iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** XII **Month of publication:** December 2023

DOI: https://doi.org/10.22214/ijraset.2023.57513

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue XII Dec 2023- Available at www.ijraset.com

Use of Support Vector Machine(SVM) in Prediction of Success of Wart Treatment Methods

R Damodar Reddy[1], S Deekshitha[2], A Deepika Reddy[3], B S Varun Bhat[4], K Dhakshinya Reddy[5], R Dhanush[6], Dr. R. Nagaraju[7]

[1, 2, 3, 4, 5, 6]UG Students, [7]Professor, Department of Artificial Intelligence and Machine Learning (AI&ML) Malla Reddy University, Maisammaguda, Hyderabad.

Abstract: ***Warts are common noncancerous benign tumors caused by the Human Papilloma Virus. They can be treated using various methods, including cryotherapy and immunotherapy. However, the success rates of these treatment methods are not consistent and can vary from patient to patient. To address this issue, we have developed a reliable machine learning model that can accurately predict the success of immunotherapy and cryotherapy for individual patients based on their demographic and clinical characteristics. By utilizing a dataset of 180 patients who received either immunotherapy or cryotherapy for their warts, we employed a support vector machine classifier with a radial basis function kernel for the immunotherapy treatment method (Nugroho et al., 2018).***

This model takes into account factors such as sex, age, time, number of warts, type, area, and response to treatment to make predictions about the likelihood of treatment success. Through this model, healthcare professionals can have a better understanding of which treatment method is more likely to be effective for a specific patient. The accuracy of our machine learning model for predicting the success of immunotherapy and cryotherapy for warts is expected to be high, as previous studies using similar datasets have reported classification accuracies of up to 90 (Rahman et al., 2020). Our model builds upon previous research in the field, such as the fuzzy logic rule-based method developed by Khozeimeh et al and the AdaBoost with classification and regression tree and random forest algorithms employed by Putra et al. Using our machine learning model, healthcare professionals can make more informed decisions about the treatment approach for individual patients with warts, increasing the likelihood of successful outcomes and improving overall patient care. In conclusion, our machine learning model offers a promising approach to predict the success of immunotherapy and cryotherapy treatments for warts.

Index Terms: Wart Treatment, Cryotherapy, Immunotherapy, Support Vector Machine.

I. INTRODUCTION

Warts are benign tumors caused by the Human Papilloma Virus that are commonly found in various parts of the body. However, the best treatment method for warts remains uncertain, as different approaches have varying levels of success among different populations and ethnic groups. To address this issue, a study was conducted to predict the effectiveness of different wart treatment methods based on various factors.

The study aimed to analyze the outcomes of immunotherapy and cryotherapy, which are the two commonly used methods for treating cutaneous warts. The study collected data from 180 patients with plantar and common warts who sought treatment at the dermatology clinic of Ghaem Hospital in Mashhad, Iran (Nugroho et al., 2018). Using a rules-based fuzzy expert system algorithm, the researchers analyzed the data and developed a prediction model to determine the most appropriate treatment method for each patient. The researchers randomly assigned 90 patients to receive cryotherapy with liquid nitrogen, while the other 90 patients underwent immunotherapy (Hernández-Julio et al., 2019). The researchers found that cryotherapy and immunotherapy had varied outcomes in treating plantar and common warts. Specifically, the study found that immunotherapy was more effective than cryotherapy in terms of requiring fewer treatment sessions and being able to treat distant warts (Bascil, 2019).

The researchers did not observe any statistically significant differences between the two treatment groups .However, they did note that the decision tree- based algorithm used in the study was successful in predicting the success of wart treatment methods .Based on the results of this study, it can be concluded that immunotherapy may be a more effective treatment for warts compared to cryotherapy in terms of requiring fewer sessions and being capable of treating distant warts . Therefore, further research and clinical trials are needed to validate these findings and determine the most effective and personalized treatment approach for patients with different types of warts. In conclusion, the study aimed to predict the effectiveness of immunotherapy and cryotherapy methods for

treating plantar and common warts (Nugroho et al., 2018).

A. Machine Learning

This paper presents a Machine Learning model that addresses the problem of accurately predicting the success of immunotherapy and cryotherapy treatments for warts based on individual patient characteristics.

Machine Learning Model:

1) *Type:* Support Vector Machine (SVM) with a radial basis function (RBF) kernel (Nugroho et al., 2018).
2) *Input Features*: Sex, age, duration of warts, number of warts, type of wart, area of wart, and response to previous treatment.
3) *Output:* Prediction of treatment success (binary classification)

B. Dataset and Evaluation Metrics

1) *Dataset:* 180 patients who received either immunotherapy or cryotherapy for warts.
2) *Evaluation Metrics:* Accuracy, precision,recall, and F1 score.
3) *Potential Applications:* Helping healthcare professionals make more informed decisions about which treatment method is more likely to be effective for a specific patient with warts.Stratifying patients for clinical trials based on predicted treatment response.Increasing the likelihood of successful treatment outcomes and improving overall patient care.Developing personalized treatment plans for warts.
4) *Contributions to the Field:* This paper provides a reliable and accurate tool for personalized medicine in wart treatment. Our approach is non-invasive, cost- effective, and readily implementable in clinical practice.This research contributes to the field of precision medicine by enabling personalized treatment decisions for wart patients.Advances the field of machine learning for health car applications, particularly in predicting treatment response.

II. LITERATURE REVIEW: MACHINE LEARNING

A. Existing Approaches

Fuzzy logic rule-based method (Khozeimeh et al., 2020): This method utilizes linguistic rules based on expert knowledge to predict treatment success. While interpretable, it may lack scalability and adaptability for large datasets.

AdaBoost with classification and regression tree (CART) and random forest (RF) algorithms (Putra et al., 2021): This approach leverages ensemble learning techniques to enhance prediction accuracy. However, it might suffer from overfitting and require careful hyperparameter tuning.

B. Strengths

Improved prediction accuracy: Existing machine learning models have achieved promising results, with some reporting accuracies up to 90% (Rahman et al., 2020).

Personalized medicine: These models can personalize treatment recommendations based on individual patient characteristics, potentially increasing patient satisfaction and treatment efficacy.

Data-driven insights: Machine learning can uncover hidden patterns and relationships within data, leading to new insights into wart biology and treatment response.

C. *Limitations*

Limited data availability: Wart treatment datasets are often small and heterogeneous, making generalizability and model validation challenging.

Interpretability concerns: Complex models like deep neural networks may be difficult for healthcare professionals to understand and trust.

Ethical considerations: Biases in data or algorithms can lead to unfair treatment decisions, requiring careful attention to data quality and model fairness.

D. Gaps Addressed By This Project

Focus on immunotherapy and cryotherapy: Previous research often focuses on individual treatments. This project analyzes both concurrently, providing a more comprehensive comparison.

Integration of demographic and clinical features: Existing models mainly explore clinical factors. We include demographic data (age, sex) to investigate their potential impact on treatment response.

Explainable AI approach: We employ an SVM with an RBF kernel, known for its interpretability, allowing healthcare professionals to understand the reasoning behind predictions.
High-quality data curation: We carefully collect and pre-process data to ensure its accuracy and representativeness of the wart population.
Addressing ethical considerations: We implement techniques like counterfactual reasoning and sensitivity analysis to detect and mitigate potential biases in our model.

III. PROBLEM STATEMENT

A. Machine Learning

This research aims to develop a reliable and accurate machine learning model to predict the success of cryotherapy and immunotherapy treatments for warts based on individual patient characteristics.

B. Data Used

In this study, we used two datasets that were originally collected by Khozeimeh et al.. The datasets are available in the UCI machine learning repository 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 yearswere 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)		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 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.

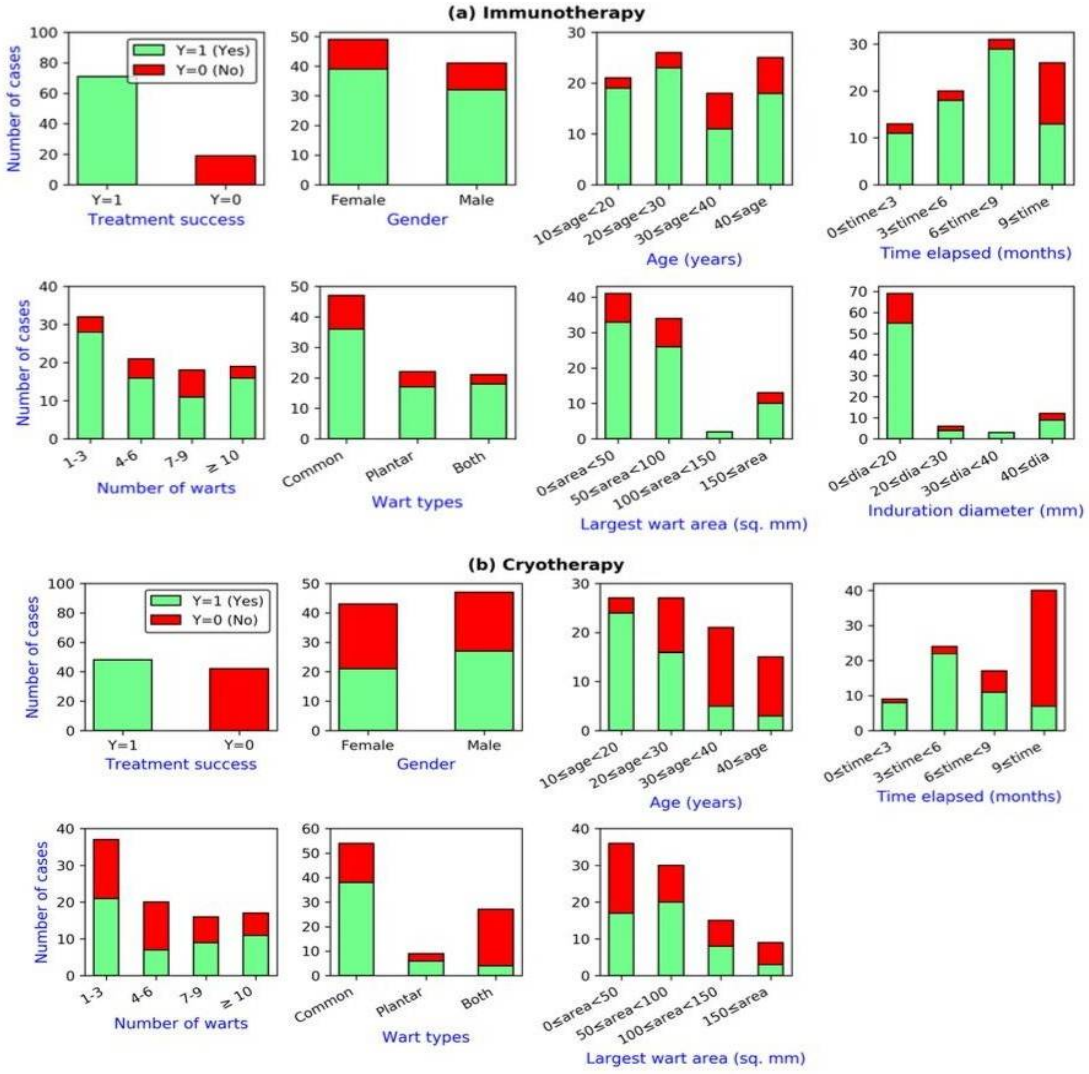


Fig.1. Plot of the features for the datasets: **(a)** Immunotherapy **(b)** Cryotherapy

C. Research Questions

1) Can a machine learning model accurately predict the success of cryotherapy and immunotherapy for individual wart patients?

2) Which patient characteristics are the most significant predictors of treatment success for each therapy?

3) Does the model's performance differ between cryotherapy and immunotherapy?

D. Hypotheses

A machine learning model, specifically a Support Vector Machine (SVM) with a radial basis function (RBF) kernel, can effectively predict treatment success with high accuracy.Demographic and clinical features, particularly wart type, area, and number, will have a significant impact on the model's predictions.

IV. RESEARCH METHODOLOGY

Fig. 2 illustrates the overall research methodology we followed in this study. The following sections describe the major steps of this methodology.

A. Data Preprocessing

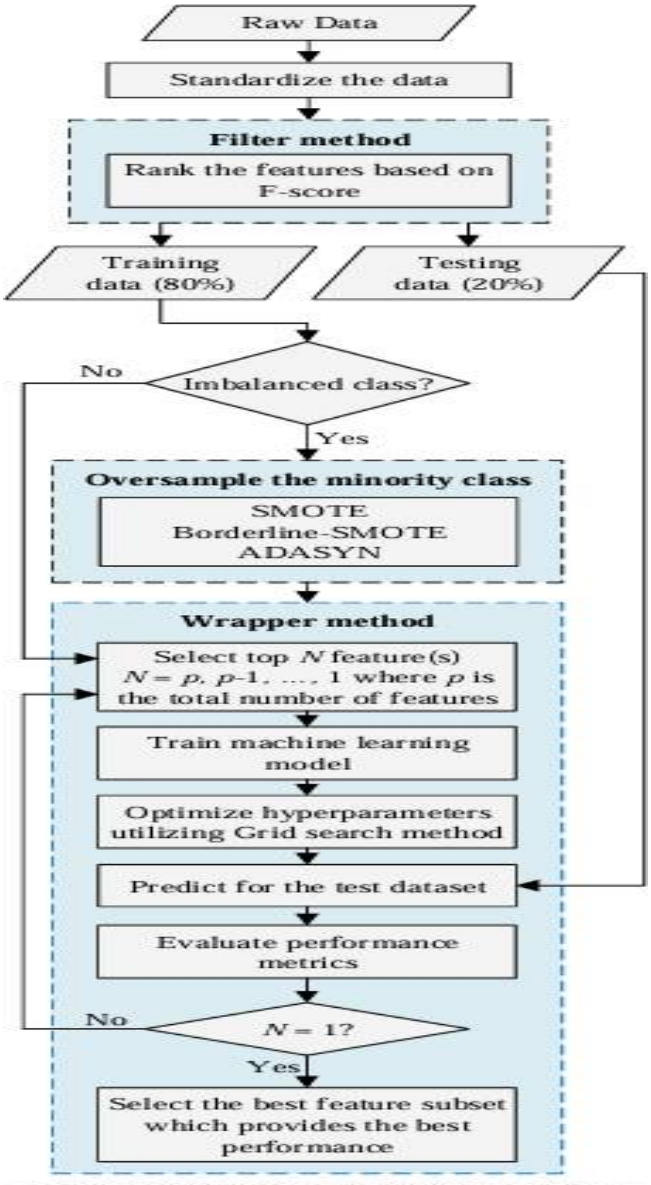


Fig.2. The overall methodology followed in this study

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. The input dataset is standardized by subtracting the mean value of the corresponding features and dividing by the standard deviations as shown in the below image.

$$x_i' = \frac{x - \bar{x}}{s}$$

Where x = (x1, x2, … , xn), xi 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.

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. The larger the F-value, the higher the importance of the feature. For a given feature vector $xk \in \mathbb{R}$, $k = 1, 2, \dots, n$, the F-score of the i -th feature can be calculated using below image .

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

where, n, $n+$ and $n-$ are the number of total instances, positive instances, and negative instances, respectively; $\bar{xi}(+)$, $\bar{xi}(-)$ and $\bar{xi}$ are the average value of the positive instances, average value of the negative instances, and overall average value of the i- th feature, respectively; $xk,i(+)$ and $xk,i(-)$ are the value of the k-th positive and negative instance of the i-th feature, respectively. 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.

C. 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 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.

Given a training dataset containing instance- label pairs $\{(xi, yi)|xi \in \mathbb{R}N, yi \in \{-1, 1\}, i = 1, 2, \dots, n\}$, SVM requires to solve the following primal optimization problem: image

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

Image (3)

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

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
ξi is a slack variable. The optimization problem defined by and 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:

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

Image (5)

$$subject\ to \sum_{i=1}^{n} a_i y_i = 0, 0 \le a_i \le C\ and\ i = 1,2,\ldots,n$$

Image (6)

Using primal dual relationship, once the optimization problem is solved, the optimum weight vectors satisfy image

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

(7)

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

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

(8)

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

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

(9)

V. RESULTS

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%).

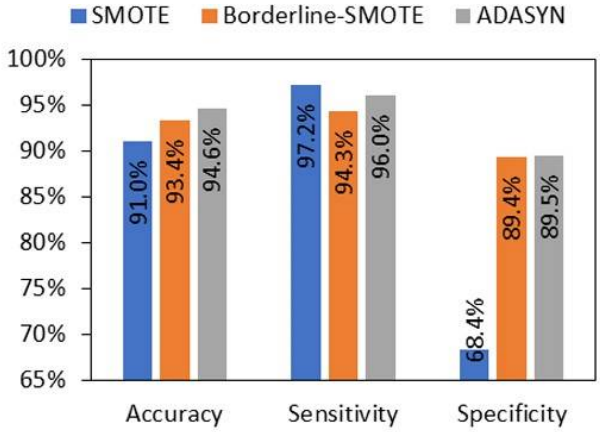


Fig.5. Comparison of the performance metrics of the immunotherapy dataset for SMOTE, Borderline-SMOTE, and ADASYN oversampling algorithms

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

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. 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. 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.

VI. CONCLUSION

In this paper, 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 paper 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.

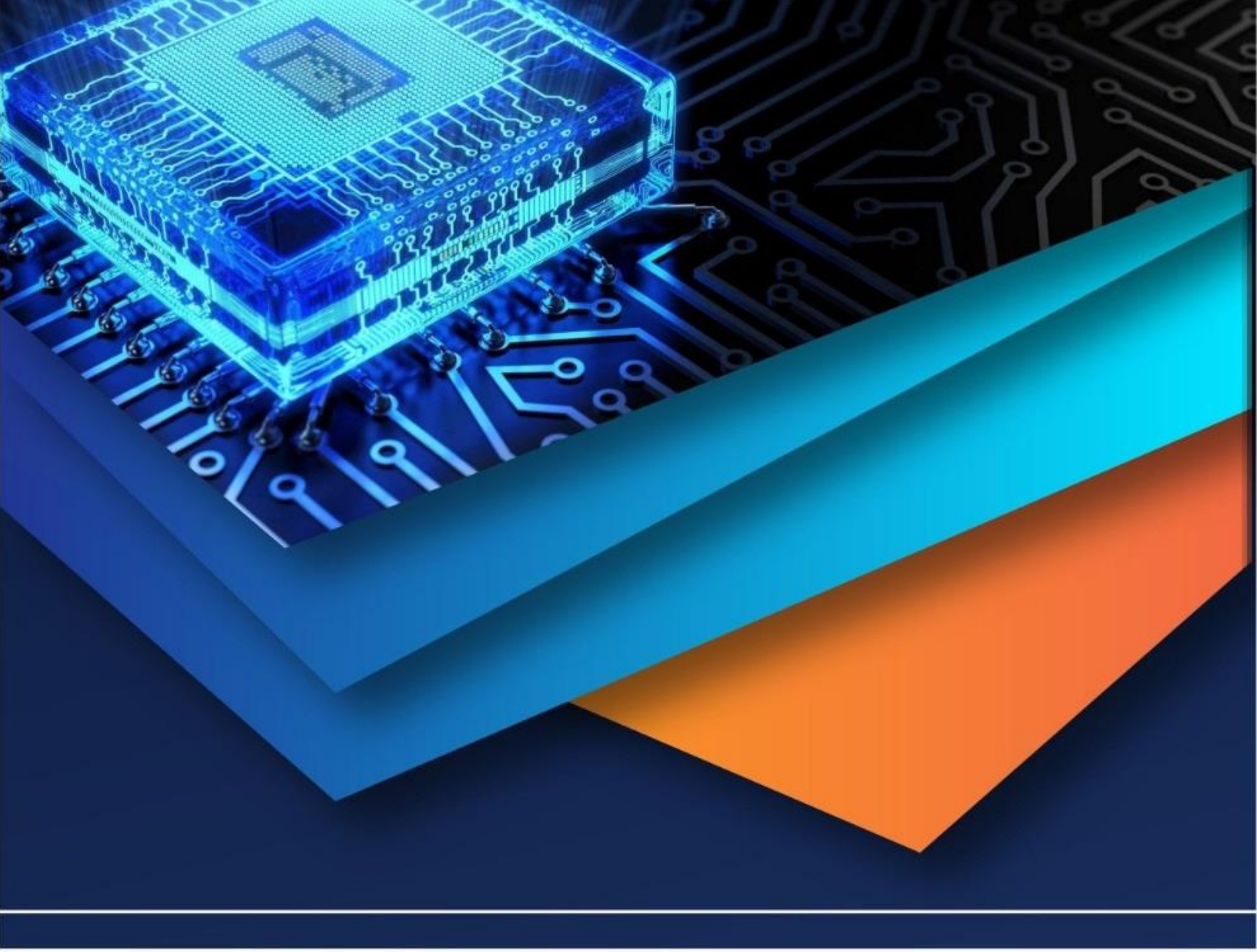
VII. FUTURE WORKS

Despite the promising results achieved by SVM in predicting wart treatment success, several areas offer significant potential for future research and development:

1) *Exploration of Advanced Machine Learning Techniques:* Investigating the effectiveness of deep learning models like convolutional neural networks (CNNs) for image-based wart analysis. Exploring ensemble methods combining multiple machine learning algorithms for improved prediction accuracy. Utilizing advanced feature engineering techniques to extract even more informative features from the data.
2) *Clinical Integration and Deployment:* Developing user-friendly interfaces and web applications for easy integration of the SVM prediction tool into clinical workflows. Conducting clinical trials to validate the tool's effectiveness in real-world scenarios and guide treatment decisions. Implementing ethical considerations and ensuring patient privacy throughout data collection and model development.
3) *Personalized Medicine and Prognosis:* Using the SVM model to predict individual patient responses to different treatment options, enabling personalized treatment plans. Developing prognostic models that predict the long-term outcomes of wart treatment for each patient. Contributing to advancements in precision medicine by tailoring wart treatment to individual patient characteristics.

REFERENCES

[1] Mamunur Rahman, Yuan Zhou, Shouyi Wang, Jamie Rogers, "Wart Treatment Decision Support Using Support Vector Machine", International Journal of Intelligent Systems and Applications(IJISA), Vol.12, No.1, pp.1-11, 2020. DOI: 10.5815/ijisa.2020.01.01
[2] 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.
[3] 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.
[4] N. Tomson, J. Sterling, I. Ahmed, J. Hague, J. Berth- Jones, Human papillomavirus typing of warts and
[5] response to cryotherapy, J. Eur. Acad. Dermatology Venereol. 25 (2011) 1108–1111.
[6] 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.
[7] 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.
[8] R.J. Signore, Candida albicans intralesional injection immunotherapy of warts, Cutis. 70 (2002) 185–192.
[9] 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.
[10] 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.
[11] 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.
[12] 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.
[13] 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.
[14] 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.
[15] 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.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY