

Detection of Autoimmune Disorders using Machine Learning

Vivek Mohali¹, Priyanka Mohali², Dr. Nishtha Hooda³

^{1,2,3}Computer Science and Technology, Chandigarh University, Mohali

Article Info

Volume 82

Page Number: 16730 - 16735

Publication Issue:

January - February 2020

Abstract

Machine learning is called future of computer science. Machine learning is being extensively used in research fields. One of such fields is Health Care sector. Machine learning algorithms are being used for prediction and detection of diseases. Autoimmune disorder is a class of diseases in which immune system attacks healthy body tissues and cause inflammation. Discoveries in the last decade have made research in autoimmune disorders one of the most promising areas [1]. These diseases are one of the most widespread diseases in modern world. These diseases are considered chronic because they are known to be aggressive to healthy body tissues. And hence, early diagnosis and treatment are considered vital for a good life. Now days, a lot of health related systems are being implemented using machine learning and artificial intelligence like fitbits, mobile apps providing a more comprehensive view of a patient than has been ever seen before. These tools along with other signs on body components like blood, organs etc. can be used to determine if patient is suffering from an autoimmune disorder or not. And hence treatment can be started at early stages. This work will help us for early detection of autoimmune disorders based on these factors. We have used GINI index decision tree for predicting whether a person is having Rheumatoid Arthritis or not.

Article History

Article Received: 18 May 2019

Revised: 14 July 2019

Accepted: 22 December 2019

Publication: 29 February 2020

Keywords; Machine learning, GINI Indexing, Rheumatoid Arthritis, Autoimmune disorders, Decision tree, Disease classification, Osteoarthritis.

I. INTRODUCTION

Autoimmune disorder is a class of diseases which contains diseases like Rheumatoid. Arthritis, Inflammatory bowel disease, Psoriasis, Multiple Sclerosis, Type 1 diabetes (juvenile arthritis) etc. In this kind of disease, immune system attacks healthy body tissues and cause inflammation. These diseases are one of the most prevalent diseases in modern world. Now a day, a lot of health related systems are implemented using machine learning and artificial intelligence like fitbits, mobile apps providing a more comprehensive view of a patient than has been ever seen before. These tools can also be applied to determine patients heading towards a diagnosis of a chronic disease, those who are misdiagnosed and how effectively patients diagnosed with an illness are managed.

The illness flares-often without any known or obvious trigger -and then gets better, and then may flare again. Few patients worsen quickly, while others can live for years before the more severe symptoms arrive. The fog of unpredictability makes it difficult for the rheumatologists to select the best treatment plans.

Because most autoimmune disorders are chronic and incurable, so, early detection can be helpful in preventing more damage to the body. These diseases are found in people of every age group and of both sexes. The autoimmune diseases depend on multiple factors, and their risk factors can be genetic as well as environmental.. Combination of these risk factors varies from one person to another, giving lot of different medical profiles.

Rheumatoid Arthritis is a systematic autoimmune disease in which immune system attacks body tissues especially near the joints due to which a person may experience chronic pain. It is considered chronic and disabling disease. Early detection is considered for lesser damage to the body because once damaged is caused, it can't be reverted. Based on physical test, inflammation symptoms, laboratory tests and imaging, we can easily identify RA. So, we have used few of these results for predicting if a person is having RA (non-inflamed or inflamed) or not.

II. RELATED WORK

Research on autoimmune disorders is not new but it is not much explored like cancer, heart diseases, etc. But recent research outcomes have made area of autoimmune research one of most promising areas for new discoveries. NIH (National Institutes of Health, USA) estimates up to 23.5 million Americans have an autoimmune disease. [2] When we compare it, cancer only affects 9 million people and heart disease nearly 22 million.[2] With increasing funding in healthcare sector, research field is evolving. Study done by Orange et al.[3] identified three subtypes on the basis of synovial gene from patients suffering with Rheumatoid Arthritis. The outcomes produced were used for designing a scoring algorithm. These scores were related to vital parameters like Sedimentation rate, C-reactive protein level (CRP), and other autoantibodies[3]. The researchers selected 14 features from given synovial samples and about 500 best expressed gene samples in 45 synovial samples. Three groups with high inflammation, low inflammation, and mixed subtype, were described based on respective genes pattern. The focus of their work was to find connection among synovial features and genomic subtype and thus put forward an approach for characterization of synovial tissue. The model's performance in separating both subtypes from each other was 0.88 and 0.71, respectively. The subtypes defined were closely

related to clinical features, as high sedimentation rate, C-reactive protein levels, rheumatoid factor, and anti-C.C.P. level in patients with high inflammation scores were observed. But this method was not reliable because it was based on Support Vector Machine which can be over fitted [4], the number of samples was very small. Also, Support Vector Machine is an unsupervised Machine Learning algorithm.

Maecker HT et al.[5] in their study, proposed new methods that may be used in near future for immune phenotyping analyses: mass cytometry, peptide and protein arrays, B-cell receptors sequencing and T-cell receptor sequencing. With the help of new methods of studying gene expression in blood samples and for deconvolution of the result data allow dividing samples in different subsets, which offered a much better approach of looking for bio markers. From these new statistical approaches, lists of more active biomarkers will indubitably be extracted that can be helpful in diagnosis, clinical sub dividing as well as for selection of best treatment plan. Since immune system is considered very complex and the high amount of activity in body, bio markers are not only for diagnosis but also for responsible for severity of disease. They recommended these techniques which may prove practical and useful.

In a study by Ismail Saritas et al.[6], artificial neural networks (ANN) machine learning method were used for diagnosis of rheumatoid arthritis (RA). For this purpose, backpropagation algorithm was examined (artificial neural networks). The data which was received from the patients having RA and normal population was classified using this method. Later on, artificial neural network's backpropagation algorithm's result and the results using Perceptron algorithm was compared for performance. ANN produced 82% accuracy using Backpropagation method while Perceptron method produced 71% with accuracy.

III. MATERIALS AND METHODS

1. Experimental setup :For this study, dataset provided by AMP Rheumatoid Arthritis phase 1 was used. This dataset consists of detailed study files.(available on ImmPort.org [7]) This original data collected consist of

- Study ID, subject ID, visit ID & site ID
- Sex
- DOB
- Date of diagnosis, and if early RA start of persistent symptoms (month/year)
- Date of visit
- Joint sampled
- Type of procedure
- Smoking status (pack-years, if applicable)
- ESR
- CRP
- RF
- Anti-CCP
- Complete blood count
- Medications
- RA-specific drugs: ever/never taken, current therapy or tried & stopped
- 28 Tender & Swollen Joint Count
- Calculated Disease Activity Measures
- DAS-28
- X-ray

Dataset :We have used dataset provide by NIH(National Institutes of Health, USA). NIH together with NHS(National Health Services, England) in 2016, collected dataset from various patients. This study was named as AMP Rheumatoid Arthritis Phase 1. The primary goal for RA

arthroplasty P1 studies are: To setup if molecular signatures and pathways identified using core AMP technologies vary in the middle of OA and RA in 20 RA surgical samples as well as 10 OA arthroplasty samples. This dataset is available on ImmPort.ord with accession number SDY998. It consist of various blood test results like clinical data, flow cytometry, immunobiology tests, CBC reports, etc. It also consist of cell RNA sequencing. We have used biological factors only.

2. Algorithms &Techniques : In our dataset, we have a lot of features. So, feature selection is very important. For feature selection, we have used Gini Indexing. Gini impurity evaluate the degree or the probability of certain variables to be wrongly categorized when it is selected. But what we actually mean by an ‘impurity’? If all of the elements be affiliated to a single class, then it is called pure. The value of Gini impurity lies between 0 and 1, where 0 signify that either every elements is from a particular class or there is only one class, and 1 signify that the all elements are randomly distributed across different classes. A Gini Index with value 0.5 denotes equal distribution of elements in some classes.

Formula for Gini Index

$$\text{Gini} = 1 - \sum_{i=1}^n (p_i)^2$$

where pi stands for probability of any object to be categorized in any specific class. While constructing a decision tree, we choose selection of attribute/quality with the minimal Gini index for root node.

IV. PROPOSED WORK

We can collect enough data to classify different diseases based on machine learning algorithms for checking patterns in problem a patient having.

Let’s take an example – A patient is having joint pain. We collected following data:

- Patient history
- Blood Test data
- Pain related issues
- Imaging data
- Wearable device data
- Genetic data
- Antibodies

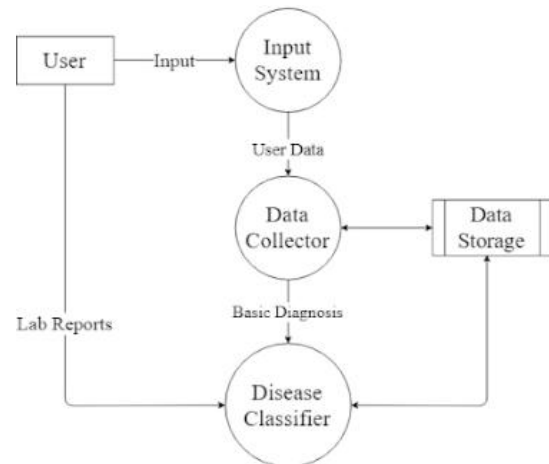
The patient input collected is as follows:

- Joint pain in almost every joint but mainly in smaller joints like knuckles, lower back, knees and neck.
- There is no physical injury that could have caused these problems.
- Pain and joint stiffness is more in the morning.
- Blood tests show signs of inflammation in body like High sedimentation rate, high CRP.
- Patient is feeling fatigue for a while.

Based on the diagnostic algorithms, the computer can easily calculate the diagnosis based on some patterns. In above case, we can see that person is suffering from a chronic illness, which is causing joint pain, this condition is most probably Rheumatoid arthritis. If we do some imaging procedures like CT scan, x-ray and a couple of more blood sample, we can easily examine for joint damage and effect on whole body. And hence, with more data we can be more accurate.

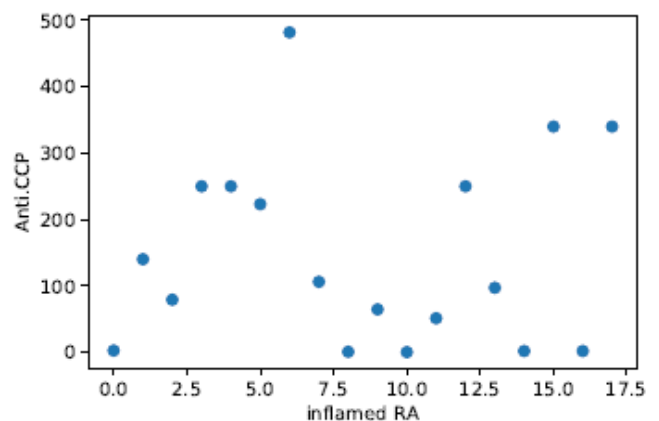
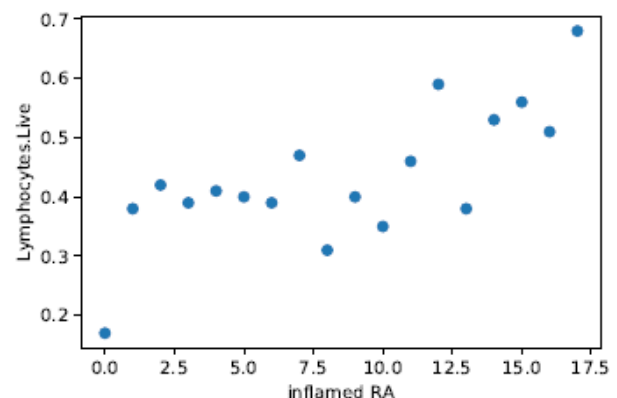
And the patient has not been seen by a physician let alone a rheumatologist (Experts in autoimmune diseases). Rheumatologists are mostly found in big cities. And number of rheumatologists in developing countries is low compared to people having these diseases. These types of machine learning algorithms are not new but they're not widely explored in a lot of fields. Various machine learning algorithms like Decision trees and pattern

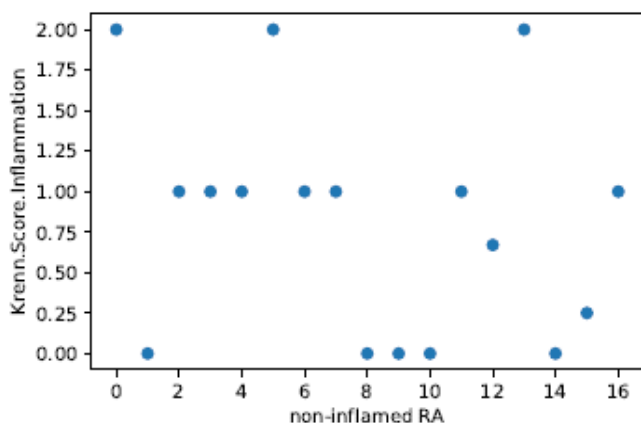
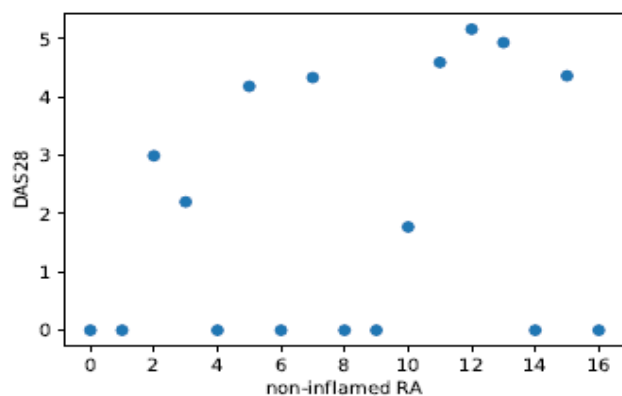
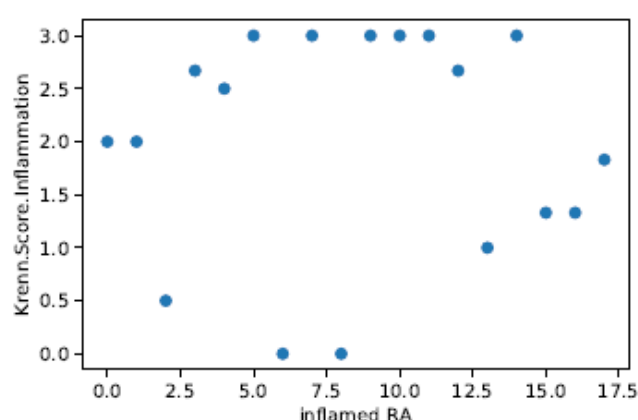
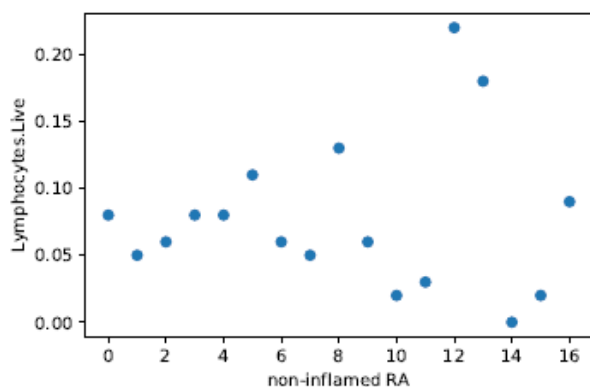
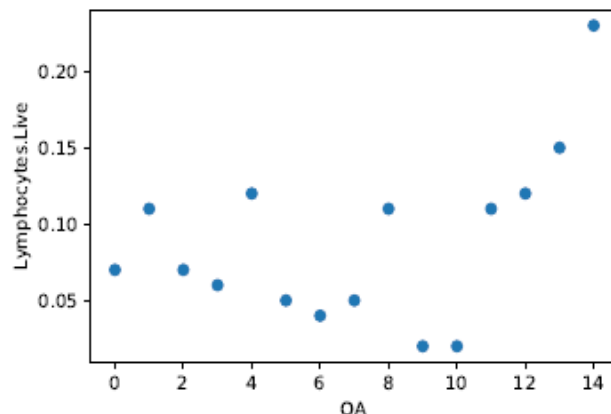
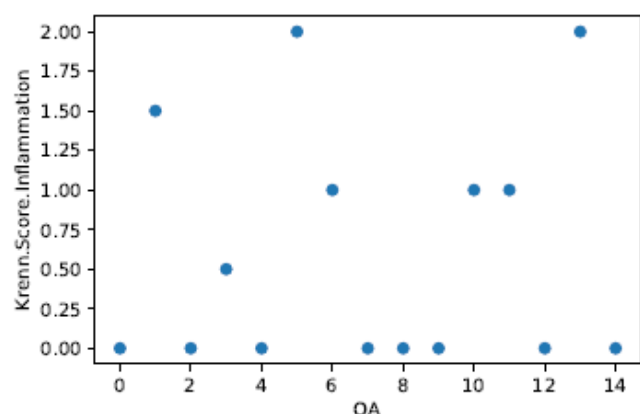
recognition. So, with good dataset and application of machine learning will make diagnosis easier, treatment much better and better life style with little damage.



V. RESULTS AND DISCUSSIONS

Using GINI Indexing, we selected the best attributes for decision making.





These plots show difference between various values in values for Inflamed RA, non-inflamed RA and OA.

Like Lymphocytes for inflamed RA are high as compared to non-inflamed RA and OA. So, this is the prime attribute for differentiating it from others. And similarly DAS28 can be used for differentiating non inflamed RA from RA.

REFERENCES

1. National Institutes of Health THE AUTOIMMUNE DISEASES COORDINATING COMMITTEE
REPORT www.niaid.nih.gov/sites/default/files/adccfinal.pdf
2. Orange DE, Agius P, DiCarlo EF, et al. Recognition of three rheumatoid arthritis disease subclassify by machine learning integration of synovial histologic features and RNA sequencing data. *Arthritis Rheumatoid*. 2018;70:690–701.

3. Noble WS. What is a support vector machine? *Nat Biotechnol.* 2006;24:1565–1567.
4. Han H, Jiang X. Overcome support vector machine diagnosis overfitting. *Cancer Inform.* 2014
5. Maecker HT, Lindstrom TM, Robinson WH, et al. New tools for classification and monitoring of autoimmune diseases [published correction appears in *Nat Rev Rheumatol.* 2012 Oct;8(10):562]. *Nat Rev Rheumatol.* 2012;8(6):317–328. Published 2012 May 31. doi:10.1038/nrrheum.2012.66
6. The Usage Of Artificial Neural Networks Method in the Diagnosis of Rheumatoid Arthritis DO - 10.18201/ijisae.2016426382
7. ImmPort [SDY998]
(<https://www.immport.org/shared/study/SDY998>)
single-cell RNA-seq, mass cytometry, bulk RNA-seq, flow cytometry, clinical and histology