number of warts''type of warts'and 'area'.
2. Response vector which contains the value of **class variable**(prediction or output) for each row of the feature matrix. In above dataset, the class variable name is 'result of treatment'.

V. PROPOSED SYSTEM

In this work, Apriori based Probability Tree Classifier (APTC) which predicts if warts can be successfully healed by the cryotherapy treatment. The prediction results gives the new patient with warts will get healed by the cryotherapy treatment or not.

1. Apriori based Probability Tree Classifier (APTC) The APTC classifier contains different level of phases which refine and model the data to make it suitable for cryotherapy wart treatment prediction. The different phases are data collection, data pre-processing, classification, association rules formation, prediction, similarity computation with the traditional Apriori algorithm and clustering of abnormal patterns.

2. Trajectory data collection from UC Irvine repository: The training and test data are separated after the data is collected. From the data collected, the activity transitions (semantic data) are considered for further processing. Since, the activity transitions are more regular than the location transitions. Here, transition denotes the patient's warts location. In the data pre-processing phase, all the four preprocessing techniques such as cleaning, integration, reduction and transformation are performed.

3. Movement pattern mining: In order to mine and extract the frequent patterns, a probability tree has to be constructed which will have the probability value of each wart location. Once the classifier is trained with the frequent patterns, the test data can be used to compute the result.

4. Association rules formation: It is a classic algorithm used in data mining for learning association rules. It is complex as it sounds on the contrast it is very simple. Suppose there are large set of patients wart treatment records which is denoted in learning association rules basically means finding the wart locations that are occurred together more frequently. Association rules are formed in this manner and the things are altered in such a way that it matches the frequently appeared wart locations.

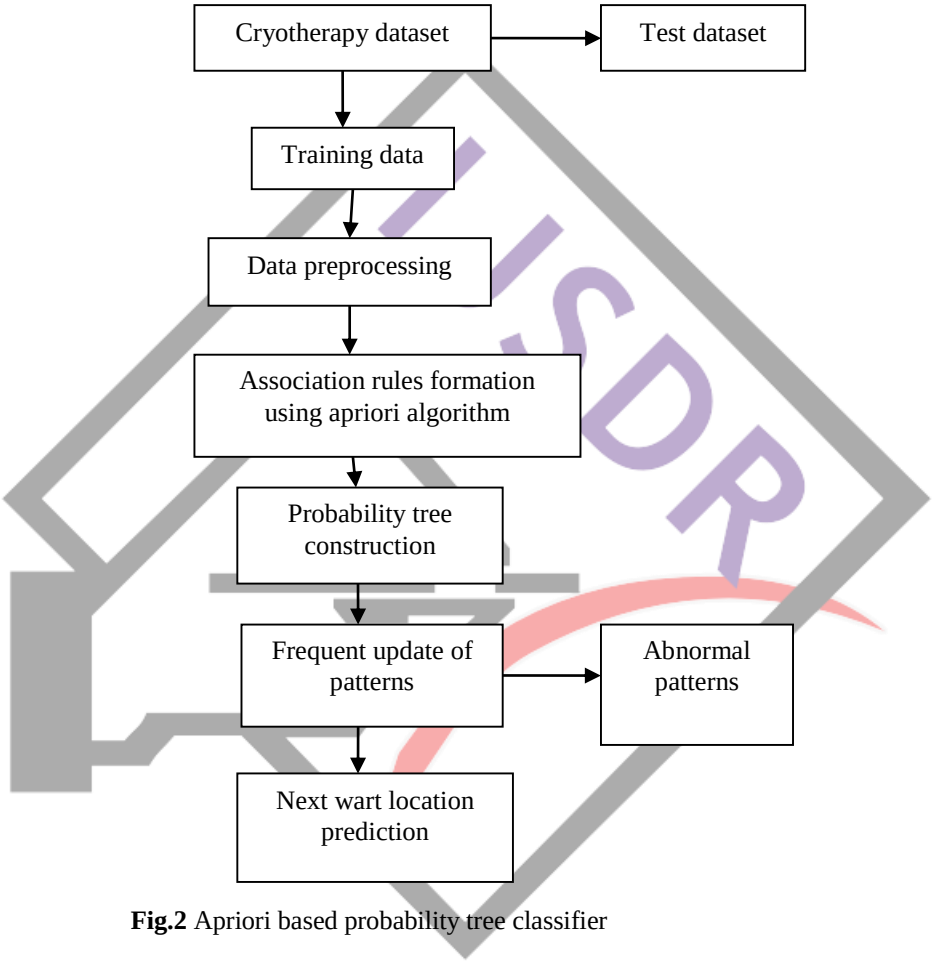

Fig.2 Apriori based probability tree classifier

Algorithm for probability tree construction Input: Set of cryotherapy patients sequence patterns i.e., the dataset containing the warts locations appeared in patients.
Output: A probability tree representing the probability value of each wart location based on the frequent appearance.
Step1: Initialize the probability value for each wart location based on the type of wart. Occurred.
Step2: For each location, calculate the probability that it was the first occurrence location in each patient.
Step3: Update the probability for every node except the root. For each of the node in the tree, if it is the root node, the probability of its child nodes is calculated using the probability values calculated in Step2; otherwise, the probability of a child node is the ratio of its support to the sum of supports of all nodes that share the same parent with it including itself.
In the, the highlighted rectangular box denotes the frequent wart location from the cryotherapy dataset. If the first location is hand, then the probability tree shows that fingers has the highest probability of being next occurrence. In a similar way, keeping track of the probability tree helps to make decision of which wart location to be occurred next. The probability tree is constructed by taking the Cryotherapy patients patterns as input and updating the probability value of each node based on its frequent wart occurrence. Each node is given a percentage based on its first location warts in the patient's body.

Detection of abnormal patterns through clustering:
By incorporating outlier detection mechanisms, the system also reports if there is any unusual warts infected in the patient's body K-means clustering algorithm is used to cluster the abnormal patterns.

APTC ALGORITHM
Input: Human movement data from UC Irvine Machine Learning Repository
Output: Wart locations and their probability values Procedure:
1. Give the cryotherapy data as input to the APTC classifier.
2. Data Pre-processing is performed over the data sets.
3. Association rules are formed using A-priori algorithm.
4. Construction of probability tree based on frequent wart occurrence. For each appearance the probability gets updated. Probability gets updated to all nodes except the root node.
5. Given a current trajectory next sequence of wart locations are predicted.
6. The sequence is verified with the results of Step3 to calculate deviation.
7. Clustering of abnormal patterns based on the probability tree constructed.
Given a current wart position of the patient, future locations are predicted by matching it with the probability tree. By considering the support values, the choice of choosing a single location from multiple locations is performed. Each node has a probability value which denotes the probability that location has been infected. All the operations are explained step by step in the Algorithm 1. Once the warts patterns are mined and stored in the database, k-mean clustering is performed to sort the abnormal patterns from it. Now, with the current wart position of the person, pattern deviation can be detected by comparing with the probability tree and the clustered data.

VI. CONCLUSION

The datasets from UC Irvine Machine Learning Repository as wart treatment cryotherapy dataset information. The classifier named Apriori based Probability Tree Classifier (APTC) which considers cryotherapy patients warts report as the input. This classifier is combined with naïve bayes approach to obtain better results. By using APTC algorithm, the final results yield a better outcome. Apriori based Probability Tree Classifier (APTC) and naïve bayes used to predict if warts can be healed by the cryotherapy treatment or not.

REFERENCES

[1]A.Khosravi, R.Alizadehsani, M.Roshanzamir,F.Khozeimeh,"An expert system forthe selecting wart treatments method," vol. 81, pg.no:167-175, 2017.
[2] I. Harvey, S.Gibbs,J. Sterling et al., "Local treatment for cutaneous warts: systematic review," BMJ, vol. 325, pg.no: 461, 2002.
[3]InternationalJournalofDermatology, F.Khozeimeh.,"Intralesionalimmunotherapy compared to cryotherapy in the treatment of warts," vol. 56, pg.no:4, pp. 474-478, 2017.
[4]Tanuja, V., 2016A Survey on Trajectory Data Mining 4 (10), 195–214.
[5]Van T., Duong, Tran, D.Q., 2015. A fusion of data mining techniques for predicting movement of mobile users. 568–581.
[6]M. Hall, E.Frank"The simple approach to the ordinal classifications," vol. 2167, 2001.
[7]UCI KDD Archive.[Online]. Available: http://archive.ics.uci.edu/ml/
[8]C.T., &JangJ.S.R"The Neuro-fuzzy &soft computing"A computational approach to learning and machine intelligence," 1997.
[9]Ali Cuvitoglu and ZerrinIsik, "Evaluation Machine-Learning Approaches for Classification of 159 Cryotherapy and Immunotherapy Datasets", International Journal of Machine Learning and 160 Computing Vol 8, No. 4 August 2018 161
[10]Machine Learning Mastery [Online]. Available:162https://machinelearningmastery.com/k-fold-cross-validation.
[11]M.M.Clifton, S.M.Johnson "Immunotherapy for recalcitrant warts in children using intralesional mumps or Candida antigens," PediatrDermatol, vol. 20, pp. 268–271, 2003.
[12]N.B.Silverberg"Squaric acid immunotherapy for warts in children," J Am AcadDermatol, vol. 42, pp. 803–808, 2000.
[13]M.Maronn, C.Salm et al., "One-year experience with candida antigen immunotherapy for warts and molluscum," PediatrDermatol, vol. 25, pp. 189–192, 2008.
[14]Laxman, S., Sastry, P.S.P., 2006.A survey of temporal data mining.Sadhana 31 (April), 173–198.
[15]Lichman, M., 2013.UCI machine learning repository.
[16]Lee, J.G., Han, J., Li, X., 2008. Trajectory outlier detection: a partition-and-detect framework. Proc. Int. Conf. Data Eng., 140–149.
[17]Andrienko, N., Andrienko, G., 2008. Basic concepts of movement data. Mobility Data Mining Privacy, 15–38.
[18] S., Azzeh, M., Nassif,Banitaan,2017. User movement prediction: The contribution of machine learning techniques.15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016, pp. 571–575.
[19]C., Passarella, Boldrini,C., 2010. "ModellingSpatial &Temporal properties of human mobility driven by users".
[20]Cho, E., Myers, S.A., Leskovec, J., 2011.User Movement in Location-Based Social Networks. International Conference on (KDD)Knowledge Discovery and Data Mining, pp. 1082– 1090.
[21]Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L., 2008. Understanding individual human mobility patterns, pp. 1–12.
[22] UCI KDD Archive.[Online]. Available: http://archive.ics.uci.edu/ 152.
[23] F. Khozeimeh, F. Jabbari Azad, Y. MahboubiOskouei, M. Jafari, S. Tehranian, R. Alizadehsani, 156 et al., 'Intralesional immunotherapy compared to cryotherapy in the treatment of warts,' International 157 Journal of Dermatology, 2017, DOI: 10.1111/ijd.13535.

[24] Najada, H. Al, Zhu, X., 2014. ISRD: Spam review detection with imbalanced data distributions. IEEE 15th International Conference on Information Reuse and Integration, pp. 553–560.
[25] Nandini, D., Seeja, K.R., 2017. "The novel path planning algorithm for visually impaired people".J. King Saud Univ. Computer. Inf. Sci.
[26] Ray, P.P., 2016. A survey on Internet of Things architectures.J. King Saud Univ. Comput. Inf. Sci.