An Innovative Approach to Chronic Insomnia Using Genetic Markers

Lucas Xavier Veneroso
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil

Marcos A. dos Santos
Department of Computer Science
Universidade Federal de Minas Gerais
Belo Horizonte, Brazil

***Abstract*—This work is a continuation of a previous one, where we found an ad-hoc method for identifying who might have chronic insomnia by certain genetic markers. In this paper we delve further into the issue and find those relevant markers with more in-depth analysis of their impact on the algorithm**

***Index Terms*—RNA microarray, Insomnia, Bio informatics, Ad-hoc, Regression algorithm, Biotechnology**

I. Introduction

In our previous work we have mentioned the importance of the issue we are tackling: sleep is essential for survival and development of humans. In that work we mentioned a guiding source behind this motivation: Outlive [1], a book by Peter Attia, well known for his contributions in the area of longevity medicine, that is focused on how to improve lifespan and quality of life, with an entire chapter dedicated to the importance of sleep, highlighting how it is connected to a myriad of other conditions that are quite serious and life threatening. We also mentioned articles such as [2] that link lack of sleep to dementia and [3] that links sleep-wake regulation to Alzheimer's.

We are still using the dataset [4] provided at the NCBI(National Center for Biotechnology Information) by the Semel Institute for Neuroscience at UCLA. This data set contains data from 43 patients, 17 of whom had the condition, each of them having data related to 33210 genetic markers.

Our previous work established a reliable algorithm, based on ad-hoc to be able to identify and properly classify someone with having or not having insomnia. Given the effectiveness of said algorithm and no positive change from modifying its base structure, we kept it the same. Our objective now is to pick up where we left off and finally delve into the genetic markers found and provide more in-depth details about the algorithm.

II. Related Work

Due to the recency of the solution, not a lot of research has been done on this specific ad-hoc variant of the algorithm, although a version of a modified logistic regression algorithm similar to the one developed can be found in [5].

Another related work is our own previous work.

III. Methodology

We still employed the use of MATLAB as our main tool for this paper and, as mentioned before, utilized the same code as our previous work as a base with certain modifications to improve upon it. One such modification can be seen on Fig. 1, where now it is more notable how separated, in a three dimensional space, our data points are.

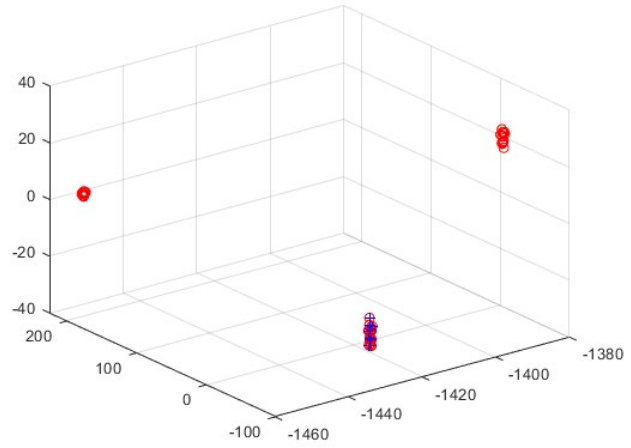

Fig. 1. 3D Plot of all the data with blue crosses on patients with insomnia

The main bulk of the advancements were in analyzing the alphas utilized to determine which markers were the most relevant. for that we first calculate the weights associated to each attribute according to our alphas and that is shown in Fig. 2.

Furthermore, after running our algorithm, we can see that the quality of our classification does not get altered as seen in Fig. 3.

Now with those results we can effectively reduce the model significantly with just the most relevant markers which can be obtained by sorting our alphas, giving us the most and least relevant markers, both being great features to characterize the dataset. We can see that sorted alpha vector in Fig. 4, with another vector of position with values corresponding to their original positions in the integral sorted alpha vector in Fig. 5.

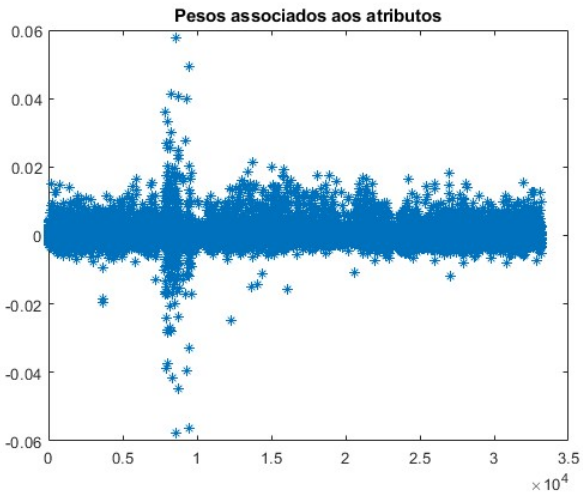

Fig. 2. Plot of the weight associated with each marker

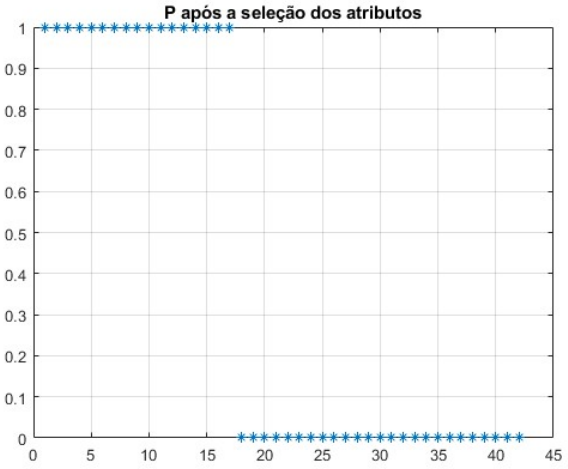

Fig. 3. Values of P after our attribute selection

We can also see the new weight distribution after our model considers only the most relevant markers in Fig. 6.

In addition to that we have identified that our most relevant markers are labeled within HS in the dataset, with some of them being: HS.501656, HS.413494, HS.534427, HS.546375 and HS.554324. Although further information about said markers was challenging to find we do know one of them, HS.413494, seems to have been classified as benign which this paper postulates should be reevaluated.

IV. Conclusion

Throughout this work spanning two papers, we managed to successfully create an algorithm based on an ad-hoc model, which already saw use in applications such as identifying bank fraud, that applies to the world of bio informatics. Given the density of the datasets within that environment, such advances are incredibly notable, with very good results, especially when it comes to the specific disease chosen.

14x1 double

	1	2	3
1	0.0439		
2	3.1733		
3	-2.4357		
4	-1.6300		
5	-0.3726		
6	0.7922		
7	-2.5875		
8	-0.0175		
9	0.3060		
10	-0.6698		
11	0.2119		
12	0.7939		
13	1.0763		
14	1.0334		
15			

Fig. 4. 14 most relevant alphas

A great knowledge of tools like MATLAB and the database at NCBI was gained and the process was indeed one that could lead to great improvements and innovations. We look forward to seeing what such knowledge may bring in the future.

References

[1] Attia, P., & Gifford, B. (. (2023). Outlive: the science & art of longevity. First edition. New York, Harmony.

[2] Robbins R, Quan SF, Weaver MD, Bormes G, Barger LK, Czeisler CA. Examining sleep deficiency and disturbance and their risk for incident dementia and all-cause mortality in older adults across 5 years in the United States. Aging (Albany NY). 2021 Feb 11; 13:3254-3268 . https://doi.org/10.18632/aging.202591

[3] Maxime Van Egroo, Justinas Narbutas, Daphne Chylinski, Pamela Villar González, Pierre Maquet, Eric Salmon, Christine Bastin, Fabienne Collette, Gilles Vandewalle, Sleep–wake regulation and the hallmarks of the pathogenesis of Alzheimer's disease, Sleep, Volume 42, Issue 4, April 2019, zsz017, https://doi.org/10.1093/sleep/zsz017

[4] Piber D, Cho JH, Lee O, Lamkin DM et al. Sleep disturbance and activation of cellular and transcriptional mechanisms of inflammatio in older adults. Brain Behav Immun 2022 Nov;106:67-75. PMID: 35953022

[5] Francielly Morais-Rodrigues, Rita Silverio Machado, Rodrigo Bentes Kato, Diego Lucas Neres Rodrigues, Juan Valdez-Baez, Vagner Fonseca, Emmanuel James San, Lucas Gabriel Rodrigues Gomes, Roselane Goncalves dos Santos, Marcus Vinicius Canario Viana, Joyce da Cruz Ferraz Dutra, Mariana Teixeira Dornelles Parise, Doglas Parise, Frederico F. Campos, Sandro J. de Souza, Jose Miguel Ortega, Debmalya Barh, Preetam Ghosh, Vasco A.C. Azevedo, Marcos A. dos Santos, Analysis of the microarray gene expression for breast cancer progression after the application modified logistic regression, Gene, Volume 726, 2020, 144168, ISSN 0378-1119, https://doi.org/10.1016/j.gene.2019.144168.

33209x1 double

	1	2
1	8566	
2	9510	
3	8713	
4	8307	
5	9324	
6	7904	
7	8015	
8	9491	
9	8063	
10	8298	
11	8216	
12	7991	
13	8165	
14	12316	
15	7908	
16	8715	
17	8179	
18	8518	
19	3679	
20	3677	
21	3678	
22	9204	
23	7825	
24	8024	
25	9610	
26	8327	
27	7970	
28	7872	
29	8702	
30	8103	

Fig. 5. Alpha positions

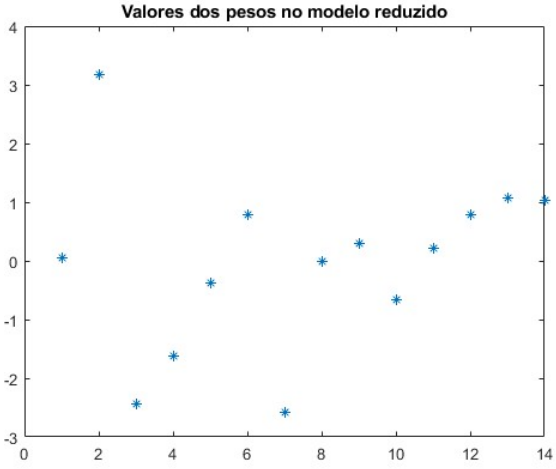

Fig. 6. Alpha positions