Association Rule-based Analysis of
Filarial Lymphoedema

Akash Nagaraj*, Mukund Sood, Bishesh Sinha, Ashok Raman, Dinkar Sitaram

Department of Computer Science and Engineering,
PES University, Bangalore
[akashn1897, mukundsood2013, bishesh.sin]@gmail.com,[ashokraman,dinkar.sitaram]@pes.edu

Abstract - **Elephantiasis has been classified as a *Neglected Tropical Disease (NTD)* which can be spread from person to person via mosquitoes. Our goal was to suggest a better treatment plan for a patient, through a thorough analysis that could lead us to results pertaining to the effectiveness of existing treatments. We also worked on trying to find the factors that majorly affect the contraction of the disease, as well as statistical proof to back the claims we have made to the factors that result in a patient contracting the disease, and their subsequent recovery period taking their conditions into account.**

Index Terms – elephantiasis, integrative, analytics, association rules.

I. INTRODUCTION

Lymphatic filariasis is also known as Elephantiasis. It is caused by parasitic worms and can spread from person to person through mosquitoes. Elephantiasis causes swelling of the scrotum, legs, or breasts.

Elephantiasis is considered a neglected tropical disease (NTD). It is more common in tropical and subtropical areas of the world, including Africa and Southeast Asia. It's estimated that 120 million people have elephantiasis.

The most common symptom of Elephantiasis is swelling of body parts. This tends to happen in the legs, genitals, breasts and arms. The legs are the most commonly affected area. The swelling and enlargement of body parts can lead to pain and mobility issues. The patient's skin may also become dry, thick, ulcerated, darker than normal and pitted. Some patients experience additional symptoms, such as fever and chills.

For our analysis, we used the data provided to us by the Institute of Applied Dermatology, situated in Kerala, India. IAD specialises in the treatment of patients suffering from *Elephantiasis*.

A few integrated treatment plans that are currently being used at IAD are Phanta Soaking, Indian Manual Lymph Drainage (IMLD) [9], Yoga exercises, Compression bandaging and following a Diet.
In addition, IAD also advocates the use of Ayurvedic medicines to augment the aforementioned practices.

II. STUDY OF ANALYSIS OF ELEPHANTIASIS

A. Motivation

Integrative Medicine (IM) is healing-oriented medicine that takes account of the whole person, including all aspects of lifestyle. It emphasizes the therapeutic relationship between practitioner and patient, is informed by evidence, and makes use of all appropriate therapies.

The Institute of Applied Dermatology (IAD) has conducted treatment by following a set of treatment procedures which involve a combination of various available treatments to obtain an optimal combination going forward. The objective of this project was to analyze the information provided with respect to the patients undergoing these treatment plans, and find which treatments are genuinely useful, and which do not have very encouraging results.

We also looked for patterns amongst patients with the disease to identify any common factors using statistical evidence to arrive at our conclusions Influence of socioeconomic status and family income were studied to identify the likelihood of contracting the disease, and the subsequent recovery periods.

The types of treatment being offered were *C-Bandage Stocking, Compression, Wrapping Cloth, Toe Compression, Sterilisation, IMLD* and *Sponge Mould Folds.* There was also a focus on the possible effects that *yoga* could have on the affected patients as well as a variety of *Ointments* prescribed to reduce swelling at the *Bacterial Entry Points.* We also tried to find patterns of *Inflammatory Episodes* before and after administering the treatment to aid the analysis being conducted on the effectiveness of a particular treatment.

We were also provided with a substantial amount of background information pertaining to each patient and using this, we found patterns within the data, for example, if particular strata of patients with a particular socio-economic background were more affected than others.

Since the compression bandages used in the treatment are expensive to be used on daily basis in

the home care setup. So, in order to analyse the effective significance of compression bandage usage in the treatment is worthwhile to recommend it for limb volume reduction.

B. *Dataset Description*

The data provided to us by the Institute of Applied Dermatology (IAD) has been collected from March 2001 to date and contains information with respect to over 3000 distinct patients.

Although the data provided to us consisted of over 100 relational tables, we worked with around 20 different tables, with data ranging from the various attributes of the patient's background to empirical measurements of the patient's limbs. The main variables that were used were ones related to *volume changes*, *patient follow-ups*, *type of treatment used* and *background information of the patient*.

The data collected while well-structured was sparse and needed to be cleaned before it could be used to conduct our analysis.

III. TECHNIQUES

A. *Statistical Analysis*

1. Analysis of Variance (ANOVA):

Analysis of Variance [10] is a collection of statistical models and their associated procedures used to analyze the differences among group means

2. Student's t-test:

The Student's t-test [11] is any statistical hypothesis test in which the test statistic follows a Student's t-distribution under the null hypothesis.

A t-test is most commonly applied when the test statistic would follow a normal distribution if the value of a scaling term in the test statistic were known. When the scaling term is unknown and is replaced by an estimate based on the data, the test statistics follow a Student's t distribution. The t-test was used, to determine if two sets of data are significantly different from each other.

3. Chi-Squared test:

A chi-squared test [12], also written as a test, is any statistical hypothesis test wherein the sampling distribution of the test statistic is a chi-squared distribution when the null hypothesis is true.

In our case, the test was applied when you have two categorical variables from a single population. It is used to determine whether there is a significant association between the two variables.

4. Repeated Measures Analysis of Variance (RMANOVA):

Repeated measures [13] design uses the same subjects with every branch of research, including the control. For instance, repeated measurements are collected in a longitudinal study in which change over time is assessed.

B. *Analytics – Association Rules*

Association rules [6] state an existing relationship between two entities. It is this very principle that is used in popular algorithms such as ID3 Decision tree [7], and Market Basket Analysis [8].

We took a novel approach to finding possible association rules. We implemented the Market Basket Analysis Algorithm and treated each combination of compressions prescribed, along with the consequent change in limb volume as a transaction, and so we ended up with a fairly large number of "transactions". We then applied the algorithm to these transactions and found association rules between large positive changes in limb volume and a few particular types of compression.

IV. RELATED WORK

Narahari, Saravu R., Terence J. Ryan, Kuthaje S. Bose, Kodimoolon each therapeutic system and a working knowledge of approach to their clinical diagnosis.e S. Prasanna, and Guruprasad M. Aggithaya. "Integrating modern dermatology and Ayurveda in the treatment of vitiligo and lymphedema in India."[3] -
A team of doctors from modern dermatology, Ayurveda, yoga therapy, and homoeopathy studied recruited patients to develop mutual orientation.

Narahari, Saravu R. "Collaboration culture in medicine."[4] -
Integrative Medicine selects best practices of public health and biomedicine. IM protocols in India. The major knowledge base of allopathy and Complementary and alternative systems of medicine is derived from the same human observation of natural events.

Narahari, Saravu R., Kodimoole S. Prasanna, and Kandathu V. Sushma. "Evidence-based integrative dermatology."[5] -
Studies conducted at the IAD have created a framework for evidence-based and integrative dermatology (ID). This paper gives an overview of the advances in ID with an example of *Lichen planus*.

V. RESULTS

1. **Socio-economic Factors**

We tried to find a relation between the socioeconomic status of the person affected to see if patients of a particular status were most likely to be affected by the disease and we found a direct relation, where patients of Status 4 are the least affected by the disease, as seen below:
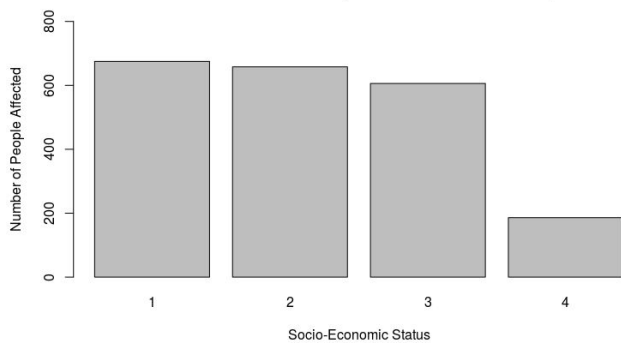
Fig. 1.   Socioeconomic Statuses of Patient data
Poor - 1, Average - 2, Good - 3, Great - 4

Examining the recovery of patients with respect to their Socioeconomic status, it can be seen that socioeconomic status 4 seems to recover the fastest as seen from the volume change in consecutive follow-ups. This could be attributed to better discipline in following therapeutic recommendations as well as better living conditions/hygiene etc.
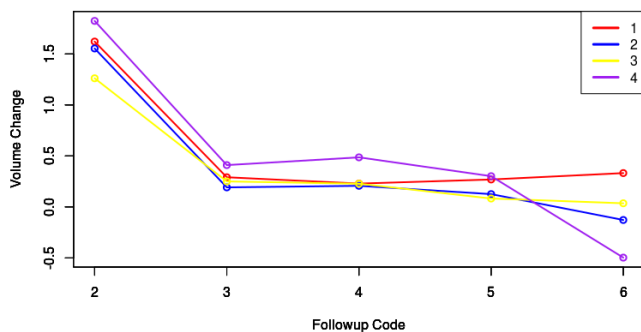

Fig. 2.   Socio-economic status vs Change in limb volume during follow-ups
Poor - 1, Average - 2, Good - 3, Great - 4

2.   C-bandage Stocking

We studied the average change in in the volume of patients who used C-Bandage Stocking and those who did not use C-Bandage Stocking

A positive value means that the volume of the limb reduced (became better), and a negative value means that the volume of the limb increased (became worse).

The initial consult is imperative in being cured. We also see that the average change at follow up 1 of patients who used C-Bandage Stocking is 2.179 units whereas those who did not use it was 1.472 units, thereby showing that it was an effective treatment.

3.   Sponge Mould Folds

We plotted a graph of change in volume per follow up of patients who used Sponge Mould Folds versus those who did not use Sponge Mould Folds.

The results show that those who used Sponge Mould Folds faired better than those who did not and this would be an advised course of action.
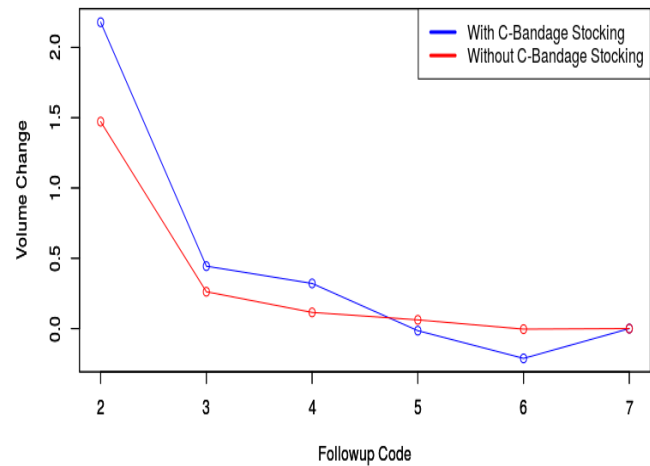

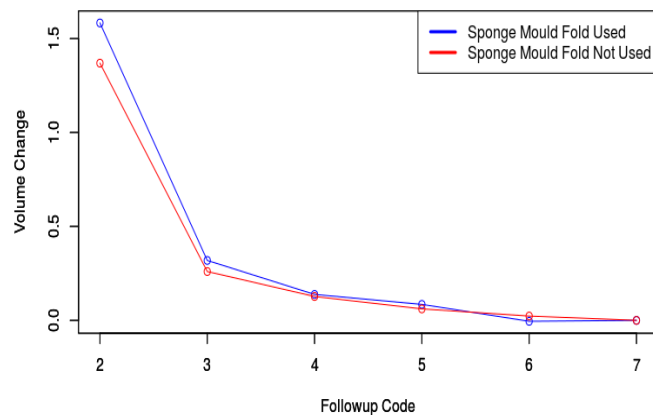Fig. 3.   Change of limb volume during follow-ups vs Usage of C-Bandage Stocking

Fig. 4.   Change of limb volume during follow-ups vs Usage of Sponge Mould Fold

4.   Demographic Analysis

We see that Karnataka and Kerala are the most affected based on the data, however, this may not be accurate as IAD is based in Kerala, so it would make sense that most of the patients the Hospital treats are from the Southern half of India.

One conclusion that we did come to however, was that the coastal districts were most affected with the exception of Gulbarga which is an Inland District. This could be attributed to various factors including proximity to IAD, access to transportation, as well as monetary reasons. In order to come to a more conclusive result, we would need to work with larger amounts of data.
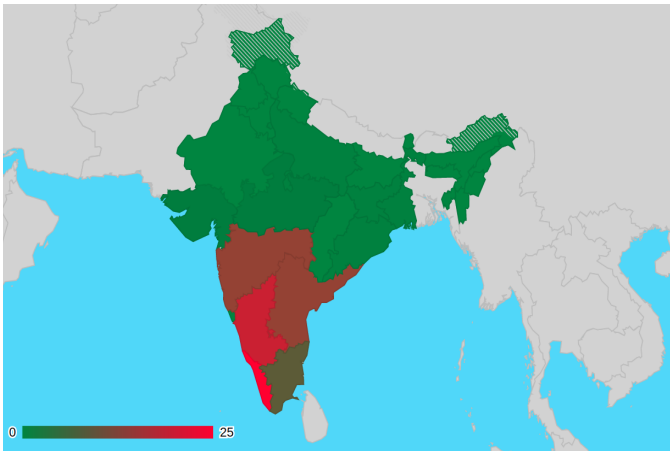
Fig. 5. Heatmap of most severely affected areas in the Indian subcontinent

Fig. 7. Change of limb volume during follow-ups vs Family Income

5. Family Income

The relation between people who have a certain family income was analyzed, to see if patients earning a certain amount were most likely to be affected by the disease and we found an apparent relation, where patients who earn an income of Level 1 are the most affected:
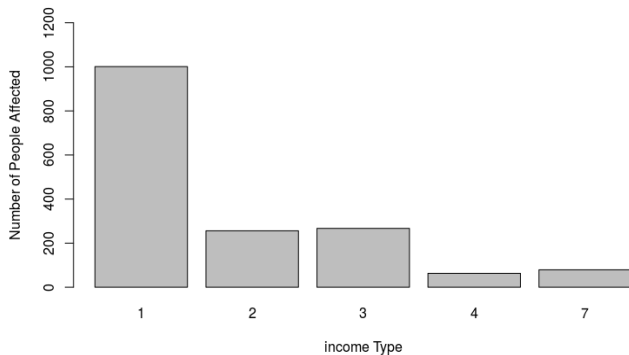

Fig. 6. Analysis of Family Income and the number of patients

The numbers represent the family income of a patient:

1 - Rs. 25,000 to 2,00,000
2 - Rs. 2,00,000 to 5,00,000
3 - Rs. 5,00,000 to 10,00,000
4 - Above Rs. 10,00,000

Looking at the recovery of patients with respect to their Family Income, it can be seen that Family Income 4 (highest income) seems to recover the fastest and 1 seems to show the least volume change as seen in consecutive follow-ups. This could be attributed to better resources, a hygienic environment, and better living conditions, etc.
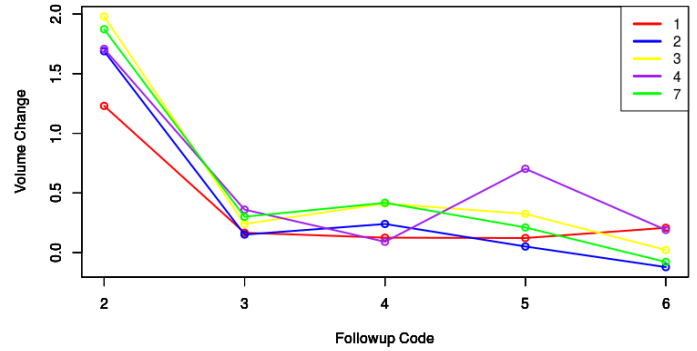
6. Effectiveness of Intertrigo Ointments

We moved on to the various types of ointments used by the patients as prescribed by Doctors.

In the data we used, Intertrigo Score was one of the independent variables associated with each patient who uses an ointment. We used the Intertrigo Score to determine Effectiveness, if the value of Intertrigo Score was 1, it was taken as effective, and a value of 2 was not effective.

This value was calculated based on the effectiveness of an ointment in reducing the intertrigo of a patient.

Ointment Name	Times Used	Times Effective	Effective%
Triben	1779	1225	68.85%
Rasnadi	677	479	70.75%
Loceryl	53	44	**83.02%**
Candid	72	59	81.94%
Tricawin	171	128	74.85%
Keto Based	111	77	69.36%
Others	537	381	70.94%

Table I. Effectiveness of the ointments used by patients

Based on Table I, *Triben* and *Rasnadi* are the most used ointments, with considerable effectiveness. On further analysis, we noticed that *Loceryl*, the most effective ointment is also the least used ointment, this is because *Loceryl* is comparatively more expensive when compared to the other ointments, and this is one of the reasons it is not used as much as the other ointments.

VI. CONCLUSION

Summarising all our results stated above we note the following three main conclusions:

- Association Rule Analysis of the types of compression showed that Wrapping Cloth, Sterilisation and Compression were the three most effective treatments to combat the disease.
- Family Income and Socio-Economic status are very important factors in determining how a person reacts to treatment as these factors indirectly affect the living conditions of a patient.
- The most used ointments currently being prescribed are *Triben* and *Rasnadi*, although empirically, the most effective ointment is *Loceryl*, we can't be sure due to the presence of a class imbalance owing to the price of *Loceryl*.

VII. METHODS FOR TRIALS

In the interest of reproducibility and open research, all the trials we have run for our analyses are publicly available in our repository [14].

VIII. REFERENCES

[1] "Integrating modern dermatology and Ayurveda in the treatment of vitiligo and lymphedema in India" Int j Dermatol 2011 Mar.

[2] Institute of Applied Dermatology - https://iad.org.in/

[3] Narahari, Saravu R., Terence J. Ryan, Kuthaje S. Bose, Kodimoole S. Prasanna, and Guruprasad M. Aggithaya. "Integrating modern dermatology and Ayurveda in the treatment of vitiligo and lymphedema in India." *International journal of dermatology* 50, no. 3 (2011): 310-334.

[4] Narahari, Saravu R. "Collaboration culture in medicine." *Indian journal of dermatology* 58, no. 2 (2013): 124.

[5] Narahari, Saravu R., Kodimoole S. Prasanna, and Kandathu V. Sushma. "Evidence-based integrative dermatology." *Indian journal of dermatology* 58, no. 2 (2013): 127.

[6] Agrawal, Rakesh, and Ramakrishnan Srikant. "Fast algorithms for mining association rules." In *Proc. 20th int. conf. very large data bases, VLDB*, vol. 1215, pp. 487-499. 1994.

[7] Quinlan, J. Ross. "Induction of decision trees." *Machine learning* 1, no. 1 (1986): 81-106.

[8] Berry, Michael J., and Gordon Linoff. *Data mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc., 1997.

[9] Gurdal, Sibel Ozkan, Alis Kostanoglu, Ikbal Cavdar, Ayfer Ozbas, Neslihan Cabioglu, Beyza Ozcinar, Abdullah Igci, Mahmut Muslumanoglu, and Vahit Ozmen. "Comparison of intermittent pneumatic compression with manual lymphatic drainage for treatment of breast cancer-related lymphedema." Lymphatic research and biology 10, no. 3 (2012): 129-135. Harvard

[10] Cohen, Yosef, and Jeremiah Y. Cohen. "Analysis of variance." *Statistics and Data with R: An applied approach through examples* (1988): 463-509.

[11] De Winter, Joost CF. "Using the Student's t-test with extremely small sample sizes." Practical Assessment, Research & Evaluation 18, no. 10 (2013)

[12] Scheffe, Henry. "The relation of control charts to analysis of variance and chi-square tests." Journal of the American Statistical Association 42, no. 239 (1947): 425-431.

[13] Potvin, Patrick J., and Robert W. Schutz. "Statistical power for the two-factor repeated measures ANOVA." Behavior Research Methods, Instruments, & Computers 32, no. 2 (2000): 347-356.

[14] Akash Nagaraj. "Analysis of Filarial *Lymphodema*". *https://github.com/grassknoted/Analysis-of-Filarial-lymphodema/* . 2018.