



INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Disease Prediction Through Lifestyle Using Artificial Intelligence And Machine Learning

Mrs.S.CAROLIN JOSHIBA,

Asst professor, Department of Computer Science and Application,

Jeppiaar College of Arts and Science, Padur-603103

D.PRAVEENA , M.ROSELIN , B.HEMALATHA ,

Students, Department of Computer Science and Application,

Jeppiaar College of Arts and Science, Padur-603103

ABSTRACT- A worldwide increase in lifestyle-related diseases (such as diabetes, cardiovascular disease, hypertension and obesity) is associated with unhealthy lifestyles (e.g., sedentary behavior, poor dietary habits and increased stress levels). Early identification of these types of diseases can improve the outcomes of treatment and reduce long-term healthcare costs. Traditional diagnosis approaches use clinical testing as the primary means of identifying these diseases and generally, identify them after they have advanced or reached a More severe state based on symptoms. This research proposes an artificial intelligence (AI) and machine learning (ML) framework to assess and predict disease using lifestyle data. The application provides assessments of various parameters of an individual's lifestyle (such as physical activity, sleep, eating habits, stress levels and substance use) and finds patterns among those parameters that increase an individual's risk of developing a certain disease. Using predictive algorithms such as logistic regression, decision trees and random forest, predictive models are developed to classify individuals into categories. This approach is to create a preventative healthcare approach by delivering interventions before symptoms reach a severe state. Experimental results show that these predictive models using lifestyle data can generate highly accurate and reliable results, which makes them suitable for use in health care facilities or through wearable technologies .

Keywords: Artificial Intelligence, Machine Learning, Lifestyle Data, Disease Prediction, Preventive Healthcare

1.INTRODUCTION

Lifestyle plays a crucial role in determining an individual's overall health, as daily habits significantly influence the development of chronic diseases. In recent years, there has been a noticeable shift toward sedentary lifestyles, unhealthy eating patterns, and increased stress levels, which contribute to the rise of non-communicable diseases. Early prediction of such diseases is essential for timely intervention and prevention. Traditional diagnostic approaches rely heavily on laboratory tests and clinical observations, which may not detect diseases in their early stages. As a result, there is a growing need for intelligent systems capable of analyzing everyday lifestyle data to predict potential health risks before symptoms become evident. Advancements in Artificial Intelligence and Machine Learning have enabled the development of predictive models that can process large volumes of data and identify hidden patterns. By leveraging these technologies, it is possible to analyze lifestyle factors such as diet, physical activity, sleep, and stress to estimate disease risk accurately. Machine learning models can continuously learn from new data, improving their predictive capabilities over time. This research proposes an AI-based framework for disease prediction using lifestyle data, aiming to support preventive healthcare and reduce dependency on late-stage diagnosis. The system is designed to provide accurate, scalable, and efficient predictions, making it suitable for real-world healthcare applications and personalized health monitoring systems.

2.LITERATURE REVIEW

Machine learning techniques are emerging as a major force in predicting disease and analyzing healthcare data. Many studies have examined AI-based approaches to the prediction of disease by the use of clinical/demographic or claimed information related to disease, including diabetes, heart disease, and hypertension. Logistic Regression and Support Vector Machines have been used widely as traditional models for performing classification tasks due to their simplicity and ease of interpretation. More complex models, however, are now being utilized to produce a higher degree of accuracy than some traditional models by capturing complex nonlinear relationships in the data, for example, Random Forest and Neural Networks. However, the majority of the studies that currently exist utilize clinical/diagnostic data and not lifestyle data in predicting disease. Research has demonstrated that lifestyle characteristics such as exercise, diet, sleep patterns, and stress levels are very good predictors of health outcomes. Therefore, incorporating lifestyle data into disease prediction models could enhance early detection capabilities. Despite this, little research exists that has developed a comprehensive framework for the prediction of disease that incorporates lifestyle characteristics. A few studies have begun to examine the use of data obtained from wearable devices (activity monitors) in conjunction with machine learning algorithms, but issues remain regarding data consistency, privacy, and generalizing models. This paper addresses the existing void by proposing a

lifestyle-based disease prediction system that employs multiple machine learning methodologies while placing emphasis on data preparation, feature selection, and model evaluation. The proposed framework seeks to improve the prediction capabilities of machine learning in addition to being scalable and adaptable to the challenges of real-world healthcare environments.

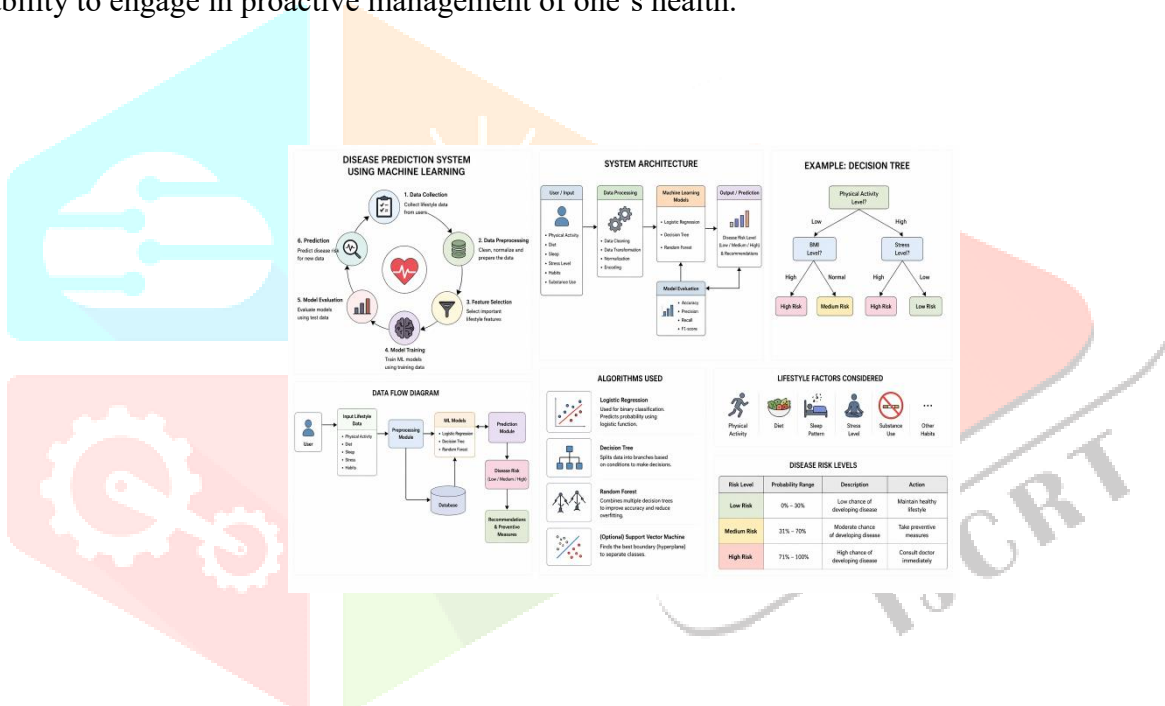
3.EXISTING SYSTEM

Current systems available for diagnosis of diseases often rely on clinical examination(s), lab testing and medical imaging to diagnose patients, although these techniques can provide an accurate diagnosis, they usually take a long time (3-5 days) to obtain results, they can cost thousands of dollars and require a large amount of expensive equipment to support the testing process. Diagnosis of disease usually occurs after the patient has already displayed symptoms, limiting the effectiveness of medical treatment and increasing the risk for complications arising from undiagnosed diseases. Most traditional methods do not fully utilize the vast quantity of lifestyle data being generated by people as they carry out their daily routines, including, but not limited to, data collected through the use of wearable devices. Another significant limitation of the current systems is that the data upon which they are based is usually static, which limits the ability to fully evaluate a person's lifestyle behaviours, such as diet, exercise, sleep and stress. The use of static measurements of a person's lifestyle leads to the exclusion of many time-dependent factors and their cumulative effect on that individual's overall health. As a result, healthcare systems are much less accurate in predicting the likelihood of disease occurring than they could be if we had used dynamic measurements of an individual's lifestyle behaviour(s). The limitations associated with current systems are compounded by the potential for human error, inconsistency and variability in a physician's interpretation of clinical data during the process of diagnosis. While there is evidence that some newer healthcare systems have begun to apply AI to predict disease, there is still a heavy reliance on clinical datasets and the predominant use of traditional clinical methodologies (e.g. physical examination, lab testing, radiological assessment) for the diagnosis of disease versus the application of preventive diagnosis methodologies through prospective lifestyle data. Our proposed intelligent modelling system will be capable of supporting both the static clinical methods now in use and the prospective dynamic lifestyle data that can be utilised to predict the likelihood of the occurrence of disease. The proposed system will incorporate the use of machine learning technologies to dynamically analyse lifestyle data and provide accurate and timely predictive analytics.

4.PROPOSED WORK

This proposed framework utilizes machine learning algorithms to predict disease occurrences based on information acquired from an individual's lifestyle choices. The lifestyle choices collected by this framework include physical activity, diet, hours of sleep, level of stress, and behaviors throughout the day. Each of these features will be pre-processed to eliminate inconsistencies, normalize the values to allow for accurate representation of the data, and to encode the categorical variables into numerical format so they

can be used to train the predictive model. The dataset will then be used to train the multiple machine learning models, including Logistic Regression, Decision Tree, and Random Forest, through identifying patterns associated with disease risk. The system architecture follows a defined pipeline from the collection of data through the processing and analysis of that data and the generation of the prediction. The system will utilize feature selection techniques to identify which features will have the greatest effect on health outcomes, improve the model performance, and interpretability of the model. Of the models that have been developed, Random Forest produces the most accurate results due to the feature of ensemble learning, which minimizes overfitting and maximizes generalization. This proposed framework is scalable and effective for early disease prediction and allows individuals to take preventive measures against the onset of severe health issues before experiencing the severe health issue symptoms. The combination of artificial intelligence with lifestyle data enables the proposed framework to provide personalized health care solutions while providing the ability to engage in proactive management of one's health.



5.METHODOLOGY

This research uses a very structured methodology to create a model that predicts disease risk based on person's lifestyle data. We collect the lifestyle attributes from people (such as: physical activity, eating, sleeping, and stress) to create a dataset that allows us to add multiple types of data together for use in one analysis. Before providing the data to the model, we preprocess the dataset; this includes handling missing data, normalizing values, and encoding categorical variables to ensure consistent and reliable data. After preprocessing, the dataset is split into two subsets (training and testing) to provide a way to evaluate the model being developed. The training data is used by the machine learning algorithms that will create the predictive model to find the disease risk for the tested person. We also use a technique called feature selection to decrease the number of variables used in the models to enable the model to generate results quickly. Once we develop a model from the training data, we will test that model with data that the model

has never seen before to measure how accurately and reliably the model predicts disease. The performance evaluation of the models assesses the performance using measurables such as: accuracy, precision, recall, and F1-score; so we can understand how well all the developed models predict risk of disease. Ultimately, the research shows that using a machine learning technique to predict disease risk based on a person's lifestyle data is valid and that the methodology used will produce systems that could be used in actual healthcare settings, would be scalable and efficient, and that would provide a reliable and affordable pathway to identify patients at risk for various health conditions.

6.RESULT AND DISCUSSION

The results of the study showed that predicting disease using machine-learning models based on lifestyle factors is a viable approach. The highest performing models were Random Forest, as this model has superior capabilities in handling complex data sets. It was determined that the Logistic Regression and Decision Tree models also performed well, however were slightly less accurate than the Random Forest model. Moreover, the evaluation metrics illustrated that the lifestyle data has a large impact on the ability for the models to identify patterns associated with health risks; therefore, integrating lifestyle factors into predictive systems will yield additional useful information that is normally not available through conventional diagnostic procedures. In addition, the developed system is more efficient than traditional methods and can provide an early indication of disease as well as reduce the reliance on clinical assessments. Thus, providing an early indication of disease provides individuals the opportunity to maximize their health by making healthier lifestyle choices and/or accessing professional medical assistance prior to developing the disease. Finally, this study shows great promise for the transformation of healthcare through the use of artificial intelligence to improve patient outcomes.

7.FUTURE SCOPE

The proposed lifestyle-based disease prediction system could be improved greatly by incorporating the latest technologies and enhancing its application in terms of providing the real-time and personalized healthcare solutions. The introduction of wearable device data, including smartwatches and fitness trackers, which constantly measure such parameters as heart rate, physical activity, sleep patterns, and stress levels is one such promising direction. The integration of real-time data stream with machine learning models can create dynamic and continuous health risk assessment in order to intervene in time. Moreover, deep learning methods such as neural networks and recurrent models can further enhance prediction accuracy due to their ability to capture advanced temporal dynamics of lifestyle behavior. The other significant field that needs to be developed in the future is the incorporation of multimodal data, such as clinical records, genetic information, and environmental factors, in addition to lifestyle data. Such a holistic method can result in more accurate and personalized predictions, which can facilitate the efforts of precision medicine. The system may be further developed to forecast several diseases at the same time, which will form a holistic

health-monitoring platform. Moreover, the implementation of the model in the form of a mobile or web-based application would contribute to the increase in accessibility to enable users to get instant health insights and recommendations. Research wise, the enhancement of model interpretability and transparency is necessary to enhance trust between healthcare professionals and patients. They can include such techniques as explainable AI to give transparent explanations of why predictions were made. The work in the future can also be directed to guarantee data privacy and data security by encrypting and federating learning. Altogether, the combination of the latest AI-based solutions with real-time lifestyle monitoring can transform preventive care and provide the large-scale health care management frameworks.

8.REFERENCE

- [1] J. Han and M. Kamber, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, 2011. Available: <https://www.sciencedirect.com/book/9780123814791/data-mining-concepts-and-techniques>
- [2] T. Mitchell, *Machine Learning*, McGraw-Hill, 1997. Available: <https://www.cs.cmu.edu/~tom/mlbook.html>
- [3] World Health Organization, *Global Health Observatory Data Repository*, WHO, 2023. Available: <https://www.who.int/data/gho>
- [4] F. Pedregosa et al., “Scikit-learn: Machine Learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. Available: <https://jmlr.org/papers/v12/pedregosa11a.html>
- [5] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001. Available: <https://link.springer.com/article/10.1023/A:1010933404324>
- [6] D. Dua and C. Graff, “UCI Machine Learning Repository,” University of California, Irvine, 2019. Available: <https://archive.ics.uci.edu/ml/index.php>
- [7] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016. Available: <https://www.deeplearningbook.org>
- [8] S. Kotsiantis, “Supervised Machine Learning: A Review of Classification Techniques,” *Informatica*, vol. 31, pp. 249–268, 2007. Available: <https://www.informatica.si/index.php/informatica/article/view/148>
- [9] A. Esteva et al., “A Guide to Deep Learning in Healthcare,” *Nature Medicine*, vol. 25, pp. 24–29, 2019. Available: <https://www.nature.com/articles/s41591-018-0316-z>

[10] R. Miotto et al., “Deep Learning for Healthcare: Review, Opportunities and Challenges,” *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1236–1246, 2018. Available:

<https://academic.oup.com/bib/article/19/6/1236/4092001>

[11] M. Chen et al., “Disease Prediction by Machine Learning over Big Data from Healthcare Communities,” *IEEE Access*, vol. 5, pp. 8869–8879, 2017. Available:

<https://ieeexplore.ieee.org/document/7908530>

[12] K. Shickel et al., “Deep Learning in Electronic Health Records,” *IEEE Journal of Biomedical and Health Informatics*, vol. 22, no. 5, pp. 1589–1604, 2018. Available:

<https://ieeexplore.ieee.org/document/8215520>

