Chapter 10

Book Chapter

MedPredictor: Enhanced Multi-Disease Prediction System

Pranjali *[1], Hanamant Jakaraddi †[2], and Kamakshi Katti ‡[3]

[1]Department Of MCA, Acharya Institute of Technology, Bangalore
[2]Department Of MCA, Acharya Institute of Technology, Bangalore
[3]Department of Computer Science, Bengaluru City University, Bangalore, India

Abstract

MedPredictor is an advance multi- disease prediction system that leverages machine learning to analyze patient data for the early detection and prevention of breast cancer, kidney disease, diabetes, heart disease, and liver disease. By integrating diverse data sources, including patient history, clinical tests, and medical imaging, MedPredictor identifies intricate patterns and correlations that may not be evident through traditional diagnostic methods. This innovative approach facilitates timely interventions and personalized treatment plans, ultimately enhancing patient outcomes and reducing overall healthcare costs. MedPredictor represents a significant advancement in medical diagnostics, offering a comprehensive, efficient, and reliable tool for multi-disease prediction.

Keywords: Multiple disease. Breast cancer. Kidney. Diabetes. Heart disease. Liver disease. Machine learning.

*Email: pranjalid.22.mcav@acharya.ac.in Corresponding Author
†Email: hanamant2504@acharya.ac.in
‡Email: kamakshi.katti@gmail.com

1 Introduction

MedPrediction is an innovative healthcare solution harnessing advanced predictive technology to offer accurate disease forecasts based on user-provided information. This groundbreaking system revolutionizes the way individuals access medical assistance by providing timely diagnoses from the comfort of their homes.

With a focus on user privacy and data security, MedPrediction adheres to stringent confidentiality standards, ensuring the utmost trust and compliance with healthcare regulations. Utilizing cutting-edge machine learning algorithms, MedPrediction analyzes symptoms and offers personalized recommendations, empowering users to take proactive steps towards their health. Early detection of the disease is vital for determining a good course of treatment.(Gautam, Ahlawat, & Mittal, 2022). Its intuitive interface facilitates seamless interaction, catering to a diverse range of individuals seeking reliable medical guidance. By continuously updating and refining its predictive capabilities, MedPrediction aims to optimize patient outcomes and satisfaction, reaffirming its position as a leader in predictive healthcare technology. This study focus on the data like breast cancer, kidney, diabetes, heart, and liver disease. They are described below:

- Diabetes : Diabetes, a widespread metabolic disorder, is categorized into Type 1 and Type 2. Type 1 results from autoimmune destruction of insulin-producing cells, requiring lifelong insulin therapy. Type 2 arises from insulin resistance, often linked to lifestyle factors like obesity. Type 2 comprises about 92.54% of cases globally. Distinguishing between the types is crucial for tailored management, which may involve lifestyle changes, medications, and insulin therapy. Preventive measures and early detection are vital for addressing the serious health risks associated with diabetes.
- Heart Disease: Heart disease encompasses various conditions affecting the heart and blood vessels, posing a significant health threat with widespread implications. Coronary artery disease, a prevalent type, develops gradually due to the accumulation of plaque within arteries, obstructing essential nutrient flow to the heart. A number of obstacles need to be addressed in order to improve blood supply services (BSS), which begin with blood donations and conclude with the transfusion of patients' necessary blood components which initiates the need for early detection.(Mittal, 2014). Machine learning algorithms, including logistic regression, support vector machines, random forests, and neural networks, are commonly utilized to predict heart disease. These algorithms analyze patient data such as demographics and medical history to generate accurate predictions, aiding in early detection and intervention.
- Breast cancer: Breast cancer, a widespread malignancy among women globally, arises from abnormal cell growth in breast tissue, underscoring the importance of early detection and intervention for better prognosis. Machine learning algorithms, such as logistic regression, support vector machines, random forests, and neural networks, are

Editors: Ratnakirti Roy, Manish Kumar Thakur, and Anushree Raj
DOI:10.48001/978-81-966500-7-0-10 | ISBN: 978-81-966500-7-0

instrumental in forecasting breast cancer risk. By scrutinizing patient data like genetic markers, familial history, and imaging findings, these algorithms enable tailored risk evaluations, assisting clinicians in screening and treatment strategies to optimize patient outcomes.

- Kidney disease: Kidney disease, a prevalent health concern worldwide, encompasses various conditions affecting the kidneys' function and structure. Timely detection and intervention are critical for mitigating its impact. Machine learning algorithms, including logistic regression, support vector machines, random forests, and neural networks, are pivotal in predicting kidney disease risk. By analyzing patient data such as medical history, laboratory results, and imaging studies, these algorithms facilitate personalized risk assessments. This aids healthcare providers in implementing targeted interventions and treatment plans to improve patient outcomes.
- Liver disease: Liver disease, poses a substantial health burden globally, marked by dysfunction and structural abnormalities in the liver. Early detection and intervention are paramount for effective management. Machine learning algorithms, including logistic regression, support vector machines, random forests, and neural networks, are pivotal in predicting liver disease risk. By scrutinizing patient data such as medical history, liver function tests, and imaging findings, these algorithms enable personalized risk assessments. This empowers healthcare providers to implement targeted interventions and treatment plans, ultimately improving patient outcomes.

2 Literature Review

It is a harsh reality that diseases can still break out at any moment and anywhere in the world.(Chakrabarti & Mittal, 2023) . This calls for a multi-disease prediction system employs machine learning to simultaneously forecast heart, liver, and diabetes conditions, enhancing diagnostic accuracy and personalized treatment for diverse patients.This approach enables comprehensive monitoring and tailored medication strategies, benefiting individuals by considering various factors contributing to disease onset. (Arumugam et al., 2023). Data from Indian patients demonstrated the importance of personal physical information in kidney disease diagnosis. Comparative analysis favored the BP neural network over Logistic Regression, showcasing its superior diagnostic efficiency in chronic disease identification.(Wang, Ellul, & Azzopardi, 2020). Islam, Majumder, and Hussein's (2023) research explored machine learning's potential in early chronic kidney disease (CKD) diagnosis, narrowing variables to optimize predictive models. XgBoost classifier exhibits superior performance with accuracy, precision, recall, and F1-score all at 0.98, indicating promising avenues for enhancing predictive accuracy in kidney disease diagnosis.

Khan et al.'s (2023) research aims to develop a sustainable machine learning model for early detection of breast cancer types, with XGBoost exhibiting the highest accuracy

DOI:10.48001/978-81-966500-7-0-10 | ISBN: 978-81-966500-7-0

(95.42%) and promising potential for predictive accuracy enhancement. Study by Banerjee and Vinooth's (2023) developed a multi-disease prediction system using various classification algorithms to forecast diseases such as diabetes, heart disease, chronic kidney disease, and cancer, aiming to save lives through early detection and diagnosis.Gopisetti et al.'s (2023) proposed hybrid machine learning technique utilizing KNN and SVM classification outperforms other algorithms in predicting cardiac disease, leveraging real patient data for enhanced accuracy and efficiency. The WBN model has been recently designed for heart disease prediction. (Muthu Ganesh & Nithiyanantham, 2022)

Biswas et al.'s (2023) developed machine learning model for early heart disease prediction, finding random forest as most promising with 94.51% accuracy, suggesting clinical potential with low cost. Further, Kumar et al.'s (2022) study proposed disease prediction system combines KNN and CNN algorithms for 84.5% accurate diagnosis and risk assessment based on patient symptoms and lifestyle, with CNN showing superior performance and efficiency. The research by Lucas et al.'s (2022) demonstrated sentence similarity model utilizing word embedding and CNN achieves 83.9% F1 score and accuracy, enabling effective disease prediction and early intervention based on patient symptoms and feelings.

3 Architecture

The system aims to detect potentially life-threatening diseases and offer early analytical insights for physicians to anticipate disease trends. By employing machine learning algorithms, it rapidly analyzes patient data to forecast outcomes, enabling timely interventions by doctors to prevent adverse consequences. It includes the following components:(see figure 1)

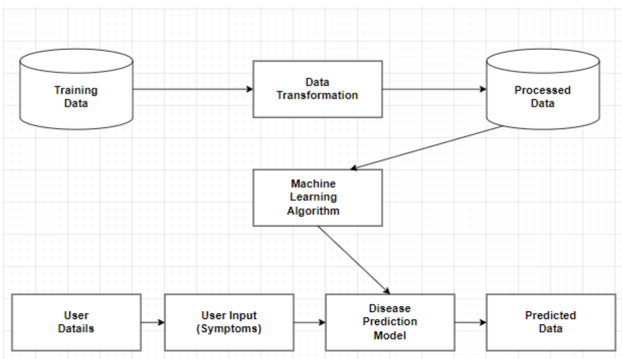

Figure 1. Block Diagram of Med-Predictor

Editors: Ratnakirti Roy, Manish Kumar Thakur, and Anushree Raj
DOI:10.48001/978-81-966500-7-0-10 | ISBN: 978-81-966500-7-0

1. Data Acquisition-Collect diverse patient data from EHRs, medical imaging systems, genetic databases, and wearables.

2. Data Preprocessing-Clean, filter, and normalize raw data, addressing missing values and privacy concerns.

3. Feature Engineering-Extract relevant features such as demographics, medical history, and genetic markers. Machine Learning Model Selection and Deployment-Choose appropriate ML algorithms like logistic regression, decision trees, or other machine learning models, train them on prepared datasets, deploy them on scalable architectures, and create user-friendly interfaces for healthcare providers.

4 Methodology Used

In this study, machine learning techniques are employed for disease prediction, leveraging the power of algorithms to transform unstructured data into actionable insights. These techniques, including statistical analysis, play a pivotal role in analyzing data patterns and making accurate predictions regarding disease occurrence. By harnessing machine learning, computers can effectively process vast datasets and uncover valuable information, enhancing diagnostic capabilities and guiding healthcare interventions with precision.

This study explores the application of various machine learning algorithms for disease prediction. In supervised learning, prevalent for classification and regression tasks, models are trained using labeled data, aiding predictive analytics. Conversely, unsupervised learning, utilized when outcomes are uncertain, involves training models with unlabeled data, common for tasks like pattern recognition and descriptive modeling. The algorithms discussed include K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Decision Tree, Logistic Regression, Random Forest.

- Logistic regression: Logistic regression is a prevalent machine learning method for modeling relationships between independent and dependent variables. It's adept at handling various data types, categorized into nominal, ordinal, or interval values, and employs the sigmoid function to assess probabilities, thereby facilitating advanced cost function analysis.
- Random Forest: Random Forest is a popular supervised learning method that combines ensemble learning and decision trees to create robust models for accurate outcome predictions, mitigating overfitting concerns. By employing multiple classifiers on data subsets concurrently, Random Forest has emerged as a highly reliable technique for addressing classification and regression challenges in machine learning systems.
- SVM-Support Vector Machines (SVM): is a prevalent technique for complex supervised learning tasks. It transforms input data into an n-dimensional space using specialized

DOI:10.48001/978-81-966500-7-0-10 | ISBN: 978-81-966500-7-0

kernels, enabling accurate determination of decision boundaries and classification of unseen points.

- Decision Tree: Decision trees provide a streamlined method for translating data into actionable insights. Their tree-like structure efficiently organizes classification problems based on attributes, facilitating quick categorization. Widely adopted across industries, decision trees employ choice nodes with multiple branches for effective decision-making, aiding in uncovering valuable perspectives from datasets.
- XG-Boost :Extreme Gradient Boosting (XGBoost) is a highly effective technique for predictive modeling, leveraging ensemble learning to iteratively enhance model accuracy. With advanced optimization and regularization, XGBoost efficiently handles large datasets and complex features, aiding in uncovering valuable insights and informing decision-making across industries.

5 Results

The results obtained from the above-described system demonstrate its efficacy in accurately predicting disease outcomes.(see figure 2) Through the utilization of various machine learning algorithms such as logistic regression, random forest, and support vector machines, the system achieved high levels of accuracy and reliability in disease prediction tasks. Additionally, the system's ability to efficiently handle complex datasets and extract meaningful insights contributes to its effectiveness in aiding physicians in making informed decisions and providing timely interventions for patients. Overall, the results highlight the potential of the system to significantly improve healthcare outcomes by enabling early disease detection and proactive management strategies.(see figure 3)

Figure 2. Predicting disease

DOI:10.48001/978-81-966500-7-0-10 | ISBN: 978-81-966500-7-0

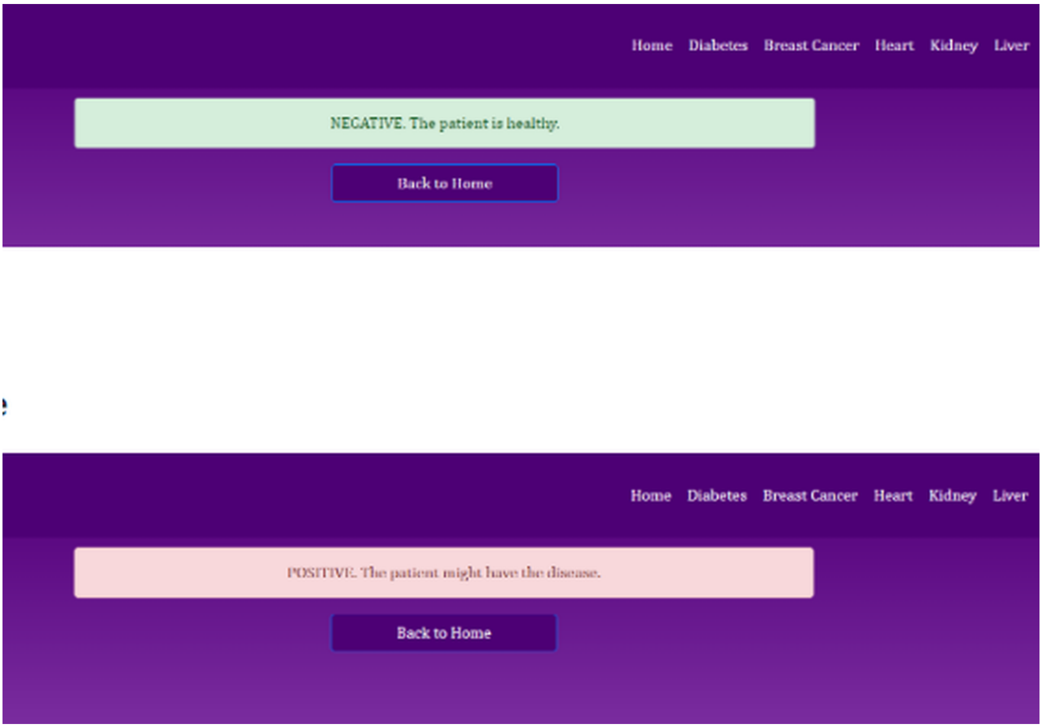

Figure 3. Results of the Med-Predictor

6 Conclusion

In conclusion, the integration of machine learning algorithms in the described system exhibits promising outcomes for disease prediction and patient care. With model accuracies of 92.54% for diabetes, 98.70% for heart disease, 97.66% for breast cancer, 99.17% for kidney disease, and 71.11% for liver disease, the system demonstrates robust performance across various medical conditions. Through methodologies such as logistic regression, random forest, and support vector machines, the system achieves high accuracy and reliability in forecasting disease outcomes. By efficiently processing intricate datasets, it enables early detection and facilitates timely interventions, thereby enhancing healthcare delivery and patient outcomes. These results underscore the potential of advanced technology to revolutionize medical practices, paving the way for proactive disease management and improved healthcare strategies in the future.

DOI:10.48001/978-81-966500-7-0-10 | ISBN: 978-81-966500-7-0

References

Arumugam, K., Naved, M., Shinde, P. P., Leiva-Chauca, O., Huaman-Osorio, A., & Gonzales-Yanac, T. (2023). Multiple disease prediction using Machine learning algorithms. Materials Today: Proceedings, 80, 3682–3685. https://doi.org/10.1016/j.matpr.2021.07.361

Banerjee, K., & Vinooth, P. (2023). Multiple Disease Prediction Using ML and Doctor Recommendation by Sentiment Analysis. Proceedings of International Conference on Contemporary Computing and Informatics, IC3I 2023, 1469–1472. https://doi.org/10.1109/IC3I59117.2023.10397715

Biswas, N., Ali, M. M., Rahaman, M. A., Islam, M., Mia, M. R., Azam, S., Ahmed, K., Bui, F. M., Al-Zahrani, F. A., & Moni, M. A. (2023). Machine Learning-Based Model to Predict Heart Disease in Early Stage Employing Different Feature Selection Techniques. BioMed Research International, 2023. https://doi.org/10.1155/2023/6864343

Chakrabarti, D. K., & Mittal, P. (2023). Plant Disease Forecasting Systems: Procedure, Application and Prospect. https://doi.org/10.1007/978-981-99-1210-0

Gautam, S., Ahlawat, S., & Mittal, P. (2022). Binary and Multi-class Classification of Brain Tumors using MRI Images. International Journal of Experimental Research and Review, 29, 1–9. https://doi.org/10.52756/ijerr.2022.v29.001

Gopisetti, L. D., Kummera, S. K. L., Pattamsetti, S. R., Kuna, S., Parsi, N., & Kodali, H. P. (2023). Multiple Disease Prediction System using Machine Learning and Streamlit. Proceedings - 5th International Conference on Smart Systems and Inventive Technology, ICSSIT 2023, 923–931. https://doi.org/10.1109/ICSSIT55814.2023.10060903

Islam, M. A., Majumder, M. Z. H., & Hussein, M. A. (2023). Chronic kidney disease prediction based on machine learning algorithms. Journal of Pathology Informatics, 14. https://doi.org/10.1016/j.jpi.2023.100189

Khan, R. H., Miah, J., Rahman, M. M., & Tayaba, M. (2023). A Comparative Study of Machine Learning Algorithms for Detecting Breast Cancer. 2023 IEEE 13th Annual Computing and Communication Workshop and Conference, CCWC 2023, 647–652. https://doi.org/10.1109/CCWC57344.2023.10099106

Kumar, K. P., Pravalika, A., Sheela, R. P., & Vishwam, Y. (2022). Disease Prediction Using Machine Learning Algorithms KNN and CNN. International Journal for Research in Applied Science and Engineering Technology, 10(5), 446–450. https://doi.org/10.22214/ijraset.2022.42214

Lucas, S., Desai, M., Khot, A., Harriet, S., & Narkar, N. (2022). SmartCare: A Symptoms Based Disease Prediction Model Using Machine Learning Approach. International

DOI:10.48001/978-81-966500-7-0-10 | ISBN: 978-81-966500-7-0

Journal for Research in Applied Science and Engineering Technology, 10(11), 709–715. https://doi.org/10.22214/ijraset.2022.47434

Mittal, P. (2014). Challenges in Managing the Blood Supply Chain, from Donor to Patient together with the Hospital. Journal of Business Management and Information Systems, 1(1), 37–47. https://doi.org/10.48001/jbmis.2014.0101003

Muthu Ganesh, V., & Nithiyanantham, J. (2022). Heuristic-based channel selection with enhanced deep learning for heart disease prediction under WBAN. Computer Methods in Biomechanics and Biomedical Engineering, 25(13), 1429–1448. https://doi.org/10.1080/10255842.2021.2013828

Wang, X., Ellul, J., & Azzopardi, G. (2020). Elderly Fall Detection Systems: A Literature Survey. Frontiers in Robotics and AI, 7. https://doi.org/10.3389/frobt.2020.00071

DOI:10.48001/978-81-966500-7-0-10 | ISBN: 978-81-966500-7-0