ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

Leveraging and Prediction of Disease Using Machine Learning Techniques

K.A.V.A.S.Mounika Varma[1], Dr.K. Venkata Ramana[2]

M.Tech Scholar, Department of Computer Science and System Engineering, Andhra University College of Engineering (A), Visakhapatnam, AP, India[1]

Department of Computer Science and System Engineering, Andhra University College of Engineering (A), Visakhapatnam, AP, India[2]

ABSTRACT: Diseases that are associated with the way a person or group of people live are known as lifestyle diseases. Healthcare industry collects enormous disease-related data that is unfortunately not mined to discover hidden information that could be used for effective decision making. Activity monitoring of a person for a long-term would be helpful for controlling lifestyle associated diseases. Such diseases are often linked with the way a person lives. An unhealthy and irregular standard of living influences the risk of such diseases in the later part of one's life. The symptoms and the initial signs of these diseases are common to the people with irregular lifestyle. In this paper, we propose a novel healthcare framework to manage lifestyle diseases using long-term activity monitoring. The framework recognizes the user's activities with the help of the sensed data in runtime and reports the irregular and unhealthy activity patterns to a doctor and a caregiver. This study aims to understand support vector machine and use it to predict lifestyle diseases that an individual might be susceptible to. Moreover, we propose and simulate an economic machine learning model as an alternative to deoxyribonucleic acid testing that analyzes an individual's lifestyle to identify possible threats that form the foundation of diagnostic tests and disease prevention, which may arise due to unhealthy diets and excessive energy intake, physical dormancy, etc. The simulated model will prove to be an intelligent low-cost alternative to detect possible genetic disorders caused by unhealthy lifestyles.

KEYWORDS: Deoxyribonucleic acid testing, healthcare industries, lifestyle diseases, support vector machine.

I. INTRODUCTION

Over the past two decades Machine Learning has become one of the mainstays of information technology and with that, a rather central, albeit usually hidden, part of our life. As a knowledge discovery process, it typically involves data cleaning, data integration, data selection, data transformation, pattern discovery, pattern evaluation, and knowledge presentation. The major dimensions of data mining are data, knowledge, technologies, and applications. It defines machine learning with respect to the knowledge discovery process. Next, data mining from many aspects, such as the kinds of data that can be mined, the kinds of knowledge to be mined, the kinds of technologies to be used and targeted applications are discussed which helps gain a multidimensional view of data mining. In recent years in healthcare sectors, machine learning became an ease of use for disease prediction. Machine learning is the process of dredge up information from the massive datasets or warehouse or other repositories. It is a very challenging task to the researchers to predict the diseases from the voluminous medical databases. A report prepared by the World Health Organization and World Economic Forum says that India will incur an accumulated loss of $236.6 billion by 2015 because of morbid lifestyles as well as imperfect diet.

Lifestyle and diet are the two main factors that are considered to influence Receptiveness to various diseases. Diseases are mainly caused by a combination of transformation, lifestyle selections, and surroundings. In addition, identifying health risks in an individual's family is one of the most crucial things an individual can do to help his/her practitioner understand and diagnose hereditarily linked syndromes like cancer, diabetes, and mental illness. Diseases that are

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

associated with the way a person or group of people live are known as lifestyle diseases. They include atherosclerosis; heart disease and stroke; obesity and type II diabetes; and smoking and alcohol-related diseases.

Diseases are mainly caused by a combination of transformation, lifestyle selections, and surroundings. In addition, identifying health risks in an individual's family is one of the most crucial things an individual can do to help his/her practitioner understand and diagnose hereditarily linked syndromes like cancer, diabetes, and mental illness. Diseases that are associated with the way a person or group of people live are known as lifestyle diseases. They include atherosclerosis; heart disease and stroke; obesity and type II diabetes and smoking and alcohol-related diseases. This study aims to understand support vector machine (SVM) and use it to predict lifestyle diseases that an individual might be susceptible to. The need for public awareness is not stressed enough, but lifestyle diseases are easy to prevent. Simply modifying an individual's lifestyle to reduce and eliminate risks can be interesting.
Monitoring a person's lifestyle over a long period of time can be helpful for the following purposes:
Early identification of the lifestyle: A person is generally at risk of a lifestyle disease when he/she lives his/her life in a certain style. Moreover, initial signs and symptoms of such diseases appear in people with irregular life patterns [5,6]. By long term activity monitoring, it would be possible to identify such styles (e.g., irregular and unhealthy) and determine the risk of the disease before the actual disease appears.

Prevention of lifestyle diseases: It would be possible to prevent a disease at the first place with the identification of the lifestyle. A healthy lifestyle and physical activity help prevent obesity, heart disease, hypertension, diabetes, colon cancer, and premature mortality [7]. A person can assess his/her lifestyle and become aware of any irregular patterns though monitoring daily activity patterns.

Management of lifestyle diseases: A person who has a lifestyle disease should make changes to assure a healthy lifestyle. For example, there is no cure for typeIIdiabetes, but it can be effectively controlled by maintaining normal blood-glucose levels. This can be achieved through weight control by following a nutritious and balanced diet, along with regular and stringent exercise.

II. RELATED WORK

Countless scholars have used ML algorithms to predict diseases in health sciences. Afew of them are explained below.Suzuki et al. [3] analysed annual health check-up data from1,546 employees. They concluded that 5% weight reductionwith succeeding weight control and daily workouts would behelpful in treating non-alcoholic fatty liver disease after asystematic health check for 12 months to evaluate changes inlifestyle with a shift in serum alanine aminotransferase(ALT). 136 subjects—who had ALT stabilization—weretracked for 24 months to assess association between lifestylemanagement and ALT levels. Their research demonstratedthe effect of change in lifestyle on non-alcoholic fatty liverdisease.

The efficacy of using ML and graph theoreticalmetrics for detecting Parkinson's disease has also beendiscussed by them. S. A. Pattsekari and A. Parveen [4]recommended an intelligent system that uses a DM techniquethat retrieves unseen data from stockpiled database andacquires user answers for predefined questions related toblood sugar, sex, age, height, etc. and compares them tostored database values, i.e., trained dataset. A. Anand and D.Shakti [5] conversed about relationship between diabetes riskprobably to be developed from an individual's daily lifestyleactivities such as his/her eating and sleeping routines,physical movement in addition to body mass index byacquiring data from questionnaires averring 75% accuracy.B. D. Kanchan and M. M. Kishor [6] used SVM, Naive Milgram et al. [8] compared renowned multiclass SVMvariations and concluded that one against all method is a bitmore accurate compared to one to one method, which is easyto train. Hossain et al. [9] scrutinized datasets using DMtechniques to predict obesity risk factor and statistical dataexploration to ascertain which obesity affected the most inBangladesh. In their work, data analysis results and class levelaccuracy evaluation technique depended on statisticalpackage for social science and Weka. They concluded that15.8% people were found to be obese. Sayali Ambekar andDr. Rashmi

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

Phalnikar [10] proposed a system that predictsdiseases a patient is susceptible to on the basis of diseasesymptoms via a decision tree.

Ashish Kumar Mishra et al.[11] proposed predictive analyses where data regulated andgenerated are again feedback to the system providingrequired decision support in realizing health condition for working professionals.There are a myriad number of lifestyle and geneticfactors that contribute toward the participation of lifestylediseases. Numerous exorbitant tests are performed and theirresults are analyzed to recognize lifestyle diseases. Doctorsscrutinize these results based on patients' previous resultswith similar conditions. Analyzing them is difficult andrequires specialized doctors.The burgeoning field of artificial intelligence, especiallyML, has made it easier to adjudicate a patient's lifestyledisease. Based on former patients' lifestyle, we can predict ifa person is suffering from a lifestyle disease or not.Computational techniques are accurate and fast, and sincethere is a need for spreading lifestyle disease awareness, wewere motivated to design a lifestyle disease predictionmodel.The literature survey carried out advocates that numerousresearch papers each directing a specific disease have beenpresented. Lifestyle and inheritances go concurrently and areinterrelated.Nevertheless, a healthy lifestyle surpasses congenitalrisk. Diseases are mainly caused by a combination oftransformation, lifestyle selections, and surroundings; hence,we propose a lifestyle disease prediction model that mutuallyforecasts lifestyle disorders that an individual might besusceptible to.

III. MOTIVATION

The Main motivation of this research work is to predict life style diseases using classification algorithms. The algorithms used in this work are Naïve Bayes and support vector machine (SVM). These classifier algorithms are compared based on the performance factors i.e. classification accuracy and execution time. the results are analysed which are given by the classification algorithms such as naïve Bayes and Support Vector Machine.

IV. AIM OF THE WORK

This study aims to understand support vector machine (SVM) and Navies Bayesian algorithms used it to predict lifestyle diseases that an individual might be susceptible to. The need for public awareness is not stressed enough, but lifestyle diseases are easy toprevent. Simply modifying an individual's lifestyle to reduce and eliminate risks can be interesting. Deoxyribonucleic acid (DNA) and genetic testing are creating a new expanse of personalized medicine.

V. PROBLEM STATEMENT

In medical research the machine learning begins with a hypothesis and results are adjusted to fit the hypothesis. This differs from standard machine learning practice, which simply starts with datasets without an apparent hypothesis. According to the doctor intuition the clinical decision are often made. The quality of service provided to patients is affected due to unwanted bias, errors and excessive medical cost. In the survey of the two machines learning algorithms are used. The Classification Accuracy should be compared for this algorithm. This work should be extended to predict the disease with reduced number of attributes. Successful diagnosis of disease like cancer is a tedious issue to look upon and hence raises a serious issue for future therapy. Various classification techniques haven been proposed for cancer diagnosis, still no proper diagnosis methodology has been developed. Traditionally all cancers are classified based on their anatomical origin which does not gives accurate results Olden techniques are economically very high.

VI. PROPOSED SYSTEM

The proposed work represents recent method that improved algorithms performance and accuracy in distributed environment. In this project we have done analysis using SVM, NB (Naive Bayesian) algorithms by applying validation measures. The paper proposes a system, with a strong prediction algorithm, which implements powerful classification steps with a comprehensive report generation module. The project aims to implement a self-learning protocol such that the past inputs of the disease outcomes determine the future possibilities of the life style disease to a

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

particular user. The proposed methodology encompasses of hybrid algorithm which contains inner and outer classification. The proposed algorithm is divided into three sections:
a) Dataset Pre-processing
b) Classification using Support Vector Machine and NB
Machine learning has ability to deal with uncertain, fuzzy or insufficient data, which fluctuate in short period of time, robust that helps to predict stock prices and returns.
1. Accuracy mostly ranges 70 to 80%.
2. Helps to classify the diseases into 3 classes: -
i. Lifestyle Disease with either positive or negative returns which gives valuable support in making decisions, but do not specify the amount of expected and expected profit.
ii. System tries to predict the Lifestyle Disease for one or more days in advance, based on the previous stock prices and on related ratios.
iii. Concerned with modeling Lifestyle Disease forecasting. Predicts future values, but also have significance estimation, sensitivity analysis and other analysis of mutual dependencies.
3. Various data models and expert system combined improves the accuracy. Among all the algorithms the Back propagation have the highest accuracy.
4. Forecasting is done by SVM, NB algorithms.

SUPPORT VECTOR MACHINES

Support Vector Machine (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In this algorithm, 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. Support Vectors are simply the co-ordinates of individual observation. Support Vector Machine is a frontier which best segregates the two classes (hyper-plane/ line).

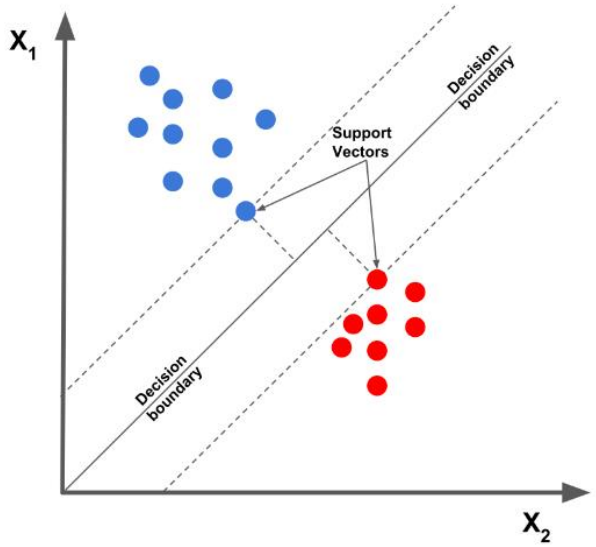

Fig: Support Vector Algorithm

Algorithm: Generate SVM

Input: Training Data, Testing Data
Output: Decision Value

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

Method:

Step 1: Load Dataset
Step 2: Classify Features (Attributes) based on class labels
Step 3: Estimate Candidate Support Value
While (instances! =null)
Do
Step 4: Support Value=Similarity between each instance in the attribute
Find Total Error Value
Step 5: If any instance < 0
Estimate

Decision value = Support Value\Total Error
Repeat for all points until it will empty
End If

VII. METHODOLOGY

Data Collection
Data will be collected from hospitals with the consent of patients who have completed their DNA test. Hospitals will provide test results and other essential factors necessary to develop the proposed system.

Data preprocessing
Data preparation requires approximately 80% of time. Once data is gathered, it needs to be preprocessed, cleaned, constructed, and formatted in a style that SVM comprehends and is able to work with. DM tools should be used to analyze collected real-time data.

Training and Testing Data
The proposed model needs to be trained and tested under various conditions by altering SVM parameters so that correctness can be obtained. In addition, we consider that the model's accuracy is maximum. From the collected data, 70:30 will be used to train and test the model, respectively. In case of necessity, there must be provisions to improvise on the algorithm being used.

Results and Analysis
The prediction results have been evaluated using following parameters:
Precision: It is the fraction of retrieved data that are useful for the query.
Recall: It is the fraction of data that are relevant for the query which is effectively retrieved.
F-measure: It is measure that sums up precision and recall.
Accuracy: It is the proximity of a computation to the true value which is calculated by taking true positive and true negative with a fraction of true positive, true negative and false positive with false negative.

VIII. CONCLUSION

ML being an essential CS application is used for predicting results given target input parameters and is being widely used for improving human lifestyle in several ways. Complex disorders—also known as polygenic—are caused by simultaneous effects of more than one gene often in a complex interaction with environment and lifestyle factors, which implies that if a parent has a particular disorder, it does not necessarily mean that a child would develop the same. However, there could be a possibility of high risk of developing the disorder (i.e., genetic susceptibility), and for such a possibility where it cannot be a sure occurrence but risk prevails, the proposed model would provide a detailed report of alterations in an individual's lifestyle such as maintaining a healthy weight, and sugar levels may be able to

ISSN(Online): 2319-8753
ISSN (Print): 2347-6710

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 8, Issue 10, October 2019

reduce risk in case of genetic predisposition known that genetic makeup cannot be altered. Further additions to the model would include when an individual enters his/her details (i.e., input to the predictive model), the model would determine his/her identity based on several inputs, show an individual's current status of his/her health contrary to a desired ideal health using graphs, let know lifestyle changes, provide balanced diet and doctor consultations, recommend exercises, etc. The model would take into account climatic conditions and pollution levels and rank cities and suburbs

with an ideal environment as to the precautionary measures that an individual could take making the model more content specific, accessible, and flexible in terms of customization. The fact that deep learning (DL) is overtaking ML algorithms in terms of accuracy would suggest the possibility of SVM being replaced by DL in the near future.

REFERENCES

[1] Sharma, M. and Majumdar, P.K., 2009. Occupational lifestyle diseases: An emerging issue. Indian Journal of Occupational and Environmental medicine, 13(3), pp. 109–112.
[2] DNA Test Cost in India, Available [Online] https://www.dnaforensics.in/dna-test-cost-in-india/ [Accessed on June 27, 2018].
[3] Suzuki, A., Lindor, K., St Saver, J., Lymp, J., Mendes, F., Muto, A., Okada, T. and Angulo, P., 2005. Effect of changes on body weight and lifestyle in nonalcoholic fatty liver disease. Journal of Hepatology, 43(6), pp. 1060–1066.
[4] Pattekari, S.A. and Parveen, A., 2012. Prediction system for heart disease using Naïve Bayes. International Journal of Advanced Computer and Mathematical Sciences, 3(3), pp. 290–294.
[5] Anand, A. and Shakti, D., 2015. Prediction of diabetes based on personal lifestyle indicators. In Next generation computing technologies (NGCT), 2015 1st international conference on (pp. 673–676). IEEE.
[6] Kanchan, B.D. and Kishor, M.M., 2016. Study of machine learning algorithms for special disease prediction using principal of component analysis. In Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC), 2016International Conference on (pp. 5–10). IEEE.
[7] Kazeminejad, A., Golbabaei, S. and Soltanian-Zadeh, H., 2017. Graph theoretical metrics and machine learning for diagnosis of Parkinson's disease using rs-fMRI. In Artificial Intelligence and Signal Processing Conference (AISP), (pp. 134–139). IEEE.
[8] Milgram, J., Cheriet, M. and Sabourin, R., 2006. "One against one" or "one against all": Which one is better for handwriting recognition with SVMs?. Tenth International Workshop on Frontiers in Handwriting Recognition, La Baule (France), Suvisoft, 2006.
[9] Hossain, R., Mahmud, S.H., Hossin, M.A., Noori, S.R.H. and Jahan, H., 2018. PRMT: Predicting Risk Factor of Obesity among Middle- Aged People Using Data Mining Techniques. Procedia Computer Science, 132, pp. 1068–1076.
[10] Sayali Ambekar and Dr.Rashmi Phalnikar, 2018. Disease prediction by using machine learning, International Journal of Computer Engineering and Applications, vol. 12, pp. 1–6.
[11] Mishra, A.K., Keserwani, P.K., Samaddar, S.G., Lamichaney, H.B. and Mishra, A.K., 2018. A decision support system in healthcare prediction. In Advanced Computational and Communication Paradigms (pp. 156–167). Springer, Singapore.
[12] Explaining the basics of machine learning, algorithms and applications, Available [Online] https://www.hackerearth.com/blog/machine-learning/explainingbasics- machine-learning-algorithms-applications/ [Accessed on June 27, 2018].
[13]https://upload.wikimedia.org/wikipedia/commons/thumb/b/b5/Svm_ separating_hyperplanes_%28SVG%29.svg/220px-Svm_separating_hyperplanes_%28SVG%29.svg.png