Wart Treatment Decision Support Using Support Vector Machine

Md. Mamunur Rahman, Yuan Zhou, Shouyi Wang, Jamie Rogers
Department of IMSE, The University of Texas at Arlington, TX, USA
E-mail: mdmamunur.rahman@mavs.uta.edu, yuan.zhou@uta.edu, shouyiw@uta.edu, jrogers@uta.edu

Abstract—Warts are noncancerous benign tumors caused by the Human Papilloma Virus (HPV). The success rates of cryotherapy and immunotherapy, two common treatment methods for cutaneous warts, are 44% and 72%, respectively. The treatment methods, therefore, fail to cure a significant percentage of the patients. This study aims to develop a reliable machine learning model to accurately predict the success of immunotherapy and cryotherapy for individual patients based on their demographic and clinical characteristics. We employed support vector machine (SVM) classifier utilizing a dataset of 180 patients who were suffering from various types of warts and received treatment either by immunotherapy or cryotherapy. To balance the minority class, we utilized three different oversampling methods- synthetic minority oversampling technique (SMOTE), borderline-SMOTE, and adaptive synthetic (ADASYN) sampling. F-score along with sequential backward selection (SBS) algorithm were utilized to extract the best set of features. For the 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%). The obtained results are competitive and comparable with the congeneric research works available in the literature, especially for the immunotherapy treatment method, we obtained 4.6% higher accuracy compared to the existing works. The developed methodology could potentially assist the dermatologists as a decision support tool by predicting the success of every unique patient before starting the treatment process.

Index Terms—Wart Treatment, Cryotherapy, Immunotherapy, Over Sampling, SMOTE, Borderline-SMOTE, ADASYN, Support Vector Machine, Machine Learning.

I. INTRODUCTION

Warts are one type of benign tumors caused by the infection of Human Papilloma Virus (HPV) and commonly found in the feet, hands, face, and genitals. Though warts are noncancerous and usually painless, patients seek treatment for complete remedy because of primarily cosmetic reasons. Even though several wart treatment methods have been developed and in practice, there is no universal consensus regarding the best one. Moreover, the outcomes of the treatment methods vary from patient to patient. Immunotherapy and cryotherapy are the two most frequently followed methods by the dermatologists for the treatment of various types of nongenital cutaneous warts. In cryotherapy, liquid nitrogen is applied to the affected area of skin with the help of cryogun, an applicator of liquid nitrogen, or cotton swab [1-3]. Consequently, the wart cells are frozen and the blood supply is disrupted locally resulting in the lesions to die off. On the other hand, intralesional immunotherapy stimulates the immune system of the body which recognizes the candida antigen and subsequently removes the HPV resulting in wart lesions to disappear [4-6]. The dermatologists choose a wart treatment method based on their personal experience. Previous studies showed that the cure rate of immunotherapy and cryotherapy are around 72% and 44%, respectively [1, 7]. Therefore, a significant portion of the patients does not get cured by these treatment methods. Choosing the right treatment method from the very beginning is highly important because there are several disadvantages of failed treatments. First, failed treatment causes various undesired side effects on the patients. Second, it jeopardizes the success of the next treatment method because the outcome of wart treatment is time sensitive. Studies showed that the success of many wart treatment methods is negatively associated with time [8]. Finally, failed treatment increases unnecessary expenses and it causes wastage of valuable resources of the hospitals.

The use of Artificial Intelligence (AI) is growing day by day and it has become ubiquitous in every aspect of life. Machine Learning (ML) is one of the most popular tools used in AI, and in recent years, it has gained significant attention in clinical studies. Many studies have been done related to the successful application of ML algorithms on disease diagnosis and treatment [9-12]. In literature, a number of studies are found related to the application of ML algorithms on wart treatment. Khozeimeh *et al.* [13] developed a fuzzy logic rule-based method to predict the success of cryotherapy and immunotherapy treatment methods. Putra *et al.* [14]

employed AdaBoost with classification and regression tree (CART) and random forest (RF) algorithms as weak learners for the prediction of cryotherapy and immunotherapy wart treatments, respectively. Akben [15] applied the ID3 decision tree algorithm for modeling wart treatment predictive models and the developed models were then converted into fuzzy informative images. Guimarães et al. [16] proposed fuzzy neural networks for the improvement of the prediction system for the immunotherapy approach. In another study, Abdar et al. [17] employed evolutionary-based diagnosis system (IAPSO-AIRS) as an improved method for wart treatment selection. The aforementioned wart treatment studies used the same data source [18, 19] for training their ML models, and the maximum reported classification accuracies are 90.0% and 96.4% for the immunotherapy and cryotherapy treatment methods, respectively [15, 20].

However, the accuracy of the prediction system can be improved, especially for the immunotherapy treatment method where the dataset is extremely imbalanced. The objective of this study is to develop highly reliable ML models to predict the success of cryotherapy and immunotherapy for the treatment of warts by utilizing patients' demographic and clinical characteristics. Another objective is to rank the important features which play a crucial role in wart treatment success. A two-stage feature selection process is employed in our methodology to find out the best set of features for developing effective ML models. In the first stage, a filter algorithm is employed for ranking the features, and in the second stage, different ML models are employed as wrapper inside a sequential backward selection algorithm. The findings of this study would potentially help the dermatologists to choose the most appropriate method for individual patients before starting the treatment process.

II. DATA USED

In this study, we used two datasets that were originally collected by Khozeimeh et al. [13]. The datasets are available in the UCI machine learning repository [18, 19] which is maintained by the University of California, Irvine, and the repository is quite popular in the machine learning community. The data were collected along two years, from January 2013 to February 2015, in a dermatology clinic in Iran. Patients who were suffering from plantar and common types of warts and aged more than 15 years were treated by either immunotherapy or cryotherapy. There were 180 patients in total who were divided randomly into two groups of equal size, group A and Group B. Group A patients were treated by immunotherapy with intralesional injection of Candida antigen while group B patients were treated by cryotherapy with liquid nitrogen. The immunotherapy treatment method involved a maximum of three sessions with a gap of three weeks between two consecutive sessions. On the other hand, the cryotherapy treatment method involved a maximum of ten sessions with a gap of one week between two consecutive sessions. The outcomes of the treatment methods along with a set of clinical and demographic attributes of the patients were recorded in the datasets.

The immunotherapy dataset contains 90 observations with 8 attributes, while the cryotherapy dataset includes 90 observations with 7 attributes. The response variable for both datasets, treatment success, is dichotomous with labels 'Yes' or 'No'. The descriptive statistics of the demographic and clinical attributes of the 180 patients are summarized in Table 1. For the numerical variables, mean values and standard deviations are presented, and for the categorical variables, absolute frequencies are presented.

Table 1. Descriptive statistics of the demographic and clinical attributes of the immunotherapy and cryotherapy datasets

Sr. #	Features	Type	Immunotherapy		Cryotherapy	
			Value (count)	Mean (SD)	Value (count)	Mean (SD)
1	Gender	Categorical	Male (41)	-	Male (47)	-
			Female (49)	-	Female (43)	-
2	Age (years)	Numerical	15-56	31.04 (12.23)	15-67	28.6 (13.36)
3	Time elapsed before starting treatment (months)	Numerical	0-12	7.23 (3.10)	0-12	7.66 (3.4)
4	Number of warts (count)	Numerical	1-19	6.14 (4.2)	1-12	5.51 (3.57)
5	Type of warts	Categorical	Common (47)	-	Common (54)	-
			Plantar (22)	-	Plantar (9)	-
			Both (21)	-	Both (27)	-
6	Surface area of the largest wart (mm^2)	Numerical	6-900	95.7 (136.61)	4-750	85.83 (131.73)
7	Induration diameter (mm)	Numerical	2-70	14.33 (17.22)	-	-
8	Success of the treatment	Categorical	Yes (71)	-	Yes (48)	-
			No (19)	-	No (42)	-

Note: SD- standard deviation; mm- millimeter

Fig. 1(a) shows the plot of the features for the immunotherapy dataset. Immunotherapy was successful for 71 patients while the treatment method was not successful for 19 patients. Consequently, the distribution of the response classes is skewed- 78.9% 'Yes' versus 21.1% 'No'. According to the plots, immunotherapy showed a better success rate for younger patients, particularly whose age are within 30 years. Besides, this treatment method worked well for the patients who started the treatment process within 9 months of getting

the disease. However, this treatment method appeared to be less successful for the patients suffering from only common type of warts compared to the patients suffering

from plantar or both types of warts. For the rest of the features, no explicit trend is observed from the plots.
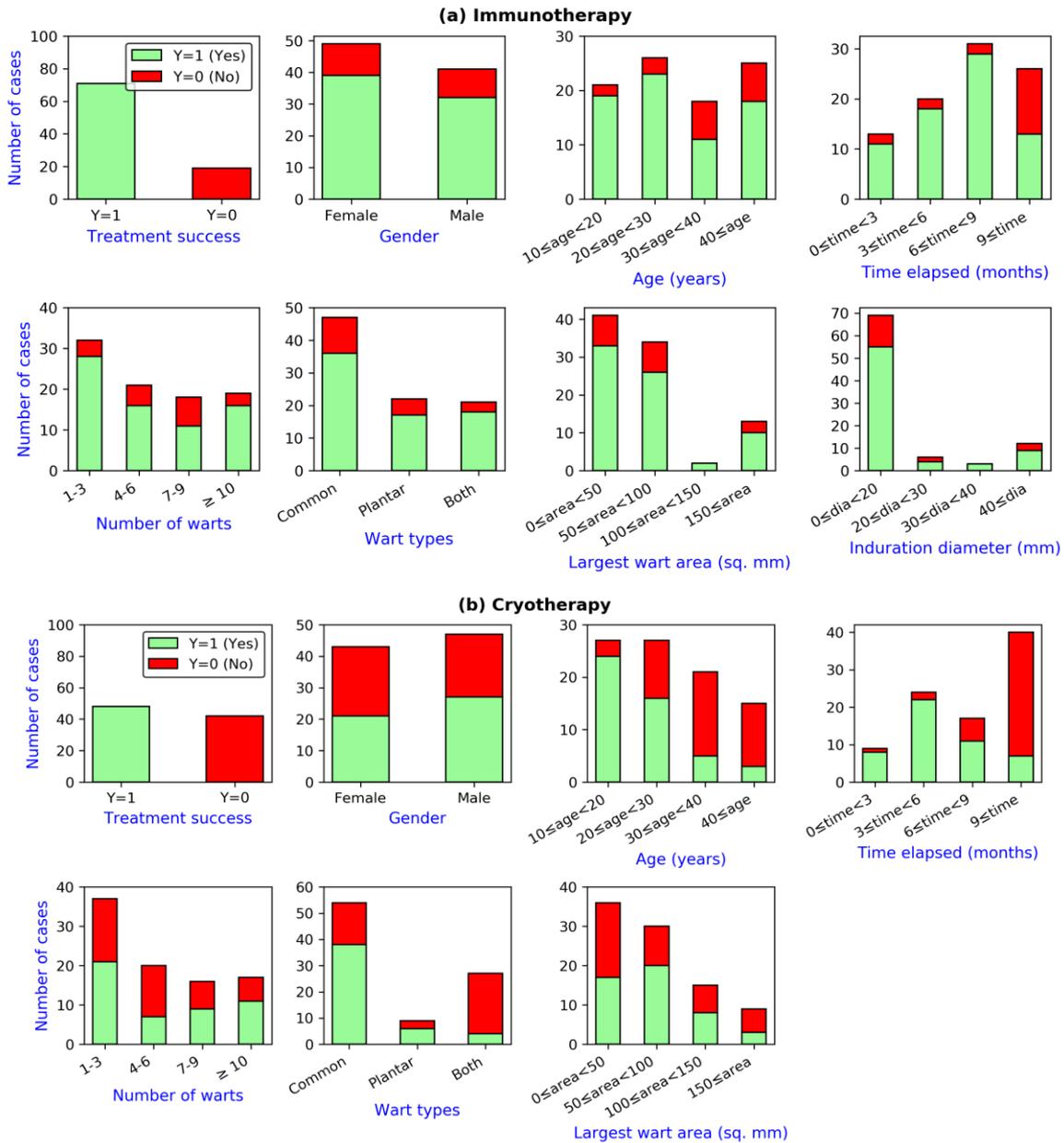

Fig.1. Plot of the features for the datasets: **(a)** Immunotherapy **(b)** Cryotherapy

Fig. 1(b) shows the plot of the features for the cryotherapy dataset. The treatment method was successful for 48 patients while the method was not successful for 42 patients. Therefore, the distribution of the response classes is fairly balanced- 53.3% 'Yes' versus 46.7% 'No'. Similar to the immunotherapy treatment method, Cryotherapy treatment showed comparatively poor performance to cure warts of the elderly patients compared to the younger patients, especially patients aged more than 30 years are significantly less likely to cure by this treatment method. Time elapsed before starting the treatment method also shows a significant impact on treatment success. For

example, patients who started the cryotherapy treatment after 9 months of getting the disease show a substantially low cure rate. In addition, the treatment method was comparatively successful for patients with the common or plantar type of warts than patients with both types of warts. For the rest of the features, no explicit trend is observed from the plots.

III. RESEARCH METHODOLOGY

Fig. 2 illustrates the overall research methodology we followed in this study. The following sections describe the major steps of this methodology.
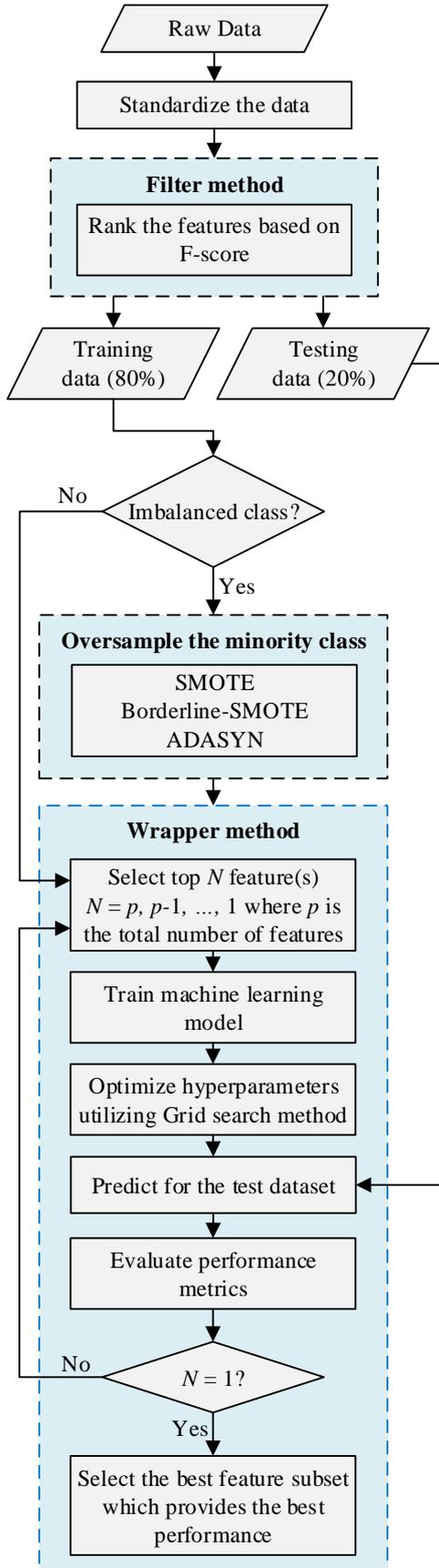
```
        ┌──────────────────────┐
        ╱      Raw Data       ╱
        └──────────────────────┘
                   ↓
        ┌──────────────────────┐
        │  Standardize the data │
        └──────────────────────┘
                   ↓
        ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐
            Filter method
        │ ┌──────────────────┐ │
          │ Rank the features │
        │ │  based on F-score │ │
          └──────────────────┘
        └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
                   ↓
     ╱Training   ╱  ╱Testing   ╱
     ╱data (80%)╱   ╱data (20%)╱
          ↓
 No    ◇ Imbalanced class? ◇
          ↓ Yes
        ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐
        Oversample the minority class
        │ ┌──────────────────┐ │
            SMOTE
        │ │ Borderline-SMOTE │ │
            ADASYN
        │ └──────────────────┘ │
        └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
        ┌ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┐
           Wrapper method
        │ ┌──────────────────┐ │
          │ Select top N      │
        │ │ feature(s)        │ │
          │ N = p, p-1, ..., 1 │
        │ │ where p is the    │ │
          │ total number of   │
        │ │ features          │ │
          └──────────────────┘
        │ ┌──────────────────┐ │
          │ Train machine     │
        │ │ learning model    │ │
          └──────────────────┘
        │ ┌──────────────────┐ │
          │ Optimize          │
        │ │ hyperparameters   │ │
          │ utilizing Grid    │
        │ │ search method     │ │
          └──────────────────┘
        │ ┌──────────────────┐ │
          │ Predict for the   │
        │ │ test dataset      │ │
          └──────────────────┘
        │ ┌──────────────────┐ │
          │ Evaluate          │
        │ │ performance       │ │
          │ metrics           │
        │ └──────────────────┘ │
     No   ◇   N = 1?   ◇
        │         ↓ Yes         │
          ┌──────────────────┐
        │ │ Select the best   │ │
          │ feature subset    │
        │ │ which provides    │ │
          │ the best          │
        │ │ performance       │ │
          └──────────────────┘
        └ ─ ─ ─ ─ ─ ─ ─ ─ ─ ─ ┘
```

Fig.2. The overall methodology followed in this study

A. Data preprocessing

In this step, the input data are shuffled first and then standardized. Data standardization is important because many ML algorithms perform efficiently and provide better results on standardized data [21]. The input dataset is standardized by subtracting the mean value of the corresponding features and dividing by the standard deviations as shown in (1).

$$x_i' = \frac{x - \bar{x}}{s} \qquad (1)$$

where $x = (x_1, x_2, ..., x_n)$, x_i' is the i-th scaled data, $\bar{x}$ and s are the mean value and standard deviation of the samples x, respectively.

B. Feature selection

Feature selection is one of the most important steps before training a model. For most of the ML models, the classification success largely depends on the selection of the right set of input features and the models might demonstrate very poor performance due to the inclusion of too many, too few, or insubstantial features. The feature selection methods can be broadly classified into two groups, i.e., filter type and wrapper type. In the filter type method, the feature selection algorithms utilize various feature evaluation functions to evaluate the importance of the features [22, 23].

On the other hand, in the wrapper method, the classifier itself is used as a fitness function to extract the important features [24-25]. In this study, we utilized a combination of filter and wrapper method to extract the best set of features for the ML algorithms. For the filter type algorithm, we utilized the F-score method which has been used successfully in many ML-based clinical studies [26-28]. The larger the F-value, the higher the importance of the feature. For a given feature vector $x_k \in \mathbb{R}$, $k = 1, 2, ..., n$, the F-score of the i-th feature can be calculated using (2) [29].

$$F_i = \frac{(\bar{x}_i^{(+)} - \bar{x}_i)^2 + (\bar{x}_i^{(-)} - \bar{x}_i)^2}{\frac{1}{n_+ - 1}\sum_{k=1}^{n_+}(x_{k,i}^{(+)} - \bar{x}_i^{(+)})^2 + \frac{1}{n_- - 1}\sum_{k=1}^{n_-}(x_{k,i}^{(-)} - \bar{x}_i^{(-)})^2} \qquad (2)$$

where, n, n_+ and n_- are the number of total instances, positive instances, and negative instances, respectively; $\bar{x}_i^{(+)}, \bar{x}_i^{(-)}$ and $\bar{x}_i$ are the average value of the positive instances, average value of the negative instances, and overall average value of the i-th feature, respectively; $x_{k,i}^{(+)}$ and $x_{k,i}^{(-)}$ are the value of the k-th positive and negative instance of the i-th feature, respectively. Fig. 3 depicts the ranking of the features and their corresponding F-scores. As per the figure, the ranking of the first four features are same for the immunotherapy and cryotherapy datasets; the ranking is different for the remaining features.
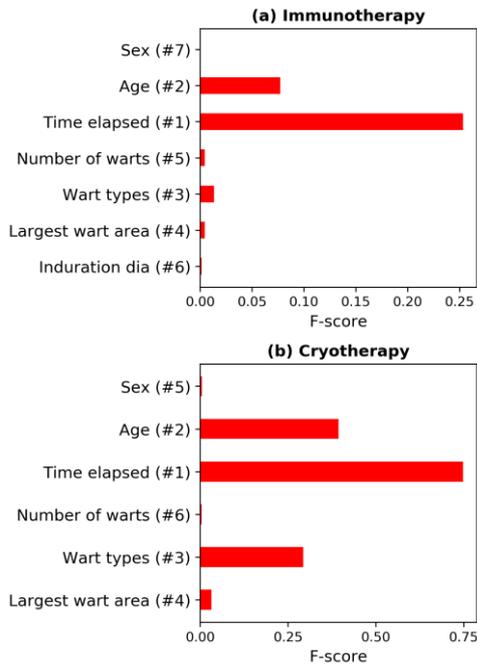
Fig.3. Feature ranking based on F-scores

C. Oversampling

In our study, the output classes of the cryotherapy dataset are fairly balanced, 53.3% 'Yes' versus 46.7% 'No'; on the other hand, the immunotherapy dataset is highly imbalanced, 78.9% 'Yes' versus 21.1% 'No'. There is a significant drawback of this imbalanced dataset. Most of the standard ML classifiers will lead to learning bias and consequently, the classification performance for the minority class ('No') will be poor [30, 31]. To overcome this problem, we balanced the immunotherapy dataset by creating synthetic samples of the minority class ('No') using three different oversampling

algorithms such as synthetic minority oversampling technique (SMOTE) [32], Borderline-SMOTE [33], and adaptive synthetic (ADASYN) sampling [34].

SMOTE is the most popular oversampling technique in literature. The minority class samples are oversampled in SMOTE algorithm along with the k-nearest neighbors of the minority class. During oversampling, all the sample points of the minority class are given equal weight. However, the samples near the borderline are more likely to be misclassified compared to the samples lying far from the borderline. To address this issue, in Borderline-SMOTE, a variant of SMOTE algorithm, oversampling is performed only near the borderline which separates the majority and minority classes. On the other hand, in ADASYN algorithm, more focus is provided on the samples which are difficult to train compared to the samples which are easy to train. Consequently, relatively more synthetic samples are generated around the difficult to train minority samples. The new balanced dataset, therefore, not only minimizes the learning bias of ML classifier but also pushes the decision boundary to the hard to train sample points.

It is noted that the oversampling techniques are applied only on the training dataset, while the testing dataset is kept completely separate so that it cannot artificially increase the testing accuracy. We split the data randomly into 80% and 20% as the training set and testing set, respectively. We used the python library 'imbalanced-learn 0.5.0' [35] for the implementation of the mentioned oversampling algorithms.

Fig. 4 shows the 3-D scatter plot of the top three features (Age, Elapsed time before starting the treatment, and Wart type) of the immunotherapy dataset before and after oversampling of the minority class by SMOTE, Borderline-SMOTE, and ADASYN algorithm.
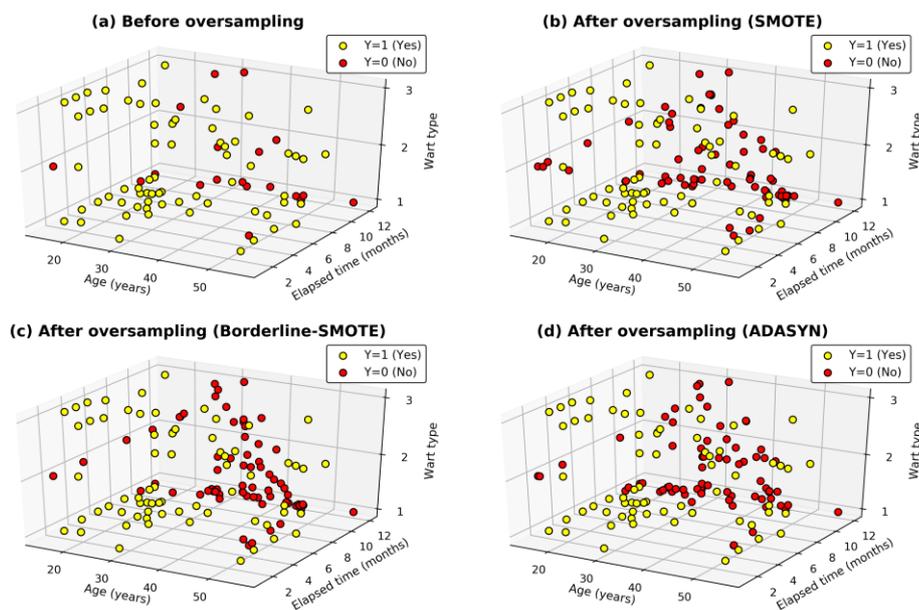

Fig.4. 3-D scatter plot of the top three features for the immunotherapy dataset before and after oversampling by three different algorithms

D. Sequential backward selection of the features

We applied the Sequential Backward Selection (SBS) algorithm [36], a wrapper type feature selection method to extract the best set of features for a given ML classifier. The SBS method works in a top-to-bottom manner. In this study, we implemented the following modified SBS algorithm.

Step 1:   Train an initial model with all the p features and evaluate the performance of the model.

Step 2:   Remove the feature with the lowest score obtained from the F-score method, train model with (p-1) features and evaluate performance.

Step 3:   Remove the feature with the 2nd lowest score, train model with (p-2) features and evaluate performance. Continue this procedure until only one feature remains in the model.

Step 4:   Select the subset of the features that achieved the highest model performance.

E. Hyperparameter optimization

Hyperparameter optimization is an important step for ML models since the classification performance greatly depends on choosing the appropriate values of the related hyperparameters. We utilized the grid search method to find out the best hyperparameter setting for our model. In our SVM classifier, the hyperparameters required to be optimized are penalty parameter: C, and kernel function parameters: d, γ. First, we considered a discrete space of the hyperparameters (C, d, γ) for the polynomial kernel, and (C, γ) for the RBF kernel, where $C, \gamma \in \{2^{-10}, 2^{-9}, ..., 2^{10}\}$, and $d \in \{1, 2, ..., 5\}$. The exponential sequence of the hyperparameters C and γ is a practical method of discretization to complete the grid search process within a reasonable duration of time. Next, we trained our SVM classifier for every possible combination of the hyperparameters. Finally, we selected the hyperparameter values which provided the maximum classification accuracy.

F. Model Building

The whole dataset is divided into training and testing set. Randomly 80% of the data are assigned as training set and rest of the 20% of the data are assigned as testing set. In this study, we utilized the Support Vector Machine (SVM) algorithm to train our ML model utilizing the training dataset. SVM was first developed by Vapnik [37] and it has been proven as a very powerful method in many clinical studies [38-41]. SVM is a non-probabilistic binary classifier which works by creating an optimal hyperplane that maximizes the margin between two classes. Besides linear classification, SVM can effectively perform non-linear classification by applying kernel trick [42].

Given a training dataset containing instance-label pairs $\{(x_i, y_i) | x_i \in \mathbb{R}^N, y_i \in \{-1, 1\}, i = 1, 2, ..., n\}$, SVM [43,44] requires to solve the following primal optimization problem:

$$\min_{w,b,\xi} \frac{1}{2} \|w\|^2 + c \sum_{i=1}^{n} \xi_i \tag{3}$$

$$subject\ to \begin{cases} y_i(w \cdot x_i + b) \geq 1 - \xi_i \\ \xi_i \geq 0\ and\ i = 1, 2, ..., n \end{cases} \tag{4}$$

Here, $C \in \mathbb{R}^+$ is the user defined penalty parameter of the error term, w is the weight vector which defines the direction of the separating hyperplane, b is the bias term, and ξ_i is a slack variable. The optimization problem defined by (3) and (4) can be solved efficiently by Lagrange multipliers method. After performing appropriate substitutions, we usually solve the below dual problem which is basically a convex constrained quadratic programming (CCQP) problem:

$$\min_{a} \frac{1}{2} \sum_{i,j=1}^{n} a_i a_j y_i y_j (x_i, x_j) - \sum_{i=1}^{n} a_i \tag{5}$$

$$subject\ to\ \sum_{i=1}^{n} a_i y_i = 0, 0 \leq a_i \leq C\ and\ i = 1, 2, ..., n \tag{6}$$

Using primal dual relationship, once the optimization problem is solved, the optimum weight vectors satisfy (7).

$$w_o = \sum_{i=1}^{n} y_i a_i x_i \tag{7}$$

According to (7), the optimal weight vector is the linear combination of input samples. The decision function for new sample points can be computed utilizing (8).

$$g(x) = \text{sgn}\left(\sum_{i,j=1}^{n} a_i y_i x_i + b\right) \tag{8}$$

If it is not possible to separate the training data by a linear hyperplane, the input vectors x_i are mapped into a higher dimensional feature space with the help of kernel functions. In this case, the decision function can be formulated as (9).

$$g(x) = \text{sgn}\left(\sum_{i,j=1}^{n} a_i y_i K(x_i, x_j) + b\right) \tag{9}$$

The following three are the most popular kernel functions:

- Linear: $K(x_i, x_j) = x_i^T x_j^T$
- Polynomial: $K(x_i, x_j) = (\gamma x_i^T x_j^T + 1)^d$, where $\gamma \in \mathbb{R}^+$ and $d \in \mathbb{Z}^+$ is the degree of the polynomial
- Radial Basis Function (RBF):

$$K(x_i, x_j) = exp(-\gamma \|x_i - x_j\|^2), \qquad \gamma \in \mathbb{R}^+$$

In our study, we deployed all the three kernel functions for the cryotherapy and immunotherapy datasets and chose the best performing ones.

G. Model assessment

For the assessment of our ML model, we predicted the dependent variable, treatment outcome, utilizing the 20% testing dataset which are completely unseen to our trained ML model. For the performance evaluation, three statistical measures are used, which include sensitivity, specificity, and accuracy [45]. Sensitivity represents the probability of correctly identifying the True Positive (TP) class, Y = 'Yes', while specificity represents the probability of correctly identifying the True Negative (TN) class, Y = 'No'. If the model predicts a class as negative while the actual class is positive, we define it as False Negative (FN). On the contrary, if the model predicts a class as positive while the actual class is negative, we define it as False Positive (FP). Overall, accuracy measures the probability of detecting the true class. Sensitivity, specificity, and accuracy were computed using (10-12).

$$Sensitivity = \frac{TP}{TP+FN} \qquad (10)$$

$$Specificity = \frac{TN}{TN+FP} \qquad (11)$$

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (12)$$

IV. Results and Discussion

Table 2 presents the performance of SVM classifier with the various subset of top features at different iterations based on feature ranking and SBS algorithm. For the immunotherapy dataset, the SVM classifier performed best when oversampled the minority class by ADASYN algorithm compared to the other two algorithms, SMOTE and Borderline-SMOTE. The

accuracy shown in the table for the immunotherapy dataset is only for the ADASYN algorithm. According to the table, the accuracy of the SVM classifier gradually increased up to iteration 5, and then gradually decreased in the subsequent iterations 6 to 7. For this dataset, the SVM classifier performed best with the top three features- time elapsed before starting the treatment, patient's age, and wart type. In a similar way, for the cryotherapy dataset, the accuracy of the classifier consistently increased up to iteration 3. For this dataset, the SVM classifier performed best with the top four features- time elapsed before starting the treatment, patient's age, wart type, and surface area of the largest wart.

Table 3 summarizes the performance of our SVM classifiers with the optimized hyperparameter values. For the immunotherapy dataset, after oversampling utilizing ADASYN algorithm, RBF kernel with the top three features performed best with 94.6% overall classification accuracy (sensitivity = 96.0%, specificity = 89.5%). On the other hand, for the cryotherapy dataset, third order polynomial kernel outperformed rest of the kernel functions. The average classification accuracy for the SVM model was 95.9% (sensitivity = 94.3%, specificity = 97.4%).
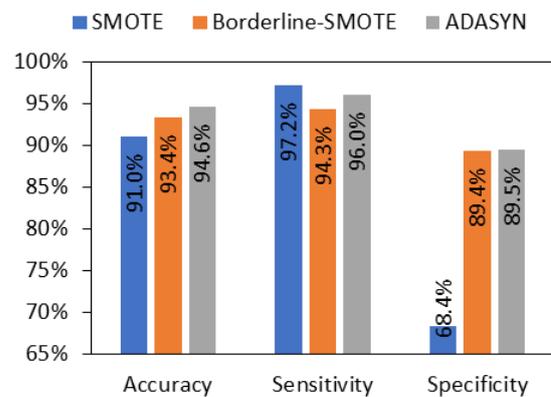

Fig.5. Comparison of the performance metrics of the immunotherapy dataset for SMOTE, Borderline-SMOTE, and ADASYN oversampling algorithms

Table 2. Performance of the feature subsets at different iterations based on feature ranking and SBS algorithm

Immunotherapy				Cryotherapy			
Model	Dimension	Selected features	Accuracy (%)	Model	Dimension	Selected features	Accuracy (%)
#1	7	3, 2, 5, 6, 4, 7, 1	79.5	#1	6	3, 2, 5, 6, 1, 4	88.9
#2	6	3, 2, 5, 6, 4, 7	81.8	#2	5	3, 2, 5, 6, 1	94.6
#3	5	3, 2, 5, 6, 4	83.1	**#3**	**4**	**3, 2, 5, 6**	**95.9**
#4	4	3, 2, 5, 6	89.0	#4	3	3, 2, 5	92.1
#5	**3**	**3, 2, 5**	**94.6**	#5	2	3, 2	87.5
#6	2	3, 2	81.1	#6	1	3	84.2
#7	1	3	80.3				

Note: Feature mapping: 1- Gender, 2- Age, 3- Time elapsed before starting treatment, 4- Number of warts, 5- Wart type, 6- Surface area of the largest wart, 7- Induration diameter

Table 3. Performance summary of SVM classifier with optimized hyperparameters

Dataset	Selected features	Kernel function	C	γ	Accuracy	Sensitivity	Specificity
Immunotherapy (ADASYN oversampling)	Time elapsed, Age, Wart type	RBF	2^{-2}	2^3	94.6%	96.0%	89.5%
Cryotherapy	Time elapsed, Age, Wart type, Wart area	Polynomial (order = 3)	2^{-4}	2^3	95.9%	94.3%	97.4%

Fig. 5 shows the comparison of the performance metrics, i.e., accuracy, sensitivity, and specificity, when the minority class of the immunotherapy dataset is oversampled by SMOTE, Borderline-SMOTE, and ADASYN algorithms. According to the figure, the overall classification accuracy is maximum for the ADASYN algorithm. Though SMOTE algorithm provided maximum sensitivity, the specificity and classification accuracy are the lowest among the three oversampling methods.

Besides SVM classifier, we employed other well-known ML models for the immunotherapy and cryotherapy datasets and the performance metrics are summarized in Table 4. According to the table, the overall performance of our proposed SVM classifier is better than the rest of the ML models. However, K-Nearest Neighbors and Random Forest classifiers provided the maximum specificity for the cryotherapy and immunotherapy treatment methods, respectively.

Table 4. Performance of proposed method with popular classification algorithms

Algorithms	Immunotherapy			Cryotherapy		
	Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity
KNN	80.9%	78.1%	**93.6%**	89.5%	86.8%	92.3%
BLR	76.3%	72.2%	80.0%	88.9%	85.7%	92.1%
LDA	75.2%	74.0%	77.8%	89.6%	86.4%	92.3%
QDA	75.2%	76.4%	72.6%	85.4%	92.9%	75.7%
CART	85.3%	85.2%	88.4%	90.8%	88.8%	92.6%
RF	92.7%	95.1%	84.8%	93.7%	90.6%	**98.0%**
Bagging	90.7%	93.2%	78.1%	91.5%	88.7%	95.2%
Adaptive Boosting	84.9%	89.5%	64.2%	93.6%	90.9%	96.4%
Gradient Boosting	87.2%	90.8%	74.9%	95.0%	**94.3%**	95.1%
Proposed method	**94.6**%	**96.0**%	89.5%	**95.9**%	**94.3%**	97.4%

Note: Bold faced numbers denote the best values; KNN- K-Nearest Neighbors; BLR- Binary Logistic Regression; , LDA- Linear Discriminant Analysis; QDA- Quadratic Discriminant Analysis; CART= Classification and Regression Tree; RF= Random Forest

Table 5. Comparison of our results with previous studies

Study and year	Dataset	Accuracy	Sensitivity	Specificity
Khozeimeh et al. (2017) [13]	Immunotherapy	83.3%	87.0%	71.0%
Akben(2018) [15]	Immunotherapy	90.0%	**97.2%**	63.2%
Nugroho et al.(2018) [46]	Immunotherapy	84.4%	91.4%	55.0%
Jain et al.(2018) [47]	Immunotherapy	88.1%	Not reported	Not reported
Basarslan et al.(2018) [48]	Immunotherapy	84.0%	Not reported	Not reported
Degirmenci et al.(2018) [49]	Immunotherapy	81.3%	Not reported	Not reported
Akyol et al.(2018) [20]	Immunotherapy	89.3%	95.7%	60.0%
Guimarães et al. (2019) [50]	Immunotherapy	88.6%	93.0%	86.0%
Abdar et al.(2019) [17]	Immunotherapy	84.4%	Not reported	Not reported
Jia et al.(2019) [51]	Immunotherapy	76.2%	Not reported	Not reported
This study	Immunotherapy	**94.6**%	96.0%	**89.5**%
Khozeimeh et al. (2017) [13]	Cryotherapy	80.0%	82.0%	77.0%
Akben(2018) [15]	Cryotherapy	94.4%	89.6%	**100.0%**
Nugroho et al.(2018) [46]	Cryotherapy	93.3%	88.5%	98.0%
Jain et al.(2018) [47]	Cryotherapy	94.8%	Not reported	Not reported
Basarslan et al.(2018) [48]	Cryotherapy	95.4%	Not reported	Not reported
Degirmenci et al.(2018) [49]	Cryotherapy	93.1%	Not reported	Not reported
Akyol et al.(2018) [20]	Cryotherapy	**96.4**%	94.4%	**100.0%**
Guimarães et al. (2019) [50]	Cryotherapy	84.3%	**97.0%**	41.0%
Abdar et al.(2019) [17]	Cryotherapy	94.4%	Not reported	Not reported
This study	Cryotherapy	95.9%	94.3%	97.4%

Note: Bold faced numbers denote the best values

Table 5 compares the results of our study with previous studies in the literature regarding the applications of ML algorithms on wart treatment. For the immunotherapy treatment method, our classification model achieved 4.6% higher classification accuracy compared to the second-best model reported by Akben [15]. Although the study by Akben achieved the maximum sensitivity of 97.2%, the achieved specificity value was only 63.2%; therefore, the model is not reliable to detect the true negative class (Y = 'No'). The author did not follow any preventive measures to balance the dataset; consequently, the trained model failed to learn to predict the minority class (Y = 'No') properly. On the other hand, for the cryotherapy dataset, our study achieved the second-highest classification accuracy, 95.9%. This value is competitive and comparable with the study by Akyol *et* al. [20] where the maximum classification accuracy of 96.4% was reported.

In our study, we followed a two-step process for finding the best set of features for training our ML model. In the first step, we employed the F-score method to rank the features and in the second step, different ML algorithms are applied. However, the ranking of the features could be different by employing other filter algorithms. In that case, the optimum set of features from the SBS algorithm could be different. In addition, to balance the minority class, we created synthetic samples of the training dataset by deploying oversampling algorithms. An alternative approach could be giving more weight values to the samples that belong to the minority class. Consequently, the cost function would be penalized more for misclassification of a minority class compared to a majority class. These alternative approaches might end up with different performance metrics.

V. CONCLUSION

In this study, we employed the SVM algorithm to develop an expert system to predict the effectiveness of immunotherapy and cryotherapy for the treatment of warts by analyzing patients' demographic and clinical information. We combined the F-score algorithm, a filter method, with the SBS algorithm, a wrapper method, to develop predictive ML models. Besides, we employed different oversampling algorithms to overcome the imbalanced dataset problem. We compared our results with state-of-the-art methodologies found in the literature. The developed SVM models showed promising classification performances, especially for the immunotherapy dataset where the distribution of the target classes was highly skewed. The developed expert system could potentially assist the dermatologists as a decision support tool to choose between cryotherapy and immunotherapy as a wart treatment method for every unique patient by predicting the success before starting the treatment process. Therefore, early prediction of a treatment method success might possibly help to reduce the undesirable side effects for the patients and save valuable resources of the hospitals by minimizing the probability of treatment failures.

It should be noted that the dataset used in this study represents a particular race. In the future, more robust and generalized ML models can be developed by obtaining additional data on different groups of patients from different races and geographic locations. Furthermore, the specificity of the immunotherapy treatment method is still below 90%; therefore, there is still room for improvement of this performance metric.

REFERENCES

[1] I. Ahmed, S. Agarwal, A. Ilchyshyn, S. Charles‐Holmes, J. Berth‐Jones, Liquid nitrogen cryotherapy of common warts: cryo‐spray vs. cotton wool bud, Br. J. Dermatol. 144 (2001) 1006–1009.

[2] N. Tomson, J. Sterling, I. Ahmed, J. Hague, J. Berth‐Jones, Human papillomavirus typing of warts and response to cryotherapy, J. Eur. Acad. Dermatology Venereol. 25 (2011) 1108–1111.

[3] S.C. Bruggink, J. Gussekloo, M.Y. Berger, K. Zaaijer, W.J.J. Assendelft, M.W.M. de Waal, J.N.B. Bavinck, B.W. Koes, J.A.H. Eekhof, Cryotherapy with liquid nitrogen versus topical salicylic acid application for cutaneous warts in primary care: randomized controlled trial, Can. Med. Assoc. J. 182 (2010) 1624–1630.

[4] M. Maronn, C. Salm, V. Lyon, S. Galbraith, One‐year experience with candida antigen immunotherapy for warts and molluscum, Pediatr. Dermatol. 25 (2008) 189–192.

[5] R.J. Signore, Candida albicans intralesional injection immunotherapy of warts, Cutis. 70 (2002) 185–192.

[6] S.M. Johnson, P.K. Roberson, T.D. Horn, Intralesional injection of mumps or Candida skin test antigens: a novel immunotherapy for warts, Arch. Dermatol. 137 (2001) 451–455.

[7] R.C. Phillips, T.S. Ruhl, J.L. Pfenninger, M.R. Garber, Treatment of warts with Candida antigen injection, Arch. Dermatology-Chicago. 136 (2000) 1274.

[8] M.M. Rahman, S. Wang, Y. Zhou, J. Rogers, Predicting the Performance of Cryotherapy for Wart Treatment Using Machine Learning Algorithms, in: IISE Annu. Conf., Institute of Industrial and Systems Engineers, Orlando, FL, 2019.

[9] E. Barati, M. Saraee, A. Mohammadi, N. Adibi, M.R. Ahmadzadeh, A survey on utilization of data mining approaches for dermatological (skin) diseases prediction, J. Sel. Areas Heal. Informatics. 2 (2011) 1–11.

[10] S.O. Olatunji, H. Arif, Identification of Erythemato-Squamous skin diseases using extreme learning machine and artificial neural network, ICTACT J. Softw Comput. 4 (2013) 627–632.

[11] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, Nature. 542 (2017) 115.

[12] M. Antkowiak, Artificial Neural Networks vs. Support Vector machines for skin diseases recognition, Neural

Networks. (2006).

[13] F. Khozeimeh, R. Alizadehsani, M. Roshanzamir, A. Khosravi, P. Layegh, S. Nahavandi, An expert system for selecting wart treatment method, Comput. Biol. Med. 81 (2017) 167–175. doi:10.1016/j.compbiomed.2017.01.001.

[14] M.A. Putra, N.A. Setiawan, S. Wibirama, Wart treatment method selection using AdaBoost with random forests as a weak learner, Commun. Sci. Technol. 3 (2018) 52–56.

[15] S.B. Akben, Predicting the success of wart treatment methods using decision tree based fuzzy informative images, Biocybern. Biomed. Eng. 38 (2018) 819–827.

[16] A.J. Guimar ães, V.J.S. Araujo, P.V. de Campos Souza, V.S. Araujo, T.S. Rezende, Using Fuzzy Neural Networks to the Prediction of Improvement in Expert Systems for Treatment of Immunotherapy, in: Ibero-American Conf. Artif. Intell., Springer, 2018: pp. 229–240.

[17] M. Abdar, V.N. Wijayaningrum, S. Hussain, R. Alizadehsani, P. Pławiak, U.R. Acharya, V. Makarenkov, IAPSO-AIRS: A Novel Improved Machine Learning-based System for Wart Disease Treatment, (n.d.).

[18] D. Dheeru, E. Karra Taniskidou, UCI machine learning repository [https://archive.ics.uci.edu/ml/datasets/Cryotherapy+Datas et+], in: University of California, Irvine, School of Information and Computer Sciences, 2017.

[19] D. Dheeru, E. Karra Taniskidou, UCI machine learning repository [https://archive.ics.uci.edu/ml/datasets/Immunotherapy+D ataset], in: University of California, Irvine, School of Information and Computer Sciences, 2017.

[20] K. AKYOL, A. KARACI, Y. GÜLTEPE, A Study on Prediction Success of Machine Learning Algorithms for Wart Treatment, Int. Conf. Adv. Technol. Comput. Eng. Sci. (2018) 186–188.

[21] L. Al Shalabi, Z. Shaaban, B. Kasasbeh, Data mining: A preprocessing engine, J. Comput. Sci. 2 (2006) 735–739.

[22] M. Dash, H. Liu, Feature selection for classification, Intell. Data Anal. 1 (1997) 131–156.

[23] L. Yu, H. Liu, Feature selection for high-dimensional data: A fast correlation-based filter solution, in: Proc. 20th Int. Conf. Mach. Learn., 2003: pp. 856–863.

[24] S. Das, Filters, wrappers and a boosting-based hybrid for feature selection, in: Icml, 2001: pp. 74–81.

[25] S. Maldonado, R. Weber, A wrapper method for feature selection using support vector machines, Inf. Sci. (Ny). 179 (2009) 2208–2217.

[26] M.F. Akay, Support vector machines combined with feature selection for breast cancer diagnosis, Expert Syst. Appl. 36 (2009) 3240–3247. doi:10.1016/j.eswa.2008.01.009.

[27] S. Güneş, K. Polat, Ş. Yosunkaya, Multi-class f-score feature selection approach to classification of obstructive sleep apnea syndrome, Expert Syst. Appl. 37 (2010) 998–1004.

[28] M.M. Rahman, Y. Ghasemi, E. Suley, Y. Zhou, S. Wang, J. Rogers, Machine Learning Based Computer Aided Diagnosis of Breast Cancer Utilizing Anthropometric and Clinical Features, (2019).

[29] Y.-W. Chen, C.-J. Lin, Combining SVMs with various feature selection strategies, in: Featur. Extr., Springer, 2006: pp. 315–324.

[30] G.M. Weiss, F. Provost, Learning when training data are

costly: The effect of class distribution on tree induction, J. Artif. Intell. Res. 19 (2003) 315–354.

[31] Y. Sun, A.K.C. Wong, M.S. Kamel, Classification of imbalanced data: A review, Int. J. Pattern Recognit. Artif. Intell. 23 (2009) 687–719.

[32] N. V Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique, J. Artif. Intell. Res. 16 (2002) 321–357.

[33] H. Han, W.-Y. Wang, B.-H. Mao, Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning, in: Int. Conf. Intell. Comput., Springer, 2005: pp. 878–887.

[34] H. He, Y. Bai, E.A. Garcia, S. Li, ADASYN: Adaptive synthetic sampling approach for imbalanced learning, in: 2008 IEEE Int. Jt. Conf. Neural Networks (IEEE World Congr. Comput. Intell., IEEE, 2008: pp. 1322–1328.

[35] G. Lema ître, F. Nogueira, C.K. Aridas, Imbalanced-learn: A python toolbox to tackle the curse of imbalanced datasets in machine learning, J. Mach. Learn. Res. 18 (2017) 559–563.

[36] S.J. Reeves, Z. Zhe, Sequential algorithms for observation selection, IEEE Trans. Signal Process. 47 (1999) 123–132.

[37] V. Vapnik, The Nature of Statistical Learning Theory ·6 {MJ New York, Springer-Verlag. 1 (1995) 995.

[38] K. Shankar, S.K. Lakshmanaprabu, D. Gupta, A. Maseleno, V.H.C. de Albuquerque, Optimal feature-based multi-kernel SVM approach for thyroid disease classification, J. Supercomput. (2018) 1–16.

[39] G. Orru, W. Pettersson-Yeo, A.F. Marquand, G. Sartori, A. Mechelli, Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review, Neurosci. Biobehav. Rev. 36 (2012) 1140–1152.

[40] B. Magnin, L. Mesrob, S. Kinkingn éhun, M. P élégrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Leh éricy, H. Benali, Support vector machine-based classification of Alzheimer's disease from whole-brain anatomical MRI, Neuroradiology. 51 (2009) 73–83.

[41] S. Pan, S. Iplikci, K. Warwick, T.Z. Aziz, Parkinson's Disease tremor classification–A comparison between Support Vector Machines and neural networks, Expert Syst. Appl. 39 (2012) 10764–10771.

[42] B. Scholkopf, A.J. Smola, Learning with kernels: support vector machines, regularization, optimization, and beyond, MIT press, 2001.

[43] C. Cortes, V. Vapnik, Support-vector networks, Mach. Learn. 20 (1995) 273–297.

[44] B.E. Boser, I.M. Guyon, V.N. Vapnik, A training algorithm for optimal margin classifiers, in: Proc. Fifth Annu. Work. Comput. Learn. Theory, ACM, 1992: pp. 144–152.

[45] A.G. Lalkhen, A. McCluskey, Clinical tests: sensitivity and specificity, Contin. Educ. Anaesth. Crit. Care Pain. 8 (2008) 221–223.

[46] H.W. Nugroho, T.B. Adji, N.A. Setiawan, Random forest weighting based feature selection for c4. 5 algorithm on wart treatment selection method, Int. J. Adv. Sci. Eng. Inf. Technol. 8 (2018) 1858–1863.

[47] R. Jain, R. Sawhney, P. Mathur, Feature Selection for Cryotherapy and Immunotherapy Treatment Methods Based on Gravitational Search Algorithm, in: 2018 Int. Conf. Curr. Trends Towar. Converging Technol., IEEE,

2018: pp. 1–7.

[48] M.S. Basarslan, F. Kayaalp, A Hybrid Classification Example in the Diagnosis of Skin Disease with Cryotherapy and Immunotherapy Treatment, in: 2018 2nd Int. Symp. Multidiscip. Stud. Innov. Technol., IEEE, 2018: pp. 1–5.

[49] A. Degirmenci, O. Karal, Evaluation of Kernel Effects on SVM Classification in the Success of Wart Treatment Methods, Am. J. Eng. Res. 7 (2018) 238–244.

[50] A. Junio Guimarâes, P. Vitor de Campos Souza, V. Jonathan Silva Araújo, T. Silva Rezende, V. Souza Araújo, 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 Cogn. Comput. 3 (2019) 22.

[51] W. Jia, Y. Deng, C. Xin, X. Liu, W. Pedrycz, A classification algorithm with Linear Discriminant Analysis and Axiomatic Fuzzy Sets, Math. Found. Comput. 2 (2019) 73–81.

Authors' Profiles

Md. Mamunur Rahman is a Ph.D. candidate at the Department of Industrial, Manufacturing and Systems Engineering, The University of Texas at Arlington, TX, USA. He received his B.S. and M.S. degrees in Industrial Engineering from Bangladesh University of Engineering and Technology, Dhaka, Bangladesh in 2011 and 2015, respectively. His research interests include data science, machine learning, statistical modeling, and discrete-event and agent-based simulation.

Yuan Zhou is an Assistant Professor of Industrial, Manufacturing, & Systems Engineering at UT-Arlington, specializing in complex adaptive system modeling, agent-based and discrete-event simulation, performance measurement and process improvement. She received a B.S. Degree in Mechanical and Electrical Engineering from Beijing Institute of Technology, Beijing, China in 2004, and a Ph.D. Degree in Industrial and Systems Engineering from University at Buffalo, The State University of New York, Buffalo, New York. Her research has been sponsored by the National Science Foundation (NSF), Agency for Health Research and Quality (AHRQ), and City of Arlington. Dr. Zhou is a member of the Institute of Industrial and Systems Engineering (IISE) and the Institute for Operations Research and Management Sciences (INFORMS). She is currently the Vice President of DFW IISE professional chapter.

Shouyi Wang received the B.S. Degree in Systems and Control Engineering from Harbin Institute of Technology, Harbin, China, in 2003, the M.S. Degree in systems and control engineering from the Delft University of Technology, Delft, The Netherlands, in 2005, and the Ph.D. degree in industrial and systems engineering from Rutgers, the State University of New Jersey, New Brunswick, NJ, USA, in 2012. He is currently an Associate Professor with the Department of Industrial, Manufacturing, and Systems Engineering, the University of Texas at Arlington (UTA), Arlington, TX, USA. He is also a core faculty member at the Center on Stochastic Modeling, Optimization, and Statistics (COSMOS) at UTA. His research interests include data mining, machine learning, applied statistics, complex system modeling, and decision analytics.

Jamie Rogers is currently a UT-System Regents' Outstanding Teacher, Professor and Associate Chair of the IMSE Department at the UT-Arlington. Dr. Rogers is serving as 2019-20 President for the Institute of Industrial & Systems Engineers (IISE) and has served as 2014-15 ABET President which now accredits approximately 4,005 programs at over 793 institutions in 32 countries worldwide. She served as Faculty Advisor for the IISE Gold Award winning UTA Student Chapter (1994-2019) and was inducted into the University of Missouri Industrial and Manufacturing Systems Engineering Hall of Fame in 2013. She has received over $3.6M in funding as PI or Co-PI, successfully supervised 33 PhD dissertations, and published over 140 papers worldwide. Dr. Rogers is a Fellow in IISE, an ABET Fellow, and a Registered Professional Engineer in Texas. Prior to joining academia, she worked at Texas Instruments in various engineering and management positions in defense electronics and semiconductor business areas.