International Journal of Innovative Technology and Interdisciplinary Sciences
www.IJITIS.org
ISSN:2613-7305
Volume 3, Issue 3, pp. 467-473, 2020
DOI: https://doi.org/10.15157/IJITIS.2020.3.3.467-473
Received July 10, 2020; Accepted August 31, 2020

Investigation of Factors Affecting Immunotherapy Treatment Results by Binary Logistic Regression and Classification Analysis

Burcu Durmuş[a1], Öznur İşçi Güneri[*b1]
[1]College Department of Statistics
Muğla Sıtkı Koçman University, 48000 Turkey
[a]burcudurmus@mu.edu.tr; *[b]oznur.isci@mu.edu.tr.

ABSTRACT

There are many factors that affect the success of immunotherapy treatment. In addition to the clinical examination, the investigation of these factors by different methods contributes to the researchers on prior knowledge and time. In this study, it was aimed to evaluate the application of logistic regression and data mining methods to evaluate the success of post-immunotherapy treatment. Bilateral logistic regression analysis with WTA and classification analysis with Weka were used to evaluate whether immunotherapy treatment was successful for warts. Decision tree structure is also discussed to determine the variables that affect classification success. According to the logistic regression result, the model is important because the probe. 0.022 <0.05. The classification result for the logistic regression model was calculated as 85.56%. This result shows that the model is successful. Data mining experiments were carried out with different classification algorithms. The best result was found in decision trees (with J48 algorithm) with 85.56% accuracy rate. According to the J48 algorithm decision tree structure, the variables that affect the outcome of the treatment were recorded as time, number of warts and age, respectively. Study results show that both methods yield parallel results. Decision tree algorithm is used as an alternative to classical statistical models. In particular, in cases where clinical research is limited, it will benefit researchers on topics such as transition to analysis, preliminary information gathering, time and effort.

Keywords: Immunotherapy, warts, generalized models, machine learning, statistical modelling.

1. INTRODUCTION

Common warts are an infection caused by the HPV virus [1]. Although it appears anywhere on the body, it is mostly seen in areas such as hands, feet, and face. It can easily spread from any part of the body to other areas. Warts can be of various sizes, colours and shapes depending on their region and type. Round, yellowish warts that appear around the hands, legs or nails are called verruca vulgaris. The warts that are buried in the skin in the feet are known as verruca plantaris and this species may be confused with callus. Small and smooth warts are verruca. They can appear anywhere on the body. Verruca acuminate in the genital area is another genetically transmitted species [2]. Warts are lesions that spread very quickly. The number of warts seen in a small area increases in many areas in a short time [3]. Therefore, it is important to treat the wart immediately after it is detected. These benign lesions have different known treatment methods. The preferred method can vary depending on the color of the wart, the area it is located, its number, type. Drugs containing salicylic acid or formic acid, cryotherapy, surgical treatment are the most used methods [4]. Besides these methods,

immunotherapy is a new treatment method used recently [5]. Despite the treatment methods mentioned above, there is no single approved method for the treatment of warts [6]. Doctors often try to guess which treatment has a better effect on a particular patient [7]. Reasoning used in medical diagnosis is similar to the functioning of machine learning algorithms. Machine learning is a statistical learning series that aims to analyze data by revealing meaningful relationships in the data [8]. Classification analysis is one of the methods discussed in machine learning. It is known that classification algorithms are used in medical science to examine the data obtained as a result of clinical examination [9]. These methods are often used in the study of disease outcomes.

Decision trees have been used in a study in which immunotherapy estimation designed to assist the doctor is a plantar or diffuse wart. It has been suggested that reducing the features taken into account to reveal the most relevant features can significantly speed up the calculation of similarity [7]. Another study focused on improving predictive accuracy of J48, a binary decision tree-based classifier, by adding genetic programming-based attributes. The genetically adjusted attribute structure not only upgraded the classification capabilities of the J48 classifier, but also expanded its knowledge area, making J48 give more precise predictions for defining a wart treatment method [10]. In Akben's study, a decision tree based method was used to determine the rules for predicting the success of wart treatment methods. According to the findings, the success rate varies between 90-95% according to the treatment method. In addition, the established decision tree rules can be converted into images to interpret the success rates of treatment methods as a function of the patient's age and visually interpret the time since the onset of the disease. This study provides a method for simple and more reasonable interpretation of the rules for medical professionals [11]. Talabani and Avcı, on the other hand, showed the effects of the Support Vector Machines in different kernel functions in order to increase the learning capacity in the treatment of warts. According to the results obtained, Polynomial Kernel and Pearson VII function based Universal Kernel were found to perform the best classification performance with an accuracy of 81.11% [12].

In this study, variables that affect the success of immunotherapy treatment with decision tree methods were tried to be determined and model reliability criteria were examined. In addition to machine learning method, classification was made by binary logistic regression analysis and the results were compared. The research generally consists of 3 main sections apart from the introduction. The methods used in the material and method section are mentioned. In the conclusion, logistic regression and classification analysis results are explained. Classification success for logistic regression analysis and J48 algorithm was found as 85.56%. This result is higher than other algorithm results. In the conclusion part, it is discussed that statistical methods and machine learning methods give similar results. With this research, it shows that choosing more successful methods to get the best result can be examined with both methods.

2. MATERIAL AND METHODS

The materials and methods used in the study are explained with the flow diagram in Figure 1 [5].

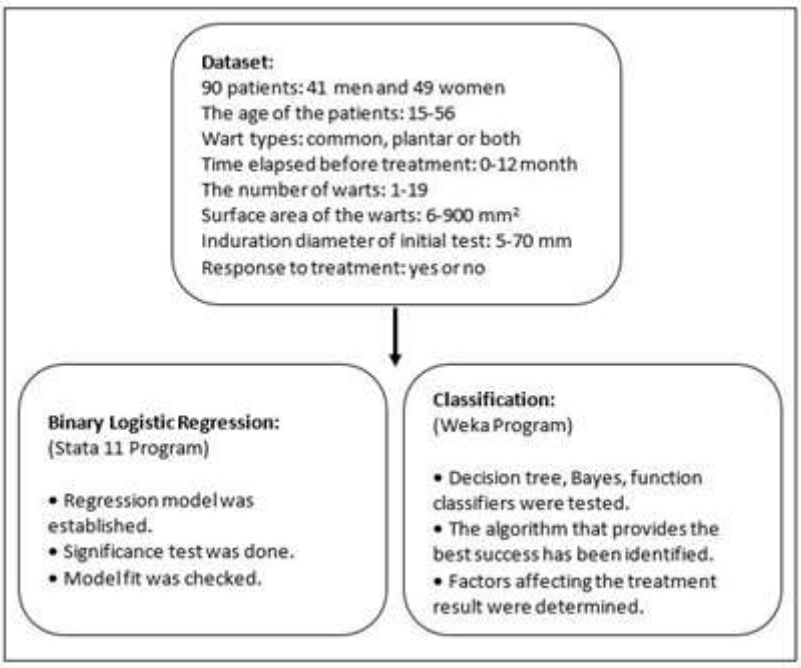

Figure 1: Flowchart of material and methods

3. RESULTS

The materials and methods used in the study are explained with the flow diagram in Figure 1 [5].

3.1 Demography

In this study, the data set collected by Khozeimeh et al. was used [5] [13]. A total of 90 patients were selected, 41 men and 49 women, to be treated with immunotherapy. The average age of patients in this group is 31.04 ± 12.23 (range: 15-56). Patients had warts on average 7.23 ± 3.10 months (range: 0-12) before treatment. Average number of warts is 6.14 ± 4.2 (range: 1-19). Wart types are observed as common, plantar and both. The surface area of the warts was measured as 95.7 ± 136.61 (range: 6-900) mm2 and induration diameter of initial test was found to be 14.33 ± 17.22 mm (range: 5-70). The patients were treated with immunotherapy by candida antigen method. Treatment of patients took three sessions with intralesional vaccine injection, and there was a three-week period between sessions. The response to treatment was recorded as yes or no.

3.2 Binary Logistic Regression Analysis

In Variables affecting immunotherapy treatment results can be determined by regression model. Here, the treatment answer is categorically given as yes and no. Other variables are numerical or categorical. Due to the logistic regression definition and the number of categories of the dependent variable (treatment response), binary logistic regression analysis was applied with the help of (two) Stata 11 programs [14].

The model is said to be significant because the result of the binary logistic regression prob. is 0.022 < 0.05. At least one variable has a significant impact on the model. When the independent variables (sex, age, time, number of warts, type, area, diameter induction) are examined, only the 'time' variable contributes significantly to the model. The most important advantage of logistic regression is that it allows the interpretation of odds ratios. According to the results in the output; women are 1.14 times more likely to respond to treatment than men. 1 unit increase in age reduces the positive response to

the treatment 1.03 times, 1 unit increase in time 1.42 times, 1 unit increase in the number of warts 1.44 times and 1 unit increase in the diameter of the induction 1.02 times. The type and area of the wart affects the response to treatment equally. When classifying with logistic regression, 77 observations were classified correctly and 13 observations were classified incorrectly. It is said that the correct classification rate is 85.56% and the model makes a good estimate, see Table 1.

Table 1: Model analysis and classification results for logistic regression.

Logistic Regression Analysis

Logistic regression				Number of obs	=	90
				LR chi2(7)	=	16.33
Log likelihood =-38.221179				Prob> chi2	=	0.0222
				Pseudo R2	=	0.1761
	Odds Ratio	Std. Error	z	P> |z|	[95% Conf.	Interval]
sex	1.14	0.66	0.24	0.81	0.36	3.57
age	0.97	0.02	-1.11	0.26	0.93	1.02
time	0.70	0.08	-2.95	0.00	0.55	0.88
number of warts	0.69	0.27	-0.91	0.36	0.31	1.51
type	1.00	0.39	0.01	0.99	0.46	2.18
area	1.00	0.00	0.29	0.77	0.99	1.00
induration diameter	0.98	0.01	-0.56	0.57	0.95	1.02

Classification Results

Classified	D	~D	Total
+	70	12	82
-	1	7	8
Total	71	19	90

Sensitivity	Pr(+| D)	98.59%
Specificity	Pr(-|~D)	36.84%
Positive predictive value	Pr(D| +)	85.37%
Negative predictive value	Pr(~D| -)	87.50%
False + rate for true ~D	Pr(+|~D)	63.16%
False - rate for true D	Pr(-| D)	1.41%
False + rate for classified +	Pr(~D| +)	14.63%
False - rate for classified -	Pr(D| -)	12.50%
Correctly classified		85.56%

3.3 Classification Analysis

Another way to analyze the response to the treatment outcome is data mining. Data mining methods can be determined according to the data and the statistical results of the model evaluated. The most commonly used method in data mining is classification analysis. There are many subfolders such as decision tree, Bayes, function, meta, rules in classification analysis. In this study, different classification analyzes (decision tree,

Bayes and function) were investigation with the help of Weka program to evaluate the response to the treatment [15].

In the classification made with decision trees, the correct classification rate was calculated as 85.56% with J48 algorithm, 83.33% with LMT, and 84.44% and 77.78% with Random Forest and Random Tree, respectively. For Bayes classifiers, 83.33% with Bayes Net, 78.89% with Naive Bayes and for function classifiers, 76.66% with logistics and 80% with SMO. The J48 algorithm with the highest classification rate was selected for detailed statistical comments. Considering other results of the algorithm; kappa statistic (0.50) is of medium value. Mean absolute error is 0.21, root mean square error is 0.3702, relative absolute error is 62.2505 and root relative square error is 90.643. According to the complexity matrix, 13 out of 90 observations were classified incorrectly and 77 correctly. This result is compatible with the classification rate obtained with binary logistic regression. In addition, when the decision tree diagram of the J48 algorithm is examined, it is seen that the main factor determining the response to treatment is 'time'. Other factors that affect response to treatment are 'number of warts' and 'age', respectively, see Table 2.

Table 2. Results of classification with datamining.

Algorithm	**Classification rate**	**Correctly classified**	**Misclassified**
Decision Trees:			
J48	85.56%	77	13
LMT	83.33%	75	15
Random Tree	84.44%	76	14
Random Forest	77.79%	70	20
Bayes:			
Bayes Net	83.33%	75	15
Naive Bayes	78.89%	71	19
Functions:			
Logistic	76.67%	69	21
SMO	80.00%	72	18

J48 Algorithm Results

	1	0	Total
1	9	10	19
0	3	68	71
Total	12	78	90

Kappa statistic	0.50
Mean absolute error	0.21
Root mean squared error	0.37
Relative absolute error	62.25%
Root relative squared error	90.65%

J48 Algorithm Decision Tree

time <= 10.25: 1
time > 10.25
| warts = 0

```
|   |   age <= 27: 1
|   |   age > 27: 0
|   warts = 1: 0
|   warts = 2: 0
```

4. CONCLUSION

This study as far as is known is the first study to evaluate the results of whether the results of immunotherapy treatment are positive by logistic regression and data mining classification methods. It is very difficult to detect factors affecting the results of immunotherapy treatment outside of clinical examinations. In this study, it was focused on the determination of the factors affecting the immunotherapy treatment result by classification analysis. Two different statistical methods were used. In binary logistic regression analysis, the model was found to be significant (probe = 0.022 <0.05). This result shows that at least one of the variables contributes significantly to the model. It was revealed that the 'time' variable from the variables included in the model has a significant contribution. When the classification results for logistic regression were examined, a success of 85.56% was achieved. Different classifiers were used for classification with data mining: decision tree, Bayes, function. It was found as 85.56% with J48 algorithm, which is the best performance decision tree method. This result is similar to the one found in binary logistic regression analysis. However, when the decision tree of the J48 algorithm is examined, it is seen that the variables 'wart number' and 'age' besides the 'time' variable are also effective in classification. For decision trees of other algorithms: Variables other than sex in the LMT algorithm and all variables in the Random Tree algorithm are effective in classification. Decision tree structures of these algorithms have been examined, but they have been ignored since they are lower than the J48 algorithm results.

The exact factors that determine the results of wart treatment are unknown. For this reason, the best wart treatment method is still being researched by doctors [16]. This study shows that data mining methods are important to identify factors that affect treatment outcomes, or at least to predict results or obtain prior knowledge. However, more research and analysis are needed to identify factors affecting treatment outcomes.

CONFLICT OF INTERESTS

The authors would like to confirm that there is no conflict of interests associated with this publication and there is no financial fund for this work that can affect the research outcomes.

REFERENCES

[1] Nammas IN, Sourvinos G and Spandidos DA. Human Papilloma Virus (HPV) Infection in Children and Adolescents. *Eur J Pediatr*. 2009; 168; 267–273.

[2] Bohlooli S, Mohebipoor A, Mohammadi S, Kouhnavard M and Pashapoor S. Comparative Study of Fig Tree Efficacy in The Treatment of Common Warts (Verruca Vulgaris) vs. Cryotherapy. *International Journal of Dermatology*. 2007; 46; 524–526.

[3] Jones O. Warts. Megan Publishing Services. 2020;12.

[4] Abeck D, Larissa T, Lüft M and Biedermann T. Extragenital Cutaneous Warts – Clinical Presentation, Diagnosis and Treatment. *Journal of German Society of*

Dermatology. 2019; 613-634.
[5] Khozeimah F, Azad FJ, Oskouei YM, Jafari M, Tehranian S, Alizadehsani R and Layegh P. Intralesional Immunotherapy Compared to Cryotherapy in Thetreatment of Warts. *International Journal of Dermatology*. 2007; 56; 474–478.
[6] Khozeimah F, Alizadehsani R, Roshanzamir M, Khosravi A, Layegh P, and Nahavandi S. An Expert System for Selecting Wart Treatment Method. *Computers in Biology and Medicine*. 2017; 81; 167–175.
[7] Saadi F, Atmani and Henni F. Integration of Datamining Techniques into the CBR Cycle to Predict the Result of Immunotherapy Treatment. International Conference on Computer and Information Sciences (ICCIS). 2019.
[8] Olson DL and Lauhoff G. Descriptive Data Mining. Computational Risk Management. Springer, Singapore. 2019.
[9] Dwivedi AK. Performance Evaluation of Different Machine Learning Techniques for Prediction of Heart Disease. *Neural Comput & Applic.* 2018; 29; 685–693.
[10] Sabita Khatri S, Arora D and Kumar A. Implicating Machine Learning towards Development of Intelligent System for Wart Treatment Therapy Identification. *International Journal of Engineering & Technology*. 2018; 7; 754–758.
[11] Akben SB. Predicting the Success of Wart Treatment Methods Using Decision Tree Based Fuzzy Informative Images. *Biocybernetics and Biomedical Engineering*. 2018; 30; 819–827.
[12] Talabani H and Avcı E. Impact of Various Kernels on Support Vector Machine Classification Performance for Treating Wart Disease. International Conference on Artificial Intelligence and Data Processing (IDAP). 2019.
[13] https://archive.ics.uci.edu/ml/index.php, Date of Access: 01.02.2020.
[14] https://www.stata.com/stata11/, Date of Access: 05.04.2019.
[15] https://www.cs.waikato.ac.nz/ml/weka/, Date of Access: 28.09.2018.
[16] Alizadehsani R, Jalali SMJ, Roshanzamir M, Khosravi A and Nahavandi S. Comparing the Performance of Feature Selection Algorithms for Wart Treatment Selection. Proceedings of The International Workshop on Future Technology FUTECH. 2019;16-18.