Journal of Engineering Science and Technology
Vol. 16, No. 2 (2021) 1030 - 1052

ENSEMBLE MODELS FOR PREDICTING WARTS TREATMENT METHODS

AMOS O. BAJEH*, HAMMID O. ADELEKE, HAMMED A. MOJEED, ABDULLATEEF O. BALOGUN, OLUWAKEMI C. ABIKOYE, FATIMA E. USMAN-HAMZA

Department of Computer Science, University of Ilorin, 240003 Ilorin, Nigeria
*Corresponding Author: bajehamos@unilorin.edu.ng

Abstract

Wart is the skin infection medically known to be caused by Human Papillomavirus. There are several methods for treating this skin illness among which are immunotherapy and cryotherapy, the most popular methods. Before applying treatment, physicians need to identify the most effective method for every individual case. Identifying the best treatment for each case can be a daunting task; data mining can be applied to existing datasets in order to discover knowledge for easy identification of suitable treatment method for individual wart case. This study examines the use of computational intelligence in the identification of suitable treatment method for individual warts case. Specifically, the ensemble approaches in machine learning which have been found to have better prediction performance are investigated. The most common types of warts, plantar and common, were studied in the data collected from 180 patients: 90 patients managed through cryotherapy and the other 90 through immunotherapy method. Bagging, Boosting and Random Forest (RF) ensemble methods were applied to predict the response to treatment by patients. It was observed that the Random Forest approach returns the best prediction accuracy of which immunotherapy and cryotherapy methods gave 86.66% and 93.33% respectively. To investigate the generalizability of the models, the models for cryotherapy were used to predict the immunotherapy dataset while models for immunotherapy were also used to predict the cryotherapy dataset. The cryotherapy-based model of RF returned 80% accuracy and 0.48 kappa statistics while that of immunotherapy-based model of boosting returned 85.5% accuracy and 0.54 kappa statistics. Thus, the cryotherapy-based RF model is better than that of bagging and boosting while immunotherapy-based boosting model is better than others. Physicians can safely apply this model to facilitate the selection of effective treatment method for warts.

Keywords: Cryotherapy, Ensemble method, Human papilloma virus, Immunotherapy, Machine learning, Warts.

1. Introduction

Techniques of machine learning and data mining are developed to explore vast repository of data-sets and discover knowledge that are hitherto out of sight from dataset [1]. In medical sciences, machine learning and data mining algorithms are also used as tools for tasks such as DNA sequencing and chemotherapy analysis [2]. These tools are also found useful in risk assessment, crime detection, weather forecast and even sales of products. Unknown patterns in large repositories are determined and analysed with the aid of these techniques.

Classification as a mining technique forms the array of vital tasks in machine learning and data mining. Research and practices in various fields of applications have shown that every classification algorithm comes with its own strengths and limitations [3]. To improve the performance of classification process, ensemble methods or frameworks are proposed to combine the strengths of different machine learning algorithms (models) by fusing several classifiers into a unit model. Some of the types of ensemble frameworks are bagging, boosting, random forest, bucket of models, stacked generalization (stacking) and voting [4]. With the widespread success recorded by ensemble methods, it would therefore be helpful to apply it to solve some trending health related issues such as skin diseases. Wart is one of the skin diseases that are currently receiving attention in dermatology.

Warts are the commonest clinical manifestation of the human papilloma-virus (HPV) infection in the skin and moist membranes that line the body [5]. The treatment methods that are medically proven to be working are Salicylic acid, freezing (Cryotherapy), Duct tape, Immunotherapy, and Zapping and cutting [6]. In spite of the various therapeutic modalities for non-genital skin warts, there is still no single method to be used as a universally approved treatment method [1, 7, 8]. Consequently, physicians need to identify the most effective treatment for each patient. They are striving to discover which treatments have better effect on specific patient. Thus, predicting the best treatment methods among various methods is a source of concern to health workers. Since machine learning techniques are also applicable in medical investigation, there is need to identify best treatment methods among various available options for warts treatment.

Among all the methods for warts treatment, immunotherapy and cryotherapy have proven to be the most effective methods for treating warts. However, studies such as [5] evidently showed that immunotherapy is better than cryotherapy. Many studies with the use of machine learning have also been done to predict the better treatment method. These studies include Khozeimeh et al. [8], Khatri et al. [9], Basarslan and Kayaalp [10], Abdar et al. [11], Putra et al. [12], Rahmat et al. [13], Pawalai and Amornsamankul [14], Rahman et al. [15], Yanik and Comert. [16], Uzun et al. [17], and others can be seen in [13, 18-28]. For instance, in the study by Khozeimeh et al. [8], single classifier of fuzzy-logic using rule-based system was used while Khatri et al. [9] used C45 decision tree (J48) as a single classifier before and after constructing new features from the original with the use of genetic algorithm (GA).

There are various studies in machine learning classification: (1) Studies with single classifiers such as Khozeimeh et al. [8] and Khatri et al. [9], (2) Studies with hybrid classifiers such as Basarslan and Kayaalp [10]; Abdar et al. [11] and Guimarães et al. [29], and (3) Studies with ensemble methods such as Putra et al. [12] and Rahmat et al. [13]. Among all these proposed approaches, none has attempted to cross the datasets against the built models with a view to generalize the models or their external validity. Hence, this research aims to formally predict response to warts treatment empirically by comparing various popular ensemble frameworks for warts treatment using immunotherapy and cryotherapy which are the famous methods, and investigate the external validity of each of the models on the dataset of the alternative treatment method.

The results of this study showed a comparable performance in which Random Forest (RF), and ensemble model of decision tree (DT), produced the best accuracy of 93.3% and 86.6% in predicting cryotherapy treatment and immunotherapy treatment methods respectively. Investigating the generalizability of the ensemble models showed that the immunotherapy-based boosting model is better than another ensemble approach irrespective of which dataset is used on the model. Also, this study showed the fact that although ensemble approaches to machine learning are generally believed to perform better than single method classifiers, some single classifiers such as support vector machine can outperform ensemble approaches.

The remaining part of this paper is organized as follows. Section 2 discusses Wart and the various treatment methods for it. Section 3 presents classification algorithms including the single classifiers and the ensemble methods used in this study. Section 4 presents the studies related to this work. The methodology including the description of the datasets, the study framework and the performance measures used in this study are presented in Section 5. Section 6 presents the results of the study and discusses them as well as compare them to the results obtainable from the literature. Section 7 draws a conclusion and presents possible future studies.

2. Warts

Warts are growths on skin caused by an infection by HPV. Generally, warts are not painful and often vanish on their own over time. Nevertheless they are unpleasant and some, such as those found on the soles of the feet, can make walking and exercising painful [6, 12]. Warts exclusively develop only in the epidermis, the upper layer of the skin. Most warts have swollen and rough surfaces. Nonetheless, some warts such as those on the face may not be swollen. The core of a wart may be flecked with dark dots which are capillaries that supply blood to it. Usually, the most effective treatments for warts are least invasive, but some such as cryotherapy are painful. Figure 1 shows the typical pictures of warts on human skin (Verruca Vulgaris).

Fig. 1 Typical warts [30].

There is an indication that about half of warts disappear on their own within a year, and two-thirds within two years [31]. However, some specialists recommend instantaneous treatment to curb the amount of virus shed into nearby tissue and possibly lower the risk of spread and recurrence. The major categories of treatments [7, 32] are presented as follows.

(i) Salicylic Acid: This is the cogent compositional active ingredient that forms aspirin, and it should usually be the first choice in treating warts [32]. In the management of a wart, it is usually dissolved in water for 10 to 15 minutes (this can be done in the shower or bath) and the acidic solution is directly applied to the warty region after the skin has been scraped with an emery board or pumice stone.

(ii) Cryotherapy: This treatment is also known as freezing and it involves spraying or wiping the warts area with liquid nitrogen [32]. This creates extreme cold (as low as –321 F) around the region and burns the skin, causing pain, redness, and usually a blister. Thereafter the burnt skin falls off, removing the warts.

(iii) Duct tape: In this treatment method, two steps are taking. Firstly, the warty skin is soaked, scraped and duct tape applied and removed immediately; the face is left open overnight. Secondly, duct tape is reapplied in the morning and left in place for six days before it is finally removed [7].

(iv) Agents: This **treatment** method involves the use of prescription drugs for treating persevering warts [7]. The topical Immunotherapy drug imiquimod (Aldara) which is a standard therapy for genital warts is also used to treat skin warts although it has not been tested in randomized trials for that purpose. Imiquimod is thought to work by causing an allergic response and irritation at the site of the wart. In an agent approach called intralesional Immunotherapy, the wart is injected with a skin-test antigen (such as for mumps or Candida) in people who have demonstrated an immune response to the antigen. Drugs (chemotherapy) such as fluorouracil, topical creams, and bleomycin are used to manage stubborn warts.

(v) Electrodessication: This is an application of an electric current to remove an unwanted skin growth [7]. This treatment is technically known as cautery and curettage.

3. Classification Algorithms

The concept of data classification or prediction comes with enormous uses in a wide variety of mining applications. Since many real-world problems could be expressed as relationships or affinity between feature and target variables, this offers a broad range of applicability of the classification algorithm. The concept of classification can be described as determining the class label or numerical score of an unlabelled instance given a set of sample tuples and their associated classes from which classification algorithms can learn. Classification techniques usually contain two discrete stages [33]:

Training Phase (Induction): the period or stage when sampling tuples are used in building classification models. This is also known as the learning phase of the technique in which the models learn patterns and associations from sample datasets

Testing Phase (Deduction): the stage when the built model is tested against new or separate dataset kept for testing and not used during the training phase. At this stage, the classifier or predictor is made to classify test tuples without prior knowledge of the class label of the tuples.

The outcomes of a classification techniques may be presented for a test instance in one of two ways:

(i) Discrete Label: when the class label is expressed in categorical value such as YES or NO.

(ii) Numerical Score: when the class label is expressed in continuous values like float or fixed number. This is also known as regression values in regression analysis.

3.1. Learning scheme

This is the method used in the construction of classification models. That is, how the models are trained. Some of the learning scheme in data mining are:

(i) Decision Tree (DT): A decision tree is a flowchart-like structure in which each internal node represents a test on a feature, the branches represent the possible results of the test, and the leaf node represents the class label associated to an instance. Some typical examples of DT are ID3, C4.5, Classification and Regression Tree (CART). Consider the sample of DT in Fig. 2, which presents the classification of people into boy, girl, man, and woman.

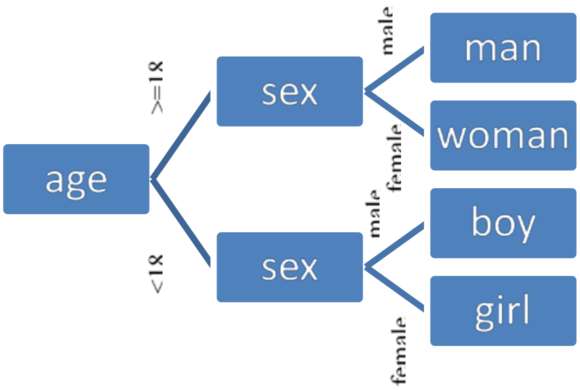

Fig. 2. Typical decision tree (DT).

(ii) Backtracking: This is also referred to as connectionist learning due to the connections between units. During the learning phase, the model learns by adjusting the weights in other not to wrongly predict the class label of the input instance. A typical example of algorithms that use backtracking learning scheme is artificial neural network (ANN). ANN comes in different varieties: single layer perceptron and multilayer perceptron; recurrent NN and convolutional NN, etc. Architecture of multilayer neural network with two hidden layers is shown in the Fig. 3. First layer is known as input layer and last layer is known as output layer; Input layers: x_1, $x_{2,}$ . . . . , x_m. Output layers: y_1, $y_{2,}$ . . . ., *yn*. Layers in between these two layers are called as hidden layers. Information passes from input layer to output layer through hidden layers in multilayer neural network [19].

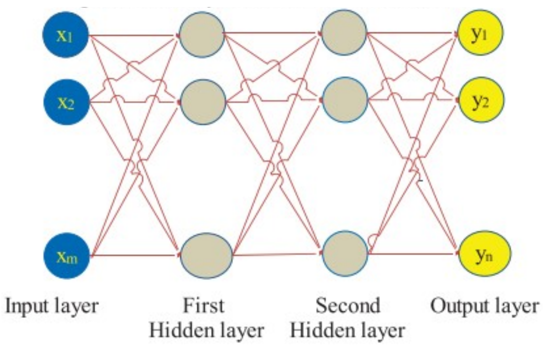

Fig. 3. Typical artificial neural network (ANN) [19].

(iii) Probability: Here, the learning scheme involves the model measuring the probability that a certain given instance belongs to a particular class, by using the feature description of the instance to be classified. Examples of algorithms that use probability learning scheme are Naïve Bayes and Bayes Belief Network.

3.2. Ensemble methods-increasing the accuracy of models

Composite methods called ensemble methods are reported as means for improving classifier and predictor performance [34]. Bagging and boosting are two such means that build a blend of models. Each joins a series of k models (classifiers or predictors), M_1, M_2, ... , M_k, (where $k \in I$) with the aim of creating an improved composite model, M. Both bagging and boosting can be used for classification as well as prediction. Other ensemble methods include but not limited to Stacking, Voting and Random Forest. The following are discussions of these ensemble methods.

(i) Bagging: Bagging (Bootstrap Aggregating) is a method to minimize the variance of classification or prediction by generating additional data for training from the original dataset using combinations with repetitions to produce multisets of the same cardinality/size as the original data. Growing the size of training set does not interpret into improving the model predictive force, but just to reduce the variance, narrowly tuning the prediction to projected outcome. The bagging method is carried out in sync with parallelism, i.e., the models are built at the same time. Each member of the ensemble is bred by a dissimilar data set. This is good for unstable models as proved by Kern [35], where small differences in the input data set yield big differences in output. Also known as high variance models. The algorithm for bagging is presented in Fig. 4.

(ii) Boosting: this is a dual-step approach, where firstly, subsets of the original data are used to produce a series of averagely performing models and then secondly boost their performance by joining them together using a particular cost function (majority vote). Unlike bagging, in the classical boosting the subset creation is not random and depends upon the performance of the previous models. Every new subset contains the elements that were (likely to be) misclassified by previous models. In this method, the models are built sequentially since the knowledge of the previously misclassified tuples are used in the subsequent models. Hence, this method reduces bias but may be susceptible to over-fitting. The algorithm for AdaBoost is presented in Fig. 5.

Input: ***E***, a set of ***d*** sampling data points;
J, the number of models in the amalgamations;
a learning system (e.g., K-Nearest Neighbor, Decision Tree, Neural Network, etc.)

Output: An ensemble model, ***M***

Method:

1. for ***i*** = 1 to ***j*** do:
2. generate bootstrap sample, ***E****i*, by sampling ***E*** with replacement;
3. utilize ***E****i* to produce a model, ***M****i*;
4. endfor

Application of the combining model on a tuple, ***Z***:

1. if classification:
2. let each of the *k* models classify ***Z*** and return the majority vote;
3. if prediction:
4. let each of the *k* models predict a value for ***Z*** and return the average predicted value;

Fig. 4. Bagging algorithm [36].

Input: ***E***, a set of *d* data points with class labels;
j, the number of iterations (one classifier is generated per iteration);
a classification learning system

Output: A combining model.

Method:

1. initialize the weight of each tuple in ***E*** to 1=*d*;
2. for ***i*** = 1 to ***j*** do
3. sample ***E*** with replacement giving credence to the tuple weights to obtain ***D****i*;
4. utilize the training set ***E****i* to generate a model, ***M****i*;
5. find *error*(***M****i*), the error rate of *Mi* (Equation 1)
6. if *error*(***M****i*) is more than ½ then
7. reinitialize the weights to 1=*d*
8. go back to step 3 and try again;
9. endif
10. for each tuple in ***E****i* that was accurately classified do
11. multiply the weight of the tuple by *equation 1*
12. normalize the weight of each tuple;
13. endfor

Application of the combining model to classify tuple, ***Z***:

1. initialize weight of each class to zero;
2. for *i* = 1 to *k* do:
3. ***W****i* = *equation 2;*
4. *f* = ***M****i*(***Z***);
5. add ***W****i* to weight for class *c*
6. endfor
7. return the class with the largest weight;

Fig. 5. AdaBoost algorithm.

The computation of the error rate of model M_i is done as follows.

The aggregate sum of the weights of each of the tuples in E_i that M_i misclassified is mathematically given as:

$$error(Mi) = \sum_{j}^{d} wj \times err(Zj) \tag{1}$$

It is desirable for classifier's error rate to go lower because it implies the classifier is becoming more accurate. Consequently, the higher its weight for voting would be. The weight of classifier M_i's vote is computed as follows:

$$Z = \log \frac{1 - error(Mi)}{error(Mi)} \tag{2}$$

For each class, f, the aggregate sum of weights of each classifier that assigned class f to Z is taken. The champion is the class with uppermost sum and is returned as the class prediction for tuple Z.

It will be observed that the initial base classifier is accomplished using weighting constants that are all the same. The weighting constants are enlarged for sampling tuples that are wrongly classified and reduced for sampling tuples that are perfectly classified in the subsequent iterations.

(i) **Random Forest (RF)**: This is an ensemble learning method for classification, regression and other tasks and it operates by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the trees. Random decision forests correct for decision trees' habit of overfitting to their training set. Random Forest works like bagging but in an enhanced way. In addition to random sampling of data points as obtained in bagging, Random Forest also performs random sampling on features thereby obviating feature engineering. It is an ensemble method that is explicitly planned for decision tree classifiers. It aggregates predictions made by decision trees under the ensemble. Each tree is produced by using bootstrap aggregation of the values of an uncorrelated set of random vectors. The random vectors are created from a static probability distribution [37]. This randomness can be aggregated in many ways as highlighted as follows.

(ii) Arbitrarily choose *F* input features to split at each node (Forest-RI).

(iii) Build linear amalgamations of the input features to split at each node (Forest-RC).

(iv) Arbitrarily choose one of the *F* best splits at each node.

RF works as follows: for each tree in the forest, bootstrap sample is chosen from S where *S(i)* symbolizes the *i*th bootstrap. Then the sample learns a method of decision-tree using a modified decision-tree learning technique. Thus, the technique is amended as follows: at each node of the tree, instead of inspecting all potential feature-splits, some subset of the features $f \subseteq F$ are indiscriminately chosen, where *F* is the set of features. The node then splits on the best feature in *f* rather than *F*. In practice *f* is lesser than *F*. The algorithm for Random Forest is presented in Fig. 6.

Precondition: A training set S: = (x1, y1). . . (xn, yn), features F, and number of trees in forest B.

```
function RandomForest(S, F)
 H ← ∅
 for i ∈ 1, . . . , B do
 S (i) ← A bootstrap sample from S
 hi ← RandomizedTreeLearn(S (i) , F)
 H ← H ∪ {hi}
 end for
 return H
 end function
 function RandomizedTreeLearn(S , F)
 At each node:
 f ← very small subset of F
 Split on best feature in f
 return The learned tree
 end function
```

Fig. 6. Random forest algorithm [38].

For prediction or classification, a new sample is pushed down the tree and assigned the label of the training sample in the terminal node that the sample ends up in. This procedure is iterated over all trees in the ensemble, and the average vote of all trees is reported as random forest prediction.

4. Related Works

Lipke [39] investigated the arrays of treatment plan for wart. It is sacrosanct that the treatment therapy be adapted to meet the needs of the patient and healthcare provider when choosing from the armamentarium of wart managements. Using randomized trial, crucial features to be considered while managing warts are patient’s age, infection site, magnitude, number and types of warts under review, the patient’s immunological status, accessibility of treatment and cost, and the patient’s desire for therapy and ability to stick to the treatment regimen. It was concluded that pending the time the ultimate treatment method would be determined, there should be an aggressive awareness campaign for HPV viral etiology and specific treatment expectations to avoid frustration of patients and healthcare providers.

Zaman [40] reported that freezing is not suitable for areas with a tendon, as tendon is susceptible to damage in the presence of robust treatment. Deformity of the nails may happen if periungual warts are managed with freezing. The researcher made this claim at the end of his randomized trial conducted. However, the study did not compare the effectiveness of freezing with other methods.

Le-Cleach et al. [32] studied comparative effectiveness of freezing and salicylic acid for plantar warts. The result from the laboratory tests authenticated that these methods are equally effective for clearing plantar warts.

Khozeimeh et al. [5] assessed intralesional immunotherapy compared to freezing in the treatment of warts. In their method, sixty patients with verruca vulgaris and plantar warts were randomly split into two clusters. T-test and chi-square tests were used for statistical analysis, and $P< 0.05$ was considered statistically significant. It was concluded from the result that Intralesional immunotherapy is an effective treatment of warts. This method has an improved healing response, needs fewer sessions, and is capable of curing remote warts.

In a related development to the study conducted in [5], Khozeimeh et al. [8] aimed at identifying the suitable management for two common types of warts, plantar and common, and to envisage the effectiveness of two of the best methods, immunotherapy and cryotherapy, for their treatment. A fuzzy logic rule-based system was proposed and implemented to guess the more effective method of warts treatment. It was observed that the prediction accuracy of freezing and immunotherapy methods were 80.7% and 83.33% respectively. This implies that immunotherapy performs better in the given scenario. The study used a single classifier, fuzzy logic rule-based system.

Putra et al. [12] aimed to increase the accuracy of wart treatment method classifiers by improving the performance of weak learner algorithm by using ensemble machine learning. In their study, AdaBoost is used as a strong learner while Random Forest (RF) is used as a weak learner. In addition, stratified 10-fold cross validation is employed to evaluate the proposed technique. The results show accuracy of 96.6% and 91.1% in cryotherapy and immunotherapy respectively.

Yanik and Comert [16]designed a decision tree based algorithm in choosing the most favourable methods for wart treatment. The results from the study show that the ROC Curve for Cryotherapy is 0.9821 while the area under ROC Curve for Immunotherapy is 0.9507, indicating that cryotherapy is better.

Pawalai and Amornsamankul [14] proposed methods to predict cryotherapy method for wart treatment. The study investigated neural network, stacked generalization, cascade generalization, complementary neural network (CMTNN), the combination of stacked generalization and CMTNN, and the combination of cascade generalization. The results from the study show that cascade generalization using complimentary neural network (MTNN) at base level and metal level returns best prediction accuracy of 98.89% using 10-fold cross validation technique.

Uzun et al. [17] proposed the technique of support vector machine to select warts treatment method and the study recorded the prediction accuracy of 85.46%. The study was to predict if the method to be chosen in the treatment of warts will be successful or not.

Nugroho et al. [25] combined the datasets of immunotherapy and cryotherapy to have a single dataset. In their study, C4.5 algorithm was used for classification while random forest feature weighting was used to pick the relevant features for the purpose of increasing the accuracy. Results from the study show that the proposed approach can increase the prediction performance. The accuracy recorded is 87.22%.

Guimarães et al. [20] presented a system of a fuzzy neural network (FNN) which is a hybrid model to predict better method between cryotherapy and immunotherapy in the treatment of warts. The results from the study showed a prediction accuracy of 84.32% for immunotherapy, and 88.64% for cryotherapy.

Rahman et al. [15] investigated Binary Logistic Regression (BLR), K-Nearest Neighbours (KNN), Naïve Bayes (NB), Classification and Regression Tree (CART), Linear Discriminant Analysis (LDA), and Support Vector Machine (SVM) in the prediction of effectiveness of cryotherapy in the treatment of warts disease. The results from the study show that The Support Vector Machine (SVM) with Radial Basis Function (RBF) kernel and K-Nearest Neighbours (KNN) algorithms were found to provide improved performance with an average prediction accuracy of 95.11% and 96.78%, respectively.

Ramana and Boddu. [26] selected some classification algorithms and examined them on some medical datasets. The classification algorithms investigated were Bagging, IBK, J48, JRip, Multilayer perceptron (MP), and Naive Bayes (NB). The results from the study showed that J48 recorded the highest prediction accuracy of 93.33% for cryotherapy while bagging recorded the highest prediction accuracy of 84.44 % for immunotherapy.

With a view to improving the prediction accuracy on immunotherapy warts treatment method, Gajendran and Vasanthi. [19] proposed a mathematical approach on Multilayer Feedforward Neural Network with Backpropagation. The results as obtained from the study showed that the approach returns better prediction accuracy of 96% for immunotherapy method for warts treatment.

Sawhney and Jain. [27] aimed to improve the classification accuracy of better warts treatment. The study leveraged on the capability of wrapper feature selection method to improve classification accuracy by proposing an alteration to a relatively new evolutionary computation method, the Binary Dragonfly algorithm (BDFA), where a penalty function was integrated for selecting optimum feature. The proposed method was applied on two treatment methods, immunotherapy and cryotherapy. The results returned by the study revealed that new penalty function seems to be vastly effective in decreasing the number of features which in turn saves computation cost and from the curse of dimensionality while not sacrificing drastically on classification accuracy.

In Khatri et al. [24] research, the objective was to choose the superlative treatment method between immunotherapy and cryotherapy in warts treatment cases. The study applied classifiers like Bayes Net, SVM, Multi-Layer Perceptron, k-NN, FURIA and Random Forest using WEKA tool. The result obtained revealed that random forest recorded highest prediction accuracy of 86% and 93% for immunotherapy and cryotherapy respectively.

Talabani and Avci [28] examined the effects of four Kernel functions of SVM: Normalized Polynomial Kernel (NP), Polynomial Kernel (PK), Radial Basis Function Kernel (RBF), and Pearson VII function based Universal Kernel (PUK) on cryotherapy and immunotherapy datasets. At the end of the study, it was discovered that each of PUK and RBF returns best performance on cryotherapy with 97.77% prediction accuracy while each of PK and PUK outputs best prediction accuracy for immunotherapy with 81.11%.

By combining the datasets of immunotherapy and cryotherapy, Abdar et al. [11] proposed the hybridization of improved adaptive particle swarm optimization (IAPSO) algorithm and artificial immune recognition system (AIRS) in predicting better treatment methods for warts disease. The experimental results obtained showed a prediction accuracy of 90%.

Guimarães et al. [29] proposed the application of a hybrid model of artificial intelligence and fuzzy logic to enhance the prediction accuracy of the expert system by creating fuzzy rules to construct a more interpretative expert system in the treatment of warts disease. The study was to investigate the success rate of immunotherapy method in the management of warts disease. The results from the study revealed that the method recorded 83.33% prediction accuracy.

Jain and Sawhney [22] proposed a method to address the problem of feature selection in medical data. The proposed scheme involved an enhanced binary version of Gravitational Search Algorithm (GSA) which is based on law of gravity and attraction of masses. This scheme pools the speed of Random Forest Classifier and optimization behaviour of GSA together. The experimental results of the study reveal substantial enhancement in the prediction accuracy.

In Hernández-Julio et al. [21] research, the objective was to develop a data-driven Mamdani-type fuzzy clinical decision support systems using clusters and pivot tables. The results from the study show that the Kappa Statistics and accuracy were close to 1.0 and 100%, respectively for cryotherapy warts treatment method, and Kappa Statistics and accuracy were 0.93% and 97.8%, respectively for immunotherapy warts treatment method.

Kehua et al. [23] presented a deep convolution neural network discriminator for distinguishing Seborrheic keratosis (SK) and flat warts (FW). The SK and FW discriminator (SFD) targeted at identifying and diagnosing the confocal laser scanning microscope images of SK and FW by deep convolution neural network. The experimental results revealed that SFD achieved almost equally compared with different dermatologists; the discriminator can be applied for the identification and diagnoses between SK and FW.

Computational intelligence approach of Adaptive Neuro Fuzzy Inference System (ANFIS) And Support Vector Machine (SVM) was proposed by Abisoye et al. [18] to predict whether immunotherapy treatment method would be successful for warts treatment or not. The experimental results showed that the prediction accuracy of ANFIS and SVM models gave 69.697% and 96.29% respectively. Therefore, it would be inferred from the study that the SVM model was considered to perform better than ANFIS in response to immunotherapy treatment of warts disease.

Rahmat et al. [13] investigated the arrays of decision tree, random forest and k-nearest neighbor in predicting better treatment method for warts disease. The experimental results obtained from the comparison showed that the best prediction accuracy on cryotherapy treatment was achieved by the k-nearest neighbor algorithm with 95.66% while the best accuracy for immunotherapy treatment was achieved by random forest algorithm with an accuracy of 88.89%. The study also performed classification by merging the two datasets into one, and k-nearest neighbor technique was found to outperform others with the accuracy of 88.03%.

Talabani and Avci [28] aimed at the development of a dependable artificial intelligent based model to perfectly foretell the success of immunotherapy and cryotherapy for individual patients. The study applied support vector machine (SVM) classifier. With a view to balancing the minority class, the study employed three different oversampling methods- synthetic minority oversampling technique (SMOTE), borderline-SMOTE, and adaptive synthetic (ADASYN) sampling. In addition, F-score along with sequential backward selection (SBS) algorithm were used to extract the best set of attributes. Results obtained from the study are thus follow. In the case of immunotherapy treatment method, SVM with radial basis function (RBF) kernel obtained an overall classification accuracy of 94.6% (sensitivity = 96.0%, specificity = 89.5%), and for the cryotherapy treatment method, SVM with polynomial kernel obtained an overall classification accuracy of 95.9% (sensitivity = 94.3%, specificity = 97.4%). However, the specificity of the immunotherapy treatment method is still below 90%; therefore, there is still room for improvement of this performance metric.

Rao et al. [30] conducted a review which aimed to provide a terse and pertinent overview to basic machine learning processes, and to unite and survey the published studies on the use of artificial intelligence in forecasting medical endings in dermatology. In the end, it was concluded that to report whether the application of artificial intelligence in predicting results is truly a worthwhile avenue for clinicians to explore, prospective randomized clinical trials are needed.

Patel et al. [31] suggested an expert system to forecast whether the designated wart treatment routine will be efficacious or not. The Multi-Layer Perceptron and the Extreme Learning Machine classification algorithms were applied in the system. Deploying 10-fold cross-validation method, the multi-layer perceptron scheme results in 78.95% of sensitivity, 98.60% of specificity, and 94.45% of accuracy to predict the success of a wart treatment method. To improve upon this study, feature selection algorithms and other well-known classifiers would need exploration.

With a view to improving on the prediction accuracy of machine learning schemes in literature, Le-Cleach et al. [32] proposed a scheme that utilizes the advantage of fuzzy rough set based feature selection (FRFS) to generate the most optimal informative feature space, which in turn makes the artificial intelligence algorithms more accurate and leads to a better prognosis. Results from the study show maximum accuracy of 96.67% and a minimum error rate of 3.33% using fuzzy rough set feature selection based Naïve Bayes method for selecting the cryotherapy treatment method. In the case of immunotherapy, a maximum accuracy of 96.43% and the minimum error rate of 3.57% was recorded. AUC measure of 97.05% using the FRFS based Naïve Bayesian classifier for the selection of the cryotherapy treatment method and 97.72% using the FRFS based CART method for the selection of the immunotherapy treatment method.

Basarslan and Kayaalp [10] applied Naïve Bayes, C4.5 decision tree, logistic regression, k- nearest neighbour classifier models with and without Correlation Based Feature Selection (CFS) as technique for feature selection. The results from the study showed that the attribute selection process increased the performance criteria of all models.

Khatri et al. [9] concentrated on boosting the predictive accuracy of J48, which is a binary decision tree based classifier by augmenting the attributes involved with the aid of genetic programming. Using WEKA tool, the result showed significant performance improvement in classification accuracy of J48 from 82.22% to 96.66% and 93.33% to 98.88% for immunotherapy and cryotherapy datasets, implemented with J48 and J48+GA respectively.

Having considered the pertinent works done so far as related to prediction of warts treatment, it is noted that the there is need to explore approaches that are not yet fully explored. Specifically, the comparison of ensemble methods has not been employed extensively. Thus, this study investigates the application of ensemble methods of Bagging, Boosting and Random Forest in the management of warts by leveraging on the two most effective treatment approaches, Immunotherapy and Cryotherapy. Also, this study investigates the external validity of the models by applying the model derived from one treatment method dataset (i.e., the immunotherapy/cryotherapy dataset model) to predict from the other dataset (i.e. the cryotherapy/immunotherapy dataset) and vice versa.

5. Methodology

This section presents the datasets and ensemble framework used in this study. The performance measures considered are also discussed.

5.1. Data Set

The dataset used was retrieved from UCI Machine Learning Repository. The two datasets used were originally gathered in the dermatology clinic of Ghaem Hospital in Mashhad from January 2013 to February 2015 [5]. The datasets were collected from patients with plantar and common warts who had referred to the dermatology clinic. The description of the features of the datasets are presented in Tables 1 and 2.

Table 1. Attributes of the cryotherapy dataset.

S/N	Features	Values	Mean ± SD
1	Therapeutic Response to treatment	48 Yes / 42 No	
2	Sex	47 Man ; 43 Woman	
3	Age (Year)	15 – 67	28.600 ± 13.361
4	Time elapsed before treatment (month)	0.25 – 12	7.667 ± 3.407
5	The number of warts	1 – 12	5.511 ± 3.567
6	Warts type (Count)	1-Common (54), 2- Plantar (9), 3 - Both (27)	
7	Exterior area of the warts (mm^2)	4 – 750	85.833 ± 131.733

Table 2. Attributes of the immunotherapy dataset.

SN	Features	Values	Mean ± SD
1	Therapeutic Response to treatment	71 Yes / 19 No	
2	Sex	Man:41 Woman:49	
3	Age (Year)	15 – 56	31.044 ± 12.235
4	Time elapsed before treatment (month)	1 – 12	7.231 ± 3.098
5	The number of warts	1 – 19	6.144 ± 4.212
6	Warts type (Count)	1. Common (47), 2. Plantar (22), 3. Both (21)	
7	Exterior area of the warts (mm^2)	6 – 900	95.700 ± 136.615
8	Induration diameter of initial test (mm)	2 – 70	14.333 ± 17.218

5.2. Study framework

This study investigates the performance of machine learning ensemble approaches for the prediction of the best treatment for warts. Three ensemble methods investigated are: Bagging, Boosting and Random Forest (RF); and the models are applied as depicted in the Fig. 7.

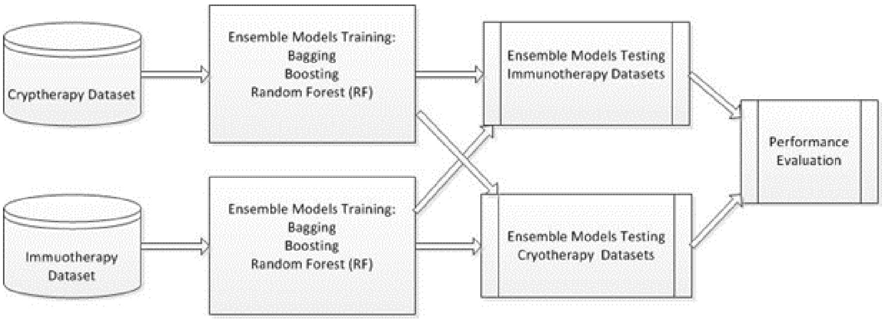

Fig.7. Study framework.

As depicted in Fig. 7, each of the methods' dataset is used to train the ensemble methods considered in this study. Two-third of each dataset is used for the training. The remaining one-third is used for the testing of the models. As also shown on the figure, the model derived from one of the datasets is tested on the other dataset. In this case, the entire dataset of one treatment method is used to train the models while the entire dataset of the other treatment method is used to test the models. The model derived from the cryotherapy dataset is referred to as the cryotherapy-based ensemble models while the models derived from the immunotherapy dataset are referred to as immunotherapy-based ensemble models. This is to examine the external validity of each of the models.

In this study, bagging uses Reduced Error Pruning (REP) Tree, a model that makes use of regression tree logic and generates several trees in different iterations. Thereafter it chooses the best one from all produced trees and appoints it as representative. REP Tree constructs a decision/regression tree with the aid of information gain as the splitting criterion and prunes it with the application of reduced error pruning method [41]. Boosting uses decision stump which is generally a one-level decision tree model. That is, it has just one internal node referred to as the root which is directly linked to its leaves called terminal node. Prediction done by decision stump is based on the value of just a single input feature as base classifier [42].

Waikato Environment for Knowledge Analysis (WEKA) is used for the implementation of the ensemble methods and experimentation.

5.3. Performance evaluation

The evaluation of the models is done using the following performance measures which are derivatives of the confusion matrix (true positive (TP), true negative (TN), false positive (FP) and false negative (FN)).

Given that this study is on binary classification into two classes: treated with cryotherapy/immunotherapy = YES or NO classes, the following performance measures can be defined as follows.

Accuracy: The accuracy value is a measure of proximity of the classification result to the actual value. This measure how often the classifier is correct. It is the ratio of the correctly classified instances to the total instances. That is,

$$Acuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Misclassification Rate: this is also known as error rate and it is a measure how often a classifier is wrong. It is simply 1 – Accuracy or computed as:

$$Error\ (misclassification)\ rate = \frac{FP+FN}{TP+TN+FP+FN} \quad (4)$$

True Positive Rate (TPR): This refers to the rate at which positive tuples are correctly labelled by the classifier. It is the proportion of actual positives (YES class) in the immunotherapy/cryotherapy datasets which are accurately classified by the classifier. It is computed as the ratio of the correctly classified YES instances to the actual total YES instances in the datasets. That is,

$$TPR = \frac{TP}{TP+FN} \quad (5)$$

This measure is the same as recall or sensitivity. The closer TPR value is closer to 1.0 the better the performance of the classifier; it implies that most often the classifier corrects predict the actual YES class.

False Positive Rate (FPR): This refers to the rate at which negative tuples are wrongly classified as positive. It is ratio of the wrongly classified NO instances to the actual total No instances in the datasets. That is,

$$FPR = \frac{FP}{FP+TN} \quad (6)$$

The smaller the value of FPR the better the performance of the classifier.

Precision: This measure how often the classifier is correct when it predicts the YES class. It is the ratio of the actual YES predicted correctly to the total YES class predicted by the classifier. That is,

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

Receiver Operating Characteristic Area (ROC): ROC Area is a curve that depicts the performance of a binary classifier system as its discrimination threshold is varied across possible values. The curve is generated by plotting the true positive rate against the false positive rate at various threshold parameter adjustments. Measurement of performance is done using the area under the ROC curve. A perfect model should have its ROC Area = 1, the closer the ROC value to 0.5 the worse it is.

Cohen's Kappa Statistics (k): Cohen's kappa statistic measures interrater reliability (sometimes called inter-observer agreement). Interrater reliability is perfect when different ratters give the same score to the same data item. This measure indicates the level of agreement or uniformity between the base classifiers of each of the ensemble methods used in this study. Equation 6 states the formula for computing the Cohen's Kappa Statistics. The closer the value to 1 the better the agreement between the ratters while the closer the value to 0 the more probable the agreement [25].

$$k = \frac{P_0 - P_e}{1 - P_e} = 1 - \frac{1 - P_0}{1 - P_e} \quad (8)$$

where:
Po = the relative observed agreement among ratters, and Pe = the hypothetical probability of chance agreement.

6. Results and Discussion

The results obtained from the experiments are presented as follow. Table 3 presents the performance of the ensemble methods on the cryotherapy dataset. Table 4 presents the performance of the ensemble methods on the immunotherapy dataset.

Table 3. Performance of the ensemble methods on cryotherapy dataset.

Performance Measures	Bagging (REPTree)	Boosting (Decision Stump)	Random Forest
Accuracy (%)	86.67	86.67	93.33
Misclassification rate (%)	13.33	13.33	6.66
Kappa Statistics	0.73	0.73	0.86
TP Rate	0.87	0.87	0.93
FP Rate	0.12	0.14	0.06
Precision	0.87	0.87	0.93
ROC Area	0.92	0.98	0.98

Table 4. Performance of the ensemble methods on immunotherapy dataset.

Performance measures	Bagging (REP-Tree)	Boosting (Decision Stump)	Random Forest
Accuracy (%)	83.33	80.00	86.67
Misclassification rate (%)	16.67	20	13.33
Kappa Statistics	0.51	0.38	0.59
TP Rate	0.83	0.80	0.87
FP Rate	0.35	0.46	0.34
Precision	0.81	0.78	0.86
ROC Area	0.90	0.80	0.88

The results of these experiments show that the model of Random Forest gives a better performance in terms of accuracy and error rates than the other ensemble methods considered.

This study goes further to generalize the models by crossing the datasets and the models involved. The cryotherapy-based models were used to predict the immunotherapy treatment dataset, and the immunotherapy-based models were used to predict the cryotherapy treatment datasets. The results obtained are presented in Tables 5 and 6. In this case, the whole 90 instances of one treatment method data set are used in training and the whole 90 instances of the other treatment method data set as well used in the prediction of treatment methods. Input mapped classifier does this on WEKA since the training and test samples of data sets are not the same.

The settings of all the three ensemble methods: bagging, boosting and random forest are maintained hitherto.

In predicting cryotherapy treatment using the immunotherapy-based models, boosting method returns the highest accuracy of 85.56 % and kappa statistics of 0.54. Dataset of Immunotherapy has 8 attributes while cryotherapy has 7 attributes. The only difference is the attribute “Induration Diameter”. This missing attribute in the cryotherapy dataset could explain the reduced performance of the ensemble models when compared to the former performance result in Table 4. However, Boosting has the best results in the performance measures except in the FPR and the ROC area where RF is better than it (boosting) and insignificantly so.

Table 5. Performance of immunotherapy-based ensemble models in predicting cryotherapy treatment method.

Performance measures	Bagging (REP-Tree)	Boosting (Decision Stump)	Random Forest
Accuracy (%)	81.11	85.56	71.11
Misclassification rate (%)	18.89	14.44	28.89
Kappa Statistics	0.37	0.54	0.40
TP Rate	0.81	0.86	0.71
FP Rate	0.48	0.35	0.33
Precision	0.80	0.85	0.79
ROC Area	0.77	0.81	0.88

In predicting immunotherapy treatment using the cryotherapy-based ensemble models, the RF model returns the highest accuracy of 80% and kappa statistics of 0.48.

Table 6. Performance of cryotherapy-based ensemble model in predicting immunotherapy treatment method.

Performance measures	Bagging (REP-Tree)	Boosting (Decision Stump)	Random Forest
Accuracy (%)	73.33	76.67	80.00
Misclassification rate (%)	26.67	23.33	20.00
Kappa statistics	0.39	0.38	0.48
TP Rate	0.73	0.77	0.80
FP Rate	0.23	0.33	0.25
Precision	0.82	0.80	0.84
ROC Area	0.79	0.76	0.82

The results showed that the ensemble approaches have good performances ranging between 80% and 93% accuracies and between 0.8 and 0.98 ROC areas. Although slightly, RF model outperformed the other ensemble models in terms of accuracy with 93% and 86% in predicting cryotherapy and immunotherapy respectively. In terms of ROC area, the three ensemble models performed at per with very insignificant difference. RF and Boosting have a better ROC area of 0.98 in predicting cryotherapy while RF and Bagging have 0.88 and 0.99 respectively in predicting immunotherapy. Thus, in this study RF models can be said to be the best prediction model for both cryotherapy and immunotherapy.

This study's results are comparable to the observations in [24] in which RF had an accuracy of 86% in predicting immunotherapy and 93% in predicting cryotherapy, and Rahmat et al. [13] in which RF had 88.89% accuracy in predicting immunotherapy. Also, in Ramana and Boddu. [26] bagging showed a comparative result of 84.44% accuracy in predicting immunotherapy to the bagging result in this study (83.33%). Although ensemble methods are expected to produce better performances given that they are assembling various single classifiers, this is not always the case; some single classifiers outperform ensemble methods. The studies in [28, 36] reported accuracies of 97.77% (for cryotherapy) and 96.29% (for immunotherapy) respectively using SVM. Tate [6] also showed SVM having 95.11% and 96.78% in predicting cryotherapy. J48 also showed better performances of 96.66% and 98.88% for immunotherapy and cryotherapy respectively in Khatri et al. [24] .

The investigation of the generalizability of the models showed that the ensemble models performed good with accuracies ranging from 70% to 80% while ROC Areas range from 0.76 to 0.88. This reduction in performances of the ensemble models in the generalization study compared to the pure prediction performances discussed earlier can be attributed to the difference in the training datasets. Recollect that the models trained using the immunotherapy dataset are used to test (predict) the cryotherapy method in the cryotherapy dataset, and vice versa. Thus, there is a difference between the training dataset and the testing dataset, and this difference is the induration diameter (in mm) of initial test attribute which is in the immunotherapy dataset but not in the cryotherapy dataset. Tables 7 and 8 present the difference in the performances of the ensemble models in terms of accuracy and ROC areas.

Table 7. Immunotherapy-based ensemble models relative performances in generalizability study.

Performance measures	Models	Crossed-prediction Performance (%)	Pure-prediction Performance (%)	Difference (%)
Accuracy (%)	RF	71.11	86.67	-15.56
	Boosting	81.11	80.00	1.11
	Bagging	85.56	83.33	2.23
ROC Area	RF	0.88	0.88	0
	Boosting	0.81	0.80	0.01
	Bagging	0.77	0.77	-0.13

Table 8. Cryotherapy-based ensemble models relative performances in generalizability study.

Performance measures	Models	Crossed-prediction Performance	Pure prediction performance	Difference (%)
Accuracy (%)	RF	80.00	93.33	-13.33
	Boosting	76.67	86.67	-10.00
	Bagging	73.33	86.67	-13.34
ROC Area	RF	0.82	0.98	-0.16
	Boosting	0.76	0.98	-0.22
	Bagging	0.79	0.92	-0.13

The reduced performance is much more pronounced in the cryotherapy models than the immunotherapy models. All the cryotherapy models experienced significant reduction in accuracy while the immunotherapy models did not experience reduction except in RF only. This trend of cryotherapy models being more reduced in performance is also seen in the ROC area. This implies that the induration diameter (in mm) of initial test is a significant attribute in the prediction of warts treatment method.

7. Conclusions

In this study, three ensemble methods (Bagging, Boosting and Random Forest) in machine learning were investigated in the prediction of suitable Warts treatment methods for patients with plantar and common warts types. A dataset of 180 patient records in the treatment of these two types of warts is collected from the UCI online machine learning repository. Results showed that RF model produced the best accuracy of 93.3% and 86.6% in predicting cryotherapy treatment and immunotherapy treatment options respectively.

The generalizability of the ensemble models was also studied by predicting cryotherapy treatment using the ensemble models from the immunotherapy dataset and predicting immunotherapy using the models from the cryotherapy dataset. The results showed that cryotherapy-based RF model performed better than bagging and boosting counterparts in predicting immunotherapy treatment method while immunotherapy-based boosting model outperforms others in predicting cryotherapy treatment method. The immunotherapy-based boosting model is better than the cryotherapy-based RF model showing that the induration diameter of initial test is a significant attribute in the prediction process.

Further studies will investigate more ensemble methods and single classifier models. Tuning of the methods will also be applied using optimization algorithms on the datasets with a view to attain better performance and identifying the best performing model(s).

References

1. Fayyad, U.; Piatetsky-Shapiro, G.; and Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37-54.
2. Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. *Artificial Intelligence in Medicine*, 23(1), 89-109.
3. Re, M.; and Valentini, G. (2011). Ensemble Methods: A Review. In *Advances in Machine Learning and Data Mining for Astronomy*, 563-594.
4. Alabdulrahman, R. (2014). *A comparative study of ensemble active learning*. University of Ottawa.
5. Khozeimeh, F.; Alizadehsani, R.; Roshanzamir, M.; Khosravi, A.; Layegh, P.; and Nahavandi, S. (2017). An expert system for selecting wart treatment method. *Computers in Biology and Medicine*, 81(17), 167-175.
6. Tate, B. (2014). Treating cutaneous warts: What are the options? *Medicine Today*, 15(10), 34-43.
7. Mulhem, E.; and Pinelis, S. (2011). Treatment of nongenital cutaneous warts. *American Family Physician*, 84(3), 283-293.

8. Khozeimeh, F.; Azad, F.J.; Oskouei, Y.M.; Jafari, M.; Tehranian, S., Alizadehsani, R.; and Layegh, P. (2017). Intralesional immunotherapy compared to cryotherapy in the treatment of warts. *International Journal of Dermatology*, 56(4), 474-478.
9. Khatri, S.; Arora, D.; and Kumar, A. (2018). Enhancing decision tree classification accuracy through genetically programmed attributes for wart treatment method identification. *Procedia Computer Science*, 132, 1685-1694.
10. Basarslan, M.S.; and Kayaalp, F. (2018). A hybrid classification example in the diagnosis of skin disease with cryotherapy and immunotherapy treatment. *2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 1-5.
11. Abdar, M.; Wijayaningrum, V.N.; Malang, P.N.; Hussain, S.; Alizadehsani, R.; Acharya, U.R.; and Makarenkov, V. (2019). IAPSO-AIRS : A novel improved machine learning-based system for wart disease treatment. *Journal of Medical Systems*, 43(220), 0-23.
12. Putra, A.; Setiawan, N.A.; and Wibirama, S. (2018). Wart treatment method selection using AdaBoost with random forests as a weak learner. *Communications in Science and Technology,* 3(2), 52-56.
13. Rahmat, G.A.; Primartha, R.; Sukemi and Wijaya, A. (2019). Comparative analysis of classification method for wart treatment method. *Journal of Physics Conference Series*, 1-9.
14. Pawalai, K.; and Amornsamankul, S. (2019). Using cascade generalization and neural networks to select cryotherapy method for warts. *International Conference on Engineering, Science, and Industrial Applications (ICESI)*, 1-5.
15. Rahman, M.; Wang, S.; Zhou, Y.; and Rogers, J. (2019). Predicting the performance of cryotherapy for wart treatment using machine learning algorithms. *IISE Annual Conference*, 1-7.
16. Yanik, H.; and Comert, M. (2019). Wart treatment selection with a decision tree-based approach. *International Artificial Intelligence and Data Processing Symposium (IDAP)*, 1-6.
17. Uzun, R.; Isler, Y.; and Toksan, M. (2018). Use of support vector machines to predict the success of wart treatment methods. *Innovations in Intelligent Systems and Applications Conference (ASYU)*, 31-34.
18. Abisoye, O.A.; Abisoye, B.O.; Lawal, K.; and Emmanuel, O. (2019). Comparism of adaptive neuro fuzzy inference system and support vector machine for the prediction of immunotherapy warts disease. *3rd International Engineering Conference (IEC 2019)*. Minna, Nigeria, 1-10.
19. Gajendran, G.; and Vasanthi, K. (2019). Multilayer feedforward neural networks with backpropagation model on immunotherapy dataset Multilayer feedforward neural networks with backpropagation model on immunotherapy dataset. *AIP Conference Proceedings*, 1-8.
20. Guimarães, A.J.; Vitor, P.; Rezende, T.S.; Araujo, V.J. .; and Souza, A.V. (2019). Pruning fuzzy neural network applied to the construction of expert systems to aid in the diagnosis of the treatment of cryotherapy and immunotherapy. *Big Data and Cognitive Computing*, 3(22), 1-20.
21. Hernández-Julio, Y.F.; Prieto-guevara, M.J.; Nieto-bernal, W.; Merino-Fuentes, I.; and Guerrero-Avendano, A. (2019). Framework for the

development of data-driven Mamdani-type fuzzy clinical decision support systems. *Diagnostics*, 9(52), 1-36.

22. Jain, R.; and Sawhney, R. (2018). Feature selection for cryotherapy and immunotherapy treatment methods based on gravitational search algorithm. *International Conference on Current Trends towards Converging Technologies (ICCTCT)*, 1-7.
23. Guo, K.; Li, T.; Huang, R.; and Kang, J. (2017). Deep convolution neural network discriminator for distinguishing Seborrheic Keratosis and flat warts Kehua. *15th International Conference on Dependable, Autonomic and Secure Computing*. Orlando, USA, 16-21.
24. Khatri, S.; Arora, D.; and Kumar, A. (2018b). Implicating machine learning towards development of intelligent system for wart treatment therapy identification. *International Journal of Engineering and Technology*, 7(3.12), 754-758.
25. Nugroho, H.W.; Adji, T.B.; and Setiawan, N.A. (2018). Random forest weighting based feature selection for C4.5 algorithm on wart treatment selection method. *International Journal on Advanced Science Engineering and Information Technology*, 8(5), 1858-1863.
26. Ramana, B.V.; and Boddu, R.S.K. (2019). Performance comparison of classification algorithms on medical datasets. *IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*, 140-145.
27. Sawhney, R.; and Jain, R. (2018). Modified binary dragonfly algorithm for feature selection in human papillomavirus-mediated disease treatment. *International Conference on Communication, Computing and Internet of Things (IC3IoT)*. 91-95.
28. Talabani, H.; and Avci, E. (2018). Impact of various kernels on support vector machine classification performance for treating wart disease. *International Conference on Artificial Intelligence and Data Processing (IDAP)*, 1-6.
29. Guimarães, A.J.; Araujo, V.J.S.; De Campos Souza, P.V.; Araujo, V.S.; and Rezende, T.S. (2018). Using fuzzy neural networks to the prediction of improvement in expert systems for treatment of immunotherapy. *In book: Advances in Artificial Intelligence - IBERAMIA 2018* , 229-240.
30. Rao, S.; Ankad, B.S.; Naidu, V.; Sampaghavi, V.V.; Vinod, V.; and Aruna, M.S. (2016). A clinical study on warts. *Journal of Clinical and Diagnostic Research*, 5(8), 1582-1584.
31. Patel, R.V.; Yanofsky, V.R.; and Goldenberg, G. (2012). Genital warts: A comprehensive review. *Journal of Clinical and Aesthetic Dermatology*, 5(6), 25-36.
32. Le-Cleach, L.; Trinquart, L.; Penso-Assathiany, D.; and Chosidow, O. (2012). Comparative effectiveness of cryotherapy and salicylic acid for plantar warts. *Archives of Dermatology*, 148(11), 1311-1313.
33. Aggarwal, C.C. (2015). *Data classification : Algorithms and applications*(1st ed.). Chapman and Hall/CRC.
34. Moniz, N.; and Branco, P. (2017). Evaluation of ensemble methods in imbalanced regression tasks. *Proceedings of Machine Learning Research*, 129-140.

35. Kern, R. (2015). Ensemble methods. knowledge discovery and data mining 2 (VU) (707.004). Retrieved April 5, 2019 from https://docplayer.net/7535139-Ensemble-methods-knowledge-discovery-and-data-mining-2-vu-707-004-roman-kern-kti-tu-graz-2015-03-05.html.
36. Gorunescu, F. (2011). *Data mining: Concepts and techniques* (Vol 12). Springer-Verlag Berlin Heidelberg.
37. Ahmad, A. (2014). Decision tree ensembles based on kernel features. *Applied Intelligence*, 41(3), 855–869.
38. Ali, J.; Khan, R.; Ahmad, N.; and Maqsood, I. (2012). Random forests and decision trees. *International Journal of Computer Science Issues*, 9(5), 272-278.
39. Lipke, M.M. (2006). An armamentarium of wart treatments. *Clinical medicine & research*, 4(4), 273-293.
40. Akram, S.; and Zaman, A. (2015). Warts and verrucas: assessment and treatment. *The Pharmaceutical Journal*, 294(7867):662-664.
41. Kalmegh, S. (2015). Analysis of WEKA data mining algorithm REPTree , simple cart and random tree for classification of Indian news. *International Journal of Innovative Science, Engineering and Technology*, 2(2), 438-446.
42. Sharma, H.; and Kumar, S. (2016). A Survey on decision tree algorithms of classification in data mining. *International Journal of Science and Research (IJSR)*, 5(4), 2094-2097.