International Journal of Aquatic Science
ISSN: 2008-8019
Vol 14, Issue 01, 2023

Smart Machine Learning Model Early Prediction of Lifestyle Diseases

Manish Patil[1], Hitesh Bhadane[2], Devdatt Shewale[3], Mahesh Lahoti[4], Dr.Anand Singh Rajawat[5], Ram Kumar Solanki[6]

[1,2,3,4]B.Tech (Scholar),School of Computer Science and Engineering,Sandip University,Nashik,India
[5]Associate Professor, School of Computer Science and Engineering,Sandip University,Nashik,India
[6]Assistant Professor, School of Computer Science and Engineering,Sandip University,Nashik,India

Email: [1]manishpatil5999@gmail.com, [2]hiteshbhadane1432@gmail.com, [3]devshewale1310@gmail.com, [4]maheshlahoti9@gmail.com, [5]anandsingh.rajawat@sandipuniversity.edu.in, [6]ramkumar.solanki@sandipuniversity.edu.in

Abstract: Lifestyle diseases like diabetes, heart disease, and obesity are a world health concern because their rates are going up and their effects are getting worse. Predicting and diagnosing these diseases early is important if you want to stop them from getting worse, lower healthcare costs, and improve patient outcomes. Modern machine learning methods are one of the best ways to find lifestyle diseases early and figure out how likely they are to happen. The summary of this study talks about how advanced machine learning models can be used to find lifestyle-related diseases early. We look into how different machine learning methods, such as decision trees, support vector machines, random forests, and deep learning models, can be used to analyse data from a wide range of sources, such as EHRs, wearables, and information about people's daily routines. Using cutting-edge data analytics methods, the sophisticated machine learning models can find trends and correlations in the data that haven't been seen before. This makes it possible to find diseases that could be life-threatening early on. These models take into account a wide range of factors, such as age, gender, health state, family history, and dietary and lifestyle preferences. Using this information, the models give each person a unique risk score, which lets doctors give more accurate diagnoses and tips on how to stay healthy.

Keywords: Lifestyle diseases, death, treatment, common lifestyle, prediction, prevention, management.

1. INTRODUCTION

Noncommunicable diseases (NCDs), which include lifestyle-related illnesses, have become a big public health issue around the world. Diabetes, heart disease, and obesity are on the rise, and a lot of it has to do with how people live their lives, like not being active, eating poorly, and smoking. Early detection and prediction are important if you want to avoid, treat, and improve the health effects of lifestyle disorders. Clinical risk assessment tools, which are often used in conventional ways to predict illness, look at things like age, family history, and regular

physical exams, among other things. But these methods might not take into account how lifestyle-related diseases change over time or how each person's unique circumstances can lead to disease. In this situation, machine learning models that are smart can be very helpful. Smart machine learning models use cutting-edge algorithms and methods to look at huge datasets and find patterns, connections, and trends that could be signs of lifestyle disease risks. By adding data from a wide range of sources, such as electronic health records, wearable devices, and apps that track your lifestyle, these models can make better guesses and make them fit each person. It takes the understanding of doctors, data scientists, and engineers all working together to make good machine learning models that can predict lifestyle diseases early on. Some of the sources of information that the models use are demographics, health data, biomarkers, genetic information, activity levels, sleep patterns, and food. Intelligent machine learning models can keep learning and changing as they get more data, which is a big plus. With this adaptive learning process, the models can improve their predictions over time and give specific advice for preventing and treating illnesses. The data from these models can give us a more full picture of how diseases develop because they can pick up on complex interactions and non-linear correlations between different risk factors. The use of powerful machine learning algorithms in healthcare systems seems to help a lot with predicting diseases and preventing them. Early identification of people at high risk for lifestyle diseases makes it possible for health care workers to do targeted interventions and give each person specific advice on how to change their behaviour. The focus can be moved from emergency care to preventive care, and these models can help make that change easy. Smart machine learning models that can predict lifestyle diseases early are important, but making them and putting them to use can be hard for a number of reasons. Some of the most important things that need to be fixed are how private and safe the data is, how easy it is to understand the models, and how widely healthcare services are available. For these models to be used in a responsible and effective way, healthcare workers, data scientists, lawmakers, and regulatory agencies must work together.
Intelligent machine learning algorithms open the door to personalised therapies and better health results by making it possible to predict lifestyle disorders early on. With the help of cutting-edge data science and algorithmic methods, these models have a great chance of making a huge difference in health care by changing the way diseases are prevented. Lifestyle diseases can still be hard to predict and prevent on an individual level, but these problems can be solved with more study and teamwork.

Problem Definition

In this platform, the people and doctors are joined and they are registered login logout and also they have early knowledge of what is happening in their area and the true information is given to this platform. About the disease which is a large amount of spearing in this area first of all the people see and their area which diseases is spread and also check their symptoms, also in bottom, and also give what the precaution do for the disease not come, also if you have any other disease then also have a chatbox to say your problem there was the best doctor, and they convey you to what to do for these diseases. And also doctors joined this platform as a sub-admin. They solve the people’s posted questions, and also add current disease information and safety tips and what precautions we do for thesediseases and also know about these doctors who give this information. Admin doctor verifies this doctor’s added information and verifies the doctor’s status and manages the site settings.

Motivation

Lifestyle diseases are common among the population today not only in India but also in almost every country. These diseases are caused by the habits that we have on a day-to-day basis. Heart disease and hypertension are examples of lifestyle diseases . Lifestyle diseases often go unnoticed and untreated due to a lack of awareness and limited access to healthcare. Many people are unaware of the warning signs and risks associated with these diseases, leading to delayed diagnosis and treatment.

To address this issue, we have developed a portal that provides early disease information and solutions or precautions. The portal is designed to help people become more aware of their health and take steps to prevent or manage lifestyle diseases. By providing early information about diseases, the portal can help people get the treatment they need before it is too late.

Liteeature Review

Our project utilizes Data Mining, a machine learning concept, to analyze vast amounts of data in order to discover valuable information and knowledge. Data Mining combines artificial intelligence, statistics, probability, machine learning, deep learning, and database system technology. The process involves collecting, selecting, cleaning, handling missing values, transforming, mining, evaluating patterns, and visualizing knowledge from the data.

In the context of the health sector, which generates a significant amount of heterogeneous and sensitive data, data mining plays a crucial role in predicting, preventing, and managing diseases effectively. Medical diagnosis heavily relies on available data, including user-entered information. However, even minor changes in the data can have a significant impact on predictions and outcomes. Therefore, data mining in healthcare is challenging but essential for uncovering hidden patterns and extracting knowledge from databases to predict potential diseases.

In our project, we are focusing on incorporating both descriptive and predictive models of data mining, specifically designed for big data scenarios. Descriptive analysis involves analyzing user data to identify patterns and relationships among variables and samples. Examples of descriptive models include the Apriori association rule, data clustering, summarization, and visualization techniques. These models help us understand the underlying structure and characteristics of the data.

On the other hand, predictive analysis utilizes historical and current data to forecast the probabilities of future lifestyle diseases or aid in the diagnosis and treatment of existing diseases. By analyzing patterns, trends, and relationships in the data, predictive models can make predictions about individuals' susceptibility to certain diseases based on their lifestyle factors, medical history, and other relevant data. However, in the case of severe symptoms, our system will always recommend consulting a doctor as soon as possible. As an additional enhancement, we plan to incorporate nearby hospitals or clinics using Google Maps, utilizing techniques such as Dijkstra's algorithm.

Commonly used predictive data models in healthcare include CART decision trees, artificial neural networks (ANN), random forests, and various regression techniques such as linear regression, logistic regression, and ridge regression. These models play a crucial role in accurately predicting and managing diseases.

Table 1: Comparative analysis

Paper	Research Topic	Methodology	Key Findings
Tan et al. (2023)	Anti-Counterfeiting and Traceability in Food Supply Chain	Consensus algorithm	Proposed a weightage-based consensus algorithm to enhance anti-counterfeiting and traceability in the food supply chain of Industry 4.0
Rajawat et al. (2023)	Real-Time Driver Sleepiness Detection	Fusion deep learning algorithm	Developed a real-time driver sleepiness detection system using a fusion deep learning algorithm
Rajawat et al. (2023)	5G-Enabled Cyber-Physical Systems for Smart Transportation	Blockchain technology	Explored the use of blockchain technology in 5G-enabled cyber-physical systems for smart transportation
Rajawat et al. (2023)	Cognitive Adaptive Systems for Industrial IoT	Reinforcement algorithm	Proposed cognitive adaptive systems for the Industrial Internet of Things (IoT) using a reinforcement algorithm
Nagaraj et al. (2023)	Secure Encryption with Energy Optimization	Pseudo algorithm	Presented an improved secure encryption scheme with energy optimization based on a random permutation pseudo algorithm in wireless sensor networks
Chouhan et al. (2022)	Position Estimation in Industry 4.0	Trilateration and RSSI	Conducted experimental analysis for position estimation using trilateration and RSSI in an Industry 4.0 context

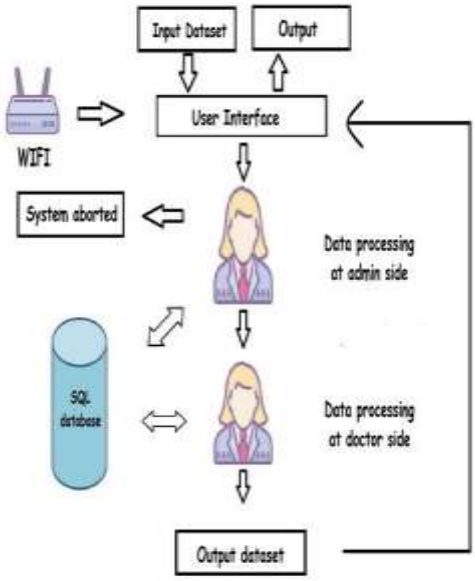
2. PROPOSED METHODOLOGY

Data mining is a powerful tool for examining large datasets with the objective of locating insights that can be put into action and easing the process of finding new information. It includes concepts from a variety of different fields, such as artificial intelligence (AI), statistics, machine learning (ML), and databases. After the data have been gathered, selected, and cleansed, the missing values have been dealt with, transformation, mining, and pattern evaluation have been performed, and now the knowledge has been visualised. Businesses and researchers will be able to acquire a more in-depth grasp of their data and make decisions that are more well-informed if these methods are followed.

The exponential increase in the quantity of data being generated in the healthcare sector. Logical regression accomplishes the same goal as linear regression does, which is to investigate the relationship between the variables using the same methods. In logistic regression, the curve is constructed not by merely assessing probabilities but rather by applying the natural logarithm of the odds to the data in order to do so. There are many different approaches to regression, but some of the more common ones are called Ordinary Least Squares (OLS), Support Vector Machines (SVM), and Generalised Linear Models (GLM).

International Journal of Aquatic Science
ISSN: 2008-8019
Vol 14, Issue 01, 2023

In the field of data mining, the Generalised Linear Model, also known as GLM, is a versatile framework that extends ordinary linear regression, often known as OLS. This allows the GLM to handle non-normal response distributions and non-linear link functions.
The study of medical data benefits tremendously from the utilisation of regression analysis due to the fact that this method makes it possible to establish correlations between a large number of variables. These connections could be of the Multiple Input Single Output (MISO) form or the Single Input Single Output (SISO) variety. Both of these terms refer to the same type of connection. In contrast, clustering means grouping things of a similar kind without resorting to a set of rigorous classifications. This is in contrast to categorising, which involves a set of rigid categories. This approach does not include any form of human supervision in any way. K-means and kernel K-means are two examples of clustering techniques. Other examples of clustering techniques include hierarchical clustering and Gaussian mixture models (GMM). In a nutshell, the use of these methodologies and algorithms is absolutely necessary for the purpose of assisting healthcare providers in gaining meaningful insights from medical data.



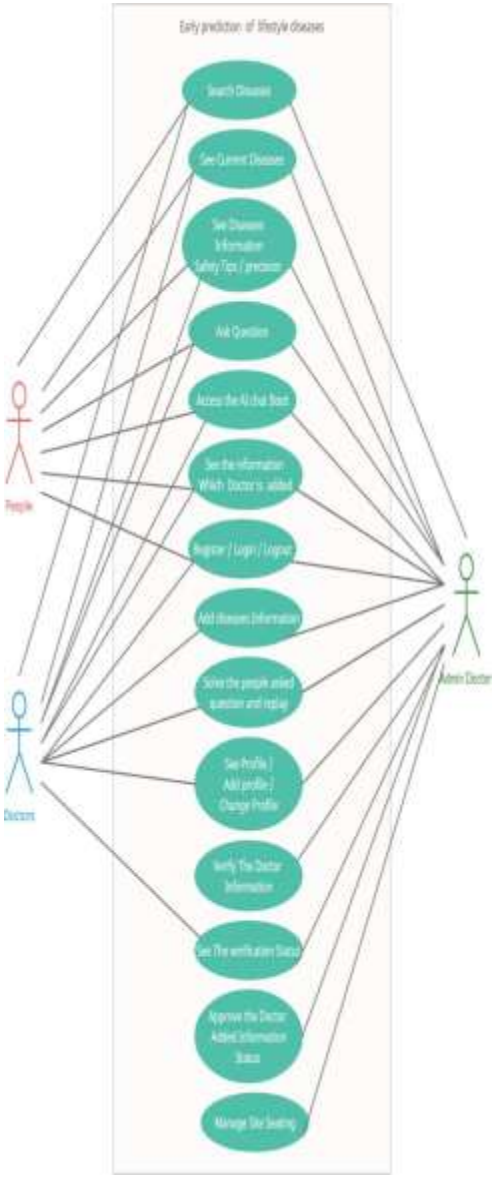
System Architecture

A mental picture of a system's architecture, which includes its physical components, operational methods, and other views, is referred to as an architectural diagram. A comprehensive and authoritative description and illustration of a system, such as one that can be found in an architecture description. In broad strokes, it describes the configuration of the portal as well as its functions. Documentation of the system's architecture provides information about its components, including how those components are connected and what they do.
how they interact with each other. It also includes information about the data that is used by the system and how it is stored and processed.

International Journal of Aquatic Science
ISSN: 2008-8019
Vol 14, Issue 01, 2023

Use Case Diagram

For the disease detection system project, the Use Case diagram has all of the usual parts. The use case diagram shows how different patients (also called people), doctors (also called physicians), and managers (also called administrative physicians) depend on each other and are linked. This is an example of how the system works as a whole.The following steps are involved in pedagogical research:



Functional Requirements

The functional needs of a software system explain what it should do and how it should do it. It describes what the system is supposed to do and how it should act. A function is a link between a certain kind of input, a certain kind of behaviour, and a certain kind of output. The functional requirements of a system or programme list the features and functions that the

system or programme must have in order to be considered useful. The goals of the system are laid out in these specs, which may include math calculations, technical details, data manipulation, processing, and other tasks.

Non-Functional Requirement

Non-functional requirements, in contrast to functional requirements, concentrate on the overall performance of a system rather than how it should behave. Functional requirements define how a system should behave. Compliance with standards, dependability, availability, security, maintainability, and portability are a few examples of desirable characteristics. Other considerations include performance, needs for the logical database, design limitations, and design portability.

3. CONCLUSION

In the end, this study shows that there are many things that make it hard for people to get regular health checks and learn about diseases. That's why we chose to do this particular thing.

We look at how the use of AI and ML in healthcare has changed the world we live in today. We build an ML-based web app with a Symptom Checker that can come up with possible diseases and information about those diseases. This can help the user get medical help right away and learn more about diseases. Also, the user can ask the doctors questions to clear up any doubts.We can now say that this project is done and has given us what we wanted.

4. REFERENCES

1. Tan, J.; Goyal, S.B.; Singh Rajawat, A.; Jan, T.; Azizi, N.; Prasad, M. Anti-Counterfeiting and Traceability Consensus Algorithm Based on Weightage to Contributors in a Food Supply Chain of Industry 4.0. Sustainability 2023, 15, 7855. https://doi.org/10.3390/su15107855
2. Rajawat, A.S. et al. (2023). Real-Time Driver Sleepiness Detection and Classification Using Fusion Deep Learning Algorithm. In: Singh, Y., Singh, P.K., Kolekar, M.H., Kar, A.K., Gonçalves, P.J.S. (eds) Proceedings of International Conference on Recent Innovations in Computing. Lecture Notes in Electrical Engineering, vol 1001. Springer, Singapore. https://doi.org/10.1007/978-981-19-9876-8_34.
3. Rajawat, A.S.; Goyal, S.B.; Bedi, P.; Verma, C.; Ionete, E.I.; Raboaca, M.S. 5G-Enabled Cyber-Physical Systems for Smart Transportation Using Blockchain Technology. Mathematics 2023, 11, 679. https://doi.org/10.3390/math11030679
4. Rajawat, A.S.; Goyal, S.B.; Chauhan, C.; Bedi, P.; Prasad, M.; Jan, T. Cognitive Adaptive Systems for Industrial Internet of Things Using Reinforcement Algorithm. Electronics 2023, 12, 217. https://doi.org/10.3390/electronics12010217.
5. Nagaraj, S.; Kathole, A.B.; Arya, L.; Tyagi, N.; Goyal, S.B.; Rajawat, A.S.; Raboaca, M.S.; Mihaltan, T.C.; Verma, C.; Suciu, G. Improved Secure Encryption with Energy Optimization Using Random Permutation Pseudo Algorithm Based on Internet of Thing in Wireless Sensor Networks. Energies 2023, 16, 8. https://doi.org/10.3390/en16010008.
6. R. S. Chouhan et al., "Experimental Analysis for Position Estimation using Trilateration and RSSI in Industry 4.0," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 904-908, doi: 10.1109/SMART55829.2022.10047276.
7. .Rajawat, A.S. et al. (2023). Real-Time Driver Sleepiness Detection and Classification

Using Fusion Deep Learning Algorithm. In: Singh, Y., Singh, P.K., Kolekar, M.H., Kar, A.K., Gonçalves, P.J.S. (eds) Proceedings of International Conference on Recent Innovations in Computing. Lecture Notes in Electrical Engineering, vol 1001. Springer, Singapore. https://doi.org/10.1007/978-981-19-9876-8_34

8. S. Rajawat, S. B. Goyal, P. Bedi, N. B. Constantin, M. S. Raboaca and C. Verma, "Cyber-Physical System for Industrial Automation Using Quantum Deep Learning," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 897-903, doi: 10.1109/SMART55829.2022.10047730.
9. S. Rajawat et al., "Security Analysis for Threats to Patient Data in the Medical Internet of Things," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 248-253, doi: 10.1109/SMART55829.2022.10047322.
10. P. Pant et al., "Using Machine Learning for Industry 5.0 Efficiency Prediction Based on Security and Proposing Models to Enhance Efficiency," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 909-914, doi: 10.1109/SMART55829.2022.10047387.
11. P. Pant et al., "AI based Technologies for International Space Station and Space Data," 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART), Moradabad, India, 2022, pp. 19-25, doi: 10.1109/SMART55829.2022.10046956
12. Rajawat, A.S.; Goyal, S.B.; Bedi, P.; Simoff, S.; Jan, T.; Prasad, M. Smart Scalable ML-Blockchain Framework for Large-Scale Clinical Information Sharing. Appl. Sci. 2022, 12, 10795. https://doi.org/10.3390/app122110795.
13. S. Rajawat et al., "Visual Cryptography and Blockchain for Protecting Against Phishing Attacks on Electronic Voting Systems," 2022 International Conference and Exposition on Electrical And Power Engineering (EPE), Iasi, Romania, 2022, pp. 663-666, doi: 10.1109/EPE56121.2022.9959765.
14. S. Rajawat et al., "Electrical Fault Detection for Industry 4.0 using Fusion deep Learning Algorithm," 2022 International Conference and Exposition on Electrical And Power Engineering (EPE), Iasi, Romania, 2022, pp. 658-662, doi: 10.1109/EPE56121.2022.9959762.
15. H. S. Yadav and R. K. Singhal, "Classification and Prediction of Liver Disease Diagnosis Using Machine Learning Algorithms," 2023 2nd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2023, pp. 1-6, doi: 10.1109/INOCON57975.2023.10101221.
16. S. Mondal, R. Maity, Y. R. Singh, S. Ghosh and A. Nag, "Early Prediction of Coronary Heart Disease using Boosting-based Voting Ensemble Learning," 2022 IEEE Bombay Section Signature Conference (IBSSC), Mumbai, India, 2022, pp. 1-5, doi: 10.1109/IBSSC56953.2022.10037445.
17. M. J. A. Junaid and R. Kumar, "Data Science And Its Application In Heart Disease Prediction," 2020 International Conference on Intelligent Engineering and Management (ICIEM), London, UK, 2020, pp. 396-400, doi: 10.1109/ICIEM48762.2020.9160056.
18. G. Shobana and K. Umamaheswari, "Prediction of Liver Disease using Gradient Boost Machine Learning Techniques with Feature Scaling," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1223-1229, doi: 10.1109/ICCMC51019.2021.9418333.
19. M. Mahyoub, M. Randles, T. Baker and P. Yang, "Effective Use of Data Science Toward

Early Prediction of Alzheimer's Disease," 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), Exeter, UK, 2018, pp. 1455-1461, doi: 10.1109/HPCC/SmartCity/DSS.2018.00240.

20. M. Gulhane and T. Sajana, "A Machine Learning based Model for Disease Prediction," 2021 International Conference on Computing, Communication and Green Engineering (CCGE), Pune, India, 2021, pp. 1-5, doi: 10.1109/CCGE50943.2021.9776374A
21. Chanchal, A. S. Singh and K. Anandhan, "A Modern Comparison of ML Algorithms for Cardiovascular Disease Prediction," 2021 9th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2021, pp. 1-5, doi: 10.1109/ICRITO51393.2021.9596228.
22. Parab, P. Gholap and V. Patankar, "DiseaseLens: A Lifestyle related Disease Predictor," 2022 5th International Conference on Advances in Science and Technology (ICAST), Mumbai, India, 2022, pp. 383-387, doi: 10.1109/ICAST55766.2022.10039533.