ORIGINAL ARTICLE

The Application of Machine Learning in Predicting Outcome of Cryotherapy and Immunotherapy for Wart Removal

Yashik Singh

Department of Telehealth, School of Nursing and Public Health, University of KwaZulu-Natal, Durban, South Africa

Background: Warts can be extremely painful conditions that may be associated with localised bleeding and discharge. They are commonly treated by cryotherapy or immunotherapy. However, each of these therapies have discomforting side effects and are no official dermatological guideline that exist that may be used to determine which of these methods would work on an individual patient. **Objective:** This study aimed at developing a machine learning algorithm that improved the prediction of the outcome of wart removing using cryotherapy and immunotherapy. **Methods:** Support vector machines, core vector machines, random forest, k-nearest neighbours, multilayer perceptron and binary logistic regression was applied on datasets in to create a model that predicted the outcome of an immunotherapy and cryotherapy treatments based on sex, age, time that has passed since last treatment, number of warts, type, area, diameter and result of treatment. **Results:** The average accuracy of the immunotherapy prediction was 88.6%±8.0% while the same measure for cryotherapy prediction was 94.6%±4.0%. The most efficient immunotherapy and cryotherapy model had an accuracy of 100%, predicating the correct treatment outcome when applied to all test cases. **Conclusion:** This study suc- cessfully created a machine learning model that improved the prediction ability of the outcome of immunotherapy and cryotherapy for wart removal. This model created a more in-depth guideline for understanding is immunotherapy would work and took a new approach to cryotherapy. **(Ann Derma- tol 33(4) 345∼350, 2021)**

-Keywords-
Artificial intelligence, Cryotherapy, Immunotherapy, Medical informatics, Warts

INTRODUCTION

The human papillomavirus (HPV) forms part of the papil- lomavirus genus that lies within the Papovaviridae family[1]. Infection with HPV may present in cutaneous warts of var- ious sizes and forms[2]. A wart is characterised as a fleshy, rough, grainy growth that appears most commonly on the hands, fingers or on the soles of the feet. Other types of warts may also appear on the other parts of the body like the face, arms or legs and sometimes grow in the genital or anal area. The two most prevalent types of warts are the common wart (verrucae vulgaris) which typically appear on the hand, and the plantar warts (verrucae plantaris) which mostly presents on the soles of the feet[3-5]. The United Kingdom and the Netherlands report an annual in- cidence of cutaneous warts between 4% and 33%, and between 9.1% and 21.7% of visits to the dermatologist are related to warts. Warts are the most common skin in- fections in children[6]. There is a 2.3% annual incidence of anogenital warts among woman in South Africa and a cor- responding prevalence of 5.7%[7]. It also has been reported that genital warts are a growing concern in Iran[8].

Warts can be an extremely painful condition that may be associated with localised bleeding and discharge. The pain associated with warts may result in the hampered ability to use the limb where the wart presents, e.g., a plantar wart may result in altering ones normal posture or gait which can result in muscle and joint pain. Warts are commonly treated by cryotherapy, which consists of freezing off the wart usually using liquid nitrogen[9-11]. However the treatment has many side-effects, is painful, and many treatment sessions are required[12,13]. Immunotherapy is newer alternative, which is based on the activation of the immune system to deal with the virus and suppress its activity. Side effects of immunotherapy include an elevation of temperature, chills, fatigue, diarrhea, headache, nausea, and vomiting (flu-like symptoms)[14].

Cryotherapy and immunotherapy do not always work and there are no official dermatological guideline that exist that may be used to determine the efficacy of these methods on an individual patient basis[15,16]. However, a first ground-breaking recent study used fuzzy logic to determine rules that can help a physician determine if treatment with one of the two therapies would result in a favourable outcome. This is very advantageous in treatment as it will provide some sort of indication to the physician whether this treatment will produce a favourable outcome or if the physician needs to change treatment altogether. This saves resources in a developing country and also ensures the patient does not endure painful procedures for no reason. However, it may be possible to improve the efficiency of the fuzzy logic algorithm. Machine learning techniques have been used in other domains, but not in the domain of predicting immunotherapy and cryotherapy treatment outcomes.

Thus this study aims at developing a machine learning algorithm that improves the prediction of the outcome of wart removing using cryotherapy and immunotherapy.

MATERIALS AND METHODS

A total of 160 tuples of publically de-identified data was obtained[12]. This data is freely available from the public UCI machine learning repository and has been completely disassociated from any identifiable characteristics of individual patients. Ethics permission was obtained through the Telehealth academic unit at the University of KwaZulu-Natal (BREC/00002203/2020). The dataset consisted of 90 tuples that were associated with patients that underwent cryotherapy and 90 tuples of data associated with patients that underwent immunotherapy. The cryotherapy data consisted of sex, age, time that has passed since last treatment, number of warts, type of wart, area and the result of

treatment. The immunotherapy dataset consisted of sex, age, time that has passed since last treatment, number of warts, type of wart, area, diameter and result of treatment. Table 1 describes the input of the machine learning algorithms in more detail. The output was a binary classification either indicating success or failure of the therapy.

Machine learning

Machine learning is an artificial intelligence technique that tries to create a mathematical model that maps inputs into outputs. There are two parts to machine learning: training, were one applies the principles of a particular machine learning model on data to create the mathematical mapping function; and the testing component, where one tests the predictive ability of the model on data with known outcomes. The dataset was divided into 50% training and 50% testing split. This means that the data used to test the algorithm was not used to train the algorithm, thus giving more reliable results.

Support vector machines (SVM)[17], core vector machines (CVM)[18], random forest (RF)[19], k-nearest neighbours (k-NN)[20], multilayer perceptron (MLP)[21] and binary logistic regression (BLR)[22] was applied on each of the two datasets in order to perform a classification. SVM is a discriminative classifier that works by separating the input space into various dimensions by finding an optimised hyperplane. CVM uses approximation algorithms as a means to create the model that maps the input and output variables together. RF is an ensemble method based of creating decision trees. k-NN is a non-parametric algorithm based on optimising a representation of distance between the input and output variables. MLP is a type of neural network learning algorithm based on the mechanisms of the synapse in the brain. BLR is a probability based algorithm that bases its output on the odds of an output occurring given the input.

Table 1. Describes the features of the dataset used for the machine learning algorithms

Dataset	Feature	Value
Cryotherapy/immunotherapy	Sex	Male
		Female
	Age (yr)	15∼67
	Time (mo)	0∼12
	No. of warts (count)	1∼19
	Type of warts	Common
		Plantar
		Both
	Area (mm^2)	4∼750
	Result	Success
		Failure
Immunotherapy	Diameter (mm)	5∼70

Statistical tests and comparisons

Accuracy, sensitivity, specificity, false-positive rate (FPR), false-negative rate (FNR), true-negative rate (TNR), true-positive rate (TPR), and $F_{measure}$ was calculated in order to determine the efficacy of the machine learning algorithms in predicting the outcome of the cryotherapy and immunotherapy treatment. Accuracy refers to the proportion of times the machine learning algorithm correctly predicted whether the cryotherapy or immunotherapy worked or did not work. Equations 1~8 show how these measures were calculated.

$$Acuuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad (1)$$

$$TPR = \frac{TP}{TP+FN} \qquad (2)$$

$$TNR = \frac{TN}{TN+FP} \qquad (3)$$

$$FPR = \frac{FP}{FP+TN} \qquad (4)$$

$$FNR = \frac{FN}{FN+TP} \qquad (5)$$

$$Sensitivity = \frac{TP}{TP+FN} \qquad (6)$$

$$Specificity = \frac{TN}{TN+FP} \qquad (7)$$

$$F_{measure} = 2 \times \frac{Sensitivity \times Specificity}{Sensitivity + Specificity} \qquad (8)$$

Where TP, FN, FP, and TN represent the number of true positives, false negatives, false positives, and true negatives values respectively.

Z-score and p-score were calculated in order to perform a statistically significant proportion test as per equation 9

and 10. This was calculated to determine if the efficiency measures obtained where statistically different from chance, and also if it is statistically different from other. Chi-squared and ANOVA tests were also performed to determine if there were statistically significant differences between the various output, inputs and classification models, with alpha set at 5%.

$$\Omega = \sqrt{\frac{P \times (1-P)}{n}} \qquad (9)$$

$$Z - Score = \frac{p-P}{\Omega} \qquad (10)$$

Where p refers to the proportion of the measure tested and P is the proportion of the average measure.

RESULTS

Table 2 and 3 describe the accuracy, precision, recall, FPR, FNR, TNR, TPR, and Fmeasure for each of the machine learning algorithms. Also to determine if the prediction results where statistically different from chance, the Z-score was calculated according to equations 9 and 10, with $P=0.5$ i.e. 50% (binary chance). Table 4 shows the average of all the measures obtained for all the machine learning algorithms with the associated Z-scores measuring the probability that the results were obtained by chance. The Z-scores all produced a $p<0.00001$, which indicated that the machine learning algorithms efficiently predicate outcome of treatment and it is not attributed to chance.

In order to compare the algorithms to each other, the average of each effectivity measure was determined. This study resulted in the creation on an effective model for the prediction of the outcome of cryotherapy and immunotherapy treatment for the removal of warts. High accuracy,

Table 2. Shows the efficacy of the machine learning algorithms in predicting Immunotherapy outcome

AI method	Accuracy	Precision	Recall	FPR	FNR	TNR	TPR	F
SVM	79.0	100	78.9	0	21.1	96.1	98.3	88.2
CVM	95.6	100	94.7	0	5.3	100	97.1	97.3
RF	100	100	100	0	0	100	100	100
k-NN	83.0	100	82.6	0	17.5	100	98.2	90.4
MLP	89.0	98.6	88.6	9.1	11.4	90.9	95.9	93.3
BLR	85.0	100	84.5	0	15.5	100	98.1	91.6

Values are presented as percentage. AI: artificial intelligence, FPR: false-positive rate, FNR: false-negative rate, TNR: true-negative rate, TPR: true-positive rate, SVM: support vector machines, CVM: core vector machines, RF: random forest, k-NN: k-nearest neighbours, MLP: multilayer perceptron, BLR: binary logistic regression.

Table 3. Shows the efficacy of the machine learning algorithms in predicting cryotherapy outcome

AI method	Accuracy	Precision	Recall	FPR	FNR	TNR	TPR	F
SVM	92.2	87.5	97.7	12.8	2.3	87.4	88.3	92.3
CVM	97.8	97.9	97.9	2.4	2.1	97.6	99.1	97.9
RF	100	100	100	0	0	100	100	100
k-NN	93.3	93.8	93.8	7.1	6.3	92.9	93.9	93.8
MLP	93.3	91.7	95.7	9.1	4.3	90.9	91.9	93.6
BLR	91.1	87.5	95.5	13.0	4.5	87.0	88.1	91.3

Values are presented as percentage. AI: artificial intelligence, FPR: false-positive rate, FNR: false-negative rate, TNR: true-negative rate, TPR: true-positive rate, SVM: support vector machines, CVM: core vector machines, RF: random forest, k-NN: k-nearest neighbours, MLP: multilayer perceptron, BLR: binary logistic regression.

Table 4. Shows the average of the efficiency measures obtained for all the models and the Z-score when testing against the models predicting outcomes based purely on chance

Dataset	Feature	Accuracy	Precision	Recall	FPR	FNR	TNR	TPR	F
Immunotherapy	Average	88.6%	0.99	0.88	0.018	0.12	0.98	0.93	0.99
	Z-score	7.403*	6.56*	8.36*	−6.53*	−8.36*	6.53*	7.42*	6.57*
Cryotherapy	Average	94.6%	0.93	0.96	0.074	0.032	0.93	0.95	0.93
	Z-score	8.56*	6.58*	8.36*	−6.53*	−8.36*	6.53*	7.42*	6.58*

FPR: false-positive rate, FNR: false-negative rate, TNR: true-negative rate, TPR: true-positive rate. *$p < 0.0001$.

sensitivity, specificity, TNR, TPR, and $F_{measure}$ were obtained for all the models created. This was coupled with low FPR and FNR which further indicated the efficiency in predicting treatment outcome. The average accuracy of the immunotherapy prediction was 88.6%±8.0% while the same measure for cryotherapy prediction was 94.6%±4.0%. High average precision and recall values were also obtained by the methods that predicted the outcomes of both the immunotherapy and cryotherapy treatments. The average precision of the immunotherapy prediction was 0.99± 0.01 while the same measure for cryotherapy prediction was 0.93±0.05. The closer the value is to 1, the more effective the model is. These results indicate that the models are correct 99.1% of the time when predicting that immunotherapy will be successful. The results also indicate the models are correct 93% of the time when they predict that cryotherapy will be successful.

The average recall of the immunotherapy prediction was 0.88±0.08 while the same measure for cryotherapy prediction was 0.96±0.02. The closer the value is to 1, the more effective the model is. These results indicate that the immunotherapy model identified 88% of all the cases that should be classified as successful and that the cryotherapy model identified 86% of all the cases that should be classified as successful.

The low false positive and false negative ratios for both the immunotherapy and cryotherapy models also indicate the efficiency of the models. The results indicate less than 6.3% of the time, the models return incorrect results. Only 12.0%±8.0% of the time does the models indicate that immunotherapy will not work, when it actually would work; and conversely only 1.2%±4.0% of the time does the models predict that immunotherapy will work when it would not work. Even lower rates were returned with the cryotherapy models. Only 3.2%±2.0% of the time does the models predict that cryotherapy will not work, when it actually would have worked; and conversely only 7.4%± 5.0% of the time does the models indicate that cryotherapy will work when it would not have worked.

The high true positive and true negatives rates for all the models indicate that the machine learning techniques cor-

rectly predicts when each of the treatments will work and when it would not. All the results obtained indicate the machine learning models created are efficient in predicting the treatment outcomes of immunotherapy and cryotherapy for wart removal.

A chi-squared test comparing the different machining learning models based on accuracy alone resulted in a $p < 0.00001$. An ANOVA test using all the efficiency measures also resulted in a $p < 0.00001$. This indicates that there is a statistical difference in predicting outcome of therapy between the models. The RF model obtained the highest values for all the efficiency measures and is the only model whose value was more than one standard deviation away from the average value for each measure. Thus it is concluded that the RF model outperformed all the other models.

DISCUSSION

It is important to compare the results obtained by this study to literature, in order to understand the efficiency of these machine learning models. A recent study showed that when feature selection fuzzy logic technique is applied to this dataset, only three features are important for immunotherapy: time, diameter and type. Age, time, type and area were the important variables when predicating the outcome of cryotherapy[12]. The study reported an average accuracy of 83.33%±6.02% and 80.00%±5.23% respectively. At 95% confidence interval, an accuracy that is greater that the upper ranger of the reported fuzzy logic study average accuracy plus 1.96 times standard deviation will prove that the machine learning method outperforms the reported fuzzy learning technique. For immunotherapy the upper range of the fuzzy logic accuracy is 83.33%+ 1.96% (6.02%)=95.1% and similarly for cryotherapy the upper range is 90.3%. The results of this study exceeded these upper range values, indicating that the models developed in this study is statistically more accurate than those reported in literature by Khozeimeh et al.[12] Fig. 1 and 2 represent the guideline for assessing the success or failure of immunotherapy and cryotherapy respectively.
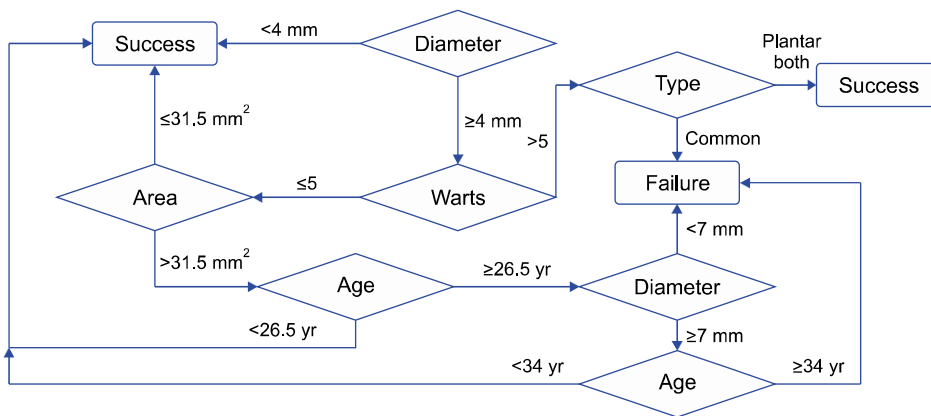
Fig. 1. Shows the guideline generated by the random tree machine learning model to predict the outcome of immunotherapy of wart removal.
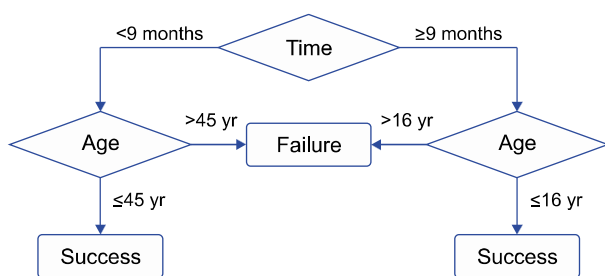

Fig. 2. Shows the guideline generated by the random tree machine learning model to predict the outcome of cryotherapy of wart removal.

Comparison of immunotherapy classification with Khozeimeh et al. fuzzy logic rules

Khozeimeh et al.[12] based the success of the immunotherapy treatment on three rules. The first simply states that if the diameter of the wart is less than 7 mm, then the immunotherapy will be successful. When applied to the dataset, this rule results in 37% of cases being misclassified. All these errors were correctly classified when the rules of this study was applied. The model described in Fig. 1 is more complicated than the previously reported rules. It states that is the diameter of the wart is less than 4 mm, then the Immunotherapy will be successful. However, unlike Khozeimeh et al.[12] model, if the diameter is greater than 4 mm, other factors affect the outcome of the therapy. If there are less than 5 warts and the area covered by the warts is less than 31.5 mm^2 or the patient is younger than 26.5 years old, then the therapy will be successful. However, older patients with fewer than seven warts lead to therapy failure.

The second rule states there is a positive treatment response if it is a plantar wart and time elapsed is greater than 6 months. When applied to the dataset it resulted in 7% of cases being misclassified. All these errors were correctly classified when the rules of this study was applied.

This study's model indicates that if there are more than 5 plantar warts and a combination of common and plantar warts, then the treatment will be successful. Failure is outcome of the therapy, when the number of common warts are greater than 5.

The third rule stated that if the time before treatment was less than 6 months then the treatment will work. However applying this rule, results in 21% of cases being misclassified. All these errors were correctly classified when the rules of this study was applied. This study model did not place the same emphasis on time for immunotherapy outcome. This study also produced a model that predicted both treatment success and failure unlike the Fuzzy logic model that did not include a pathway that may result in failure with regards to immunotherapy.

Comparison of cryotherapy classification with Khozeimeh et al. fuzzy logic rules

Khozeimeh et al.[12] based the success of the cryotherapy treatment on a few rules. The first simply states that if the warts is both common and plantar then cryotherapy does not work; however, if is a plantar wart where the time elapsed is greater than 6 months, then the therapy is responsive. However, analysing the data resulted in 15% of cases that are both common and plantar but the therapy works. Similarly, 33% of cases where found where plantar wart with a time elapsed greater than 6 months resulted in non-responsive therapy. All these errors were correctly classified when the rules of this study was applied.

The second rule stated that if the time before treatment was less than 6 months then the treatment will work. However applying this rule, results in 21% of cases being misclassified. All these errors were correctly classified when the rules of this study was applied. The third rule states there is a treatment response if it is a plantar wart and time elapsed is greater than 6 months and when applied it resulted in 7% of cases being misclassified. All

these errors were correctly classified when the rules of this study was applied. The model shown in Fig. 2 shows a very different approach to cryotherapy outcome prediction. Like Khozeimeh et al's model[12], age and time played an important role; but type of wart and time elapsed played no role in this study's model.

This study successfully created a machine learning model that improved the prediction ability of the outcome of immunotherapy and cryotherapy for wart removal. The model correctly classified all the cases that the previously published fuzzy logic model incorrectly classified.

This study will be extended by using other machine learning algorithms and performing feature selection and more in-depth data pre-processing. Also, it will be very beneficial to validate the study in a clinical setting.

CONFLICTS OF INTEREST

The author has nothing to disclose.

FUNDING SOURCE

None.

DATA SHARING STATEMENT

This data is freely available from the public UCI machine learning repository (https://archive.ics.uci.edu/ml/datasets/ Immunotherapy + Dataset).

REFERENCES

1. Gerlero P, Hernández-Martín Á. Treatment of warts in children: an update. Actas Dermosifiliogr 2016;107:551-558.
2. Leung L. Recalcitrant nongenital warts. Aust Fam Physician 2011;40:40-42.
3. Bavinck JN, Eekhof JA, Bruggink SC. Treatments for common and plantar warts. BMJ 2011;342:d3119.
4. Buttaravoli P, Leffler SM. Warts: (common wart, plantar wart). In: Buttaravoli P, Leffler SM, editors. Minor emergencies. 3rd ed. Philadelphia: Elsevier Saunders, 2012:752-755.
5. Esterowitz D, Greer KE, Cooper PH, Edlich RF. Plantar warts in the athlete. Am J Emerg Med 1995;13:441-443.
6. Bruggink SC, Eekhof JA, Egberts PF, van Blijswijk SC, Assendelft WJ, Gussekloo J. Warts transmitted in families and schools: a prospective cohort. Pediatrics 2013;131:928-934.
7. Chikandiwa A, Kelly H, Sawadogo B, Ngou J, Pisa PT, Gibson L, et al. Prevalence, incidence and correlates of low risk HPV infection and anogenital warts in a cohort of women living with HIV in Burkina Faso and South Africa. PLoS One 2018;13:e0196018.
8. Ramezanpour A, Fallah R. Genital warts in North-West of Iran. World Appl Sci J 2012;18:1326-1328.
9. Bertolotti A, Dupin N, Bouscarat F, Milpied B, Derancourt C. Cryotherapy to treat anogenital warts in nonimmunocompromised adults: systematic review and meta-analysis. J Am Acad Dermatol 2017;77:518-526.
10. Uyar B, Sacar H. Comparison of cryotherapy session intervals in the treatment of external genital warts. Dermatol Sin 2014;32:154-156.
11. Zouboulis VA, Zouboulis CC. Small cryotherapy devices for the treatment of skin warts. J Dermatolog Treat 2017;28:745-750.
12. Khozeimeh F, Alizadehsani R, Roshanzamir M, Khosravi A, Layegh P, Nahavandi S. An expert system for selecting wart treatment method. Comput Biol Med 2017;81:167-175.
13. Izadi Firouzabadi L, Khamesipour A, Ghandi N, Hosseini H, Teymourpour A, Firooz A. Comparison of clinical efficacy and safety of thermotherapy versus cryotherapy in treatment of skin warts: a randomized controlled trial. Dermatol Ther 2018;31:e12564.
14. El-Khalawany M, Shaaban D, Aboeldahab S. Immunotherapy of viral warts: myth and reality. Egypt J Dermatol Venerol 2015;35:1-13.
15. Hogendoorn GK, Bruggink SC, Hermans KE, Kouwenhoven STP, Quint KD, Wolterbeek R, et al. Developing and validating the cutaneous WARTS (CWARTS) diagnostic tool: a novel clinical assessment and classification system for cutaneous warts. Br J Dermatol 2018;178:527-534.
16. Micali G, Lacarrubba F. Standardized classification tools in dermatology: time to focus on cutaneous warts. Br J Dermatol 2018;178:330.
17. Liu L, Chu M, Gong R, Peng Y. Nonparallel support vector machine with large margin distribution for pattern classification. Pattern Recognit 2020;106:107374.
18. Afzal AL, Asharaf S. Deep multiple multilayer kernel learning in core vector machines. Exp Syst Appl 2018;96:149-156.
19. Katuwal R, Suganthan PN, Zhang L. Heterogeneous oblique random forest. Pattern Recognit 2020;99:107078.
20. Pan Z, Wang Y, Pan Y. A new locally adaptive k-nearest neighbor algorithm based on discrimination class. Knowl-Based Syst 2020;204:106185.
21. Zhang F, Sun K, Wu X. A novel variable selection algorithm for multi-layer perceptron with elastic net. Neurocomputing 2019;361:110-118.
22. Arabameri A, Pradhan B, Lombardo L. Comparative assessment using boosted regression trees, binary logistic regression, frequency ratio and numerical risk factor for gully erosion susceptibility modelling. CATENA 2019;183:104223.