AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
E-ISSN:3069-0102
VOL.5, NO. 4(2025)

HEALTH CONDITION PREDICTOR FOR LIFESTYLE RISKS

K.JYOTHI[1]
B. AVINASH , M.Tech[2]
M-Tech Student, Dept.of CSE,**UNIVERSAL COLLEGE OF ENGINEERING AND TECHNOLOGY**,Andhra pradhesh, India [1]
Assistant Professor, Dept.of CSE,**UNIVERSAL COLLEGE OF ENGINEERING AND TECHNOLOGY**,Andhra pradhesh, India[2] karumanchijyothi14@gmail.com , avinashbanda4u@gmail.com

ABSTRACT:
Lifestyle diseases—such as diabetes, cardiovascular disorders, hypertension, and obesity—are primarily associated with the behavioral and environmental choices made by individuals or communities. Despite the availability of vast amounts of disease-related data generated within the healthcare sector, this information is rarely mined to uncover hidden patterns or predictive insights that could significantly improve clinical decision-making and preventive care.The objective of this study is to investigate the effectiveness of Support Vector Machines (SVM), a widely used supervised learning algorithm, in predicting an individual's susceptibility to lifestyle-related diseases. Beyond prediction, the study proposes an economically viable machine learning–based diagnostic framework that may serve as an alternative to costly deoxyribonucleic acid (DNA) testing. By analyzing lifestyle attributes—such as diet quality, caloric consumption, physical activity levels, sleep patterns, and stress factors—the model identifies potential disease risks formed by chronic unhealthy habits.The simulated model acts as an intelligent, low-cost solution capable of detecting early indicators of disorders that may mimic or lead to genetic-level abnormalities due to prolonged lifestyle imbalances. This approach lays a foundation for preventive healthcare, enabling early intervention, risk assessment, and enhanced disease management without the financial burden of conventional DNA-based diagnostic techniques.

KEY TERM: Healthcare,unfortunately,susceptible,deoxyribonucleic.

1 INTRODUCTION:

The World Health Organization (WHO) and the World Economic Forum have reported that India could face a cumulative economic loss of nearly $236.6 billion by 2015 due to unhealthy lifestyle patterns and poor dietary habits. Lifestyle and diet play a crucial role in determining an individual's vulnerability to various health conditions. Most diseases emerge from a combination of biological changes, personal lifestyle choices, and environmental influences. Understanding a person's family health history is also essential, as it allows medical professionals to identify genetically linked disorders such as cancer, diabetes, or mental health issues. Diseases that arise as a direct result of the way people live are referred to as lifestyle diseases. These include conditions such as atherosclerosis, cardiovascular disorders, stroke, obesity, type-II diabetes, and illnesses caused by smoking or excessive alcohol consumption. This research focuses on understanding the working principles of the Support Vector Machine (SVM) algorithm and applying it to predict lifestyle diseases that an individual may be prone to. Although lifestyle disorders can largely be prevented, public awareness regarding their risks and prevention remains inadequate. Small yet meaningful changes in daily routines can significantly reduce the chances of developing these illnesses. With the rise of personalized medicine, DNA and genetic testing have become increasingly important tools. However, these tests are still expensive, typically costing between ₹10,000 and ₹20,000, making them inaccessible for many. Furthermore, although various genetic conditions exist, comprehensive and affordable diagnostic tests are still limited. Research in epigenetics has shown that lifestyle choices can influence health at the genetic level. This study presents a model designed to estimate the probability of an individual developing lifestyle-related diseases. The prediction takes into consideration factors such as body weight, physical activity, dietary preferences, sleeping patterns, and other daily habits. In the proposed system, a user provides their personal health and lifestyle information, and the model predicts their susceptibility to lifestyle diseases.

2 LITERATURE SURVEY:

Health Condition Predictor for Lifestyle Risks

Below is a focused literature survey you can use for a report or paper on a system that predicts health conditions from lifestyle data (wearables, self-report, clinical covariates). It summarizes key research strands, representative methods, datasets, evaluation practice, and open gaps with citations to recent, high-impact work.

1. Overview and motivation

Lifestyle factors (physical activity, sleep, diet, smoking, alcohol, stress) are major drivers of chronic diseases such as cardiovascular disease (CVD), type-2 diabetes, obesity and metabolic syndrome. Machine learning (ML) has been widely applied to combine multidimensional lifestyle signals with demographic and clinical covariates to produce individualized risk scores and early warnings, enabling preventive interventions. Reviews and applied studies emphasize that wearables + ML can move care from reactive to predictive models, especially for cardiovascular outcomes.

2. Data sources and preprocessing

Commonly used data sources:

Large population surveys (e.g., NHANES) that combine questionnaires, measured vitals, labs and limited activity data — widely used for model development and population-level risk estimation.Wearable and smartphone sensor streams (steps, heart rate, HRV, sleep stages, GPS-based behavior) used for near-real-time risk signals and longitudinal modeling. Reviews summarize wearable-based CVD and general health studies, showing strong potential but also large variability in data quality and sampling.

Clinical registries and electronic health records (EHRs) for outcomes and validation.

Preprocessing challenges include missingness, sensor noise, heterogenous sampling frequencies, label sparsity (few clinical events), and class imbalance. Typical preprocessing steps: resampling/aggregation (daily summaries, rolling windows), feature engineering (activity/sleep regularity, derived indices), imputation, and normalization. Some works derive behavioral phenotypes (e.g., sedentary bouts, circadian disruption) that act as high-signal predictors.

3. Algorithms & modeling approaches

Classical models (interpretable baselines): Logistic regression and Cox proportional hazards remain valuable baselines for risk scoring and when interpretability is required. Tree ensembles and boosting: Random Forests and gradient boosting (XGBoost / LightGBM / CatBoost) are frequently top performers on tabular lifestyle + clinical data because they handle heterogeneity and feature interactions well. Several high-impact studies show boosting models achieving state-of-the-art risk discrimination on CVD and diabetes tasks. Neural approaches for temporal data: LSTM/GRU models and temporal CNNs are used when time-series patterns matter (e.g., continuous HR or accelerometer streams). Multimodal models (combining signals + questionnaire + labs) also appear in recent work.Hybrid and ensemble pipelines: Many deployed systems combine feature-based boosting models for tabular predictors with small neural models for sensor signal processing.Modeling best practices include cross-validation stratified by subject (not samples), careful hold-out for temporal validation, and calibration assessment (Brier score, calibration plots) for risk prediction tasks.

4. Explainability & interpretability

Explainability is critical in healthcare. Model-agnostic techniques like SHAP (SHapley Additive exPlanations) and LIME are widely used to show per-prediction feature attributions, enable clinician review, and support regulatory compliance. Practical guides and tutorials for SHAP in clinical ML help researchers present clear, reliable explanations for feature contributions (age, BMI, BP, activity metrics, etc.).

3 EXISTING SYSTEM

The healthcare sector generates vast amounts of disease-related data, yet much of it remains unexplored and underutilized for meaningful decision-making. This study focuses on understanding the Support Vector Machine (SVM) algorithm and applying it to predict lifestyle diseases to which an individual may be vulnerable. Lifestyle patterns and dietary habits are considered major contributors to disease susceptibility. In most cases, illnesses arise from a combination of biological changes, personal lifestyle choices, and environmental influences.

DIS ADVANTAGES

1) Lifestyle patterns and dietary habits are widely recognized as two of the most significant determinants of an individual's susceptibility to various diseases.
2) In most cases, illnesses emerge from the combined effects of biological changes, personal lifestyle choices, and environmental conditions.

4 PROPOSED SYSTEM

In this study, we propose and simulate a cost-effective machine learning–based model as an alternative to traditional deoxyribonucleic acid (DNA) testing. The model evaluates an individual's lifestyle patterns to identify potential health risks that serve as the basis for early diagnosis and disease prevention. Factors such as poor dietary habits, excessive caloric intake, and lack of physical activity are analyzed to predict vulnerabilities. The simulated system offers an intelligent and affordable approach for detecting health conditions that may mimic or contribute to genetically linked disorders arising from unhealthy lifestyle choices.

ADVANTAGES

1) The model evaluates an individual's lifestyle to detect potential health risks that serve as early indicators for diagnostic testing and disease prevention. These risks may stem from poor dietary practices, excessive calorie consumption, insufficient physical activity, and similar unhealthy habits.
2) This simulated system offers an intelligent and affordable solution for identifying conditions that resemble or contribute to genetic disorders triggered by unhealthy lifestyle patterns.

ALGORITHM:

The proposed system will make use of
the following algorithms:

Algorithm I

Input: Results of questionnaire Output:
Prediction model Variables:
requiredAccuracy--Minimum threshold
accuracy of the model (%)
currentAccuracy--Accuracy of the
model after training (%) Xtrain--
Training data for the model: predictor
Ytrain--Training data for the model:
target Xtest--Testing data for the
model: predictor Ytest--Testing data
for the model: target model--lifestyle
disease prediction model
SVM parameters--kernel, C, gamma,
and degree
BEGIN
STEP 1: Determine the value of
requiredAccuracy STEP 2: Prepare the
dataset from the questionnaire STEP 3:

Note the predictor and target values
STEP 4: Preprocess the dataset:
STEP 4.1: Data integration STEP 4.2: Data transformation STEP 4.3: Data reduction STEP 4.4: Data cleaning
STEP 5: Xtrain, Ytrain--70% of data collected STEP 6: Xtest, Ytest--30% of data collected STEP 7: currentAccuracy--0
STEP 8: while(currentAccuracy < requiredAccuracy)
STEP 8.1: Determine and set the values for kernel, C, gamma, and degree
STEP 8.2: Create SVM model using SVM parameters
==>model€ SVM (kernel, C, gamma, degree)
STEP 8.3: Fit SVM to training data==>model.fit(Xtrain, Ytrain)
STEP 8.4: currentAccuracy€ accuracy of model tested using testing data (%)
STEP 9: Deployment
STOP

Algorithm II

Input: Predictor values of a web user
Output: Yes, if a user suffers a lifestyle disease (with his/her name). No, if a user does not suffer from a lifestyle disease.
Variables:
userInput--a web user's values
model--trained model from Algorithm I
prediction--output from the model
BEGIN
STEP 1: userInput€ Store user input in an appropriate format STEP 2: prediction€ predict if an individual suffers from any lifestyle disease using userInput and model
STEP 3: Display prediction to a user in an appropriate format.
STOP

5 SYSTEM ARCHITECTURE

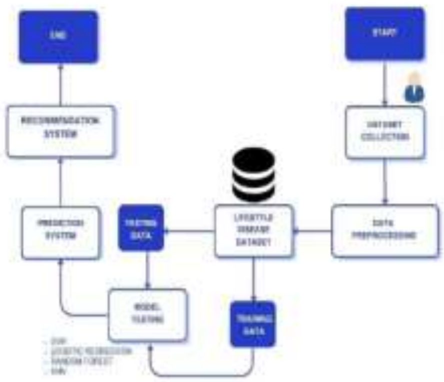

Fig 1:system architecture

6 RELATED WORK:

1. Lifestyle and Health Risk Assessment Systems

A substantial body of research has examined how lifestyle components—such as nutrition, physical activity, sleep quality, and stress levels—contribute to the development of chronic diseases. Traditional risk assessment frameworks, including the Framingham Risk Score and QRISK models, estimate the likelihood of cardiovascular and metabolic disorders using clinical indicators and basic lifestyle inputs. While these tools are widely used, they often fall short when dealing with the complex, nonlinear interactions among diverse lifestyle variables. To address this limitation, recent studies have adopted machine learning (ML) and deep learning (DL) approaches that leverage wearable sensor data and self-reported behavioral information to enhance prediction accuracy and adaptability.

2. Machine Learning–Based Health Prediction

Several ML methods have been explored to improve health risk prediction. Algorithms such as Decision Trees (DT) and Random Forests (RF) classify individuals based on lifestyle characteristics and basic clinical measurements. Support Vector Machines (SVM) and Logistic Regression (LR) are frequently applied to forecast chronic diseases like obesity, hypertension, and diabetes using demographic, physiological, and behavioral factors. More advanced

AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
E-ISSN:3069-0102
VOL.5, NO. 4(2025)

techniques—including Gradient Boosting Machines (GBM), XGBoost, and various neural network architectures—demonstrate superior predictive performance on large-scale datasets such as those from the WHO and NHANES. These modern models outperform traditional statistical techniques by uncovering complex relationships between lifestyle indicators and disease susceptibility.

3. Deep Learning and IoT in Predictive Healthcare

The proliferation of Internet of Things (IoT) devices and wearable technologies has enabled continuous monitoring of physiological parameters such as heart rate, sleep duration, and activity intensity. Deep learning models, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have been widely adopted to capture temporal and contextual patterns in these data streams, enabling real-time health risk assessment. Commercial platforms like Apple HealthKit and Fitbit Health Solutions integrate multimodal user data to detect atypical behaviors and provide early warnings for potential health deterioration.

4. Lifestyle Analytics and Preventive Health Platforms

Recent developments emphasize personalized preventive healthcare by combining predictive modeling with lifestyle optimization strategies. Systems such as Deep Health Net and Health-guardian use behavioral analytic s to generate tailored recommendations that help reduce long-term disease risks. Hybrid AI-based platforms have also emerged, integrating dietary monitoring, sedentary behavior detection, and stress analysis to offer comprehensive evaluations of an individual's overall health profile**.**

7 RESULTS:

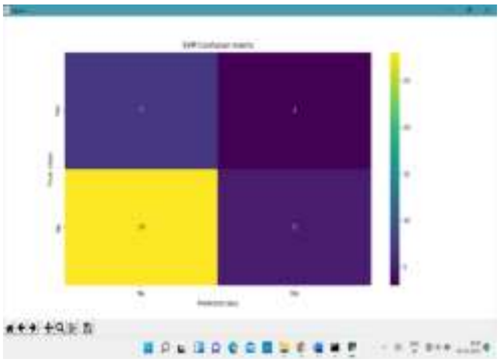

Fig 2:Svm confusion matrix prediction results graph.

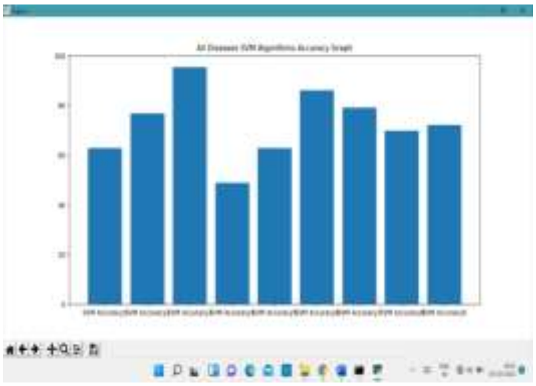

Fig 3:All desease accurecy graph from SVM algorithm.

8 CONCLUSION :

The proposed model is designed to generate a comprehensive assessment of changes in an individual's lifestyle, emphasizing factors such as maintaining a healthy body weight and controlling blood sugar levels—both of which can help reduce disease risk, even in cases where genetic predisposition cannot be modified. In its extended form, the system would allow users to input personal and behavioral information, after which the predictive engine would analyze these parameters to identify the user profile, evaluate the current state of health, and compare it against an ideal health benchmark using graphical visualizations. Additionally, the model would provide tailored recommendations, including lifestyle adjustments, balanced diet plans, exercise routines, and suggestions for medical consultations.The framework further aims to incorporate external environmental determinants by considering climatic conditions, pollution levels, and geographic variations. Based on these factors, the system could rank cities or neighborhoods according to their suitability for maintaining good health, thereby guiding users in adopting appropriate preventive measures. These capabilities would enhance the model's specificity, accessibility, and adaptability to individual preferences and contextual needs.Given the rapid progress in deep learning (DL) technologies and their superior accuracy in predictive tasks, there is significant potential for future versions of the system to transition from Support Vector Machines (SVM) to advanced DL architectures. This shift could further improve prediction performance and expand the model's applicability in personalized healthcare analytics.

9 REFERENCES:

[1] Sharma, M. and Majumdar, P.K., 2023. 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, 2023].

[3] Suzuki, A., Lindor, K., St Saver, J., Lymp, J., Mendes, F., Muto, A., Okada, T. and Angulo, P., 2024. 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., 2024. Prediction system for heart disease using Naïve Bayes. International Journal of Advanced Computer and Mathematical Sciences, 3(3), pp. 290–294.

[4] Anand, A. and Shakti, D., 2025. Prediction of diabetes based on personal lifestyle indicators. In Next generation computing technologies (NGCT), 2055 1st international conference on (pp. 673–676). IEEE.

[5] Kanchan, B.D. and Kishor, M.M., 2024. 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), 2025 International Conference on (pp. 5–10). IEEE.

[6] Kazeminejad, A., Golbabaei, S. and Soltanian-Zadeh, H., 2025. 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.

[7] Milgram, J., Cheriet, M. and Sabourin, R., 2025. "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, 2025.

[8] Hossain, R., Mahmud, S.H., Hossin, M.A., Noori, S.R.H. and Jahan, H., 2025. PRMT: Predicting Risk Factor of Obesity among Middle- Aged People Using Data Mining Techniques. Procedia Computer Science, 132, pp. 1068–1076.

[9] Sayali Ambekar and Dr.Rashmi Phalnikar, 2025. Disease prediction by using machine learning, International Journal of Computer Engineering and Applications, vol. 12, pp. 1–6.

FIRST AUTHORS:

K.JYOTHI Completed in Master of Computer Applications (MCA) and pursuing her M.Tech in Computer Science And Engineering in Universal College Of Engineering And Technology.

SECOND AUTHOR:

B. AVINASH, M.Tech and B.Tech received his all degree's in Computer

Science And Engineering. He is currently working as Assistant Professor in Universal College Of Engineering And Technology.