

Applications of data mining in predicting the stability of Vitiligo

Gagandeep Singh^{1*}, Kavita Rathi²

^{1,2}Dept. of Computer Science and Technology, Deenbandhu Chhotu Ram University of Science and Technology, India

*Corresponding Author: gagandeep1661@gmail.com, Tel.: +91-8802114770

DOI: <https://doi.org/10.26438/ijcse/v7i8.7073> | Available online at: www.ijcseonline.org

Accepted: 10/Aug/2019, Published: 31/Aug/2019

Abstract— Vitiligo is growing at a good speed among the population and people have to go repeated surgeries to get rid of this disease. Though it's not easy to define the stability, but it's indispensable in the treatment of vitiligo. There have been many cases where people had gone for skin replacement surgery, but after sometime, white patches redeveloped on the skin. So the treatment goes on forever and patients get disheartened. The aim is to help people to identify the saturation of the disease before seeking the remedy which is skin transplantation. In this paper, improved J48 algorithm is used to predict the stability of vitiligo which gives optimal results. This algorithm uses the medical history of patients, Koebner phenomenon and VIDA score of sample data to feed into the systems and draw patterns to predict stability in the patients. We use various algorithms of data mining to extract useful information from data and check the accuracy of their medical history. The data includes the vitiligo patients, healthy people, the ones who have undergone surgery and the patients who haven't undergone skin replacement and are still experiencing growth in their patches. With the prediction of various parameters, an optimal target value is predicted. In the end, we conclude with the most optimal algorithm which can be used to determine the stability of this disease and help the doctors and patients to determine the precise time of surgery.

Keywords— J48 algorithm, Vitiligo, White patches, Patch development, data mining, prediction

I. INTRODUCTION

In the problem of vitiligo, the growing patches of the skin give up their original colour and this becomes long-term. Any age-group can be affected by it irrespective of gender or ethnic group. There's a substance within the skin, known as melanocytes which are meant for the colour of the skin pigment. These eventually die off and make the patches appear on the skin. So there's no layer which maintains the colour of the skin and protects it from the sun's UV rays. There are approximately 0.5 to 2 percent of people affected by vitiligo the world over.

Facts about Vitiligo:

- There is no bar of age, gender or ethnicity which could be invulnerable to this disease[1]
- The condition persists for the whole life without treatment
- It's uncertain to define the prominent cause of vitiligo, though an autoimmune disorder or some virus can play a significant role[2]
- The exact cause is unknown, but it may be due to an autoimmune disorder or a virus
- Vitiligo can't be spread from one person to another
- Various skin specialists choose a variety of procedures in its treatment which includes exposure to UVA or

UVB light and removing skin pigments in critical cases[3]

As illustrated, the specific cause of this disease is hard to define but there are many factors which contribute to it. These are listed below:

- Immune system itself attacks on melanocytes and the condition can be termed as autoimmune disorder
- a genetic oxidative stress imbalance
- a stressful event
- A severe sunburn might cause harm to the pigmentation of skin
- A physical injury, like a cut
- a neural cause
- It passes on in genes, so heredity can be a determining factor
- a virus



Fig. 1

Stability in Vitiligo:

We can't certainly define stability when the condition is settled while it becomes indispensable in the diagnosis planning of regimens and also in predicting the condition of the patient. Some interesting points are:

- Stability becomes the deciding factor when the dermatologist plans for any transplantation technique of vitiligo[4]
- There's no transparency in defining this concept and is way beyond doubt
- It could be misleading to increase dependency on KpH or TG, though these two levels reveal the stability at clinical levels[5]
- While at molecular level, the picture can be altogether different from the clinical diagnosis
- Surgical treatments highly rely on the stability of the disease[6]

Parameters of assessment of Stability:

The below points discuss the points which assess the stability:

- **History**
 - In unilateral vitiligo, old lesions would be seen with diminishing effects within the past two years while if we look at the other case of bilateral vitiligo, it could take several years to establish stability
 - New lesions will stop developing anymore[7]
- **Koebner phenomenon**
 - Absence of recent Koebner phenomenon, i.e., depigmentation of skin at the sites of a specifically traumatized skin
- **VIDA score**
 - This score grades the degree of activity of disease in vitiligo patients individually off any therapies for at least 6 months

II. RELATED WORK

- The clinical approach and quantitative approach doesn't match in most cases[8]
- The researcher doesn't consider ground truth of medical images[9]
- The algorithm results are sometimes far away from actual medical condition
- The pigmentation degradation is not considered in most of the papers[10]

Min Chen (2017) initiated a technique known as data mining, which is a way of innovating a new pattern of

information from a good amount of data. Patterns are identified to transform the irrelevant data into sensible information. The information which is collected is further fed into various classifiers, for example, decision trees, KNN or SVM where every step is recognized for a deterministic task. These techniques are used to predict the skin diseases at an early stage. If such diseases can be predicted at an early stage, proper preventive steps could be taken to reduce the chances.

Priyanka N B.E., et. al (2017) talked about two ways of data mining that is decision trees and Naive Bayes for prediction of chronic diseases at an early stage. This strategy could be used where the resources are deficient in medical and health region. The study of this paper concludes that both the methods are useful in different fields, can be cost effective and accurate at times.

Koushik Lahiri (2014) cited the concept of stability of vitiligo. It's crucial to know when this disease is stable to start the transplantation technique to treat it. This paper looked into the minimum span of stability in different studies. Three indicators were emphasized which determine the stability: History, Koebner phenomenon and test grafting. This paper bridges a big gap between the medical reasons and the prediction of severity and stability of this disease.

Davinder Parsad (2013) discussed algorithms to predict the stability of vitiligo. Various biochemical parameters are defined which affect the severity of this disease. A pattern has been drawn from all the contributing factors. With the various decision trees, it's shown that medical history of skin diseases and the size of the patches can be detrimental in predicting the stability.

III. METHODOLOGY

Data mining is identifying patterns and information with a huge amount of data sets. Associative Rule and Cluster Analysis are its main aspects. Different rules and techniques have an indispensable role in this process[11]. The testing and training are two steps in the learning process. With the huge data, machine is made to identify the patterns and get trained[12]. Thereafter, a test information is used to check how well the machine has learnt.

- Using a computational approach to predict the stability of vitiligo
- Understanding the assessment factors and quantify their impacts[13]
- Devising an improvement in J48 algorithm
- Implementing it using WEKA wrapper in Python environment

Vitiligo affects a notable portion of the population. There are various stages of this disease. The treatment can bear good

results if the stability of the disease is observed[14]. A quantitative approach using data mining would be very useful to detect the stability and start treating the patients. This would minimize the doctor’s efforts and the patient’s disease could be cured without letting the case to go out of control[15].

Basic steps in the algorithm:

(i) Data mining is identifying patterns and information with a huge amount of data sets. Associative Rule and Cluster Analysis are its main aspects. Different rules and techniques have an indispensable role in this process. The testing and training are two steps in the learning process.

(i) If the instances are related to the same class, the tree represents a leaf so the leaf is returned back by labelling with the same class[16].

(ii) The possible information is calculated for all the attributes, given by a test on the attribute. Then the increase in information is taken out that would result from a test on the attribute[17].

(iii) Then the most appropriate attribute is found on the basis of the current selection criteria and that attribute which is used for branching[18].

(iv) J48 algorithm utilises pruning method to create decision tree branches as well as leaves. Therefore, it reduces the complexity and enhances the performance of classification.

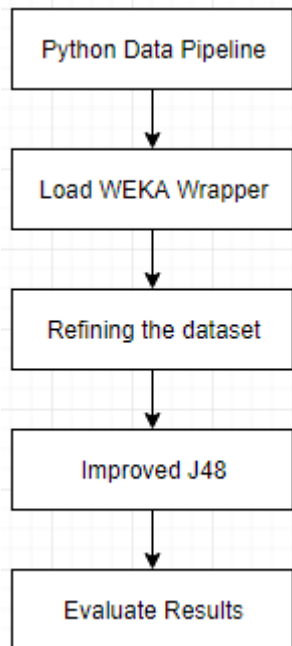


Figure 2 - Proposed Setup

Features of the Algorithm:

(i) J48 algorithm is a decision tree created based on ID3 (Iterative Dichotomiser 3) by Weka after C4.5 Decision tree

begins with database (training set), then by recursive decrease, it segregates the data into lighter and lighter sets to thereby terminate with child roots and nodes spread over various levels.

(ii) Both the discrete and continuous values are taken care by this algorithm. Continuous attributes are handled by a threshold value which is decided by C4.5. This value segregates the data list into those who have their values of attributes below the threshold and those having larger or similar value[19].

(iii) Is the training data is missing, it’s handled by the algorithm[20].

(iv) Once the tree is completely constructed, this algorithm runs the pruning of the tree. C4.5 after its construction moves back through the tree and thereafter, challenges to remove branches which are not supporting in reaching the leaf nodes[21].

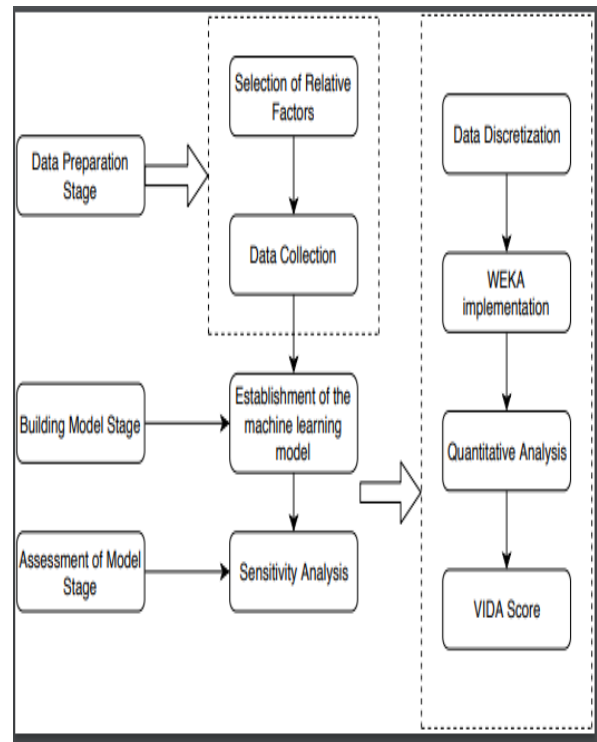


Figure 3 - Steps in Algorithm

Performance Evaluation:

Let’s represent Total Accuracy as TA, True Positive as TP, True Negative as TN, False Positive as FP, False Negative as FN and Random Accuracy as RA[22]. We calculate accuracy with the following formula:

$$TA = (TN+TP) / (TN+TP+FN+FP)$$

$$RA = ((FP+TP)*(FN+TN)+(TP+FN)*(TP+FP)) / Total$$

IV. RESULTS AND DISCUSSION

We collected data from a number of patients suffering from this disease for at least more than two years and observed for a period of three months. On applying algorithm and evaluating the results, we computed the success rate for predicting the stability of vitiligo.

For the proposed algorithm, accuracy percentage = 95.19% and error percentage = 4.81%. The accuracy rate of over 95% implies that this algorithm can be really useful in further treatment of this disease. The proposed algorithm is an efficient approach for predicting the stability of vitiligo.

V. CONCLUSION AND FUTURE SCOPE

In our work, we proposed a model that improves the vitiligo stability prediction mechanism. Depending upon the presence of a number of factors, patients get more accurate results or prediction and hence, they can better decide to adopt the method to cure it. According to which the best prediction rate and risk of failure of surgery is minimized. To achieve this goal, we built up a pipeline on a set of data. The main aim of this pipeline is to manage the patient's health and history efficiently. Thus, it's best to predict the stability of vitiligo before going for skin replacement surgery so as to have a successful treatment.

The proposed algorithm is introduced to have an optimized prediction of vitiligo patients which is important for the patients as well as doctors. However, the work can be extended further with more modifications of data in the pipeline before feeding it into algorithm, so that the error rate could be decreased. Of course, the sample size to construct the pipeline could be increased as well.

REFERENCES

- [1] Witten, I. H. & Frank, E. (2005). Data Mining: Practical machine learning tools and techniques, (2ndEd.) San Francisco: Morgan Kaufmann.
- [2] WEKA: Data Mining Software in Java.
- [3] Falabella R. Surgical treatment of Vitiligo: Why, when and how. J Eur Acad Dermatol Venereol. 2003;17:518–20.
- [4] Yusuf Perwej, Md. Husamuddin, Fokrul Alom Mazarbhuiya, "An Extensive Investigate the MapReduce Technology", International Journal of Computer Sciences and Engineering (IJCSSE), E-ISSN : 2347-2693, Volume-5, Issue-10, Page no. 218-225, Oct-2017, DOI : 10.26438/ijcse/v5i10.218225
- [5] M. Mohammad, "Performance Impact of Addressing Modes on Encryption Algorithms", In the Proceedings of the 2001 IEEE International Conference on Computer Design (ICCD 2001), Indore, USA, pp.542-545, 2001.
- [6] H.R. Singh, "Randomly Generated Algorithms and Dynamic Connections", International Journal of Scientific Research in Network Security and Communication, Vol.2, Issue.1, pp.231-238, 2014.
- [7] Rao A, Gupta S, Dinda AK, Sharma A, Sharma VK, Kumar G, et al. Study of clinical, biochemical and immunological factors determining stability of disease in patients with generalized vitiligo undergoing melanocyte transplantation. Br J Dermatol. 2012;166:1230–6.
- [8] A. Mardin, T. Anwar, B. Anwer, "Image Compression: Combination of Discrete Transformation and Matrix Reduction", International Journal of Computer Sciences and Engineering, Vol.5, Issue.1, pp.1-6, 2017.
- [9] H.R. Singh, "Randomly Generated Algorithms and Dynamic Connections", International Journal of Scientific Research in Network Security and Communication, Vol.2, Issue.1, pp.231-238, 2014.
- [10] Ines D, Sonia B, Riadh BM, Amel el G, Slaheddine M, Hamida T, et al. A comparative study of oxidant-antioxidant status in stable and active vitiligo patients. Arch Dermatol Res. 2006;298:147–52.
- [11] V. Krishnaiah, G. Narsimha, N. Subhash Chandra, "Heart Disease Prediction System using Data Mining Techniques and Intelligent Fuzzy Approach: A Review", International Journal of Computer Applications, February 2016.
- [12] Pattekari SA, Parveen A, "Prediction system for heart disease using naive bayes", International Journal of Advanced Computation.
- [13] R. Alizadehsani, J. Habibi, B. Bahadorian, et al., "Diagnosis of coronary artery stenosis using data mining", J MED Signals Sens, vol. 2, pp. 153-9, 2012.
- [14] Upasana Juneja et. al., "Multi Parametric Approach Using Fuzzification on Disease Analysis", IJESRT, Juneja et al., 3(5) ISSN: 2277-9655, Page No.492-497, 2014.
- [15] Hann SK, Shin HK, Park SH, Reynolds SR, Bystry JC. Detection of antibodies to melanocytes in vitiligo by western immunoblotting. Yonsei Med J. 1996;37:365–70.
- [16] Thirumal, P. C., & Nagarajan, N. (2015). Utilization of data mining techniques for diagnosis of diabetes mel-litus - A case study. ARPN Journal of Engineering and Applied Sciences, January, 10(1), 8-13.
- [17] Naughton GK, Reggiardo D, Bystry JC. Correlation between vitiligo antibodies and extent of depigmentation in vitiligo. J Am Acad Dermatol. 1986;15:978–81.
- [18] Baharav E, Merimski O, Shoenfeld Y, Zigelman R, Gilbrund B, Yecheskel G, et al. Tyrosinase as an autoantigen in patients with vitiligo. Clin Exp Immunol. 1996;105:84–8.
- [19] Xie Z, Chen D, Jiao D, Bystry JC. Vitiligo antibodies are not directed to tyrosinase. Arch Dermatol. 1999;135:417–22.
- [20] Hann SK, Park YK, Lee KG, Choi EH, Im S. Epidermal changes in active vitiligo. J Dermatol. 1992;19:217–22.
- [21] Kumar R, Parsad D, Kanwar AJ. Role of apoptosis and melanocytorrhagy: A comparative study of melanocyte adhesion in stable and unstable vitiligo. Br J Dermatol. 2011;164:187–91.
- [22] Ahn SK, Choi EH, Lee SH, Won JH, Hann SK, Park YK. Immunohistochemical studies from vitiligo: Comparison between active and inactive lesions. Yonsei Med J. 1994;35:404–10.

Authors Profile

Mr. Gagandeep Singh is pursuing Master of Technology in Computer Science from Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Haryana. He has a great interest in machine learning and algorithms.

Ms Kavita Rathi is Assistant Professor at the Department of Computer Science, Deenbandhu Chhotu Ram University of Science and Technology. She has guided numerous M.Tech students in their research and is actively known for her contribution in various conferences and technical research papers.