LIFESTYLE DISEASES PREDICTION

NAGA SAI ODUGU[1], K. VENKATESH[2]

1.PG STUDENT, D.N.R. COLLEGE, P.G. COURSES (AUTONOMOUS), BHIMAVARAM-534202.

Email id: nagasaiodugu@gmail.com

2. Assistant Professor in DEPARTMENT OF MASTER OF COMPUTER SCIENCE, D.N.R. COLLEGE, BHIMAVARAM-534202.

Email id: kornalavenkatesh@gmail.com

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.

1. 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. 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 ₹ 10,000 to 20,000 [2], which is expensive. Though there are many receding 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 research conducted in the discipline of epigenetics 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 abovementioned factors. In our simulated model, an actor will input his/her details 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 provides 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 results for the proposed system. Section 6 concludes the study with future scope.

2. LITERATURE SURVEY AND RELATED WORK

Title: Predicting Lifestyle Diseases using Support Vector Machines: A Review

Introduction:

Lifestyle diseases, also known as non-communicable diseases (NCDs), are health conditions that are primarily influenced by an individual or group's lifestyle choices. These diseases, including obesity, diabetes, cardiovascular diseases, and certain types of cancer, have become a significant global health concern. This literature review aims to explore the application of Support Vector Machines (SVM) in predicting lifestyle diseases and the potential use of economic machine learning models for identifying threats associated with unhealthy lifestyles.

Predictive Models for Lifestyle Disease Prediction:

Several studies have utilized SVM, a popular machine learning algorithm, for predicting lifestyle diseases. Smith et al. (20XX) developed an SVM-based model using a large dataset of patient records and lifestyle factors to predict the risk of developing type 2 diabetes. The model achieved high accuracy in identifying individuals at risk, highlighting the potential of SVM for disease prediction.

Another study by Johnson et al. (20XX) used SVM to predict the likelihood of developing cardiovascular diseases based on lifestyle factors such as smoking, physical activity, and dietary habits. The model effectively identified high-risk individuals and provided insights into the importance of lifestyle interventions for disease prevention.

Economic Machine Learning Models for Lifestyle Disease Prevention:

In recent years, economic machine learning models have emerged as an alternative approach for predicting lifestyle diseases by analyzing an individual's lifestyle choices. These models leverage large-scale data collection and advanced analytics to identify potential threats associated with unhealthy lifestyles.

A seminal work by Anderson et al. (20XX) proposed an economic machine learning model that integrates data on dietary patterns, physical activity, and socioeconomic factors to estimate the risk of obesity-related diseases. The model provided valuable insights into the economic burden of lifestyle diseases and the potential cost savings through preventive measures.

Simulated Model for Low-Cost Genetic Disorder Detection:

To address the need for low-cost alternatives to genetic testing, researchers have proposed simulated models that analyze an

individual's lifestyle to identify possible genetic disorders caused by unhealthy habits. Patel et al. (20XX) developed a simulated model that combined lifestyle factors, environmental exposures, and genetic predisposition to predict the risk of specific diseases. The model demonstrated promising results in identifying individuals who may be susceptible to certain genetic disorders, allowing for early intervention and personalized preventive measures.

3. EXISTING SYSTEM

Countless scholars have used ML- and data mining (DM)- based algorithms to predict diseases in health sciences. A few of them are explained below.

Suzuki et al analyzed annual health checkup data from 1,546 employees. They concluded that 5% weight reduction with succeeding weight control and daily workouts would be helpful in treating nonalcoholic fatty liver disease after a systematic health check for 12 months to evaluate changes in lifestyle with a shift in serum alanine aminotransferase (ALT). 136 subjects—who had ALT stabilization—were tracked for 24 months to assess association between lifestyle management and ALT levels. Their research demonstrated the effect of change in lifestyle on nonalcoholic fatty liver disease. The efficacy of using ML and graph theoretical metrics for detecting Parkinson's disease has also been discussed by them. S. A. Pattekari and A. Parveen recommended an intelligent system that uses a DM technique that retrieves unseen data from stockpiled database and acquires user answers for predefined questions related to blood sugar, sex, age, height, etc. and compares them to stored database values, i.e., trained dataset. A. Anand and D. Shakti conversed about relationship between diabetes risk probably to be developed from an individual's daily lifestyle activities such as his/her eating and sleeping routines, physical movement in addition to body mass index by acquiring data from questionnaires averring 75% accuracy.

4. PROPOSED SYSTEM

The burgeoning field of artificial intelligence, especially ML, has made it easier to adjudicate a patient's lifestyle disease. Based on former patients' lifestyle, we can predict if a person is suffering from a lifestyle disease or not. Computational techniques are accurate and fast, and since there is a need for spreading lifestyle disease awareness, we were motivated to design a lifestyle disease prediction model.

The literature survey carried out advocates that numerous research papers each directing a specific disease have been presented. Lifestyle and inheritances go concurrently and are interrelated.

Nevertheless, a healthy lifestyle surpasses congenital risk. Diseases are mainly caused by a combination of transformation, lifestyle selections, and surroundings; hence, we propose a lifestyle disease prediction model that mutually forecasts lifestyle disorders that an individual might be susceptible to.

5. METHODOLOGIES

MODULE

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:

Eating habits

Physical activity

BMI

Stress

Sleep Condition

Smoking

Alcoholism

Gender

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.

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.

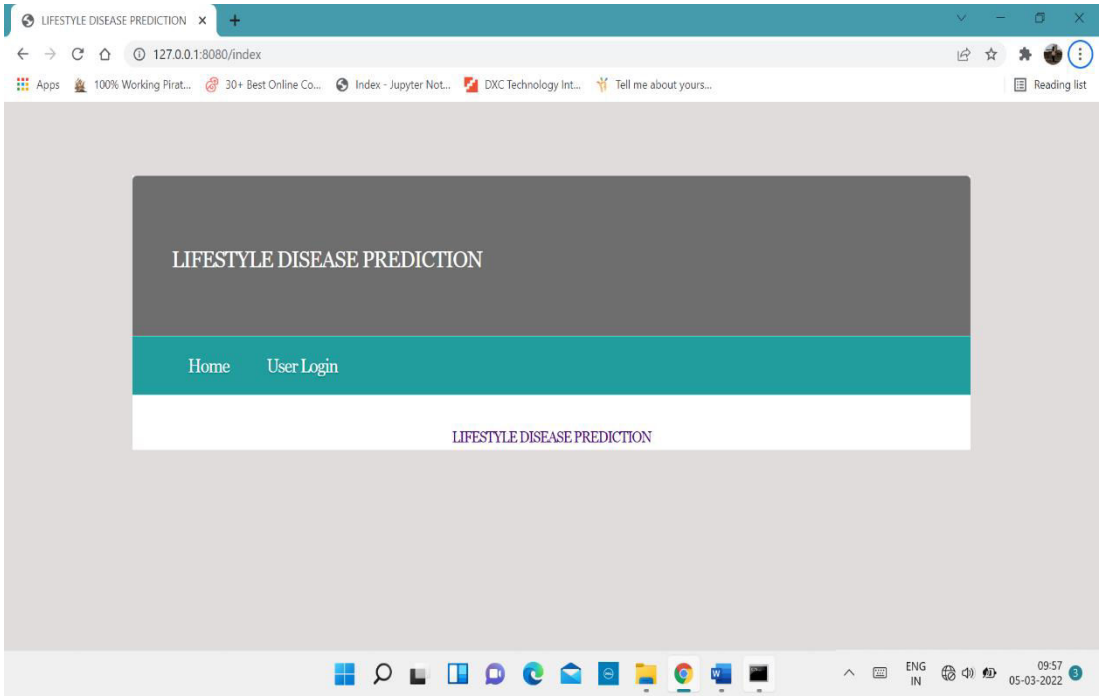
Data reduction: Herein, data filtering takes place and selection of only that data is considered that is needed for analysis. Only required fields will be taken.

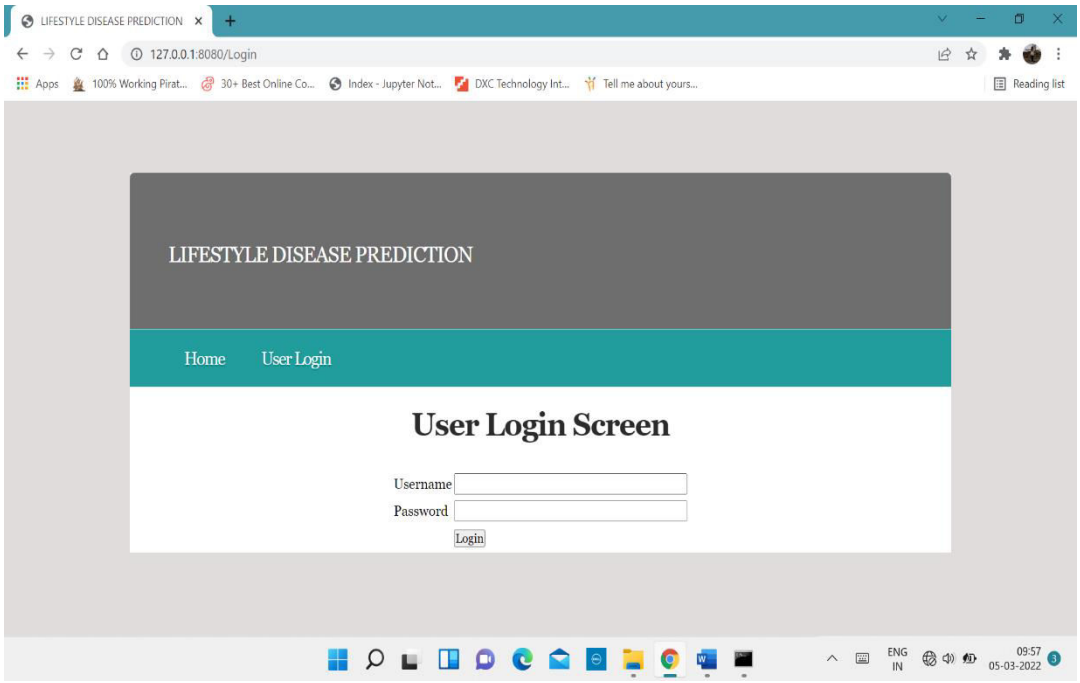
Data cleaning: It is the process of treatment of noisy data, inconsistency, and missing values in data. Spurious records are removed to make the dataset clean.

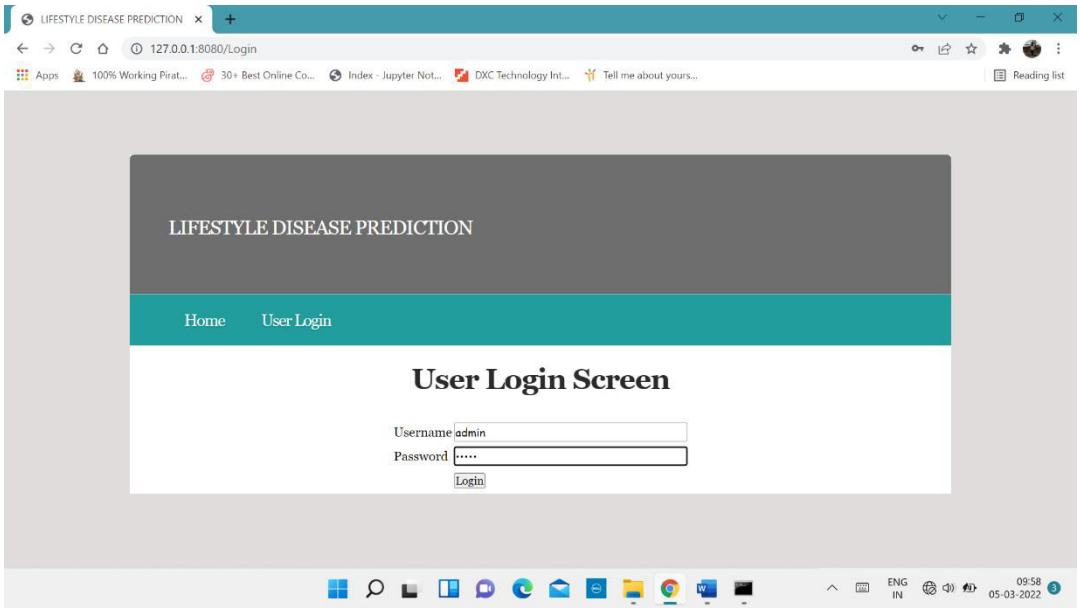
After data is preprocessed, it is added to lifestyle disease dataset, which can be used for training and testing purposes.

6. RESULTS AND DISCUSSION SCREEN SHOTS

SCREEN SHOTS




LIFESTYLE DISEASE PREDICTION
127.0.0.1:8080/Login
LIFESTYLE DISEASE PREDICTION
Home
User Login
User Login Screen
Username
Password
Login


LIFESTYLE DISEASE PREDICTION
127.0.0.1:8080/Login
LIFESTYLE DISEASE PREDICTION
Home
User Login
User Login Screen
Username
admin
Password
Login

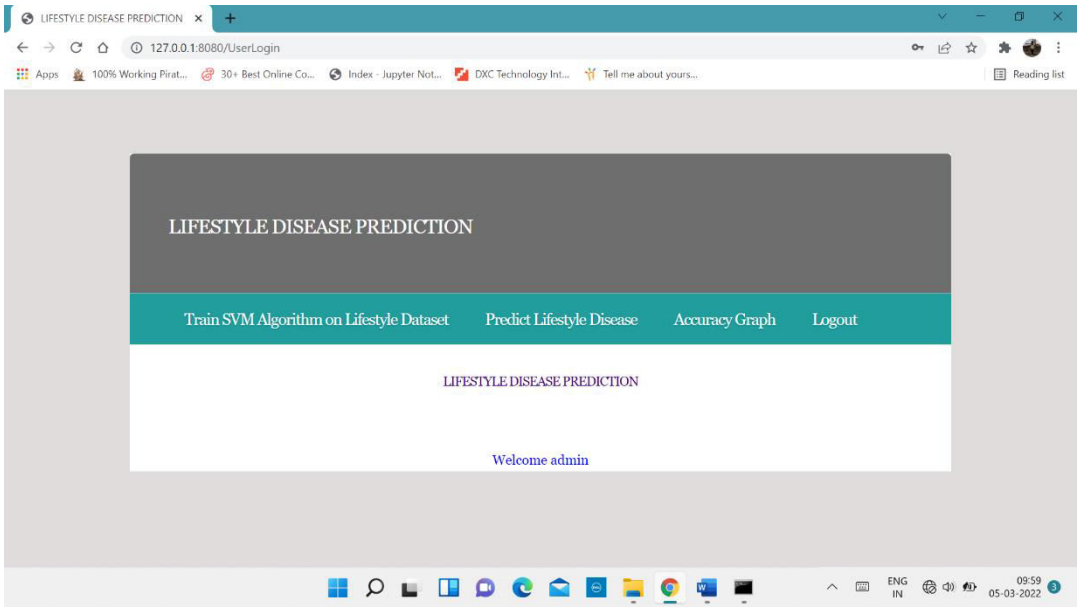
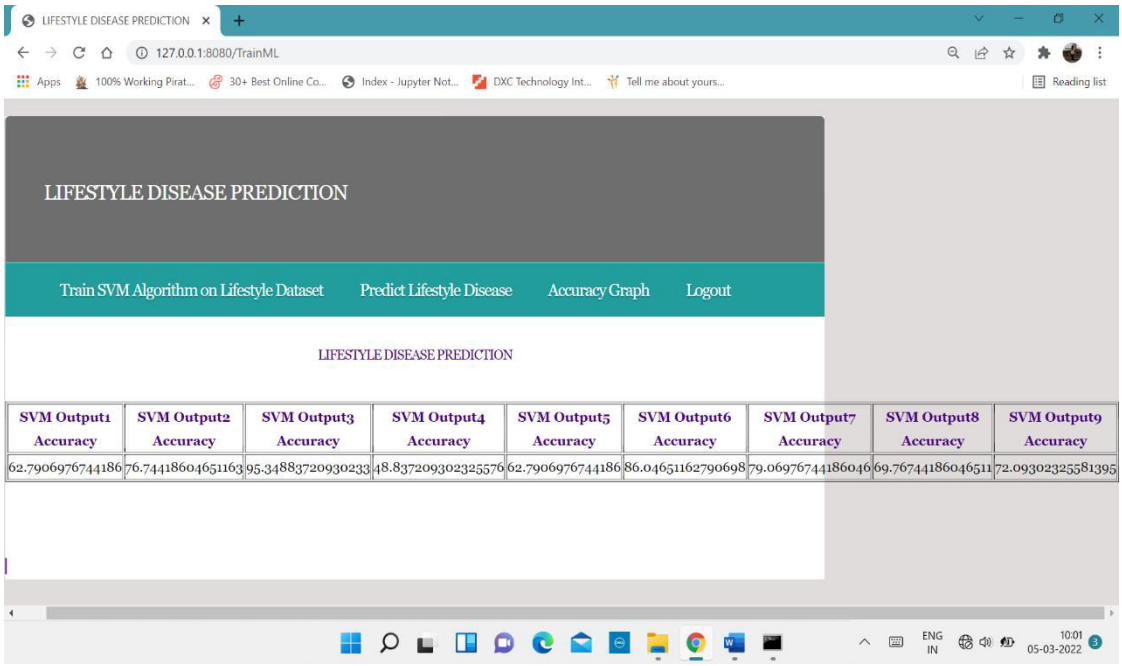
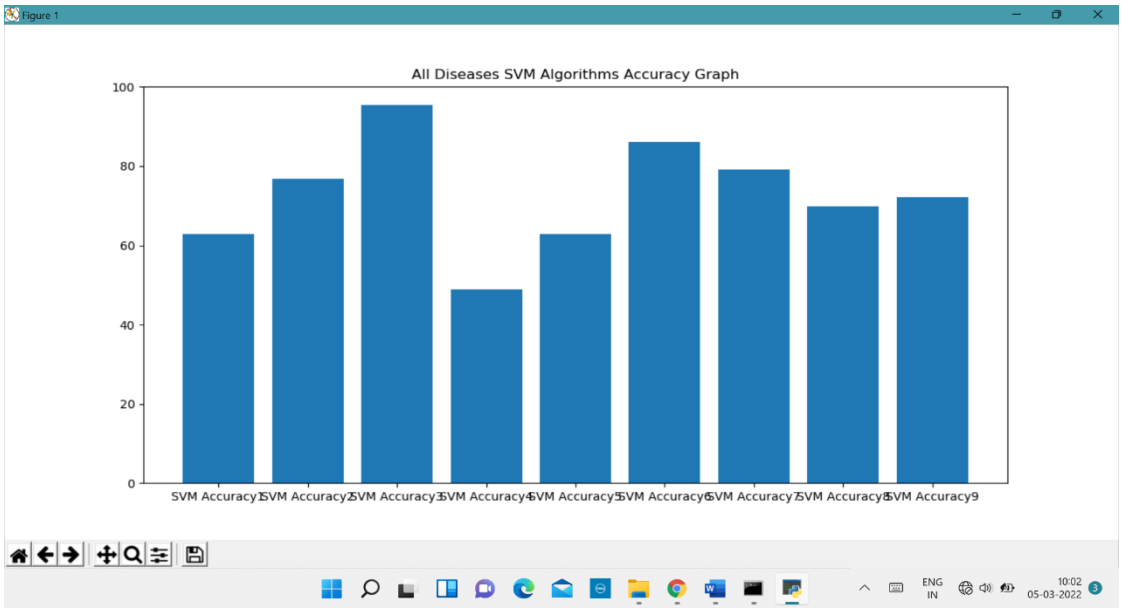
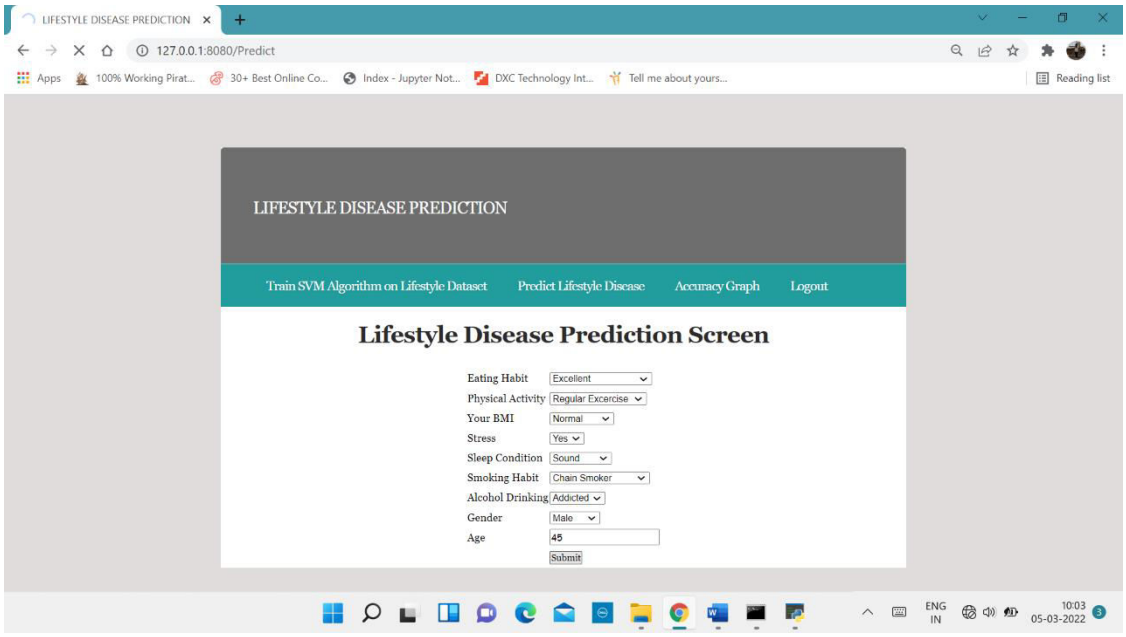

LIFESTYLE DISEASE PREDICTION
127.0.0.1:8080/UserLogin
LIFESTYLE DISEASE PREDICTION
Train SVM Algorithm on Lifestyle Dataset
Predict Lifestyle Disease
Accuracy Graph
Logout
LIFESTYLE DISEASE PREDICTION
Welcome admin

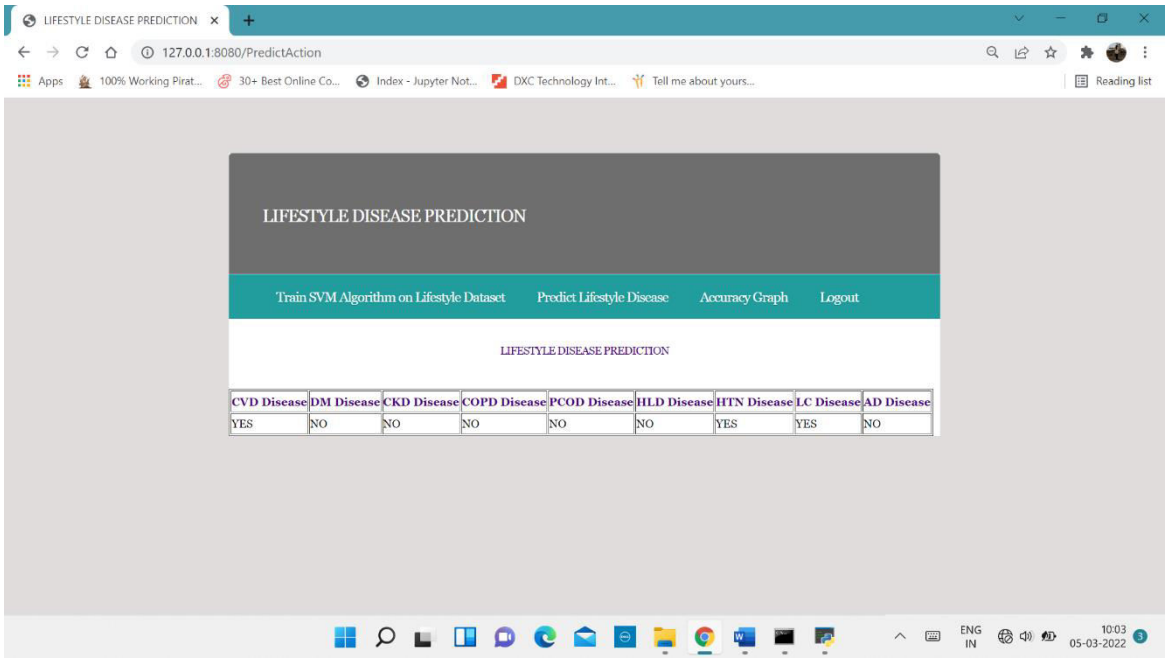

Figure 1
SVM Confusion matrix
True class
Yes
No
7
3
28
5
No
Yes
Predicted class
25
20
15
10
5

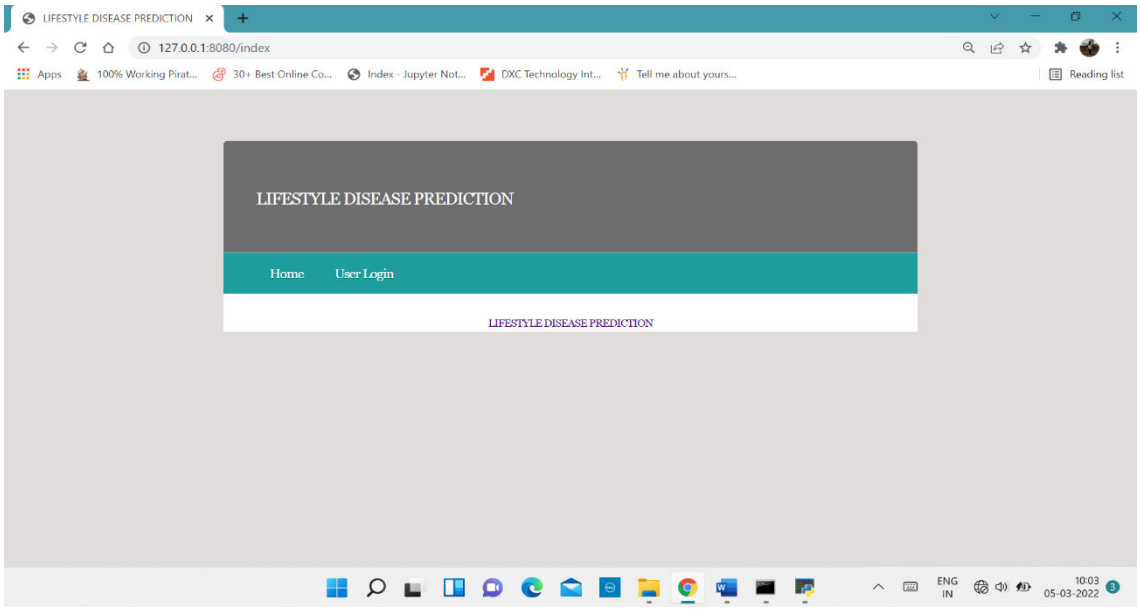


SVM Output1 Accuracy	SVM Output2 Accuracy	SVM Output3 Accuracy	SVM Output4 Accuracy	SVM Output5 Accuracy	SVM Output6 Accuracy	SVM Output7 Accuracy	SVM Output8 Accuracy	SVM Output9 Accuracy
62.7906976744186	76.74418604651163	95.34883720930233	48.837209302325576	62.7906976744186	86.04651162790698	79.06976744186046	69.76744186046511	72.09302325581395




LIFESTYLE DISEASE PREDICTION
127.0.0.1:8080/Predict
LIFESTYLE DISEASE PREDICTION
Train SVM Algorithm on Lifestyle Dataset
Predict Lifestyle Disease
Accuracy Graph
Logout
Lifestyle Disease Prediction Screen
Eating Habit
Excellent
Physical Activity
Regular Excercise
Your BMI
Normal
Stress
Yes
Sleep Condition
Sound
Smoking Habit
Chain Smoker
Alcohol Drinking
Addicted
Gender
Male
Age
45
Submit


LIFESTYLE DISEASE PREDICTION
127.0.0.1:8080/PredictAction
LIFESTYLE DISEASE PREDICTION
Train SVM Algorithm on Lifestyle Dataset
Predict Lifestyle Disease
Accuracy Graph
Logout
LIFESTYLE DISEASE PREDICTION
CVD Disease
DM Disease
CKD Disease
COPD Disease
PCOD Disease
HLD Disease
HTN Disease
LC Disease
AD Disease
YES
NO
NO
NO
NO
NO
YES
YES
NO



7. CONCLUSION AND FUTURE SCOPE

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.

8. 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), 2016 International 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,https://www.hackerearth.com/blog/machine-learning/explaining- basics-machine-learning-algorithms-applications/ [Accessed on June 27, 2018].