An Improved Machine Learning-driven Patient's Sickness or Health Status Prediction System

A Project Work Synopsis

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE WITH SPECIALIZATION IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Submitted by:

Shreya Jadon 20BCS6769

Kunal 20BCS6278

Hitesh Kumar 20BCS6157

Under the Supervision of:

Siddharth Kumar (E12853)

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413, PUNJAB

March, 2023

Abstract

In the healthcare industry, the ability to accurately forecast a patient's health status is critical. With the growing availability of patient data and the advancement of machine learning techniques, there is a growing interest in using machine learning algorithm to predict patient outcomes. We propose an improved machine learning-driven patient sickness or health status prediction system in this project.

The ability to predict a patient's health status properly is essential in the healthcare sector. The use of machine learning algorithms to forecast patient outcomes is becoming more popular as a result of the expansion of patient data availability and the development of machine learning methodologies. In this study, we provide an enhanced machine learning-driven method for predicting patient illness or health condition.

The results show that our system beats existing methods in terms of accuracy and efficiency. The proposed system is tested on a sizable dataset of patient records. The system is made to be scalable and adaptable to various healthcare settings, making it an important tool for healthcare providers to use in patient outcome prediction.

In conclusion, our suggested machine learning-driven system for predicting a patient's illness or health state is a potential strategy for enhancing patient outcomes in the healthcare sector. We can give medical personnel insightful information about patient health by utilising machine learning and data analytics, empowering them to make wiser decisions and deliver more efficient care.

Table of Contents

Abstract	ii
1. Introduction	1
1.1 Problem Definition	1
1.2 Project Overview	2
1.3 Hardware Specification	3
1.4 Software Specification	3
2. Literature Survey	5
2.1 Existing System	5
2.2 Proposed System	7
2.3 Literature Review Summary	9
3. Problem Formulation	10
4. Research Objective	12
5. Methodologies	13
6. Experimental Setup	14
7. Conclusion	15
8. Reference	16

1. INTRODUCTION

1.1 Problem Definition

Accurately forecasting patients' health state is a serious difficulty for the healthcare business. Accurately forecasting patient outcomes can be difficult because to the complexity of human physiology, the wide range of medical diseases, and the unpredictability of patient data. Despite advances in medical science and technology, predicting patient outcomes remains a difficult task.

The dependence on basic statistical models for predicting patient outcomes is one of the key causes behind this problem. These models may be insufficient to capture the intricate interactions between patient variables and health outcomes. Given the rising quantity of data collected in healthcare, more advanced machine learning models that can capture the complexities of patient data are required.

Another problem in forecasting patient outcomes is developing an efficient feature engineering approach. Feature engineering is the process of choosing and altering key characteristics from a dataset to improve the performance of our machine learning models. Finding the correct collection of traits that are most relevant to the work at hand can be a time-consuming procedure. The development of automated feature selection and engineering approaches can relieve this problem and boost prediction efficiency.

Notwithstanding these obstacles, the proposed approach has the potential to greatly enhance patient outcomes in the healthcare business by providing medical personnel with accurate and timely forecasts of patient health

status. We can solve the issues of forecasting patient outcomes and give useful insights into patient health by utilising the capabilities of machine learning and data analytics. We can enhance patient outcomes and have a beneficial influence on the healthcare business by using the correct combination of powerful machine learning algorithms and efficient feature engineering approaches.

1.2 Problem Overview

We may go over the components and possible benefits of the suggested machine learning-driven patient illness or health status prediction system in further depth.

To begin, the system would need a big dataset of patient health records, which would include medical history, laboratory test results, vital signs, and demographic information. To guarantee accuracy and completeness, this data would need to be properly selected and cleansed. In healthcare, data availability is critical, yet getting and integrating data from many sources may be difficult.

After the data is gathered, the system will employ machine learning algorithms to discover patterns and links between patient characteristics and health outcomes. At this stage, the model selection procedure is critical since various models may perform better on different types of data. To guarantee generalizability and reduce overfitting, the chosen model would be trained and evaluated using cross-validation procedures.

Feature engineering is also an important component of system development. It entails picking important characteristics and translating them into a format that machine learning algorithms can understand. Choosing features may be difficult since it takes domain expertise and an

awareness of which characteristics are most relevant to the situation at hand.

There are various advantages to such a system. It might offer early warning indicators of future health concerns to medical practitioners, allowing for early intervention and preventative actions. It might also help in triaging patients, prioritising those who are most vulnerable to negative consequences. Furthermore, by streamlining treatment regimens and shortening hospital stays, the technology might help lower healthcare expenses.

Finally, the proposed machine learning-driven patient illness or health status prediction system has the potential to transform healthcare by offering useful insights into patient health and empowering medical practitioners to make better decisions. To achieve accuracy and utility, however, great consideration must be given to data availability, model selection, and feature engineering.

1.3 Hardware Specification

- Processor – 64-bit eight-core,
- 2.5GHz per core
- RAM – Minimum 4GB required
- Hard Disk – SSD or HDD minimum 40GB free space required
- Webcam – HD Webcam supporting 720p video recording

1.4 Software Specification

- Edition - Windows 10 Home Single Language
- Python installed – version 3.7 to 3.10
- Steamlit.
- Spyder
- Numpy
- Pandas
- Matplotlib

2. LITERATURE SURVEY

2.1 Existing System

There are numerous options that may be considered to enhance the quantity of material and overcome the limits of classic statistical methods for forecasting patient sickness or health status:

- Deep Learning Models: Deep learning models like convolutional neural networks (CNNs) and recurrent neural networks (RNNs) have demonstrated promising outcomes in a variety of medical applications. These models may learn complicated patterns and correlations from large-scale datasets without requiring considerable feature engineering. Moreover, they can be trained on massive volumes of data and handle high-dimensional input, making them ideal for real-world settings.
- Ensemble Methods: Ensemble methods such as random forests and gradient boosting can help classic statistical models increase their accuracy and dependability. These approaches produce predictions by combining numerous models, lowering the danger of overfitting and boosting generalisation performance.
- Transfer Learning is training a model on a big dataset and then fine-tuning it on a smaller dataset for a specific purpose. This method can save time and money by using information from a huge dataset and adapting it to a new issue area.
- Methods for Feature Selection: The most significant characteristics for forecasting patient health status can be identified using feature selection approaches. This method can lower the complexity of the dataset while improving the performance of older methods.

Overall, these approaches can be used to improve the accuracy and reliability of patient health prediction systems and overcome the limitations of traditional statistical methods.

Proper illness diagnosis is critical to the successful treatment and care of patients all over the world. Unfortunately, because of the complexities of disease causes and symptoms, early identification and successful therapy are frequently difficult. By offering reliable diagnosis tools, machine learning (ML), a branch of artificial intelligence (AI), has the potential to assist address these obstacles. This article

provides an overview of how machine learning is being used to help with illness identification.

The paper is founded on relevant research and is subjected to a bibliometric analysis utilising data from the Scopus and Web of Science (WOS) databases. This study highlights the most productive authors, countries, organisations, and papers in the field of machine-learning-based illness diagnosis (MLBDD).

The paper then analyses the most recent developments and techniques in MLBDD, taking algorithm, ailment categories, data type, application, and assessment measures into account. The study focuses on how various ML algorithms, such as supervised, unsupervised, and deep learning, are used to diagnose diseases in a variety of categories, including cancer, neurological disorders, cardiovascular diseases, and infectious diseases.

Additionally, the study delves into the usage of several data types in MLBDD, including as imaging, genomics, and electronic health records (EHRs), as well as the problems connected with data quality, interoperability, and privacy. Furthermore, the paper includes many assessment measures for measuring the effectiveness of machine learning models for illness diagnosis, such as accuracy, sensitivity, specificity, and area under the curve (AUC).

Finally, the article summarises major findings and highlights future trends and prospects in the MLBDD field, such as integrating ML models into clinical workflows, developing explainable and transparent ML models, and emphasising the importance of ethical considerations in the use of ML for disease diagnosis.

Finally, this paper gives an in-depth look at the present level of machine-learning-based disease diagnosis and its potential to enhance early identification and treatment of a variety of ailments.

Datasets from the medical and healthcare industries are essential for medical research, diagnosis, and treatment. Given the growing volume of medical data, an effective and accurate system for organising and analysing these datasets is required. For dealing with medical or healthcare datasets, several programmes are available, including electronic health record (EHR) systems, clinical decision support systems (CDSS), and illness prediction systems.

The accuracy of illness prognosis is one of the most difficult difficulties in disease prediction. Several illness prediction systems use machine learning techniques, such as the Optimized Naive Bayes classifier, to overcome this difficulty. To simulate the system, the TensorFlow simulator is usually utilised, and the input dataset is used to train the deep belief network. The deep belief network is built, and the accuracy

index is computed using it. This accuracy index indicates how effectively the system anticipates sickness.

Our proposed approach was tested and found to have a good accuracy index when compared to current techniques. We recognise, however, that illness classification is not a one-size-fits-all approach, and that several classifiers may be required to properly anticipate certain disease categories. As a result, we want to provide several classifiers in the future to precisely estimate particular illness categorization.

Finally, for medical research, diagnosis, and treatment, an efficient and accurate system for handling medical and healthcare datasets is critical. Machine learning algorithms used in illness prediction systems are an excellent technique to improving disease prognosis accuracy.

According to the study's findings, AI techniques in the healthcare system, particularly for disease diagnosis, are critical. The current study has been divided into several sections that cover the diagnosis of Alzheimer's, cancer, diabetes, chronic diseases, heart disease, stroke and cerebrovascular disease, hypertension, skin disease, and liver disease, with the goal of illuminating how machine and deep learning techniques work in various disease diagnosis areas.(reference from reasearh paper “

2.2 Proposed System

We can develop on the provided deep learning feedbackward model and its possible applications in the healthcare business to boost the content size of the given passage.

Deep learning feedbackward models have received a lot of interest in recent years because of their potential to increase the accuracy of machine learning systems over time. This is accomplished through the use of feedback loops to update the system's parameters and modify its predictions in response to fresh input.

The suggested approach might have numerous uses in the field of healthcare. It may, for example, be used to estimate the chance of a patient contracting a specific disease based on their medical history and demographic data. This might allow for earlier diagnosis and intervention, potentially saving lives and lowering healthcare expenditures.

Another possible application is in the realm of customised medicine. The technology might decide which medicines are most beneficial for a certain patient by assessing patient data, taking into account aspects such as genetic profile, medical history, and

lifestyle. This might result in more tailored and effective treatment strategies, better patient outcomes, and lower healthcare expenditures.

Building on the suggested approach, we may investigate the benefits it could provide to both healthcare facilities and patients.

One significant advantage of the system is its capacity to manage large-scale datasets. This implies it may be used in a variety of healthcare contexts, ranging from individual patient diagnosis to community health management. The system might spot patterns and trends that people would find difficult or impossible to detect by processing and analysing massive volumes of data, allowing for more accurate forecasts and better-informed judgements.

The system's data visualisation features would also be a great resource for healthcare practitioners. Physicians might easily identify patients who require immediate treatment or additional inquiry by providing patient data in a simple and easy-to-use interface. This might enhance patient outcomes and minimise healthcare workers' workloads, allowing them to focus on providing better care to patients.

The suggested method also has the advantage of employing machine learning techniques such as deep neural networks and support vector machines. These strategies are well-known in the field of artificial intelligence and have been shown to be useful in a variety of applications, including healthcare. The suggested system might create accurate and trustworthy forecasts by utilising these methodologies, allowing healthcare practitioners to make better informed judgements.

Apart from the potential to change the way healthcare professionals approach patient care, the suggested system may also benefit patients. Patients may become more engaged in their own care and make better educated health decisions if they were given real-time information about their health state and prospective treatment alternatives. This might result in better results and a higher quality of life.

Overall, the proposed system has the potential to transform healthcare by providing a powerful tool for evaluating patient data, making accurate predictions, and giving real-time information to both healthcare professionals and patients. As the system develops and refines its algorithms, it has the potential to become a vital resource for improving patient outcomes and providing personalised treatment.

2.3 Literature Review Summary

Year	Author	Tools/ Software	Technique	Source	Evaluation Parameter
2022	Md Manjurul Ahsan	ML	SVM, Random forest	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8950225/	Accuracy, *Precision*
2022	Yogesh Kumar, Apeksha Koul	AI,ML	SVM, KNN	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8754556/	Accuracy, *Precision*
2021	Naresh Kumar, Nripendra Narayan Das	Jupyter Notebook	supervised learning	https://www.hindawi.com/journals/jhe/2021/9983652/	Accuracy, *Precision*

3. PROBLEM FORMULATION

The suggested deep learning feedbackward paradigm provides a viable remedy to these restrictions. The system may raise its accuracy over time and adapt to fresh patient data by employing feedback loops to adjust the system's settings and modify its predictions based on new data. This might allow doctors to make more accurate diagnoses and provide more effective treatments, eventually improving patient outcomes and lowering healthcare costs.

Furthermore, the utilisation of data visualisation tools in the suggested system should promote more efficient and effective communication among healthcare professionals, patients, and carers. The technology might help healthcare practitioners to swiftly identify individuals that require immediate treatment or additional inquiry by providing easy-to-use interfaces that deliver real-time information. Patients may become more involved in their own treatment as a result, better knowing their health state and the alternatives accessible to them.

The suggested system's capacity to manage large-scale datasets is another potential advantage. The capacity to collect and evaluate massive volumes of data is becoming increasingly vital as electronic health records and other digital health data become more widely available. The suggested method might be used in a variety of healthcare settings, starting with individual patients.

It is also worth noting that the suggested system makes use of machine learning techniques such as deep neural networks and support vector machines. These strategies have been shown to be useful in a variety of areas, including healthcare, and are well-suited to dealing with large datasets. The suggested system might create accurate and trustworthy predictions by utilising these methodologies, allowing healthcare practitioners to make better informed decisions about patient treatment.

Therefore, the suggested deep learning feedbackward model provides a potential approach for forecasting patients' illnesses or health condition in real time. The system might enhance patient outcomes and lower healthcare costs by using the most recent developments in machine learning and data visualisation, thereby benefiting healthcare institutions, healthcare professionals, and patients alike.

Besides from enhancing patient outcomes, the suggested approach offers a number of additional potential advantages. It might, for example, assist healthcare facilities in optimising their resources and lowering healthcare costs by enabling more effective

utilisation of healthcare services. The method might assist healthcare personnel prioritise treatment and prevent wasteful interventions by identifying individuals who require immediate attention or additional inquiry.

The suggested method might also help healthcare organisations uncover trends and patterns in patient data that could guide public health actions, allowing for more accurate population health management. The system may, for example, be used to detect risk factors for certain diseases or disorders, or to follow the development of infectious diseases in real time.

Furthermore, by utilising machine learning techniques such as deep neural networks and support vector machines, the suggested system may be able to find previously unexplored correlations between patient data and health outcomes. This might lead to new insights into disease pathophysiology as well as possible treatment targets.

The suggested system must be rigorously built, tested, and validated using rigorous scientific procedures to verify its effectiveness and reliability. To maintain patient data security, the system will also need to be connected with existing healthcare systems and comply with appropriate privacy and security requirements.

Finally, the suggested deep learning feedbackward model has the potential to change the way healthcare practitioners approach patient care by providing real-time information that can help them make more accurate diagnoses and implement tailored treatment regimens. The system might improve patient outcomes, lower healthcare costs, and open the way for new insights into disease causation and therapeutic intervention by utilising the latest breakthroughs in machine learning and data visualisation.

4. OBJECTIVES

A properly built system that employs advanced machine learning algorithms and data visualisation tools to enable healthcare workers to make educated decisions regarding patient care will be required to achieve these goals. To manage large-scale information and evolving healthcare demands, the system must be scalable, versatile, and adaptive.

To meet the first goal, the system must be built to capture and track patient data over time. This might entail creating a user-friendly interface for patients to enter their data and interacting with existing healthcare systems to acquire pertinent medical information.

To meet the second goal, the system must be able to monitor patient data in real-time and warn healthcare personnel if any deviations or anomalies are detected. This might entail creating algorithms to detect patterns and trends in patient data and alerting healthcare practitioners when specific thresholds are breached.

To reach the third goal, the system must be capable of predicting patient ailments or health status based on their data. This might entail creating powerful machine learning models that can learn from patient data and generate accurate real-time predictions. The system must also be capable of making individualised suggestions for patient treatment based on anticipated diseases or health condition.

Overall, the suggested system has the potential to change the way healthcare professionals approach patient care by providing real-time data that can help them make more accurate diagnoses and implement personalised treatment regimens. The system might improve patient outcomes, lower healthcare costs, and open the way for new insights into disease causation and therapeutic intervention by utilising the latest breakthroughs in machine learning and data visualisation.

5. METHODOLOGY

To generate accurate and reliable predictions, the proposed system will employ a combination of feature engineering and deep learning algorithms. The system will be built to handle large datasets and will employ a variety of machine learning techniques such as deep neural networks and support vector machines. To ensure the accuracy and reliability of the predictions, the system will also employ various data preprocessing and visualisation techniques.

The proposed system will utilize a deep learning feedbackward deep learning model to predict the sickness or health status of patients in real-time. The system will be trained on historical medical records of patients, including symptoms, diagnosis, and treatment outcomes. The data will be preprocessed and transformed into a format that can be inputted into the deep learning model.

The deep learning feedbackward model will be trained using a large dataset of historical patient data. The model will be trained to predict a patient's sickness or health status based on their medical history, which includes symptoms, diagnosis, and treatment outcomes.

The deep learning feedbackward model will enable the system to learn from its predictions and improve its accuracy over time. As new data becomes available, the model will continue to refine its algorithms and provide more accurate predictions. This approach will allow the system to adapt to new patient data and improve its accuracy in real-time.

The system will be designed with data visualization tools to provide healthcare professionals with an intuitive interface that displays patient data and predictions. The interface will enable clinicians to quickly identify patients who require immediate attention and make informed decisions about patient care.

The proposed methodology for this system has the potential to revolutionize the way healthcare professionals approach patient care, providing them with real-time insights that can help them make more accurate diagnoses and deliver personalized treatment plans. The use of deep learning feedbackward deep learning models will enable the system to learn from its predictions and improve its accuracy over time, leading to better patient outcomes and reduced healthcare costs

6. EXPERIMENTAL SETUP

The suggested system would use feature engineering and feature learning approaches to improve the system's accuracy and dependability even more. Feature engineering entails identifying and manipulating the most important characteristics in the dataset, whereas feature learning entails utilising deep learning algorithms to discover significant features from raw data automatically. By identifying the most relevant elements that impact patient health status, this combination of methodologies will allow the system to produce more accurate forecasts.

The suggested system would leverage data visualisation tools such as Matplotlib and Seaborn, in addition to machine learning techniques, to give healthcare practitioners with an easy-to-use interface for studying patient data and making educated choices. These visualisation tools will allow healthcare practitioners to easily spot patterns and trends in patient data, allowing them to influence treatment decisions and enhance patient outcomes.

The system will be built to run on cloud-based platforms such as Amazon Web Services or Microsoft Azure to ensure that it can manage large-scale datasets. This will allow the system to scale up or down in response to changing computational demands, as well as ensuring that it can manage the massive volumes of data created by current healthcare systems.

Overall, the proposed approach has the potential to enhance patient outcomes dramatically by giving healthcare practitioners with real-time insights into patient health status and allowing them to make better educated treatment decisions. The technology might alter the way healthcare professionals approach patient care by utilising the latest breakthroughs in machine learning and data visualisation, paving the path for fresh insights into disease prevention and treatment.

7. CONCLUSION

Furthermore, the system will provide a user-friendly interface for healthcare practitioners, making it easier for them to obtain patient data and make educated patient care decisions. Data visualisation capabilities will be included in the interface, allowing healthcare practitioners to discover trends and patterns in patient data and create individualised treatment plans based on this information. Early diagnosis of health problems results in prompt treatments, which can avoid the advancement of chronic diseases, resulting in improved patient outcomes and lower healthcare expenditures.

Additionally, the suggested system may be connected into current healthcare systems, giving healthcare providers remote and real-time access to patient data. This can result in greater care coordination and better patient outcomes, especially for patients with complicated medical disorders that need continual monitoring and treatment.

Ultimately, the suggested system has the potential to completely change how healthcare professionals approach patient care. The system can enhance the quality of care, save healthcare costs, and ultimately save lives by delivering precise and reliable forecasts of patient health status.

8. REFERENCES

[1] Improving Disease Prediction by Machine Learning by Dinesh Bhagwan Hanchate

[2] Machine-Learning-Based Disease Diagnosis: A Comprehensive Review Md Manjurul Ahsan, Shahana Akter Luna, and Zahed Siddique

[3] Intelligent health risk prediction systems using machine learning: A review Santosh Shinde

[4] Intelligent Health Risk and Disease Prediction Using Optimized Naive Bayes Classifier Latifah Alamer

[5] McPhee S.J., Papadakis M.A., Rabow M.W., editors. *Current Medical Diagnosis & Treatment.* McGraw-Hill Medical; New York, NY, USA: 2010

[6] Ahsan M.M., Ahad M.T., Soma F.A., Paul S., Chowdhury A., Luna S.A., Yazdan M.M.S., Rahman A., Siddique Z., Huebner P. Detecting SARS-CoV-2 From Chest X-ray Using Artificial Intelligence. *IEEE Access.* 2021;9:35501–35513. doi: 10.1109/ACCESS.2021.3061621.

[7] Coon E.R., Quinonez R.A., Moyer V.A., Schroeder A.R. Overdiagnosis: How our compulsion for diagnosis may be harming children. *Pediatrics.* 2014;134:1013–1023. doi: 10.1542/peds.2014-1778.

[8] Balogh E.P., Miller B.T., Ball J.R. *Improving Diagnosis in Health Care.* National Academic Press; Washington, DC, USA: 20

[9] Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda by Yogesh Kumar , Apesksha Koul , Ruchi