Human Immune Prediction using Machine Learning

Jayant Deshmukh [1], Sanket kaspate [1], Umesh Khot [1], Rajan Saroj [1], Ms. Sunita Nandgave [2]

[1] *UG Students,* [2] *Assistant Professors, Computer Engineering Department,*
G H Raisoni College of Engineering and Management, Pune

ABSTRACT

Machine Learning appears as one of the key features to derive information from corporate operating datasets. Machine Learning in medical health care evolving as an enormous research field for delivering deeper understanding on medical data. Most methods of machine learning depend on several features defining the behavior of the algorithm, influencing the output, and thus the complexity of the resulting models either directly or indirectly. Many Machine Learning methods are utilized within the past to detect diseases. Prior detection of disease or low immune can help person to reinforce immune so on guard our body from harmful germs and viruses. Process is completed with the assistance of things like RBC counts, WBC, HB, MCV, MCH, MCHC, Neutrophils and platelet count. Observing and identifying of immune isn't an easy task to do manually because it takes a lot of time and effort. So it's easier to predict immune with an automatic processing and Machine Learning System. Different types of knowledge processing and Machine Learning algorithms are available to make best algorithm like rectilinear regression, Logistic regression and KNN algorithm. Various machine learning algorithm are often used to predict, so that the immune can often predicted easily and clearly because accuracy is vital in machine learning algorithm. This paper presents a study of various methods of predicting human immune.

Keywords—Machine learning, KNN, Immune, Logistic regression

I. INTRODUCTION

Immunity is resistance to disease, specifically communicable diseases. Human system is collection of cells, tissues, and molecules that mediate resistance to the infections, and therefore the coordinated reaction of those cells and molecules to infectious microbes comprises an immune reaction. Immunology is the study of the immune, including its responses to microbial pathogens and damaged tissues and its role in disease. The importance of the system for health is shown by the frequent observation that individuals with defective immune responses are too serious, often life-threatening infections. Predicting immune responses against through machine learning algorithm is the best method for shielding individuals against infections. After spread of diseases like aids and covid-19 humans are conscious about their health. And now due to corona pandemic immunity is at the top most priority within the health sector as finding vaccine is difficult till that keeping immune can help people to live longer. The human immune contains multi layered mechanisms of responding to cellular stress and tissue damage to make sure defense from pathogens, maintenance of tissue homeostasis, and integrity of the holobiont. In body every cell plays different role, but some of the white blood cells are particularly more important for immunity.

Immunology might be a sophisticated subject to review for several reasons. There are many details, complicated study and usually these details get within the approach of understanding the ideas and it's difficult to know the deep detail convictions. As learning biological things for common people are often difficult task. This is often an enormous issue else immunologists love these exceptions; as a result they work on clues and furnish theme. Another issue is there associated with the new studies which are becoming out on the regular basis. Some fact which is stated correct today is often claimed false tomorrow which is challenging due to this continuous updating is required. Something stated correct today are often wrong tomorrow that's why continue updations with perfect knowledge is must.

Prediction is important from medical aspect, that's why immune must be predicted. During approach, immune of the person is predicted using machine learning algorithm has been given. An scientist or hematologists might subdivide this further, there are some essential factors which we will be use for classification. In physical body proteins are conduit through which cells interchange with one another. Receptors are proteins blind ligands which work as receptors for the opposite cells. Cells had differing types of receptors on other cells. They perform work like transporting glucose into the cells. The receptors associated with the system square measure usually are involved interrogating the atmosphere for proof of danger, infection, and abnormal death. Within the course of associate in nursing immune response, cells follow a way, overall outcome maximizes the probability of living and eliminating the infection. Complications arise once the system doesn't work properly. Some problems arise unit smaller, like spore allergy, whereas others arise unit vital, like genetic disorders that wipe out the presence. This disorder is handled by analyzing the immune and dealing consequently.

II Literature Survey

Lot of research is happening regarding system. Now as of due to corona pandemic it's important to know our immune and work consistent with that. There are some factors through which we will predict immune. We will use those factors and form machine learning module which may perfectly predict health system of person. By studying the various research papers we

get some information about - three sorts of immune diseases: (a) allergies, (b) autoimmune diseases, and (c) infectious diseases. We compared the representative machine learning algorithms respectively from three different categories, namely decision tree learning, instance-based learning, and maximum margin learning machine learning techniques to construct accurate classification models for 3 sorts of immune diseases, allergy, autoimmune disorder and communicable disease , caused by different protein antigens et al [1] ORTALLER is a web allergen classifier supported allergen family featured peptide (AFFP) dataset and normalized BLAST E-values, which establish the featured vectors for support vector machine (SVM)[2].After lot of research ready to we'll state that algorithms are laid out in different papers but there's need of actual implementation then only we will be able to use the web applications.

II. PROPOSED SYSTEM

METHODOLIES TO RECOGNIZE IMMUNE PREDICTION

Data set description:

Data is key concept in machine learning without data it's impossible to predict or implement any machine learning algorithm. Immune prediction datasets are also available. The Immune Detection Dataset (IDD) is the only public dataset to the best of our knowledge. The data set consist of RBC, WBC, HB, MCV, MCH. As for machine learning algorithm 75% of data from dataset is used for training purpose and rest of 25% of data is used for testing purpose.

Block Diagram Methodology :

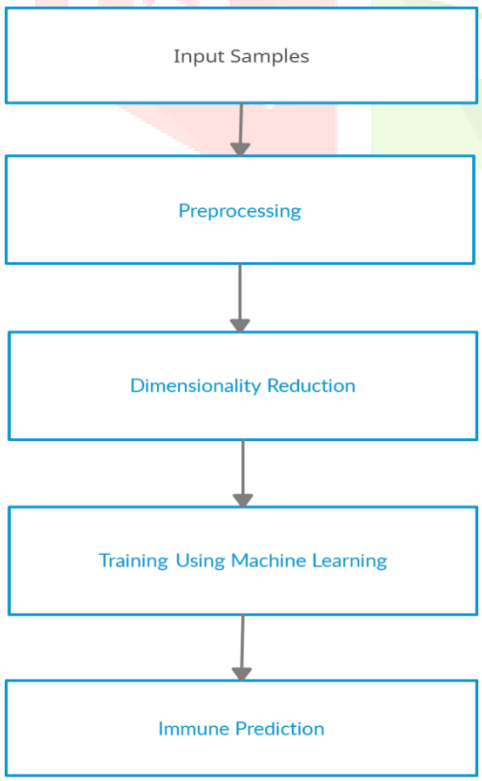

Fig 1. Block Diagram for Immunitizer (Application)

A. Input Samples

Parameter includes RBC count, WBC count, HB, MCV, MCH, MCHC, Neutrophils and platelet count which are given by user by anlysing blood report.

B.Preprocessing

Data preprocessing is one of the necessary steps in Machine learning as quality info and therefore the helpful information that may be derived from it directly affects the flexibility of our model to find out, we must tend to preprocess our knowledge before feeding it into our model. Preprocessing includes Handling Null Values, standardization, Handling construct variables, and multiple regressions. Universal dataset consists of some null values. It doesn't very matter whether or not it's regression, classification, or the other reasonably drawback, no model will handle these null values on its own thus we'd like to intervene.

Standardization is another integral preprocessing step. In standardization we tend our values such as average ought to be zero, and deviation ought to be one.

C.Machine learning algorithms

1) **Logistic Regression:**

In terms of its name, logistic regression is not a regression model but a linear classification. Logistic regression is recognized in literature as logistical regression. In this model, a logistic function is used to predict the probabilities that characterize the future consequence of a single test. The logistic regression in scientific learning can be approached from the Logistic Regression framework. For optional L2 or L1 regularization this design will work in a multi-class (one vs rest) logistic regression. In order to optimize the logistic regression penalized to binary class L2, the following risk function minimizes:

$$\min_{\omega,c} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{n}\log(exp(-y_i(X_i\omega + c)) + 1$$

Likewise, the L1 regularized logistic regression addresses the following problem of optimization:

$$_N \min k\omega k_1 + C\,\mathrm{X}\log(exp(-y_i(X_i\omega + c)) + 1)\ _{\omega,c\ i=1}$$

2) **K-means Clustering:**

In order to separate samples in n classes that have the same variance, the K- Means algorithm cluster results, minimizing the inertia criterion or using a square number. The number of clusters to be defined is needed for this algorithm. This applies well to a huge number of samples and is used in several different areas of use. The k-means algorithm partitions a set of observations N through K disjoint clusters C. Each is

the average sample of μ_j in the cluster. The mean square often referred to as" centroids;" usually they are not X points, although they reside in the same space. The K-means algorithm attempts to identify centers within a cluster sum that reduces inertia or squared criteria.

3) K-Nearest Neighbors (KNN):

kNN is used for classification and inference by a non-parametric approach of pattern recognition. The input is the closest example of k in the feature space in both cases. kNN implies an instance-based learning or lazy learning that only approximates the function locally and delays all calculations until classification is completed. Also for grouping as well as for estimation, weight can be assigned to the neighbors' inputs so that the near neighbors create a more reasonable contribution than the farther ones.

4) **support-vector machines :**

In machine learning, support-vector machines unit supervised learning models with associated learning algorithms that analyze information for classification and statistical procedure. Developed at AT&T Bell Laboratories by Vladimir Vapnik with colleagues , SVMs unit one in every of the foremost robust prediction methods, being supported applied mathematics learning frameworks or VC theory projected. Given a set of work examples, each marked as happiness to one of two categories, Associate in Nursing SVM work formula builds a model that assigns new examples to one category or the alternative, making it a non-probabilistic binary linear classifier .prediction strategies, being supported applied math learning frameworks or VC theory projected.. SVM maps coaching examples to points in area thus on maximise the dimension of the gap between the 2 classes. New examples square measure then mapped into that very same area and foreseen to belong to a class supported that facet of the gap they fall.

In addition linear classification, SVMs will efficiently perform non-linear classification to victimization that is known as kernel trick, implicitly mapping their inputs into high-dimensional feature areas..

When knowledge square measure unlabeled , supervised learning isn't doable, Associate in Nursing an unsupervised learning approach is needed, that makes an attempt to search out natural cluster of the info to teams, and so map new knowledge to those fashioned teams. The support-vector clustering[2] algorithmic rule, created by Hava Siegelmann and Vladimir Vapnik, applies the statistics of support vectors, developed within the support vector machines algorithmic rule, to reason unlabelled knowledge, and is one amongst the foremost wide used cluster algorithms in industrial applications.

D.Evaluation Metrics

The classifiers will produce the percentage of the immune type. The routine assessment of the result dependent on accuracy values.

E.Design and Implementsion of user interface

After perfectly implementation of machine learning module. Next step can be creating a user interface so that user can use machine learning module without any technical knowledge for that various possibilities are there. Developer can develop web based application or desktop application according to the requirement. For the web based application front end scripting knowledge is must. Further developer can use bootstrap to make it easier. Framework such as django and flask are also required for zipping all different modules. Various vendors are available for deployment of project in the final stage.

III. RESULT & DISCUSSION

Machine learning recognition systems are used to identify immune in patients and healthy people. In the present paper, we address briefly some popular classification algorithms and their theoretical context.

Here is comparison of Machine learning Modules with their prediction percentages as of dataset

.

Classifier	Accuracy
Kmean	13.13
svm_sigmoid	44.44
KNN_1	98.98
KNN_3	97.97
KNN_5	97.97
KNN_7	97.97
svm_rbf	67.67
Kmeans	27.27
SGD	67.67
MLP	67.67
svm_poly	81.81
svm_linear	97.97

Fig.2 Algorithm with Accuracy

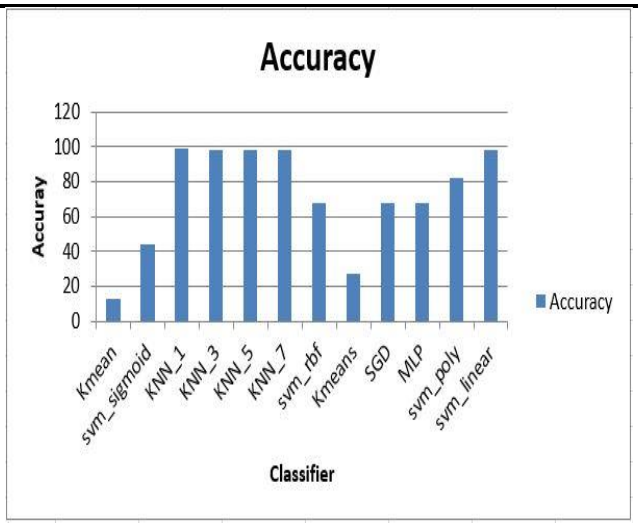

Fig.3 Algorithm with Graphical representation

IV. Conclusion

In this paper, the machine learning algorithms for Immune prediction has been reviewed. It is observed that only one dataset i.e. Immune Detection Dataset is available publicly for research. There is plenty need for developing and updating immune based dataset. The most of the existing work were implemented using machine learning algorithms. Mostly implementation work is pending due to dataset sufficiency. Recently, it is observed that deep learning algorithms are used which are much more accurate.

In future, there is a scope for development of dataset for immune detection all around the globe. Also, there is need of implementation of highly accurate classifiers which can be invariant in term of predicting perfect immune.

References

[1] Kuan-Hui Lin, Yuh-Jyh Hu, "Application of Machine Learning to Immune Disease Prediction", International Journal of Engineering and Innovative Technology (IJEIT) Volume 7, Issue 11, May 2018

[2] Lida Zhang, Yuyi Huang, Zehong Zou, Ying He, Ximo Chen, Ailin Tao, "SORTALLER: predicting allergens using substantially optimized algorithm on allergen family featured peptides", *Bioinformatics*, Volume 28, Issue 16, 15 August 2012, Pages 2178–2179.Barbedo, J. G. A., & Godoy, C. V. (2015). Automatic classification of soybean diseases based on digital images of leaf symptoms. In Embrapa Soja-Artigo em anais de congresso (ALICE). In: CONGRESSO BRASILEIRO

[3] Xu, J., Jo, J. Immunological Recognition by Artificial Neural Networks. *J. Korean Phys. Soc.* **73,** 1908–1917 (2018).Muhammad Thaqif bin MohamadAzmi and Naimah Mat Isa, "Orchid disease detection using image processing and fuzzy logic," International Conference on Electrical, Electronics and System Engineering, IEEE, 2013.

[4] H Ai JShen JWang, "An Improved Artificial Immune System-Based Network Intrusion Detection by Using Rough Set", *conference of Communications And Network Published Online*, February 2012.

[5] M Middlemiss, "Framework for Intrusion Detection Inspired by the ImmuneSystem", 2005,

[6] J Chen, D Yang and M Naofumi, "A Study of Detector Generation Algorithms Based on Artificial Immune in Intrusion Detection System", *WSEAS TRANSACTIONS on BIOLOGY*, vol. 4, no. 3, pp. 1109-9518, March 2007.

[7] L. N. De Castro and F. J. Von Zuben, *Artificial immune systems: Part I basic theory and applications*, 1999

[8] D. Dasgupta and F. González (Gonzalez), "An immunity-based technique to characterize intrusions in computer networks",

[9] J. Y. Le Boudec and S. Sarafijanovic, *An artificial immune system approach to misbehavior detection in mobile ad hoc networks*, Sep. 2003.

[10] Learning Ensembles of Neural Networks by Means of a Bayesian Artificial Immune SystemPablo A. Dalbem Castro;Fernando José Von Zuben

[11] P. A. Heidenreich, J. G. Trogdon, O. A. Khavjou, J. Butler, K. Dracup, M. D. Ezekowitz, E. A. Finkelstein, Y. Hong, S. C. Johnston, A. Khera, et al., "Forecasting the future of cardiovascular disease in the united states: a policy statement from the american heart association," Circulation, vol. 123, no. 8, pp. 933–944, 2011

[12] S. S. Yadav and S. M. Jadhav, "Machine learning algorithms for disease prediction using iot environment," International Journal of Engineering and Advanced Technology, vol. 8, no. 6, pp. 4303–4307., 2019.

[13] S. Ghwanmeh, A. Mohammad, and A. Al-Ibrahim, "Innovative artificial neural networks-based decision support system for heart diseases diagnosis," Journal of Intelligent Learning Systems and Applications, vol. 5, no. 3, p. 176, 2013.

[14] "Human Immune Prediction using Machine Learning ", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.8, Issue 1, page no.121-123, January-2021

[15] S. S. Yadav, S. M. Jadhav, S. Nagrale and N. Patil, "Application of Machine Learning for the Detection of Heart Disease," 2020 2nd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, India, 2020, pp. 165-172, doi: 10.1109/ICIMIA48430.2020.9074954.