

Modern Lifestyle Disease Prediction Using Proposed Navie Based Alogrithm

Dr.D.J.Samatha Naidu¹, K.Mahesh²

Principal, Annamacharya PG college of computer studies, Rajampet.
MCA Student, APGCCS, Rajampet.

Date of Submission: 05-10-2022

Date of Acceptance: 15-10-2022

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. 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

I. INTRODUCTION

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 [1]. 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 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. Deoxyribonucleic acid (DNA) and genetic testing are creating a new expanse of personalized medicine. However, on an average, DNA testing may incur \square 10,000 to 20,000 [2], which is expensive. Though there are many preceding diseases and tests, they are erratically tested because they are costly, and factual tests have not been developed yet. Our lifestyles are imperative in increasing or decreasing risks of various diseases. According to some conducted in the discipline of epigenetics, it determines that an individual's lifestyle selections can modify his/her well-being at genetic level. This study discusses about a model that can predict the probabilities of an individual obtaining a lifestyle disease. Lifestyle diseases depend on factors like heaviness, workout, and food likings and thus have a strong association with the above-mentioned factors. In our simulated model, an actor will input his/her like fatness, sleeping habits and will discover the likelihood of suffering from lifestyle diseases. The remainder of this manuscript is organized as follows. Section 2 a brief summary about related work in machine learning (ML) domain. Section 3 focuses on ML and SVM (linear and multiclass) algorithm. Section 4 explains the proposed system (block diagram and working) for lifestyle disease prediction. Section 5 presents simulation for the system. Section 6 concludes the study with future scope.

Objective

To develop machine learning algorithms in to detect possible genetic disorders caused by unhealthy lifestyles.

Motivation

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.

Existing work

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.

Limitations

The quality of service provided to patients is affected due to unwanted bias, errors. Excessive medical cost

Proposed Work

The proposed work represents recent method that improved algorithm performance and accuracy in distributed environment. In this project we have done analysis using SVM, NB (Naive Bayesian) algorithms by applying validation measures. 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 particular user.

Contributional Work

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 c) NB Machine learning has ability to deal with uncertain and insufficient data.

II. RELATED WORK

With a sharp increment in AI advancement, there has been an exertion in applying machine learning and deep learning strategies to recommender frameworks. These days recommender frameworks are very regular in the travel industry, e-commerce, restaurant, and so forth. Unfortunately, there are a limited number of studies available in the field of drug proposal framework utilizing sentiment analysis on the grounds that the medication reviews are substantially more intricate to analyze as it incorporates clinical wordings like infection names, reactions, a synthetic names that used in the production of the drug [8]. The study [9] presents Galen OWL, a semantic-empowered online framework, to help specialists discover details on the medications. The paper depicts a framework that suggests drugs for a patient based on the patient’s infection, sensitivities, and drug interactions. For empowering Galen OWL, clinical data and terminology first converted to ontological terms utilizing worldwide standards, such as ICD-10 and UNII, and then correctly combined with the clinical information. Leilei Sun [10] examined large scale treatment records to locate the best treatment prescription for patients.

a). SYSTEM ARCHITECTURE

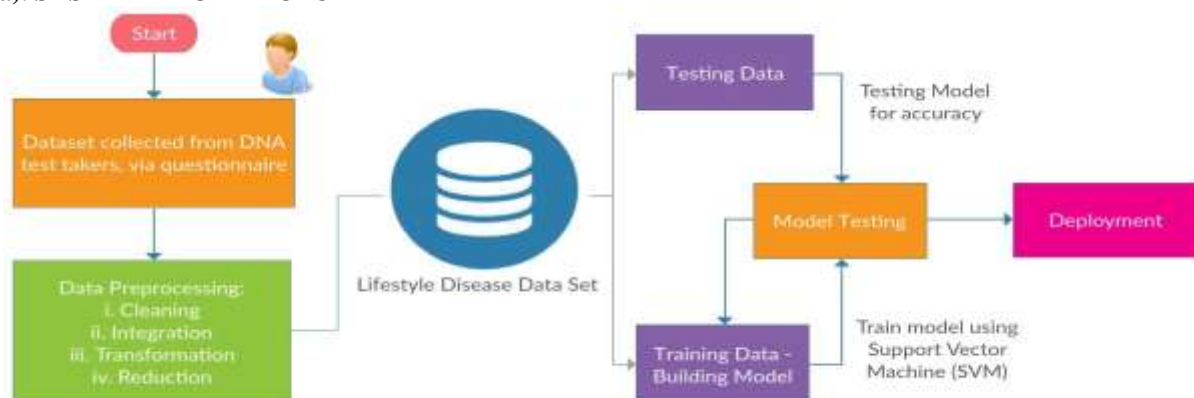


Fig.2.1. System Architecture

1. 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. Dataset shall contain following patient attributes:

1. Unhealthy eating habits (1–5)
2. Lack of physical activity (1–5)
3. Obesity (yes/no)
4. Stress and anxiety (1–5)
5. Poor sleep (1–5)
6. Smoking (daily, sometimes, or never)
7. Alcoholism (daily, sometimes, or never)
8. If an individual is suffering from any other lifestyle disease (yes/no)
9. Family history of a lifestyle disease (yes/no)
10. Gender (male/female) (Grading: 1=Excellent, 2=Good, 3=Average, 4=Bad, and 5=Very bad)
11. age

The aforementioned attributes have confirmed to be linked to majority of lifestyle diseases. As per [5], lifestyle disease diabetes can take place due to insalubrious eating habits, genetics, and absence of physical activity. Note that the attributes will be input variables for the simulated system.

The abovementioned attributes are daily records averaged over a week. For the simulated system to work efficiently, data was assembled from people who suffer a lifestyle disease and from those who fail to clear DNA test. Lifestyle diseases should be epitomized correspondingly so that the simulated system is unbiased toward a certain disease.

2. 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. There is a possibility that real-time data might hold mislaid values, they need to be replaced with a median. Herein, data has to be comprehensively reconnoitred and patterns or similarities in data need to be recognized.

Data preprocessing includes the following steps:

Data integration:

It is the process of combining significant data from different sources. Data will be obtained from the questionnaire that patients fill after their DNA test.

Data transformation:

It is the process of applying different algorithms on integrated data for significant outputs using data analysis.

Data reduction:

Herein, data filtering takes place and selection of only that data is considered that is needed for analysis. Only required fields will be taken.

Data cleaning:

It is the process of treatment of noisy data, inconsistency, and missing values in data. Spurious records are removed to make the dataset clean.

After data is preprocessed, it is added to lifestyle disease dataset, which can be used for training and testing purposes.

3. 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. Furthermore, the model must adapt to new changes made in the dataset as dataset size will constantly increase. Data preprocessing would be needed to be performed to newly added data and include it to previously collected results. The model can then be retrained and checked for efficiency.

4. Working of the Model (Model Testing):

An individual who desires to know whether he/she is exposed to a lifestyle disease can make use of the model as a replacement for DNA test. A questionnaire will be provided on a web application asking an individual to rate his/her eating habits, physical activity, anxiety, sleep, etc. Once an individual submits his/her questionnaire, the data collected via the questionnaire will act as an input to the model working behind the web application on a cloud service like Amazon web service or Microsoft Azure. The model will quickly respond with predicted results, which will be shown to the individual.

The obtained results should specify whether a person is susceptible to a disease or not. Also, it should display graphs, charts showing an individual the probability of him/her suffering a disease. Results should also advise an individual medicines or exercises and motivate him/her to live a healthy lifestyle. Compared to DNA test, the model will prove to be faster, cheaper, and easier to predict an individual's chances of suffering from a lifestyle disease. Moreover, there is provision for an individual to change his/her input parameters and check for prediction.

5. Deployment:

Once the model is tested thoroughly, the web application will be deployed for users.

III. MODULES:

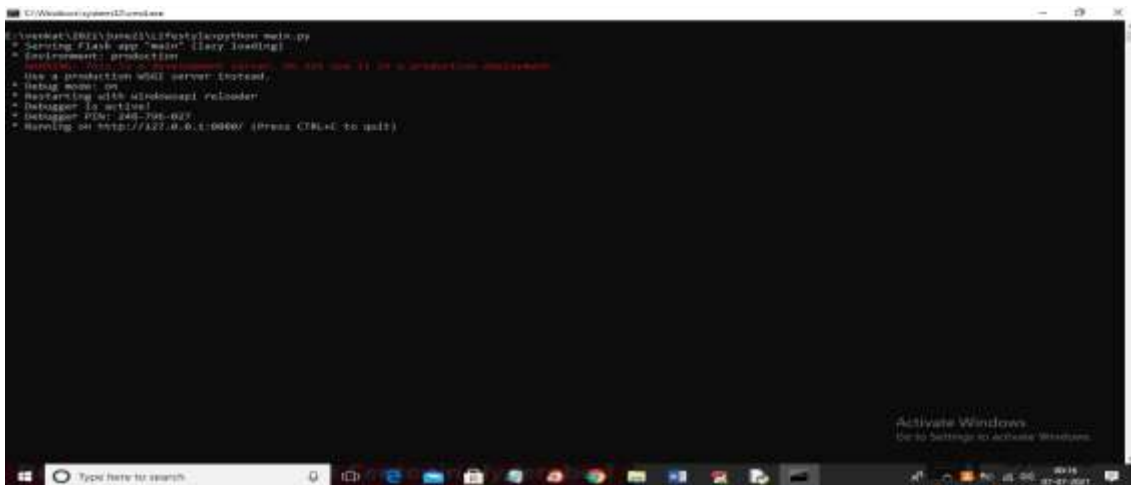
Data preprocessing includes the following steps:
Taking User Data
Applying SVM
Applying life style filter to get minute points
Applying person daily works to detection
Applying Key points extraction using Harris method
This key points and original Data will be used to compute Euclidean Distance

Extract 30 points from Data rows and then normalize those points to get 0 and 1 binary values
Convert binary number to decimal number to get master key

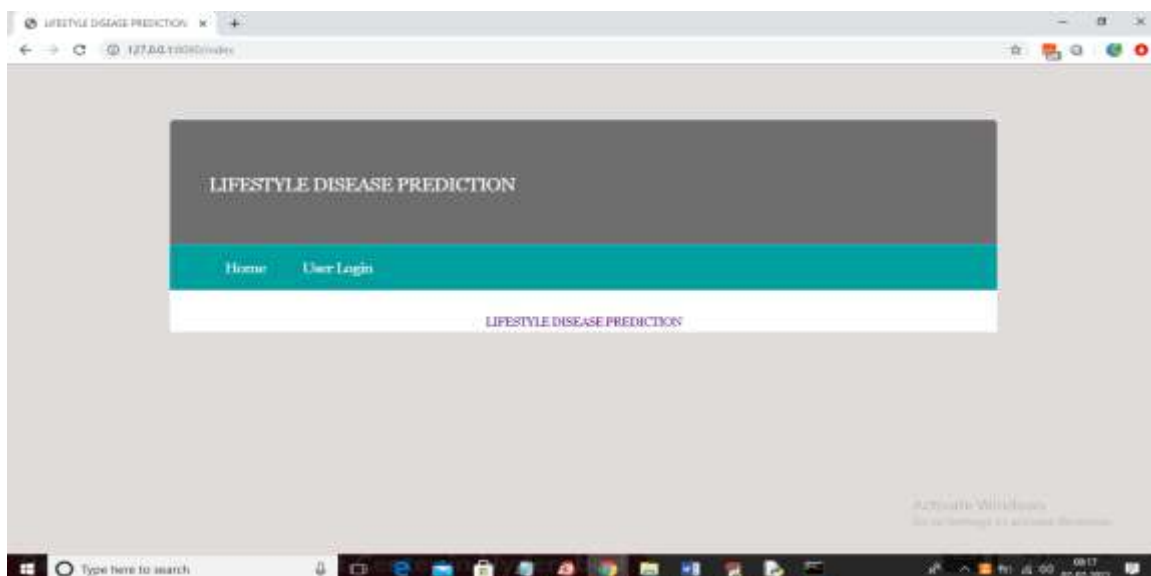
IV. RESULTS:

LIFESTYLE DISEASE PREDICTION

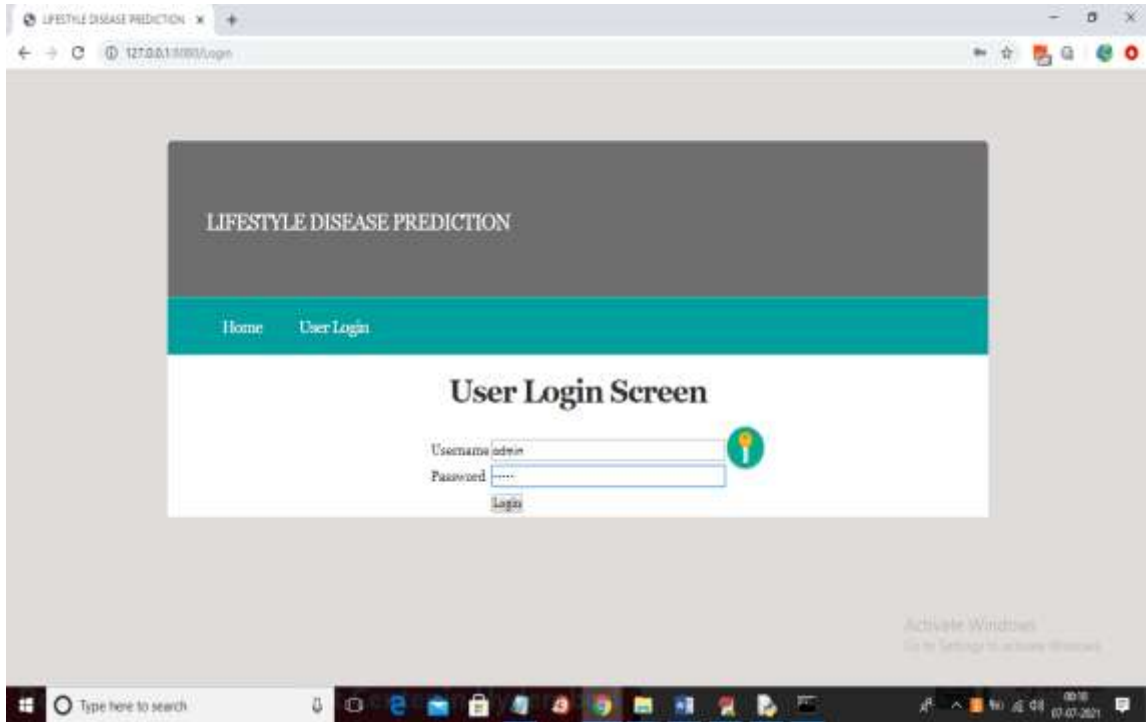
We have used same dataset given by you and we have used SVM algorithm to train life style dataset and to run project double click on 'run.bat' file to get below screen



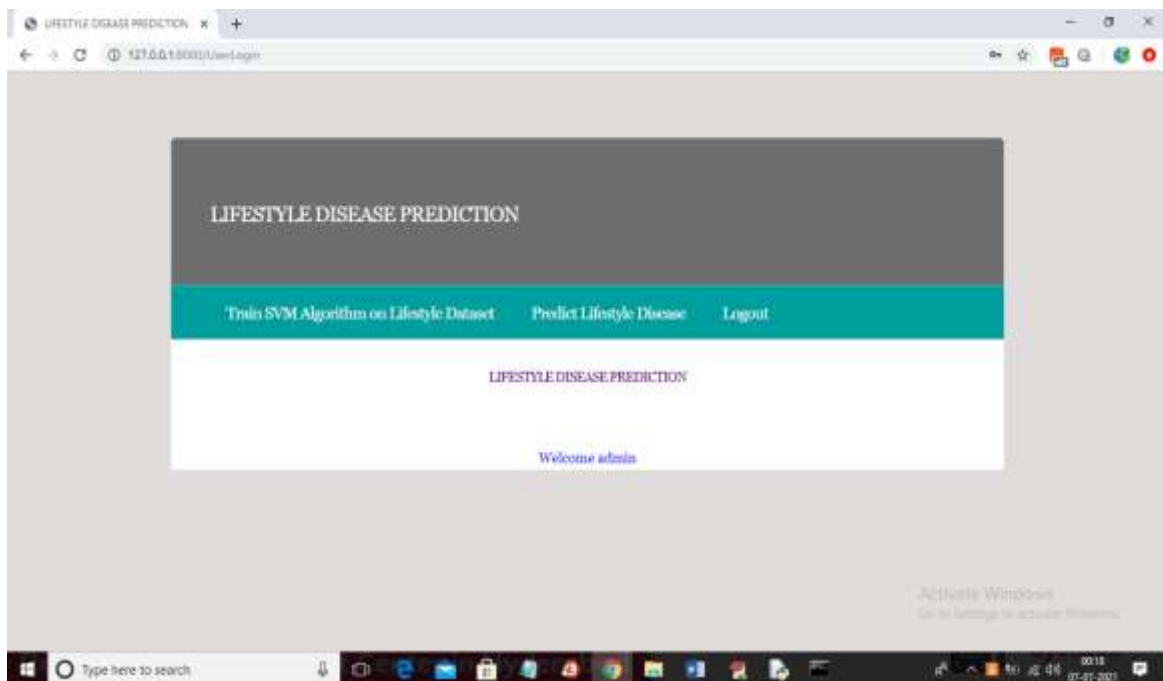
Screen 1: In above screen flask server started and now open browser and enter URL as 'http://127.0.0.1:8080/index' and press enter key to get below page



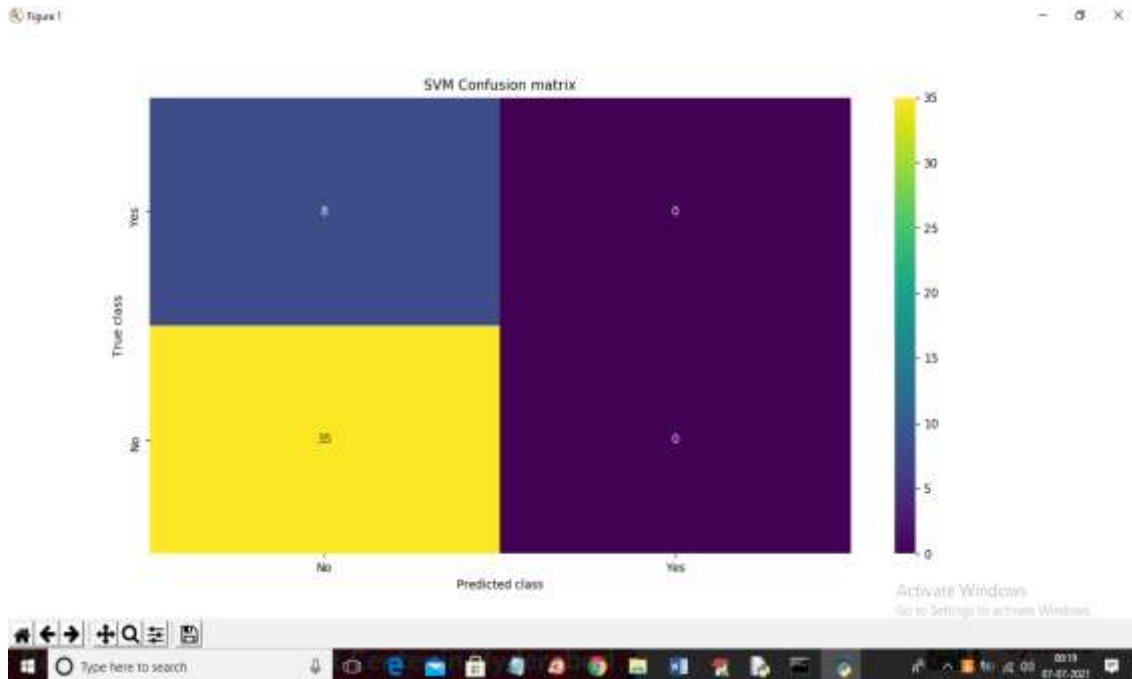
Screen 2: In above screen click on user login link and then enter username and password as 'admin' and 'admin'



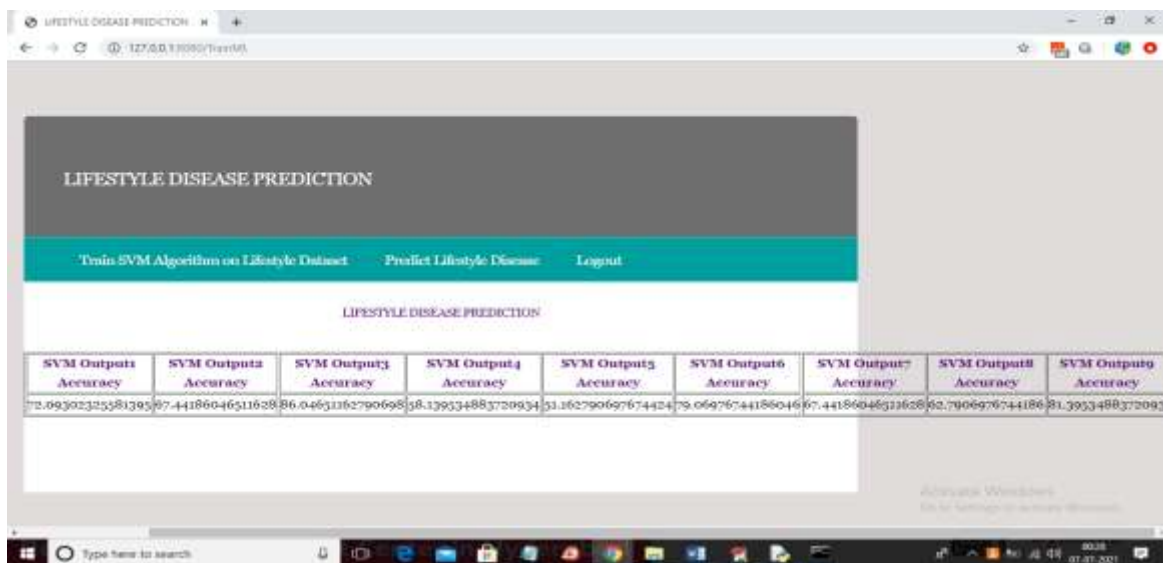
Screen 3: Now in above screen click on 'Login' button to get below screen



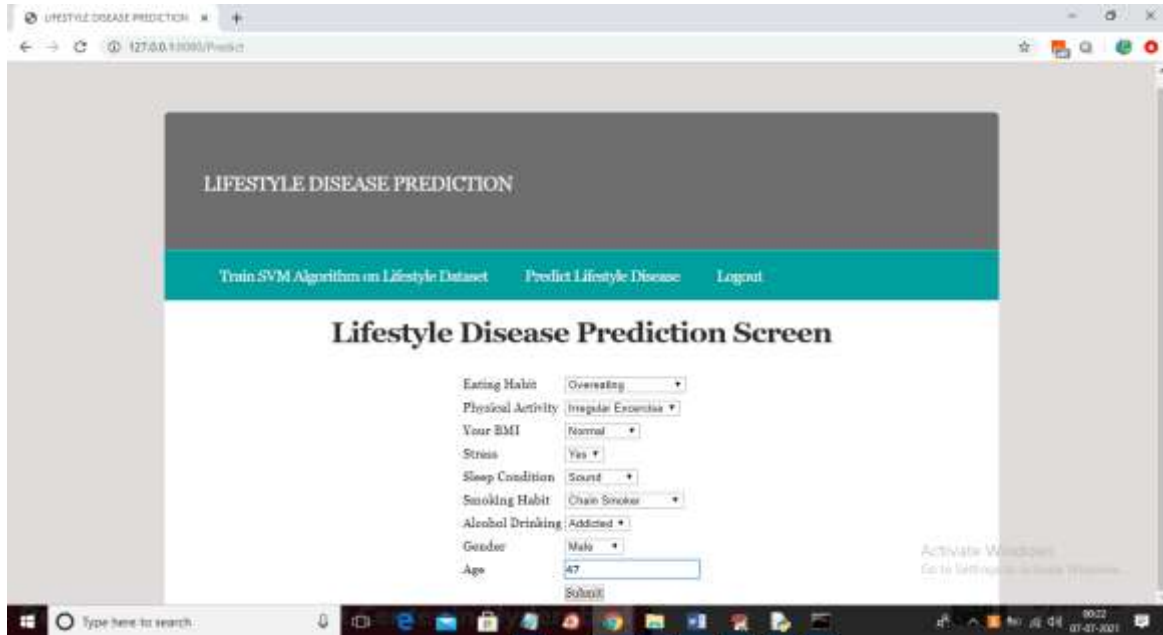
Screen 4: In above screen click on 'Train SVM Algorithm on Lifestyle Dataset' link to train dataset with SVM algorithm and then calculate accuracy and confusion matrix



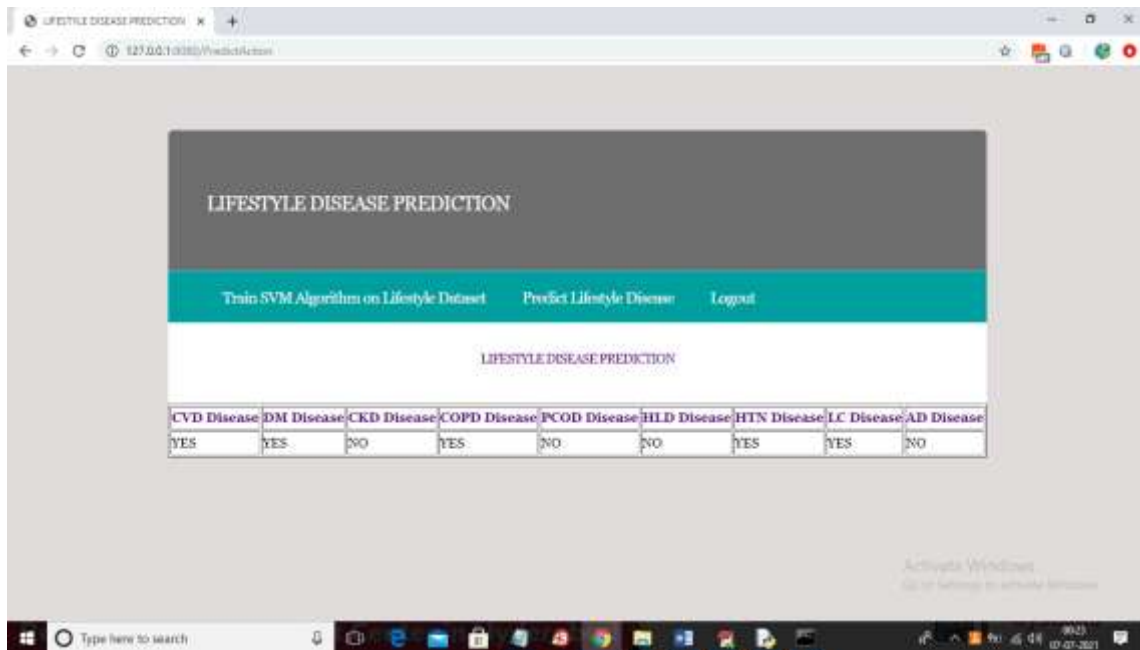
Screen 5: In above screen we can see confusion matrix generated from SVM algorithm after training dataset and now close above graph to get below screen



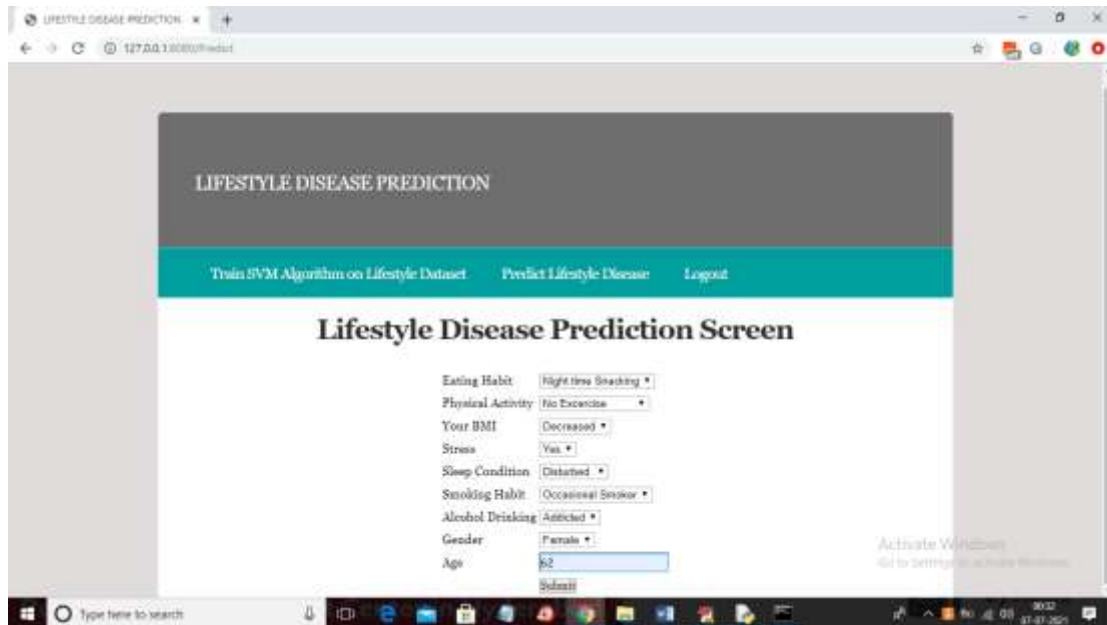
Screen 6: In above screen we can see SVM trained on 9 diseases available in dataset so we are getting accuracy for each disease prediction and the disease names are given in dataset like 'CVD', 'DM', 'CKD', 'COPD', 'PCOD', 'HLD', 'HTN', 'LC', 'AD'. In above screen for each disease we got accuracy and now SVM model is ready and now click on 'Predict Lifestyle Disease' link to get below screen



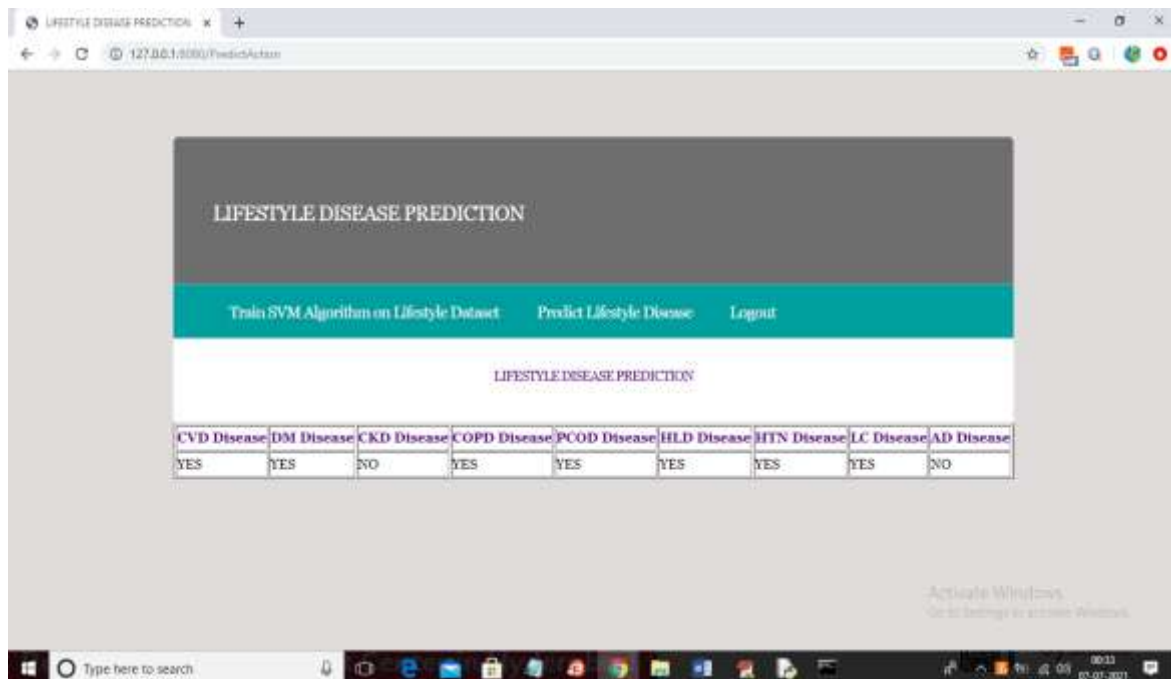
Screen 7: In above screen select desired values and then click on ‘Submit’ button to get below prediction result



Screen 8: In above screen we can see the disease prediction for selected values and similarly you can select various values to get different results. You need to train only one and predict N number of times.



Screen 9: For above selection will get below result



Screen 10: In above screen we can see the disease prediction for selected values and similarly you can select various values to get different results. You need to train only one and predict N number of times.

V. 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 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 component analysis. In *Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)*, 2016 International Conference on (pp. 5–10). IEEE.
- [7]. Kazeminejad, A., Golbabaie, 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., Hossain, 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/explaining-basics-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/220pxSvm_separating_hyperplanes_%28SVG%29.svg.png