Webology (ISSN: 1735-188X)
Volume 18, Number 6, 2021

Prediction And Detection Model Of Systemic Lupus Disease By Using Machine-Learning And Artificial Intelligence Along With Jupyter Anaconda Navigator Simulation

Payel Saha*

[1]Department of Computer Science, North Bengal St. Xavier's College, West Bengal, India.

Mail id:payel17.10@gmail.com

Abstract: Artificial neural networks are intelligent systems that have been successfully used for prediction in different medical data fields. Medicinal Data is the way in the direction of removing concealed examples from therapeutic data. This paper shows the advancement of a crossover model for ordering Indian lupus database (ILDD). The model comprises of two phases. In the primary stage, the machine learning algorithms bunching is utilized to distinguish and take out erroneously grouped examples. The nonstop data is changed over to all out frame by rough width of the coveted interims, in light of the conclusion of restorative master. In the second stage an adjusted arrangement is finished utilizing artificial intelligence algorithm RNN by taking the accurately bunched event of first stage. Test comes about imply the fell ML grouping and AI based RNN has upgraded arrangement precision of ANN. Additionally, administers produced utilizing fell Artificial intelligence with jupyter simulation along with python with clear cut data are less in numbers and simple to translate contrasted with principles created with RNN alone with persistent data. The proposed fell model with all out data got the arrangement precision of 93.33 % when contrasted with exactness of 73.62 % utilizing ANN alone for ILDD Indian lupus disease dataset. The well-trained model will fully have qualified to assist healthcare providers to make timely and accurate decisions.

Keywords – Medical data set, machine learning, artificial intelligence, prediction model, for casting, testing and training method, Jupiter simulation, RNN, etc.

1. Introduction

Human beings are the most composite organisms on this globe. It is hard to envision how billions of microscopic parts, each one with its own identity, work together in a planned manner for the profit of the total being. A system is a union of various organs arranged together so that they can carry out complex functions for the body. Body functions are the physiological or psychological functions of body systems. Survival is the body's most important ambition and it depends on the body's maintaining homeostasis. Homeostasis is a situation of relative constancy of body's internal environment and it depends on the body's carrying out many actions in a structured manner continuously. Its major actions or functions are responding to changes in the body's environment, exchanging materials between the environment and cells, metabolizing foods, and integrating all of the body's diverse actions. In case there is any transform in the homeostasis diseases set in . A disease is an abnormal situation affecting any part of the body. Diseases are roughly classified into communicable and non-communicable disease.

Medical Data for Healthcare Management (DMHM) is a promising field where researchers from both academia and industry have acknowledged the potential of its contact on improved healthcare by inventing patterns and trends in large amounts of complex data generated by healthcare transactions. Medical Data also helps to discover attractive business insights to help make business decisions that can sway cost efficiency and yet maintain a high quality of care. Healthcare management has received grand deal of attention in current times and application of data mining techniques to this field is ahead increasing popularity.

The knowledge-rich nature of the healthcare domain has made it a perfect atmosphere, where knowledge on data mining should also have to be extended further for the increasing need. Though, the abstract nature of tacit healthcare knowledge has resulted in the under-utilization of such a fundamental component of the general healthcare delivery system. There are many algorithms for these problems, but they are not strict and accurate. Patient Reported Outcomes (PROs) in clinical studies have gradually improved in frequency, for their importance in evaluating therapies and budding treatment plans. Due to the extraordinary growth-rate of Health Care data, which is being collected, loaded and stored through the World Wide Web and accessed electronically in almost all fields of human endeavor, there is an vital need for sophisticated tools and techniques that can lever extremely large multiple data.

Foundational Lupus Erythematosus (SLE) is a persistent immune system illness that makes the invulnerable framework assault the body's own connective tissues and organs. It is a significant and possibly dangerous disease that influences about 5,000,000 individuals all over the planet (Lupus Foundation of America, 2018). People experience issues anticipating SLE side effect seriousness levels due to the intricate collaborations of sickness trigger openness levels over the long run. SLE side effect seriousness levels ceaselessly vary in light of the people's awareness and openness to illness triggers, for example, action and feeling of anxiety, UV openness, measure of rest, ailment and injury. To deal with their side effects, SLE patients should comprehend and restrict their openness to these triggers. On account of their complicated communications over the long haul, notwithstanding, this is anything but a basic errand.

For instance, an individual might have the option to securely endure three hours of UV openness when very much refreshed and loose. Assuming they have other intensifying trigger openings, in any case, something very similar about of UV openness might make them foster a rash. Furthermore, a similar measure of trigger openness can cause various responses relying upon how much openness encountered the earlier days. This isn't just valid for a similar trigger, yet others too. This implies that similar three hours of UV openness when loose could cause a response assuming the earlier day the individual was incredibly anxious.

This intricacy can cause SLE to feel capricious, unmanageable and startling to victims. "With side effects that go back and forth, infection flares and abatements, and the vulnerability of what every day will bring, it's generally expected to encounter sensations of despondency, dissatisfaction, outrage, or misery" (Lupus Foundation of America, 2018). This enthusiastic tole can become incapacitating. Concentrates on show that somewhere in the range of 15 and 60 percent of individuals with an ongoing sickness like SLE will encounter clinical wretchedness throughout their illness (Lupus Foundation of America, 2018).

A device that can foresee SLE side effect seriousness levels with a more noteworthy precision than people is one method for diminishing this weight. By distinguishing and enlightening examples in trigger openness and side effect seriousness levels, such a device would likewise further develop SLE patients' capacity to deal with their side effects. This would make significant and wide-going impacts. Better side effect the board has been displayed to increment patient personal satisfaction, diminish medical services costs, and further develop sickness anticipation (Lupus Foundation of America, 2018).

II. RELATED WORK:

AI is an arising field in software that looks to tackle complex issues like the one presented by SLE. Since there are many trigger openness levels that associate to bring about numerous side effect seriousness levels, the issue is appropriate to a particular kind of AI design known as a counterfeit brain organization (Fausett, 1994).

Counterfeit brain networks are a sort of figuring framework motivated by the natural brain networks that comprise creature minds (Fausett, 1994). They are involved layers of hubs with weighted associations between them that are intended to immitate neurons (Fausett, 1994). In their least difficult

execution, information is stacked into the info layer hubs, sifted through the organization, and outputted by means of the result layer hubs (Fausett, 1994).

There are numerous exceptional structures that fake brain organizations can take to work on their presentation in various issue areas. In one construction, called a recursive brain organization (RNN), the past advance's result values are reappeared into the net as info values for the following stage (Fausett, 1994). This is valuable for issues like SLE side effect expectation in which past results impact future forecasts.

Setting units are one more kind of design that permit brain organizations to represent history-delicate difference (Fausett, 1994). An exceptional sort of info hub, these units accept their information from the past advances' secret layer hubs and reinsert it into the net during the following stage. This infuses recorded input information into the situation (Fausett, 1994).

We estimated that a counterfeit brain organization could effectively foresee SLE side effect levels since it has been utilized effectively for other infection side effect forecast issues (Wu, Roy, and Stewart, 2010). The particular type of a RNN with setting units was chosen because of its demonstrated capacity to represent chronicled info and result information in an assortment of issue spaces (Fausett, 1994). This permits it to represent the impact of recorded trigger openness and side effect seriousness levels in its forecasts.

Existing clinical AI studies have utilized a comparative time series approach, like Wiens, Guttag, and Horvitz's (2012) work anticipating patient gamble of emergency clinic gained disease. The current assortment of work is restricted, notwithstanding, in its application to SLE.

Existing AI studies tending to SLE treatment are centered around bunching issues, for example, distinguishing the qualities answerable for SLE (Armañanzas et al., 2009) and risk factors for explicit unfavorable SLE results (Ravenell et al., 2012). These examinations use quantitative information, for example, DNA, that is gathered and broke down in a lab setting. We observed no SLE concentrates on that used or anticipated emotional patient information, for example, weariness and agony levels.

This shortfall is probable because of the hardships related with the precise elicitation and

standardization of abstract wellbeing information for use in PC calculations. While get-together day to day side effect level reports, we have found that elements, for example, passionate state and character significantly sway revealed values. This adversely influences the prescient capacity of models that utilization the information. The failure to address for this issue is a critical shortcoming in the current assemblage of information, since it significantly restricts what parts of SLE can be successfully contemplated.

III. PROBLEM SPECIFICATION:

To start to fill the current information holes, we have planned, carried out, prepared and tried a RNN for SLE side effect forecast in view of noteworthy sickness trigger openness and side effect seriousness levels. We conjectured that this was conceivable in light of the fact that AI has been utilized effectively for other illness side effect expectation issues (Wu, Roy, and Stewart, 2010). The model created shows that, however people battle to anticipate SLE side effects, they depend on an example of trigger openness connections that is machine learnable.

The model can retain an extended time of fake SLE information to a generally 20% level of precision. This demonstrates that, however people battle to foresee SLE side effects, they depend on an example of trigger openness connections that is machine learnable. Strangely, the model remembers real quiet informational indexes more precisely than it does the fake informational collections. This proposes that there might be clashing information in the fake sets that was incidentally infused in the development cycle. Assuming this theory is valid, it implies that the issue is much more appropriate to an AI arrangement than is clear by our outcomes.

The model delivered by the standard train/approve process, in any case, is too unbending to even consider performing great on different and fluctuating SLE introductions. SLE incorporates trademark sickness flare and abatement cycles, and the reduction time frames were being over-prepared however the flares were under - prepared. This drove us to develop a second, novel preparation strategy for the RNN that is more performant in this sickness area. It persistently turns preparing on and off in view of the most extreme mistake every day. When set off on, the RNN retrains throughout recent days prior to keeping on preparing new approaching information. Thusly, the RNN can become familiar with the SLE flare movement without overtraining the reduction action.

Since SLE is a quickly moving illness, the new model's pace of progress is likewise altogether expanded, permitting it to respond rapidly to a gradually approaching informational collection. In our review, models prepared in this style performed 3.5 to 5% preferable on normal over those prepared through standard remembrance. Albeit a combined preparation informational index was utilized as a result of information access constraints, the models produced were precise on both little genuine patient informational collections. Later on, we might want to repeat these outcomes with a huge SLE preparing informational collection.

IV. RESEARCH METHODLOGY:

M-tree is a supervised approach to classification. A M-tree is a simple tree structure where all non-terminal nodes denotes tests on one or more attributes and terminal nodes reflect M outcomes. The basic M-tree induction algorithm has been enhanced. The JUPYETR classifier package has its own version of known as ANN. Information gain and gain ratio measures are used by as splitting criterion respectively.

M-tree is a dynamic access method appropriate to index generic "metric spaces", where the function is used to calculate the distance between any two objects satisfies the positivity, symmetry, and triangle inequality postulates. The M-tree design fulfills distinctive necessities of multimedia applications, where objects are indexed using difficult features, and similarity queries can require application of time-consuming distance functions.

Steps of M-tree Algorithm:

1. Choose an attribute that best differentiates the output attribute values.

2. Create a separate tree branch for each value of the chosen attribute.

3. Divide the instances into subgroups so as to reflect the attribute values of the chosen node.

4. For each subgroup, terminate the attribute selection process if:

(a) The members of a subgroup have the same value for the output attribute, terminate the attribute selection process for the current path and label the branch on the current path with the specified value.

(b) The subgroup contains a single node or no further distinguishing attributes can be determined. As in (a), label the branch with the output value seen by the majority of remaining instances.

5. For each subgroup created in (3) that has not been labeled as terminal, repeat the above process.

V. SIMULATION WORK DETALS:

The RNN fabricated is made out of a five-hub input layer with two recursive information sources and ten setting units, a ten-hub stowed away layer, and a two-hub yield layer (see Appendix A). The patient's everyday anxiety, movement level, UV openness, long stretches of rest, and sickness or injury level are placed into the organization. The model results anticipated weakness and rash levels. The real upsides of these side effects are then placed and involved the next day as the recursive information sources. Assuming that it is in preparing mode, blunder is determined and back propagated through the RNN.

Preparing an AI calculation requires huge informational collections (Fausett, 1994). Due to restricted admittance to patient information, we assembled huge preparation and testing informational indexes utilizing misleadingly built patient information. The precision of these sets was checked on by a SLE patient and confirmed in the model through their prescient limit on two little SLE patient informational collections.

The first idea for this venture was an iOS application for SLE patients that would permit them to enter trigger openness levels and get anticipated side effect levels. Inside the primary long stretches of the undertaking, in any case, we understood that this expectation must be achieved by means of AI because of the perplexing connection of trigger openness levels over the long haul. Never having concentrated on AI, we chose to execute the RNN without outside AI libraries so we could all the more likely learn and comprehend what it was doing and to oblige our clever preparation model.

Outside AI libraries, for example, Tenser Flow were thought of and might be utilized in ensuing models. Freely assembling the model, notwithstanding, had critical scholarly worth. It permitted us to apply the new data we were learning in an involved manner, and it uncovered marks of disarray. For instance, a few false impressions were made clear when we

endeavored to execute the backpropagation technique, including how the qualities were being determined and where the signs were being sent. These could never have become known and been rectified on the off chance that we had essentially settled on a capacity decision to an outer library.

VI. RESULT ANALYSIS:

The following table shows the assignment of cluster using ANN and M-tree. The LUPUS

DISEASE dataset contains 768 attributes which divides into two cluster which has been shown in the table. The experimental outcome of proposed M-tree based LUPUS prediction model are as follows:

Table 1.0 Cluster Assignment using M-tree

Division of Objects	Cluster 1 477	Cluster 2 291
Percentage	62%	38%

In order to assess the performance of prediction of lupus DISEASE the outcome of M-tree has been compared with the outcomes of clustering algorithms like ANN. The result has been compared in terms of Sum of Squared Errors, Number of Iterations, Execution Time, Goodness of Fit(accuracy).

Table 1.1 Comparison of Sum of Squared Errors, Number of Iterations, Execution Time, Goodness of Fit (accuracy).

Performance	**ANN**	**M-tree**
Sum of Squared Errors	255.000	69.00
Number of Iterations	9	5
Execution Time (in Sec)	0.05	1.19
Goodness of fit (in %)	67.318	92.125

Now, we will show the result of M-tree algorithm of predictive value of LUPUS data set the details off result shown below.

Table 1.3 Analysis report of M-tree model for lupus DISEASE data set

0 test negative	1 test positive	Test done by model
454	46	Negative
23	245	Positive test

The following tables show the analysis of the negative and positive test of the patients, unpredicted value of the instances and accuracy of the M-tree algorithm on lupus IDSEASE dataset. Further, we will discuss of ANN approach of lupus dataset for prediction of positive and negative record set of patients, unpredicted value of the instances and accuracy of the M-tree algorithm on LUPUS DISEASE dataset.

In these tables,

0 represents negative test.

1 represents positive test.

Table 4.6 M-tree and k-means model detect for unpredicted value of instances and accuracy

Factors	Sum of errors/	Percentage of errors	%(accuracy)
	Incorrected instances		
M-tree	69	8.999	92.125
ANN	255	33.023	66.977

This table shows the final result of our analysis. It shows sum of errors/Incorrected instances, Percentage of errors and accuracy.

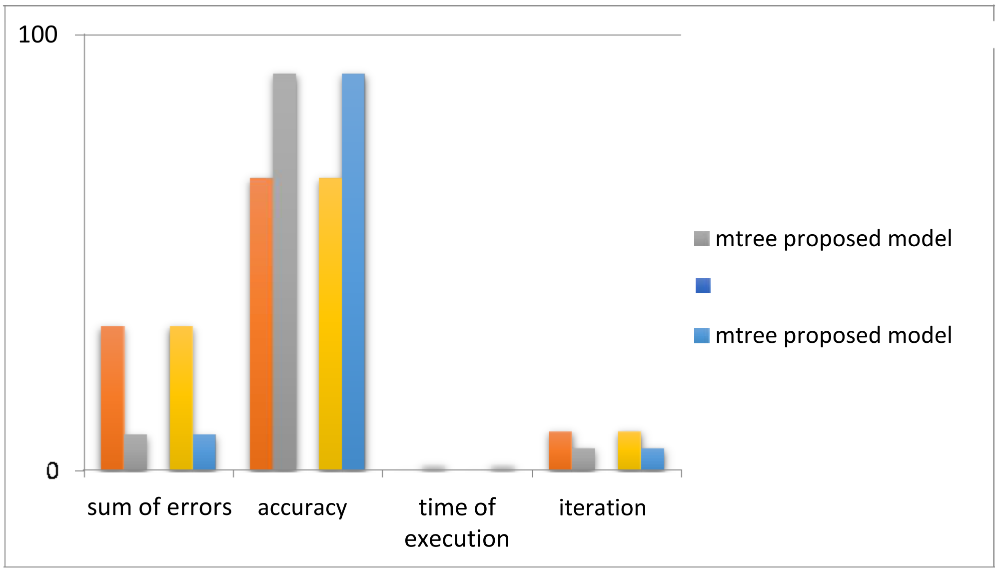

FIG 1.1 PLOTTED GRAPH REPRESENTATONF PREDICTION OF LUPUS DISEASE ACCURACY BY PROPOSED MODEL

This venture is an underlying model into a complex and beforehand neglected issue space. It confronted significant time and information accessibility restrictions, and it was a stretch of our scholastic capacities. All things considered, the outcomes are huge for quite a long time.

To start with, the model's capacity to retain the fake and SLE informational indexes shows that SLE side effect seriousness levels depend on an example of trigger openness collaborations that is machine learnable. This makes the way for additional AI investigation into their relationship and expectation. This disclosure had the option to be made in spite of restricted admittance to SLE patient information since we invested impressive energy building and testing fake informational collections that could be utilized to prepare and test the RNN.

Second, we had the option to develop a clever preparation strategy that is 3.5-5% more viable than the standard technique since we invested in some opportunity to build the RNN without any preparation as opposed to utilizing existing libraries. This permitted us to advance precisely the way that the RNN was working (regardless of being new to AI research), which uncovered the issues with the standard model. Further work is expected to diminish the normal mistake however much as could reasonably be expected.

At long last, the outcome addresses a significant measure of self - coordinated autonomous learning. Undertaking it was our first openness to AI ideas, but we had the option to develop a functioning model that is sensibly precise inside the space of a solitary semester.

Assuming an adequate number of patient information had been accessible to prepare and test the RNN without counterfeit informational collections, and similar outcomes were accomplished, this would be distribution commendable work. It pushes the limits of SLE AI research and demonstrates that further investigation is probably going to be productive. In particular, it gives a functioning prescient model which SLE patients can use to deal with their side effects.

VII. CONCLUSION WORK:

The most important result of the model is the noteworthy data it gives to doctors and patients. However still in its earliest stages, it as of now shows guarantee for side effect the executives. The following stage is repeating the examination just utilizing genuine patient information. Enormous volumes of information from a wide assortment of SLE patients is important to work on the model's precision and pliancy. We propose utilizing information from something like 50 different SLE patients. Empowered perceive illness flares. It is essential to likewise test the full scope of infection triggers and side effects in the RNN to frame a total certifiable model. Such information is accessible through existing clinical and drug associations right now occupied with SLE research .

As well as working on the model, we are attempting to advance it to anticipate results as well as immediate patients utilizing change in behavior patterns criticism, directing them into an ideal illness state. Fox, Ang, Jaiswal, Pop-Busui, and Wiens' (2018) as of late demonstrated in a comparable illness side effect direction issue that AI is compelling for multi-step guaging, showing that this is logical workable for SLE too. Utilizing this sickness way expectation, future projects could guide patients to keep away from explicit trigger openings when they are wandering from the ideal illness course.

If effective, this exploration would be groundbreaking for patients with SLE and comparable problems. Social training for persistent sickness taking care of oneself is monetarily unattainable for most patients. An available computerized adaptation, for example, an AI fueled application, would take into consideration more noteworthy admittance to this kind of care and further developed results.
Such an application wouldn't be valuable, in any case, on the off chance that it didn't accompany a pretrained model. Further examination ought to consider the manners by which a model or set of models can be developed for introductory use, then, at that point, sharpened to the particular illness show of every client. This would almost certainly include AI research analyzing the conceivable distinguishing proof of one of a kind SLE subtypes.

Future work ought to likewise consider manners by which the requirement for manual information section can be decreased without lessening the model's prescient limit. This might incorporate the programmed transferring of clinical trial results, action screens, and related wellbeing data. Positioning the convenience of information factors and taking out those with the least impact may likewise assist with lessening the weight on clients.

At last, this undertaking uncovered critical constraints brought about by the utilization of emotional wellbeing information, for example, weariness and agony levels, in AI. Further work ought to analyze the precise elicitation and standardization of emotional wellbeing information for use in PC calculations. Identifying and representing these issues could work on the prescient capacities of models that depend on the information, similar to the one portrayed here.

VIII. REFERENCES:

1. Tsokos, G. C. Systemic lupus erythematosus. N. Engl. J. Med. 365, 2110–2121 (2011).

2. Tsokos, G. C., Lo, M. S., Reis, P. C. & Sullivan, K. E. New insights into the immunopathogenesis of systemic lupus erythematosus. Nat. Rev. Rheumatol. 12, 716– 730 (2016).CAS Article Google Scholar

3. Barturen, G. & Alarcón-Riquelme, M. E. SLE redefined on the basis of molecular pathways. Best Pract. Res. Clin. Rheumatol. 31, 291–305 (2017). Article Google Scholar

4. Baechler, E. C., Gregersen, P. K. & Behrens, T. W. The emerging role of interferon in human systemic lupus erythematosus. Curr. Opin. Immunol. 16, 801–807 (2004). CAS Article Google Scholar

5. Bennett, L. et al. Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. J. Exp. Med. 197, 711–723 (2003).CAS Article Google Scholar

6. Panousis, N. I. et al. Combined genetic and transcriptome analysis of patients with SLE: distinct, targetable signatures for susceptibility and severity. Ann. Rheum. Dis. 78, 1079–1089 (2019). CAS Article Google Scholar

7. Zhao, M. et al. Increased 5-hydroxymethylcytosine in CD4+ T cells in systemic lupus erythematosus. J. Autoimmun. 69, 64–73 (2016).

8. Scharer, C. D. et al. ATAC-seq on biobanked specimens defines a unique chromatin accessibility structure in naïve SLE B cells. Sci. Rep. 6, 27030 (2016).

9. Rao, S. S. P. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680 (2014).

10. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin domains: the unit of chromosome organization. Mol. Cell 62, 668–680 (2016).

11. Beagrie, R. A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. Nature 543, 519–524 (2017).

12. Denas, O. et al. Genome-wide comparative analysis reveals human-mouse regulatory landscape and evolution. BMC Genomics 16, 87 (2015).

13. Delaneau, O. et al. Chromatin three-dimensional interactions mediate genetic effects

on gene expression. Science (80-) 364, eaat8266 (2019).

14. Ernst, J. & Kellis, M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nat. Biotechnol. 28, 817–825 (2010).

15. Kharchenko, P. V. et al. Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. Nature 471, 480–485 (2011).

16. Rennie, S., Dalby, M., van Duin, L. & Andersson, R. Transcriptional decomposition reveals active chromatin architectures and cell specific regulatory interactions. Nat. Commun. 9, 487 (2018).

17. Dumitru Dan Burdescu, Marian Cristian Mihaescu, Bogdan Logofatu, Costel Marian Ionascu "Improving the Quality of Knowledge Discovery Process by Information Gain Computing" (2012).

18. Rada-Iglesias, A., Grosveld, F. G. & Papantonis, A. Forces driving the three-dimensional folding of eukaryotic genomes. Mol. Syst. Biol. 14, e8214 (2018).

19. Pombo, A. & Dillon, N. Three-dimensional genome architecture: players and mechanisms. Nat. Rev. Mol. Cell Biol.12 (2015).

20. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. Cell 174, 744-757.e24 (2018).

21. Tsochatzidou, M., Malliarou, M., Papanikolaou, N., Roca, J. & Nikolaou, C. Genome urbanization: clusters of topologically co-regulated genes delineate functional compartments in the genome of Saccharomyces cerevisiae. Nucleic Acids Res. 45, 5818–5828 (2017).

22. Lercher, M. J., Urrutia, A. O. & Hurst, L. D. Clustering of housekeeping genes provides a unified model of gene order in the human genome. Nat. Genet. 31, 180– 183 (2002).

23. Ebisuya, M., Yamamoto, T., Nakajima, M. & Nishida, E. Ripples from neighbouring transcription. Nat. Cell Biol. 10, 1106–1113 (2008).

24. Coolen, M. W. et al. Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity. Nat. Cell Biol. 12, 235–246 (2010).

25. Letourneau, A. et al. Domains of genome-wide gene expression dysregulation in Down's syndrome. Nature 508, 345–350 (2014).

26. Liao, Y., Smyth, G. K. & Shi, W. FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923– 930 (2014).

27. Bombardier, C. et al. Derivation of the sledai: a disease activity index for lupus patients. Arthr. Rheum. 35, 630–640 (1992).

28. Ran, D. & Daye, Z. J. Gene expression variability and the analysis of large-scale RNA-seq studies with the MDSeq. Nucleic Acids Res. 45, e127 (2017).

29. Chaussabel, D. et al. A modular analysis framework for blood genomics studies: application to systemic lupus erythematosus. Immunity 29, 150–164 (2008).

30. Li, S. et al. Molecular signatures of antibody responses derived from a systems biology study of five human vaccines. Nat. Immunol. 15, 195–204 (2014).

31. Zyla, J. et al. Gene set enrichment for reproducible science: comparison of CERNO and eight other algorithms. Bioinformatics 35, 5146–5154 (2019).

32. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinform. 9, 559 (2008).

33. Shin, H. et al. TopDom: an efficient and deterministic method for identifying topological domains in genomes. Nucleic Acids Res. 44, e70–e70 (2016).

34. Chen, B., Khodadoust, M. S., Liu, C. L., Newman, A. M. & Alizadeh, A. A. Profiling tumor infiltrating immune cells with CIBERSORT. in Methods in Molecular Biology, vol. 1711. 243–259 (NIH Public Access, 2018).

35. Armañanzas, R., Calvo, B., Iñaki, I., López-Hoyos, M., Martínez-Taboada, V., Ucar,

E., …& Zubiaga, A. (2009, May 9). Microarray analysis of autoimmune diseases by machine learning procedures. IEEE Transactions on Information Technology in Biomedicine, 13(3), 341-350.

36. Fausett, L. (1994). Fundamentals of Neural Networks: Architectures, Algorithms, and Applications. Upper Saddle River, NJ: Prentice Hall. Fox, I., Ang, L., Jaiswal, M., Pp-Busui,

37. R., & Wiens, J. (2018). Deep multi-output forecasting: Learning to accurately predict blood glucose trajectories. 24th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Lupus Foundation of America. (2013, July 31). National Resource Center on Lupus. Retrieved from Lupus.org: https://www.lupus.org/resources

38. Ravenell, R., Kamen, D., Fleury, T. J., Spence, D., Hollis, B. W., Janech, M. G., …&

Almeida, J. S. (2012). Premature Atherosclerosis Is Associated With Hypovitaminosis D and AngiotensinConverting Enzyme Inhibitor Non-use in Lupus Patients. The American Journal of the Medical Sciences, 268-273.

39. Wiens, J., Guttag, J., & Horvitz, E. (2012). Patient risk stratification for hospital-associated c. diff as a time-series classification task. Proc. Advances in Neural Information Processing Systems.

40. Wu, J., Roy, J., & Stewart, W. (2010). Prediction modeling using EHR data: Challenges, strategies, and a comparison of machine learning approaches. Medical Care, 48(6), S106-S113.

41. Cheng B, Ning Y, Liang C, Li P, Liu L, Cheng S, et al. Genome-Wide Association Analysis Identified ANXA1 Associated with Shoulder Impingement Syndrome in UK Biobank Samples. G3: Genes, Genomes, Genetics. 2020; 10:3279-84.

42. Verma A, Ranga V. Machine learning based intrusion detection systems for IoT applications. Wireless Personal Communications.

43. Lundberg S, Lee S-I. A unified approach to interpreting model predictions. arXiv preprint arXiv:170507874. 2017.

44. Jung LC, Wang H, Li X, Wu C. A machine learning method for selection of genetic variants to increase prediction accuracy of type 2 diabetes mellitus using sequencing data. Statistical Analysis and Data Mining: The ASA Data Science Journal. 2020; 13:261–81.

45. Lu H, Zhang J, Jiang Z, Zhang M, Wang T, Zhao H, et al. Detection of Genetic Overlap Between Rheumatoid Arthritis and Systemic Lupus Erythematosus Using GWAS Summary Statistics. Frontiers in genetics. 2021; 12:389.

46. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, Behrens TW, et al. STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. New England Journal of Medicine. 2007; 357:977–86.

47. Dietterich TG. Ensemble methods in machine learning. International workshop on multiple classifier systems;2000: Springer;2000. p. 1-15.

48. Scott D, Wolfe F, Huizinga T. Rheumatoid arthritis. Lancet [Internet]. Elsevier Ltd 2010.

49. Tsokos GC. Systemic lupus erythematosus. N Engl J Med. 2011 Dec 1; 365:2110–21.

50. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS Genet. 2011;7: e1002254.

51. Cui Y, Sheng Y, Zhang X. Genetic susceptibility to SLE: recent progress from GWAS. J Autoimmun. 2013; 41:25–33.