Advancing Autoimmune Disease Treatment with AI-Assisted Gene Expression Analysis

Sarah Cross

Texas A&M Science and Engineering Fair

Table of Contents

Abstract

Over 50 million Americans have an Autoimmune Disease (AD). Diagnosis of symptoms is challenging and the available treatments are invasive and costly. As chronic diseases, ADs often significantly impact patients for the entirety of their lives, with treatments remaining costly, invasive, and frequently uncovered or under-funded by healthcare providers.

In this work, the researcher will show how state-of-the-art computational biology techniques combined with modern AI can be used to diagnose systemic lupus erythematosus (SLE) with 99.1% validation accuracy using a patient's gene expression profile. The primary genes used by the model to determine a patient's diagnosis (XIST, USP18, ALMS1, and AVIL) have been associated strongly in previous studies with lupus or cytokines directly tied to lupus, suggesting that the model's findings corroborate with existing research.

Existing traditional methods of detecting AD require 5 or more tests at significant cost. The advances in modern DNA gene expression panels that are used in this work reduce costs dramatically, and our proposed model is efficient enough to be run on almost any laptop. This work not only has the ability to significantly improve the affordability and availability of AD diagnosis on its own, but it also highlights a multitude of genes correlated to lupus, serving as a starting point by which to design targeted treatments.

Background

The cases of autoimmune diseases are rising, and we don't know why.

We don't know what factors have culminated to create an epidemic of malfunctioning immune systems, trigger-happy granulocytes crying wolf on a healthy cell system and inciting inflammation.

We don't fully know why up to 75% of cases are women, or how the environment and our habits are involved. Autoimmune diseases are shrouded in mystery, a mystery that can only be uncovered through intensive research and completely reorienting our perspective on inflammatory illnesses.

Fundamental challenges in the healthcare system result in autoimmune diseases (ADs) being under-prioritized, as research on ADs receives less funding than "killer diseases" such as cancer (Martin). ADs also often prevent themselves from common symptoms and are associated with feelings of ridicule, preventing patients from getting properly diagnosed until the disease becomes potentially life-threatening. Fad diets exploit patients uncertain of their symptoms, promising a fast fix to reduce inflammation when in reality inflammation is triggered by an underlying process that differs largely across patients and different ADs. Several websites such as the Gluten-Free Society go further, suggesting that a patient's diet may have resulted in the onset of AD, while experts such as Dr. Patrick Nachman strongly refute such statements, arguing that there is not enough evidence to imply causality and that such statements promote the "self-blame mindset," when patients believe that they brought their disease upon themselves through their own actions (Falk). Within such sites,

association and causation are blurred; for example, IBD significantly affects the body's ability to properly digest food and absorb nutrients, making patients that contract the disease more likely to suffer from malnutrition and vitamin deficiencies, regardless of their diet (Malnutrition). However, websites such as the Gluten-Free Society argue these deficiencies caused the onset of IBD, and adopting a nutrient-rich diet would have prevented the patient from becoming sick. On top of the disinformation, self-blame, and mistreatment of autoimmune disease patients, particularly IBD patients, existing medications to mitigate symptoms produce significant side effects, such as reducing bone density or lethargy. Although most, if not all, medications produce side effects, autoimmune treatments look to suppress the immune system, rather than remedying the core problem. As a result, the immune system continues to malfunction, although its responses are suppressed by medications, and patients are left unable to fight external viruses and diseases. Patients with autoimmune disease are often unable to take standard vaccines that use a weaker strain of a virus; what would be a simple task of recognizing and creating antibodies towards the virus for a healthy immune system becomes potentially dangerous for those with immune systems paralyzed by autoimmune medications.

Kurtis Baute, an environmental scientist turned full-time YouTuber, recently opened up about his diagnosis with Crohn's disease, explaining not only his symptoms but also the difficulties he faced that transcended his symptoms, and spoke of a larger problem within the healthcare system that prevented him from being diagnosed at an earlier stage. Before becoming diagnosed, he ate a strictly vegan diet and was able to

run marathons, contradicting the concept promulgated by sites that benefit from the self-blame mindset of those with ADs (Baute). Crohn's disease is a type of inflammatory bowel disease in which the gastrointestinal tract becomes inflamed, resulting in severe abdominal pain and distension. Although when looking back, he experienced signs as early as college, his symptoms began to increase in severity following a stress crisis as he became concerned about the environment and the crushing reality of climate change, punctuated by increasingly frequent visits to the ER in which he was told simply that he was experiencing an anal fissure and that it would eventually go away, which it did not. He was unable to receive the colonoscopy necessary for a diagnosis until several months after his symptoms became severe, not only because of COVID-19 being prioritized but also because the Canadian health system as a whole tends to disregard chronic illnesses (Baute).

Clinical literature, financed by companies and individuals, has trended towards researching acute illnesses, as they pass quickly and involve significantly less uncertainty. Since the 1950s, clinical literature has been unable to follow the increasing trends of chronic and autoimmune diseases, producing a healthcare crisis by which patients with chronic diseases are neglected by the healthcare system (Holman). Because research towards mitigating symptoms or finding a cure is inadequately financed, little work has gone into improving or replacing current medications, which produce significant side effects. For those with severe autoimmune diseases, the only options available are immunosuppressants, which prevent the immune system from reacting appropriately to any given threat. Much like chemotherapy,

immunosuppressants protect the body from internal threats, at the cost of critical functions such as preventing the body from outside threats such as viruses.

Ultimately, although there are several theories as to why autoimmune diseases are becoming more prevalent, only generalities have been found. Because around 20% of the population is currently estimated to have an autoimmune disease, significantly more than in the late 1900s, and the proportion of youth with ANAs, or antinuclear antibodies (a common biomarker correlated with autoimmunity), has grown even larger, it is likely that changes in lifestyle and the environment on a national level have catalyzed this exponential growth. While genetics play a role in the risk factor of an individual, they do not vary largely across one or two generations, making them unlikely candidates for precipitating the growth in AD ("Autoimmunity May Be Rising"). As a result, the National Institute of Environmental Health Sciences (NIEHS) has focused its research on lifestyle and the environment.

Another theory largely debunked as a major cause of autoimmune diseases is the chemical theory of molecular mimicry, first posed by Raymon T. Damian in the early 1960s. The theory states that pathogens, such as throat infections with the *Streptococcus pyogenes* bacterium, with similar epitopes as the host, resulting in the host's immune system learning to form antibodies for these epitopes and accidentally attacking similar epitopes that exist in the human body. Although the molecular mimicry theory is conceptually sound, infection with *Streptococcus pyogenes* is common, while complications are very rare.

Currently, the gap in funding is resolved by observing associations between well-funded, well-known diseases such as cancer, and autoimmune diseases. For example, Tiphaine Martin, a Bioinformatics and modeling researcher at the Mount Sinai School of Medicine, works with TwinsUK to analyze both cancer and autoimmune data. TwinsUK is a database containing 14,274 twins and extensive “clinical, physiological, and behavioral and lifestyle data, including biochemical and genetic data, available to researchers, as well as hundreds of phenotypes related to common diseases” (TwinsUK). Using the data available from TwinsUK, Dr. Martin found that various biomarkers should be used when identifying autoimmunity. She highlighted TPOAb responses, dysregulated immune features, and germline variants together aided in most accurately finding a positive NK-cell-mediated immune response. She also hypothesized that women are the most susceptible to AD, particularly autoimmune thyroid disease, because of their hormonal cycle and pregnancy. Pregnancy, by definition, is when a foreign body is being raised within one’s own body; as a result, the immune system’s response is subdued to prevent the rejection of a baby as a foreign substance.

The National Institute of Health (NIH) sponsors several large-scale research studies by the NIEHS and National Toxicology Program (NTP) to obtain what is likely to be the most critical aspect of finding a cure or effective diagnosis for autoimmune diseases: big data. Because the environment has been observed as a significant factor in the vast spread of ADs, finding hotspots where habitants experience autoimmune disease most frequently is essential to finding what about the environment has led to such a large jump in the number of cases. The NTP looks at associations such as the

correlation between organic mercury and AD, agricultural chemicals and rheumatoid arthritis, and genetic factors in autoimmune muscle disease, and has found that white blood cells, lifestyle, and hormones may have an influence on the onset and severity of AD, in addition to genetic and environmental factors. The NIEHS researches how the environment affects the development of autoimmune diseases, and highlights within this article several risk factors of developing AD: solvent exposure can lead to systemic sclerosis, smoking can lead to rheumatoid arthritis, UV radiation may lead to juvenile dermatomyositis, and genetic factors such as HLA DRB1*03:01 alleles. The NIEHS also has found associations between AD and other seemingly unrelated issues; for example, childhood poverty, measured by lower socioeconomic status and education level, was found to be positively associated with higher levels of rheumatoid arthritis. Using big data and large-scale research studies, associations can be made surrounding what genes tend to be up and downregulated amongst those with ADs, or, on a more general scale, between the environment that a patient lives in and the risk that it poses (Ramos-Casals). There is one omics database available for autoimmune diseases (compared to the hundreds available for cancer), called ADeX. The publicly available database contains 5,609 samples and 82 previous studies on transcriptomics and methylation for some of the most common autoimmune diseases. Upon analyzing the data, the ADeX researchers found that there were higher expressions of IFN (interferon) regulated genes in patients with AD compared to healthy people; however, the association was highly variable and should not be used as a factor for diagnosis. The researchers link ADs through their risk factors and molecular mechanisms, and

describe analyses that could be performed with their database, ADeX, for biomarker discovery; for example, they found that the Type II IFN has a key role in AD pathogenesis. The database offers additional resources for viewing simple information the database, exploring gene expression and methylation through gene queries, finding gene sets that fit the research being conducted, and several additional resources for meta-analysis (Martorell-Marugán).

The target protein that has proven to be the largest success in mitigating autoimmune diseases is tumor necrosis factor-alpha (TNF-alpha). TNF-alpha is a cell signaling protein involved in systemic inflammation, and its inhibition through immunosuppressants has proven to be highly successful. The most popular drug in curing ADs is infliximab, an antibody that binds and neutralizes TNF-alpha, the most widely known signaling protein involved in systemic inflammation. Baute found that, of the seven different medications he's been placed on, infliximab, paired with pain-reducing drugs such as CBD and Tylenol, has resulted in the most significant reduction in the severity of his symptoms.

The inhibition of TNF-alpha has also been suggested by several researchers to hold promise in creating an "anti-aging drug"; for example, Charles Mobbs is a neuroscience researcher and professor at the Mount Sinai School of Medicine, who founded a company, Gilga-Med, solely for the purpose of identifying small molecules that would be able to cross the blood-brain barrier and reduce inflammation in the brain to cure Alzheimer's disease by inhibiting TNF-alpha. The identification of small molecules to up or downregulate genes is a relatively unexplored field for autoimmune

diseases, and likely holds a lot of promise in finding an oral drug that produces fewer side effects than an immunosuppressant.

In conclusion, there are three features of autoimmune diseases that should be given significantly more focus: community outreach and awareness, accurate and simple diagnosis methods, and insurance coverage.

Wider insurance coverage is a major issue for autoimmune diseases. Kurtis Baute only received his diagnosis after dozens of visits to the ER, at which point standard treatments no longer worked and he was placed on the biologic therapy infliximab, which, because it is classified as a specialty drug, costs around $8,000 a year (“Remicade”); before the Affordable Care Act (AMA) and without insurance, the infusions used to cost up to $69,000 a year, which left most patients in an absurd amount of debt (Pinder). Biologic therapy compounds have alternatives, called biosimilars, that are chemically and functionally almost identical to the original compound; however, they tend not to be covered by insurance, although they aid in providing competition towards the original compound, driving profits down to a manageable rate. For example, infliximab has an alternative biosimilar, Avsola, that has dropped the original cost of the compound down to 33% of its original cost (Hagen). The creation of biosimilars not only creates a cheaper alternative, it prevents monopolies on treatments and reduces the original massively inflated prices to a manageable level for patients, making it more likely that biosimilars will eventually be covered by insurance as they are proven to be viable substitutes for the original compound.

Community outreach is another critical factor in creating awareness for vastly underdiagnosed illnesses, such as Crohn's disease and IBD. Diseases that involve areas considered taboo or less talked about, such as the colon, stomach, or bladder, tend to be associated with feelings of ridicule, preventing many from talking openly about their symptoms. Baute is one of a select few IBD patients willing to openly establish a dialogue with an audience online surrounding the difficult parts of his condition, such as the procedure he went through to drain a large abscess in his colon ("Surgery"); through placing more funding into institutions that create events to spread awareness and create a community, more patients, like Baute, are likely to open up about their symptoms, and more people with correlated symptoms are likely to seek help earlier. Informing the public further about the issue at hand is also at both the best interests of future patients and investment companies; diagnosis at an earlier stage prevents patients from being placed on more expensive medications.

Given estimates surrounding the large proportion of the population with an autoimmune disease, creating an environment by which the diagnosis of AD is no longer a last option decision is crucial to preventing patients from being diagnosed too late and placed on extremely expensive and dangerous medications. What makes autoimmune disease typically a last option decision for doctors is the lack of available methods to accurately diagnose AD. Crohn's disease, for example, requires a colonoscopy for diagnosis, which is typically not covered by insurance when used to diagnose Crohn's, and with the pandemic, was no longer considered an essential procedure, even in countries such as Canada. The symptoms themselves, bloodwork,

and additional procedures (varying on the type of autoimmune disease) are typically all required for diagnosis. Blood tests currently search for the levels of ANAs, which, while highlighted as a biomarker for autoimmunity, is not solely enough for diagnosis.

The truth is that we suspect what causes autoimmune disease. We know genes such as XIST contribute to females being disproportionately affected by AD. The problems lie not in what we don't know about AD, but in our unwillingness to go further in our exploration.

AD is unable to get proper funding because enterprises are unwilling to invest in a complex problem that will take years to solve, and could lead nowhere. While there is no simple answer, what is clear is that going from concepts to treatments is happening significantly slower than it needs to, perhaps suggesting the need for investment by a public entity less influenced by short-term demand.

AD diagnosis requires several tests, particularly during early diagnosis, when symptoms are less severe. Creating a simple, cost-effective diagnosis using not just a singular biomarker, but a collection of biomarkers analyzed using artificial intelligence methods would inform patients sooner and more effectively. Our perspective surrounding autoimmune disease needs to be shifted. Like research and diagnosis in this field, we fall dangerously far behind in our perception of AD, as the statistics attempt to make the seemingly impossible seem real.

Rationale

Up to 20% of adults are estimated to have an autoimmune disease (AD) or symptoms associated with AD. Existing diagnosis tools for AD involve invasive, costly procedures that many patients with symptoms cannot or should not undergo until their symptoms are severe. Fundamental challenges in the healthcare system result in AD treatment being under-prioritized, as research receives less funding than “killer diseases” such as cancer (Martin). Another significant aspect of severe AD symptoms that have recently come to light since the COVID-19 pandemic is patients’ inability to use standard vaccines, as what would be a simple task of recognizing and creating antibodies towards the virus for a healthy immune system becomes potentially dangerous for those with immune systems paralyzed by autoimmune medications. Several ADs, such as those regarding bowel movements, are associated with feelings of ridicule, another barrier preventing patients from getting properly diagnosed, and fad diets continue to exploit patients uncertain of their symptoms, promising a fast fix to reduce inflammation when, in reality, inflammation is triggered by an underlying process that differs largely across patients and different ADs. Ultimately, ADs are:

a) much less likely to have a “quick fix” compared to acute illnesses, reducing the number of researchers interested in the area and the amount of funding research receives,

b) associated with ridicule and shame with online sources such as the Gluten-Free Society promoting a “self-blame mindset” in which patients believe they brought the disease upon themselves,

c) much harder to treat, with current medications producing significant side effects such as reducing bone density and lethargy, as they often involve suppressing the immune system, preventing it from producing inflammation but leaving the patient unable to fight external viruses and diseases,

d) deprioritized by health systems, as healthcare insurance tends not to cover invasive diagnosis methods until symptoms become severe and potentially dangerous, even within highly ranked healthcare systems such as that of Canada,

e) very expensive, as biosimilars that drive down prices to manageable levels tend not to be covered by insurance. For example, the biologic therapy infliximab, which inhibits the production of TNF-alpha (the most widely known signaling protein involved in systemic inflammation), is classified as a specialty drug and costs around $8,000 a year; before the Affordable Care Act (AMA), and without insurance, the infusions used to cost up to $69,000 a year.

While there is no simple solution to the growing problem of autoimmune diseases, what is clear is that going from concepts to treatments is happening significantly slower than it needs to without pressure coming from investment entities or supportive healthcare systems. Ultimately, autoimmune diseases tend to be very complex, with hundreds of genes involved in the process of inflammation, making it difficult for standard research

methods to pinpoint a single gene associated with the onset of AD. Within this paper, the researcher proposes a method of computationally identifying genes correlated with an AD using neural networks, a promising, relatively new technology designed to capture the nuance of issues such as that of ADs, and using those genes to identify potential drug candidates.

Purpose

Is it possible to identify a series of genes correlated with lupus, and, if so, can those genes be used to identify drug candidates to mitigate symptoms of lupus?

Hypothesis

Potential drug candidates will be identified to mitigate symptoms of lupus, based on a gene profile analysis exploring the upregulation of downregulation of genes in healthy and sick patients. If successful, this process may hold important implications in approaching drug identification for other diseases.

The Code

The code for this project is available at https://github.com/html1101/SLE-Diagnosis.

Procedure

While the complexity of this project stems from being unable to pre-determine what implementation will best result in accurate results, the following general steps will be used when creating, analyzing, and testing models:

1. Collect gene expression profiles online across the following databases:

    a. Ensure that both healthy patients and patients diagnosed with lupus are accurately represented within the data.

    b. Clean the data, filtering out genes that were not measured across all databases.

    c. Transform data into a one-dimensional array, with a binary output (the patient was or was not diagnosed with lupus).

2. Create a neural network, inputting the one-dimensional representation of the gene expression data and outputting a prediction of whether or not a patient had lupus.

3. Prune the model, reducing the number of inputted genes to several biomarkers identified by the model and enforced by existing research as important in diagnosing lupus.

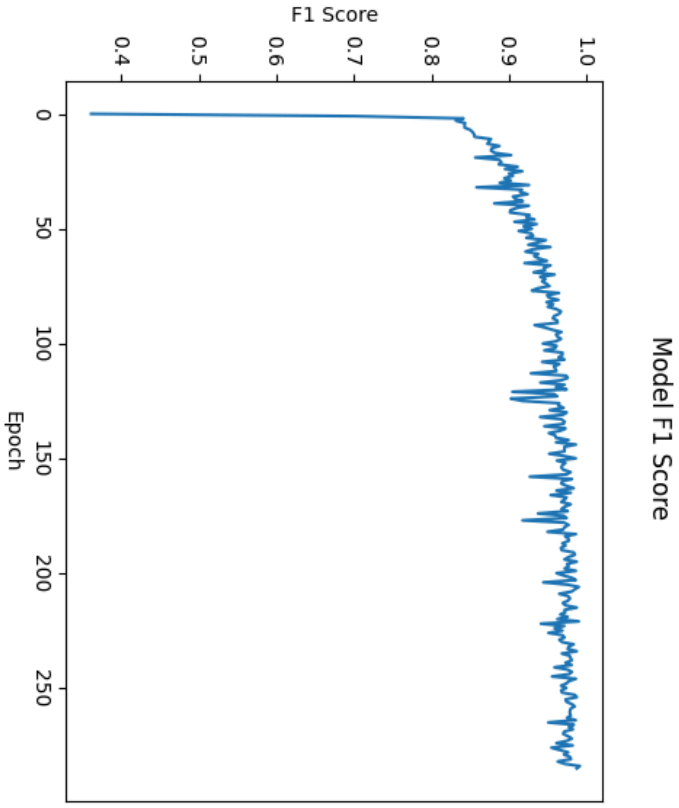
4. Explore genes identified as important, ensuring that there is existing research indicating a correlation and determining whether the gene is upregulated or downregulated in those with lupus.

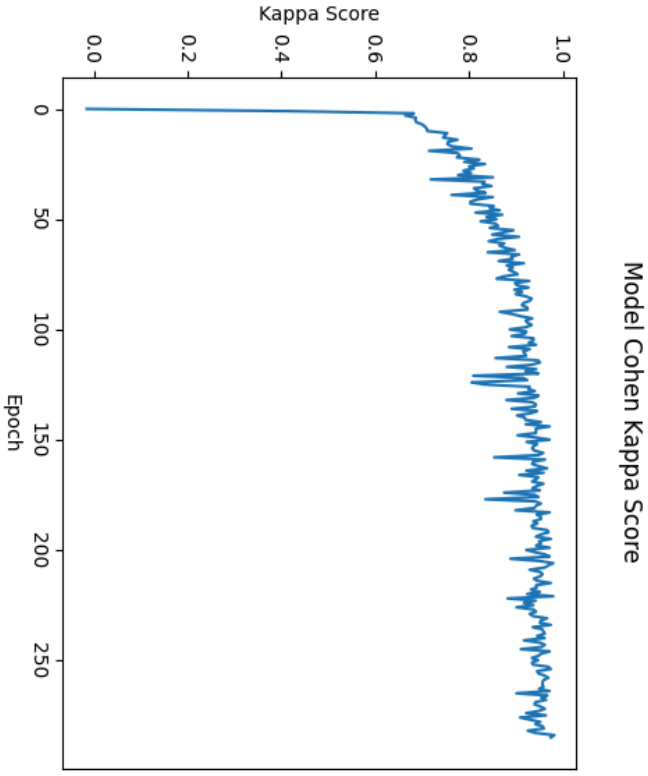
5. Input genes found into Clue to find small molecules that regulate the genes indicated.

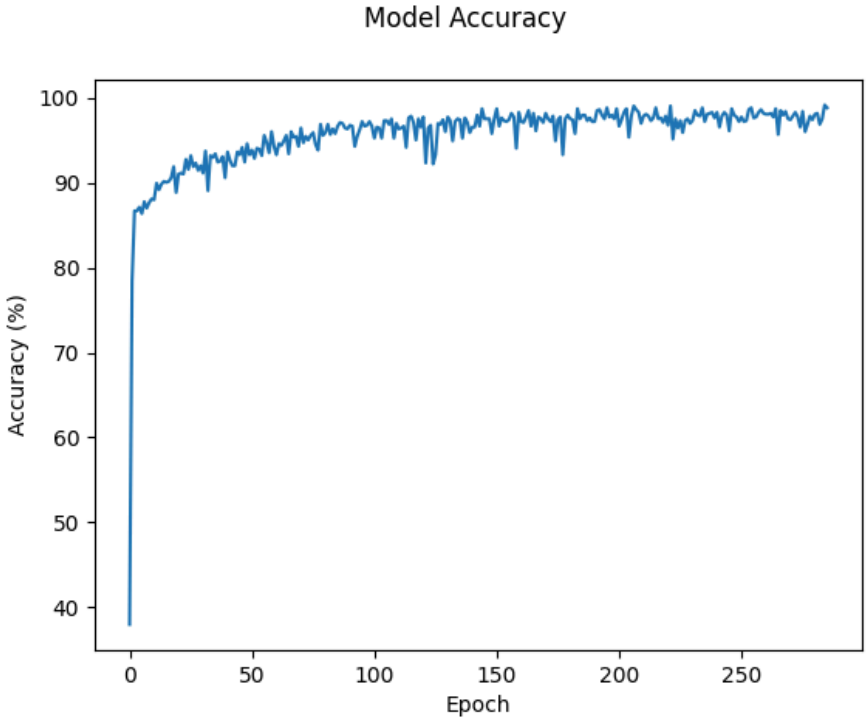
Materials

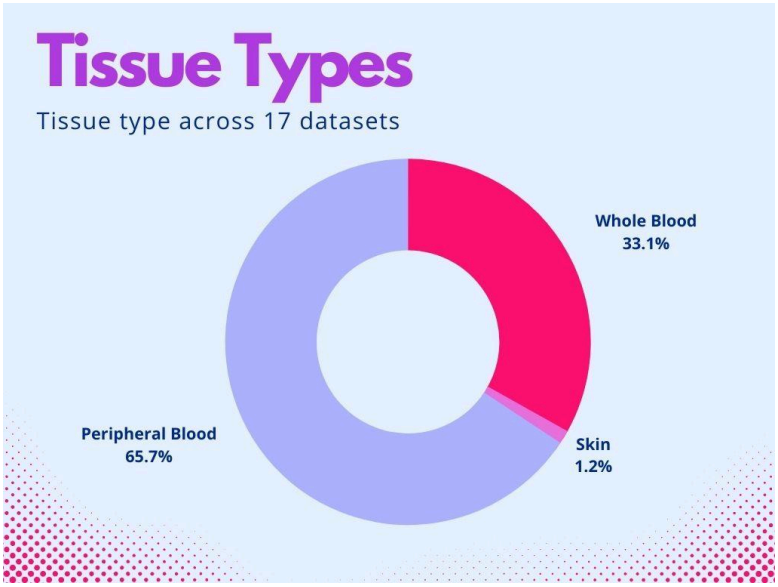
- Laptop or desktop computer (the researcher used a Ryzen 7 5800x for training). A minimum of 16GB RAM is recommended when training the model.
- Python (version 3.10.2) - See GitHub repository for list of libraries necessary; libraries can be installed by passing requirements.txt into Pip.

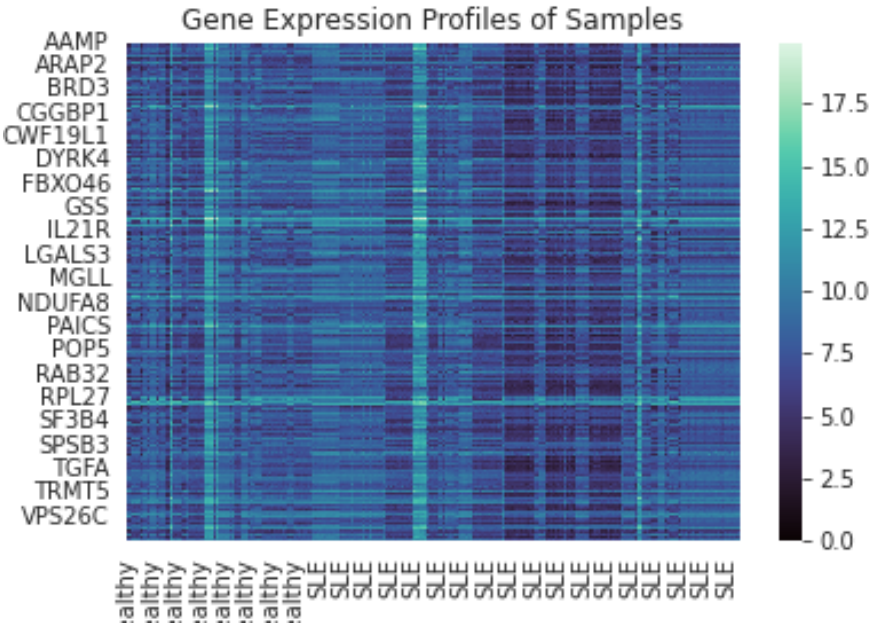
Data

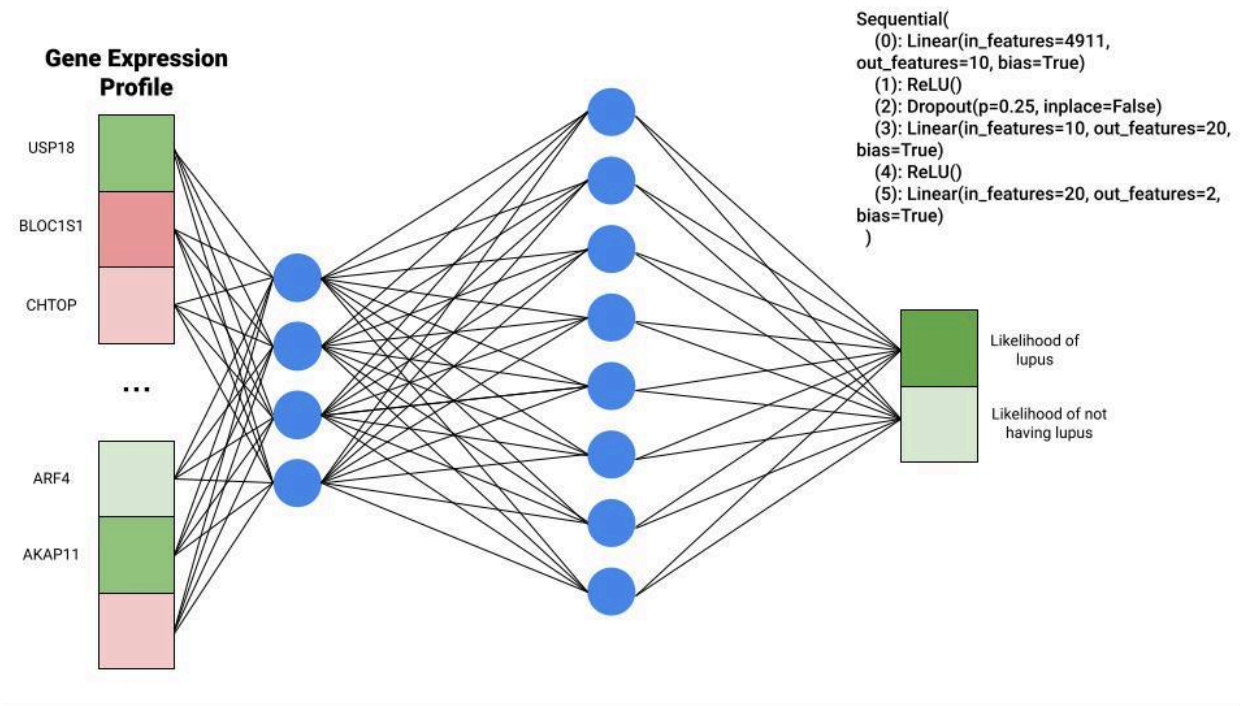

F1 Score
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Epoch
0
50
100
150
200
250
Model F1 Score


Kappa Score
0.0
0.2
0.4
0.6
0.8
1.0
Epoch
0
50
100
150
200
250
Model Cohen Kappa Score


Model Accuracy
Accuracy (%)
100
90
80
70
60
50
40
0
50
100
150
200
250
Epoch


Tissue Types
Tissue type across 17 datasets
Whole Blood
33.1%
Peripheral Blood
65.7%
Skin
1.2%


Gene Expression Profiles of Samples
AAMP
ARAP2
BRD3
CGGBP1
CWF19L1
DYRK4
FBXO46
GSS
IL21R
LGALS3
MGLL
NDUFA8
PAICS
POP5
RAB32
RPL27
SF3B4
SPSB3
TGFA
TRMT5
VPS26C
17.5
15.0
12.5
10.0
7.5
5.0
2.5
0.0
SLE



Model formatting:

```
  (linear_relu_stack): Sequential(
    (0): Linear(in_features=4911, out_features=10, bias=True)
    (1): ReLU()
    (2): Dropout(p=0.25, inplace=False)
    (3): Linear(in_features=10, out_features=20, bias=True)
    (4): ReLU()
    (5): Linear(in_features=20, out_features=2, bias=True)
  )
)
```

Conclusion

There are four features of autoimmune diseases that should be given significantly more focus: community outreach and awareness, insurance coverage, accurate and simple diagnosis methods, and more effective drugs to mitigate symptoms. Using a four-layer neural network, this project suggests that predicting SLE from a simple DNA microarray gene expression panel can be done with astounding accuracy. After training on 1,023 samples for 300 epochs using a cross entropy loss function, an Adam optimizer, and a Cosine Annealing Warm Restarts scheduler (serving to reduce the learning rate as the loss approaches a local minima), the model reached an accuracy of **99.13%**, with a Cohen Kappa score of 0.979 (suggesting that the model was making successful predictions, correcting for the probability of random chance). After being trained to diagnose SLE, the researcher tested the genes being used by the model for diagnosis. The genes that the model marked as most significant in lupus diagnosis were identified by predicting the chance of lupus in fake patients which would have only the gene in question significantly upregulated. The following three genes were found to most directly correlate with a diagnosis of lupus:

1. USP18 - A type I interferon receptor repressor and an isopeptidase; without USP18, mice were found to be hypersensitive to interferon signaling proteins released in the presence of viruses. Significantly higher levels of USP18 have been found in SLE patients compared to healthy controls, suggesting that the model successfully found a significant biomarker for diagnosing lupus (Demirkaya).

2. XIST - A non-protein-coding gene that maps to the X chromosome inactivation center, xic. The gene controls x chromosome inactivation and miRNAs in females, and is suspected to be a key factor for the sex bias in SLE. As a result of XIST overexpression on miRNAs, proinflammatory and proliferative pathways are upregulated, increasing the likelihood of contracting SLE (Bost).
3. ISG15 - A gene that encodes for an interferon-stimulated secreted protein. It stabilizes the levels of USP18, and has been found in significantly higher levels within SLE patients compared to healthy controls (Demirkaya).

The top 10 genes identified by the model have literature published surrounding their correlation with SLE, providing strong evidence that the model appropriately weighs genes associated with SLE during prediction. The data and scripts are available for public use at https://github.com/html1101/SLE-Diagnosis. Essentially, using preexisting data collected in gene expression profiles, a model was trained to diagnose lupus with a 99.13% accuracy, prioritizing genes already suspected to be involved in lupus without prior knowledge of such genes and using a set of algorithms to perform a more complex analysis than simply performing linear correlation. Not only does this highlight key biomarkers in autoimmune data, it creates a method to accurately leverage the expression of several genes for a significantly more accurate diagnosis than simply looking at a single gene. Using the genes identified, pertabagens that are capable of regulating the genes selected by the model could be identified and provide a potential oral medication to cure or mitigate the symptoms of SLE. This study also serves as a

basis by which methods to diagnose other autoimmune diseases can be developed in the same manner.

Practical Applications

This project could be applied in order to potentially diagnose patients suspected to have lupus in a single DNA microarray panel test, compared to the several tests currently needed in order to narrow down the problem to lupus. The process used within this project can very easily be expanded to other chronic illnesses whose symptoms stem largely from the increase in inflammatory cytokines caused by gene regulation. Not only does this highlight key biomarkers in AD data, it creates a method to accurately leverage the expression of several genes for a significantly more accurate diagnosis than simply looking at a single gene. Using the genes identified, perturbagens capable of regulating the genes selected by the model could be identified and provide a potential oral medication to cure or mitigate the symptoms of SLE.

Works Cited

"Autoimmunity May Be Rising in the United States." *National Institute of Environmental Health Sciences*, U.S. Department of Health and Human Services, May 2020, factor.niehs.nih.gov/2020/5/papers/autoimmunity/index.htm.

Bost, Chloé, et al. "Long Non-Coding RNA Xist Contribution in Systemic Lupus Erythematosus and Rheumatoid Arthritis." Clinical Immunology, Volume 236, 2022, 108937, ISSN 1521-6616, doi.org/10.1016/j.clim.2022.108937.

Demirkaya, Erkan et al. "New Horizons in the Genetic Etiology of Systemic Lupus Erythematosus and Lupus-Like Disease: Monogenic Lupus and Beyond." *Journal of clinical medicine* vol. 9,3 712. 5 Mar. 2020, doi:10.3390/jcm9030712.

Hagen, Tony. "Amgen 2020 Biosimilar Report Provides Market Share." *The Center For Biosimilars*, Center For Biosimilars, 19 Oct. 2020, www.centerforbiosimilars.com/view/amgen-2020-biosimilar-report-provides-market-share-asp-info.

Falk, Ron, and Patrick Nachman. "Getting Diagnosed with an Autoimmune Disease." *UNC School of Medicine*, www.med.unc.edu/medicine/news/chairs-corner/podcast/autoimmune-diagnosis. Accessed Jan. 2022.

Kegerreis, B., Catalina, M.D., Bachali, P. *et al.* Machine learning approaches to predict lupus disease activity from gene expression data. *Sci Rep* 9, 9617 (2019). doi.org/10.1038/s41598-019-45989-0

"Malnutrition and IBD." *Crohn's & Colitis Foundation*, www.crohnscolitisfoundation.org/diet-and-nutrition/malnutrition-and-ibd.

Martorell-Marugán, Jordi, et al. "A Comprehensive Database for Integrated Analysis of OMICS Data in Autoimmune Diseases." *BMC Bioinformatics*, vol. 22, no. 1, 2021, doi.org /10.1186/s12859-021-04268-4.

Osborne, Peter. "What Is the Root Cause of Autoimmune Disease?" *Gluten-Free Society*, 6 Dec. 2018, www.glutenfreesociety.org/what-is-the-root-cause-of-autoimmune-disease.

Pinder, Jeanne. "The True Cost of Crohn's Disease." *Clear Health Costs*, 10 Feb. 2020, clearhealthcosts.com/blog/2019/03/how-much-does-remicade-cost-the-true-cost-of-crohns-disease.

Ramos-Casals, Manuel, et al. "Google-Driven Search for Big Data in Autoimmune Geoepidemiology: Analysis of 394,827 Patients with Systemic Autoimmune Diseases." *Autoimmunity Reviews*, vol. 14, no. 8, 2015, pp. 670–679, doi.org/10.1016/j.autrev.2015.03.008.

"Remicade Prices, Coupons and Savings Tips." *GoodRx*, https://www.goodrx.com/remicade.

"Surgery for Crohn's Disease and Ulcerative Colitis." *Crohn's and Colitis Foundation*, Crohn's and Colitis Foundation, www.crohnscolitisfoundation.org/sites/default/files/legacy/assets/pdfs/surgery_brochure_final.pdf.

Tiphaine Martin. “Autoimmune Disease.” 25 Feb. 2022.

Works Consulted

Autoimmune Diseases and Your Environment. www.niehs.nih.gov/health/materials/autoimmune_diseases_and_your_environment_508.pdf.

“Causes of Autoimmune Diseases.” *Johns Hopkins Pathology*, pathology.jhu.edu/autoimmune/causes.

“Connectivity Map.” *Clue*, 1.1.1.43, Clue, clue.io.