

Disease And Drug Prediction Using Different Machine Learning Algorithms

Ranjith Katta¹, Nasapu Aarti², Likitha sunkavalli³, Killi venkata karthik⁴

#1##2,#3,#4 Student, Department of CSE, GITAM (deemed to be University),
Gandhi nagar Rushikonda Visakhapatnam530045 Andhra Pradesh, INDIA

Abstract

Most people in our digital age are prone to ailments as a result of a lack of good food, adequate sleep, and daily exercise. It is critical to know if we are afflicted with sickness at an early stage rather than discovering it later. As a result, illness prediction systems play a significant role since they predict diseases based on symptoms. This disease prediction system employs the Machine Learning algorithm. This approach also recommends medications that are widely used to treat the illness.

Indexed Terms- Diseases, Drugs, Machine Learning, Prediction, And Random Forest.

1. INTRODUCTION

Since the arrival of advanced computing, doctors still require the technology in various possible ways like surgical representation process and x-ray photography, but the technology perceptually stayed behind. The method still requires the doctor's information and experience due to alternative factors starting from medical records to weather conditions, atmosphere, blood pressure, and numerous alternative factors. The huge numbers of variables are considered as entire variables that are required to understand the complete working process itself, however, no model has been analysed successfully. To tackle this drawback, Medical decision support systems must be used. This system is able to assist the doctors to make the correct decision. The medical decision support system refers to both the process of attempting to determine or identify possible diseases or disorders and the opinion reached by this process. The diagnostic opinion in this sense, indicates either degree of abnormality on a continuum or a kind of abnormality in a classification. It's influenced by non-medical factors such as power ethics and financial incentives for patients or doctors.

It can be a brief summation or an extensive formulation, even taking the form of a story or metaphor. It might be a means of communication such as computer code through which it triggers payment, prescription, notification, information, or advice. Indication of medical diagnostic includes knowledge of what is normal and measuring of patient's current condition. Automated decision support systems are rule-based systems that are automatically providing solutions to repetitive management problems.

Medical decisions could be an extremely specialized and difficult job due to alternative factors or in the case of rare diseases. The alternative factors include stress; tired misdiagnosis might vary from ignorance of doctors and incomplete information. The standard algorithm may go through the entire variables like prevailing conditions history of medical records, history of family records and various factors relating to the patient records, the sheer magnitude of obtainable hidden factors. Differential diagnosis methods can be used to identify the presence of an entity where multiple alternatives are possible and also refers to including the candidate alternatives. This method needs a process of elimination or

obtaining information that shrinks the probability of candidate conditions to negligible levels. It contains four steps: 1) The doctor gathers all information about the patients and creates a symptoms list. 2) The doctor should make a list of all possible causes of symptoms. 3) The doctor should prioritize the list by which is the most dangerous possible cause of symptoms put at the top of the list. 4) The doctor should rule out or treat the possible causes beginning with the most urgently dangerous conditions. "Rule Out" in the sense to use the test method or other scientific method. If there will be no such diagnosis means removing the diagnosis from the list and using tests that should have distinct results, depends on which diagnosis is correct. This can be done based on the doctor's knowledge and experience. This method is very easy to implement.

To reduce a large number of variables and find the most probable diseases by using the K-Means algorithm. This algorithm is more suitable to cluster a greater number of diseases. K-Means is one of the unsupervised learning algorithms which are used to solve the clustering problem. The main idea is to determine the k centroids, one for each cluster. Different tests performed on the patients will serve as an attribute for clustering. By using this algorithm, it reduces the number of iterations, boundaries of clusters are well defined without overlapping, to produce the accurate result for each and every diagnosis. This system uses Service-oriented architecture (SOA), which anyone can access with internet connections and LAMSTAR Network can be used to calculate the weight, increase the accuracy of the algorithm, overall speed test, and produce the better result.

2. LITERATURE SURVEY

2.1 Campillos, M., Kuhn, M., Gavin, A.C., Jensen, L.J., Bork, P.: Drug target identification using side-effect

similarity. Science 321(5886), 263–266 (2008)

Targets for tablets have so ways been expected on the foundation of molecular or mobile features, for example, by means of exploiting similarity in chemical shape or in recreation throughout mobile phone lines. We used phenotypic side-effect similarities to infer whether or not two tablets share a target. Applied to 746 marketed drugs, a community of 1018 facet effect-driven drug-drug family members grew to become apparent, 261 of which are shaped through chemically numerous pills from exclusive therapeutic indications. We experimentally examined 20 of these surprising drug-drug family members and validated thirteen implied drug-target members of the family via in vitro binding assays, of which eleven disclose inhibition constants equal to much less than 10 micromolar. Nine of these have been examined and established in cell phone assays, documenting the feasibility of the use of phenotypic data to infer molecular interactions and hinting at new makes use of of marketed pills

2.2 Chiang, A.P., Butte, A.J.: Systematic evaluation of drug–disease relationships to identify leads for novel drug uses. Clinical Pharmacology & Therapeutics 86(5), 507–510 (2009)

Drug repositioning refers to the discovery of alternate uses for medications—uses that differ from the original purpose of the drugs. One issue in this approach is determining which indications a medicine of interest could be explored in the future. In order to overcome this difficulty, we conducted a thorough evaluation of a medication treatment-based view of disorders. A "guilt by association" strategy was used to generate suggestions for novel drug usage. The recommended novel drug applications generated by this approach were significantly enriched in comparison to previous and ongoing clinical studies

when compared to a control group of drug uses.

2.3. Hu, G., Agarwal, P.: Human disease-drug network based on genomic expression profiles. PloS one 4(8), e6536 (2009)

Drug repositioning presents the opportunity for quicker improvement instances and decreased dangers in drug discovery. With the fast improvement of high-throughput applied sciences and the ever-increasing accumulation of complete genome-level datasets, a growing variety of ailments and tablets can be comprehensively characterized through the adjustments they set off in gene expression, protein, metabolites, and phenotypes.

We carried out a systematic, large-scale evaluation of genomic expression profiles of human ailments and tablets to create a disease-drug network. A community of 170,027 large interactions used to be extracted from the ~24.5 million comparisons between ~7,000 publicly reachable transcriptomic profiles. The community consists of 645 disease-disease, 5,008 disease-drug, and 164,374 drug-drug relationships. At least 60% of the disease-disease pairs had been in the identical disorder region as decided via the Medical Subject Headings (MeSH) disorder classification tree. The final can pressure a molecular stage nosology by means of discovering relationships between reputedly unrelated diseases, such as a connection between bipolar ailment and hereditary spastic paraplegia, and a connection between actinic keratosis and cancer. Among the 5,008 disease-drug links, connections with poor ratings advise new warning signs for current drugs, such as the use of some antimalaria capsules for Crohn's disease, and a range of current tablets for Huntington's disease; whilst the fantastic scoring connections can be useful resource in drug facet impact identification, such as tamoxifen's

undesired carcinogenic property. From the ~37K drug-drug relationships, we discover relationships that resource in goal and pathway deconvolution, such as 1) KCNMA1 as a viable molecular goal of lobeline, and 2) each apoptotic DNA fragmentation and G2/M DNA injury checkpoint legislation as attainable pathway ambitions of daunorubicin.

3. PROPOSED SYSTEM

Reduce the number of variables and classify the most possible diseases using the K-means algorithm. This algorithm is best adapted for grouping further pathogens. K-Mean is one of the unattended learning algorithms that address the issue of clustering. The key concept is to create the k centroids, one per cluster. Various checks on the patients are used as a clustering attribute.

3.1 ALGORITHMS

3.1.1 SUPPORT VECTOR MACHINE

“Support Vector Machine” (SVM) is a supervised machine learning algorithm that can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot). Support Vectors are simply the co-ordinates of individual observation. The SVM classifier is a frontier which best segregates the two classes (hyper-plane/ line).

3.1.2 LOGISTIC REGRESSION

Logistic Regression was used in the biological sciences in early twentieth century. It was then used in many social science applications. Logistic Regression is used when the dependent variable(target) is

categorical. Consider a scenario where we need to classify whether an email is spam or not. If we use linear regression for this problem, there is a need for setting up a threshold based on which classification can be done. Say if the actual class is

malignant, predicted continuous value 0.4 and the threshold value is 0.5, the data point will be classified as not malignant which can lead to serious consequence in real time.

4.RESULTS AND DISCUSSION

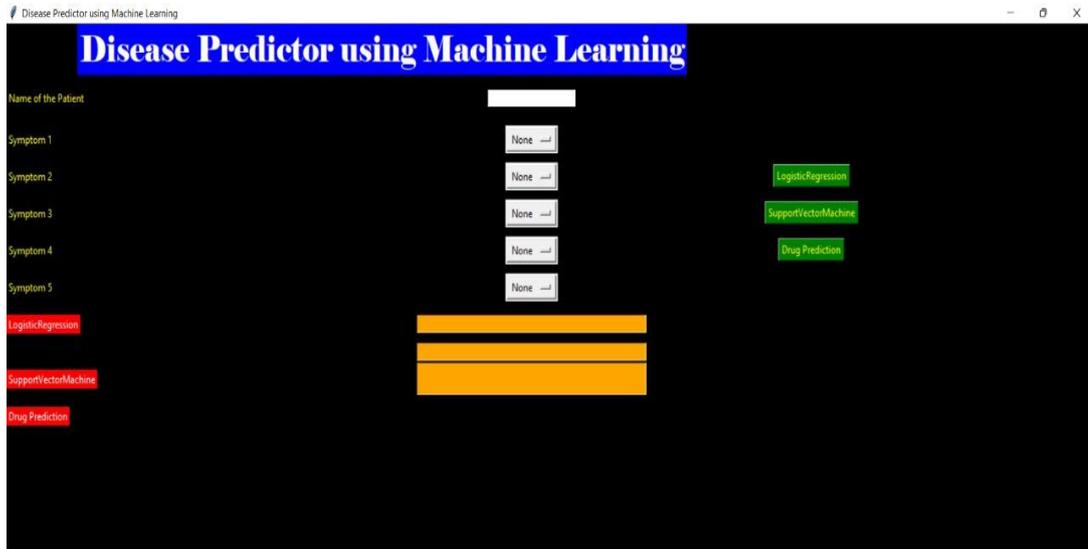


Fig 1: Enter the Name of patient and enter the symptoms of the patient to prediction the disease. And then click on algorithm from which you want to predict.

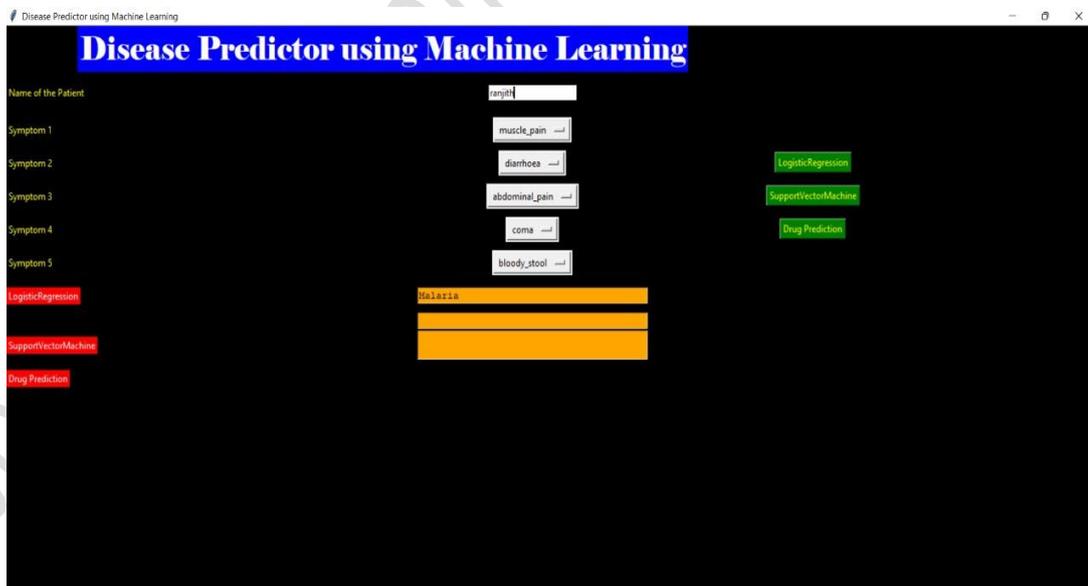


Fig 2: From the above figure for given symptoms it predicted Malaria by using Logistic regression

Now test for SVM also.

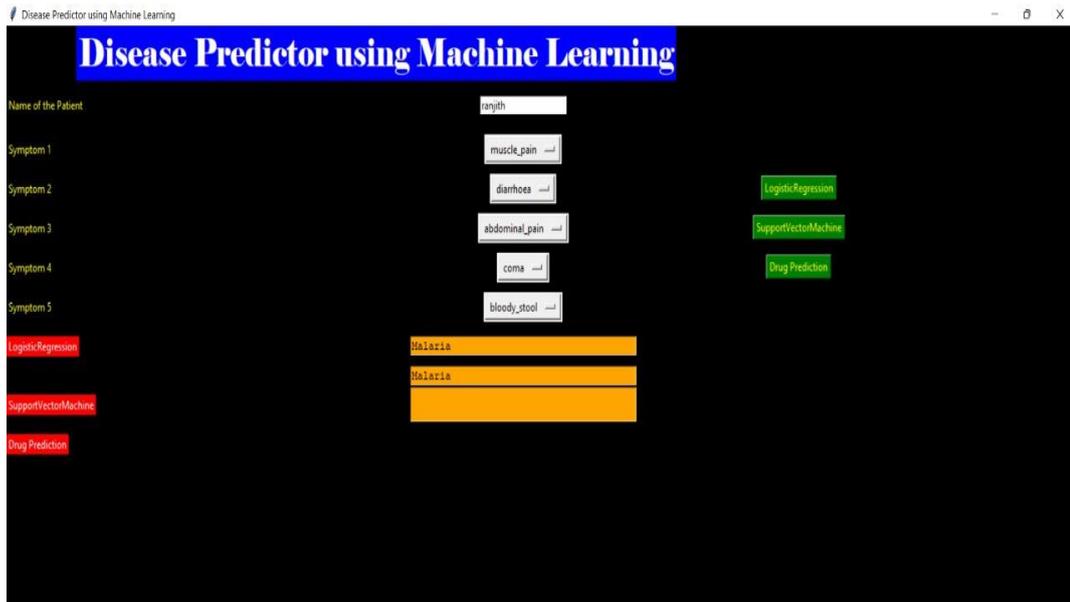


Fig 3: For SVM also for the given symptoms it predicted Malaria. Now predict the Drug for the disease.

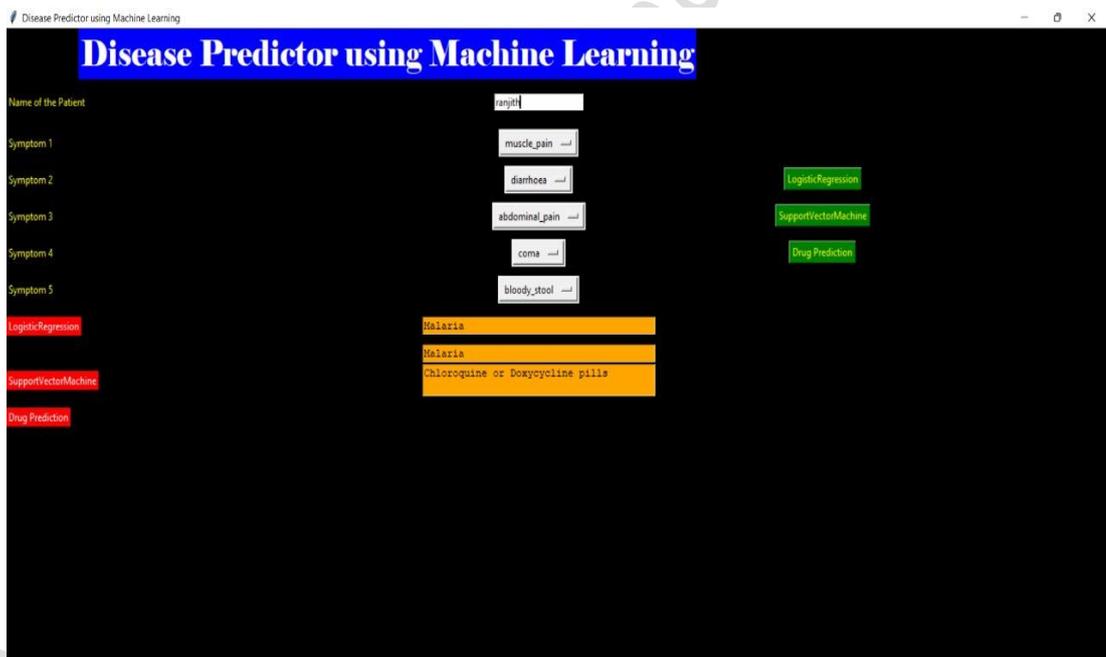


Fig 4: For Malaria it suggested 2 drugs for relieve the pain.

5.CONCLUSION

In this paper, algorithm used to predict the disease based on symptoms is discussed. Various symptoms are provided in the dropdown menu, out of which user selects any five of them and using algorithm the disease is predicted. The drugs that are

commonly prescribed for a particular disease can also be suggested in this system. The main aim is to predict the disease at the early stage and lead to early diagnosis. This system can also be used by doctors to avoid confusion while predicting the disease. This system can provide assistance to doctors

BIBLIOGRAPHY

1. Brown, A.S., Patel, C.J.: A standard database for drug repositioning. *Scientific Data* 4, 170029 (2017)
2. Callahan, A., Cruz-Toledo, J., Ansell, P., Dumontier, M.: Bio2rdf release 2: improved coverage, interoperability and provenance of life science linked data. In: *Extended Semantic Web Conference*.pp. 200{212. Springer (2013)
3. Campillos, M., Kuhn, M., Gavin, A.C., Jensen, L.J., Bork, P.: Drug target identification using side-effect similarity. *Science* 321(5886), 263{266 (2008)
4. Chiang, A.P., Butte, A.J.: Systematic evaluation of drug{disease relationships to identify leads for novel drug uses. *Clinical Pharmacology & Therapeutics* 86(5), 507{510 (2009)
5. Gottlieb, A., Stein, G.Y., Rupp, E., Sharan, R.: Predict: a method for inferring novel drug indications with application to personalized medicine. *Molecular systems biology* 7(1), 496 (2011)
6. Guney, E.: Reproducible drug repurposing: When similarity does not succeed. In: *PACIFIC SYMPOSIUM ON BIOCOMPUTING 2017*. pp. 132{143 (2017)
7. Hay, P.J., Claudino, A.M.: Bulimia nervosa: online interventions. *BMJ clinical evidence* 2015 (2015)
8. Hu, G., Agarwal, P.: Human disease-drug network based on genomic expression profiles. *PloS one* 4(8), e6536 (2009)
9. Kuhn, M., Letunic, I., Jensen, L.J., Bork, P.: The SIDER database of drugs and side effects. *Nucleic acids research* 44(D1), D1075{D1079 (2015)

10. Lamb, J., Crawford, E.D., Peck, D., Modell, J.W., Blat, I.C., Wrobel, M.J., Lerner, J., Brunet, J.P., Subramanian, A., Ross, K.N., et al.: The connectivity map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* 313(5795), 1929{1935 (2006) 1

Author's Profile



Name : Ranjith Katta

University: Gitam University

Address: Gandhi nagar, rushikonda, Gitam University, Visakhapatnam 530045



Name : Nasapu Aarti

University: Gitam University

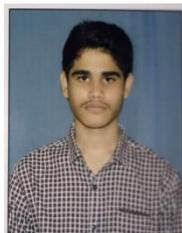
Address: Gandhi nagar, rushikonda, Gitam University, Visakhapatnam 530045



Name : Likitha sunkavalli

University: Gitam University

Address: Gandhi nagar, rushikonda, Gitam University, Visakhapatnam 530045



Name : Killi venkata karthik

University: Gitam University

Address: Gandhi nagar, rushikonda, Gitam
University, Visakhapatnam 530045

Journal of Engineering Sciences