WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Intelligent Disease Condition Prediction System Victimisation K-Means Cluster

Mr. HariHaran. M[1], [2]Dr. Alwin Pinakas James. M.Phil., (Ph.D.)

[1]Master of Computer Application (MCA), KGISL Institute of Information Management, Thudiyalur Road, Saravanampatti, Coimbatore – 641035, Tamil Nadu, India

[2]Assistant Professor & Head of the Department, Master of Computer Application, KGISL Institute of Information Management, Thudiyalur Road, Saravanampatti, Coimbatore – 641035, Tamil Nadu, India

ABSTRACT

Clinical errors contribute to numerous deaths worldwide, highlighting the need for intelligent decision support systems in healthcare. This study proposes a disease condition prediction system using K-Means clustering, a data mining technique that groups similar patient data to uncover hidden patterns. Despite the abundance of medical records, much of this data remains underutilized. By applying clustering methods, our system enhances disease diagnosis accuracy, reducing misdiagnosis risks and improving patient safety. Leveraging data mining in medical diagnosis can transform healthcare data into actionable insights, aiding clinicians in making informed decisions.

Keywords: Intelligent System, Disease Prediction, K-Means Clustering, Patient Diagnosis ,Clinical Decision Support System (CDSS)

INTRODUCTION

One promising approach to improving medical diagnosis is **K-Means clustering**, a machine learning technique that groups similar patient data to identify patterns associated with diseases. By applying this method, healthcare providers can enhance diagnostic accuracy, optimize data storage, and reduce misdiagnosis risks. This study focuses on developing a disease condition prediction system using **K-Means clustering**, leveraging patient attributes such as age, symptoms, medical history, and test results to categorize disease conditions more effectively. The integration of such techniques into medical information systems can significantly improve healthcare decision-making and patient outcomes.

LITERATURE REVIEW

There are three different supervised machine learning algorithms for heart disease prediction. They are Naïve Bayes, K-nearest neighbour, and Decision tree. These algorithms have been used for analyzing the heart disease. Tanagra is the data mining tool used for classifying these medical data and these data are calculated using 10 fold cross validation. Naive Bayes algorithm performs well when compared to other algorithms. Genetic algorithm has been used in, to reduce the definite data size to obtain the best possible subset of attribute which is essential for heart disease prediction. Classification is supervised learning method to extract models relating main classes of data. Decision Tree, Naïve Bayes and Classification via clustering are the three classifiers used to analyze the occurrence of heart disease for the patients. Shekar et al proposed new algorithm to mine association rules from medical data based on digit sequence and clustering for heart attack prediction the entire data base is divided into partitions of equal size, each partition will be called cluster. This approach reduces main memory requirement since it consider only a small cluster at a time and it is scalable and efficient

SYSTEM SPECIFICATION

Hardware Requirements:

- Processor: Intel Core i5
- RAM: Minimum 8GB
- Storage: At least 256GB SSD
- GPU (Optional): NVIDIA GTX

- Network: Stable internet connection

Software Requirements:

- Operating system: Windows 10.
- Coding Language: PHP
- DataBase: MYSQL.

EXISTING SYSTEM

Existing research on heart disease prediction has largely relied on artificial neural networks and supervised data mining techniques such as decision trees and Naïve Bayes. While these methods have shown promising results, they require labeled datasets and extensive training, making them computationally expensive and less adaptable to large-scale, unstructured medical data. Recent advancements have introduced Intelligent Disease Prediction Systems (IDPS) that leverage neural networks and decision trees, but these approaches often struggle with discovering hidden patterns in healthcare data. To address these limitations, unsupervised learning techniques like K-Means clustering offer a promising alternative by grouping patients based on similarities without needing predefined labels. This enhances disease prediction accuracy, risk assessment, and early diagnosis, making clustering-based models an essential area of research in healthcare analytics.

DISADVANTAGES

- Accuracy Issues: A manual system alone does not ensure accuracy, and the warehouse data is only as good as the data entry that created it.
- The system is not fully automated; it needs data from user for full diagnosis.
- Major challenge is how to extract the information from these data because the amount is very large so some data mining and machine learning techniques can be used.

PROPOSED SYSTEM

The proposed system follows a **structured methodology** to preprocess medical data, apply clustering techniques, and analyze disease patterns for better prediction and diagnosis. The key steps involved are:

Data Collection

Medical datasets are gathered from hospitals, healthcare organizations, or publicly available sources.The data includes patient details such as age, symptoms, medical history, test results, and lifestyle factors.

Data Preprocessing

Handling Missing Values: Attributes with missing values exceeding **5%** are imputed instead of being removed to avoid data loss.

Normalization & Scaling: Medical attributes (e.g., blood pressure, cholesterol levels) are scaled to ensure uniformity in clustering.

Application of K-Means Clustering

Optimal K Selection: The Elbow Method or Silhouette Score is used to determine the best number of clusters (K).

Cluster Formation: Patients are grouped based on similarity in medical attributes, identifying low-risk, moderate-risk, and high-risk categories.

Disease Risk Analysis and Prediction

Cluster Interpretation: Each cluster is analyzed to identify patterns and correlations between patient attributes and disease risks.

Medical Insights: Doctors and healthcare professionals use clustering results to improve early diagnosis and treatment recommendations.

Performance Evaluation

Accuracy Assessment: Clustering effectiveness is validated using intra-cluster similarity and inter-cluster distance.

Comparison with Other Techniques: The results are compared with supervised models like Decision Trees and Naïve Bayes to assess the efficiency of K-Means in disease prediction.

Advantages

- User can search for doctor’s help at any point of time.
- User can talk about their Disease and get instant diagnosis.

- Doctors get more clients online.
- Very useful in case of emergency.

Modules

Patient Module

Patient register:

Patients register their details like name, email, phone, username and password.

Patient Login:

Patient logs into the system with registered username and password.

Patient disease history:

Patient enter their disease history attribute details like age, gender, smoking habit, drinking habit, fast food intake habit, whether they are overweight, whether they are mentally stressed, , whether they have pin in chest, blood pressure detail, blood sugar details.

View Disease Report:

Patient disease history attribute values are analysed with k-means algorithm and finally Disease prediction is show to user. If blood glucose sugar level is very high above 300, patient will be remained to go to hospital and take insulin.

Doctor Module

Doctor register:

Doctor registers their details like name, email, phone, specialist in, username and password.

Doctor Login:

Doctor logs into the system with username and password.

Doctor view patient history:

In this module doctor can view patient disease history attribute details like age, gender, smoking habit, drinking habit, fast food intake habit, whether they are overweight, whether they are mentally stressed, whether they have pin in chest, blood pressure detail, blood sugar details.

Doctor Analyse Report:

In this module doctor can view overall clustered report. The cluster report will say large percentage of heart attack disease is due to smoking habit or drinking habit or fast food intake habit or stress factor, etc. The report says whether the patient has the high risk of heart disease, diabetes or cancer.

Performance Report:

In this module doctor can analyse the disease based of the set of attributes like how many patients have cancer attack who have their age greater than 65, whether sugar level greater than 180 may cause diabetes, cholesterol level greater than 120 may cause heart disease.

SCREENS

CONCLUSION

The Intelligent Disease Prediction System using K-Means Clustering enhances disease diagnosis by grouping patients based on medical attributes, enabling early detection and timely intervention. K-Means ensures efficient and scalable classification, aiding healthcare professionals in analyzing disease patterns. Despite challenges like sensitivity to initial partitions, the system improves medical decision-making. Future enhancements, including hybrid models and deep learning, can further optimize accuracy. Overall, this system reduces diagnostic errors, improves patient care, and supports data-driven healthcare decisions

FUTURE SCOPE

The **Intelligent Disease Prediction System using K-Means Clustering** can be further enhanced with advanced technologies to improve accuracy, scalability, and real-time medical insights. Some key future developments include:

1. **Hybrid Clustering Models:** Combining K-Means with **Hierarchical or DBSCAN clustering** to improve classification accuracy and stability.
2. **Deep Learning Integration:** Using **neural networks and AI models** to enhance disease prediction with higher precision.
3. **Real-Time Data Processing:** Implementing **IoT-based health monitoring systems** for continuous patient data updates.
4. **Cloud-Based Deployment:** Developing a **scalable, cloud-integrated platform** for real-time accessibility by healthcare professionals.
5. **Personalized Predictions:** Enhancing the system with **patient-specific recommendations** based on genetic data and medical history.

REFERENCES

1. Han, J., Kamber, M., & Pei, J. (2011). Data Mining: Concepts and Techniques. Morgan Kaufmann.
2. MacQueen, J. (1967). Some Methods for Classification and Analysis of Multivariate Observations. Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, 1(14), 281-297.

3. World Health Organization (WHO). (2023). Global Health Statistics on Cardiovascular Diseases. [Online] Available: www.who.int

4. Ramesh, D., & Sathya, M. (2018). Heart Disease Prediction Using Machine Learning Techniques. International Journal of Computer Applications, 182(1), 22-28.

5. Tan, P. N., Steinbach, M., & Kumar, V. (2005). Introduction to Data Mining. Pearson Education.

6. Shekar, B. H., & Ramani, S. (2019). A Novel Approach to Medical Data Clustering Using K-Means Algorithm. Journal of Healthcare Informatics, 15(4), 112-125.

7. Patil, S. B., & Kumaraswamy, Y. S. (2009). Intelligent and Effective Heart Attack Prediction System Using Data Mining and Artificial Neural Networks. European Journal of Scientific Research, 31(4), 642-656.

8. Chaurasia, V., & Pal, S. (2014). Data Mining Techniques: To Predict and Resolve Breast Cancer Survivability. International Journal of Computer Science and Mobile Computing, 3(1), 10-22.

9. Das, R., Turkoglu, I., & Sengur, A. (2009). Effective Diagnosis of Heart Disease Through Neural Networks Ensembles. Expert Systems with Applications, 36(4), 7675-7680.

10. Dey, L., & Rautaray, S. S. (2014). Study and Analysis of Data Mining Algorithms for Healthcare Decision Support System. International Journal of Computer Science and Information Technologies, 5(1), 470-477.