Machine Learning-Based Healthcare Guidance System

Kumar P
Department of CSE
Rajalakshmi Engineering College
Chennai, India
kumar@rajalakshmi.edu.in

Yashini P
Department of CSE, REC
Chennai, India
220711002@rajalakshmi.edu.in

2024 Second International Conference on Advances in Information Technology (ICAIT) | 979-8-3503-8386-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICAIT61638.2024.10690330

Abstract— Hospitals are increasingly utilizing online recommender systems to assist in the medical treatment of patients. Nowadays, many people research potential prescriptions online before consulting their doctors for various medical issues. Medical recommendation systems can be beneficial, especially during pandemics or natural disasters when resources are limited. These systems provide more precise, dependable, and accurate clinical predictions using fewer resources. The patient receives trustworthy information from the medication recommendation system regarding the medication, dose, and potential side effects. The system considers the patient's symptoms when selecting the appropriate medication, which is then delivered based on the user profile. The system employs K-means clustering algorithms to analyze patient data to provide personalized medical recommendations. These algorithms use patient demographics, medical histories, and symptoms gathered from a vast dataset of medical records to produce precise suggestions. By providing fast, accurate recommendations tailored to each person's specific needs, this system aims to enhance patient outcomes

Keywords: Learning Algorithm, Medical Recommender systems, Clustering algorithm..

I. INTRODUCTION

The healthcare[1,2] industry has seen a significant increase in data collected from various sources, such as wearable devices, electronic health records, and medical imaging[3,4]. To provide personalized and efficient medical recommendations, this data can help patients and healthcare providers make informed decisions about diagnosis, treatment, and overall healthcare management. One of the innovative uses of technology in healthcare is the Medical Recommendation System, which uses data analysis and computer algorithms. This system incorporates many data sources, such as clinical guidelines, medical literature, and patient records, to generate tailored and evidence-based suggestions. This study presents a new medical recommendation system methodology using the K-means clustering algorithm. Traditional medical recommendation systems typically use rule-based or collaborative filtering techniques. Although these methods have shown effectiveness, they may only sometimes capture the complex patterns in patient data.

The K-means clustering algorithm, commonly used in data mining and pattern recognition, presents a viable alternative by effectively grouping similar data points into clusters. This methodology aims to use unprocessed patient data, including demographic details, medical records, and diagnostic reports, and subject it to comprehensive preprocessing techniques to address missing values and establish a standardized format. The selection of pertinent qualities is based on their influence on patient well-being and data accessibility to the healthcare device in people. Next, the preprocessed data is applied to the K-means clustering method. This unsupervised learning technique identifies intrinsic patterns within a dataset and groups patients into clusters based on their similarity. Each cluster represents a unique patient profile. After the clustering process, the system analyzes each cluster to identify significant traits and shared attributes among patients in the respective group. This study provides valuable insights into the probable health issues, lifestyle variables, or treatment responses commonly observed among the identified clusters.

II. LITERATURE SURVEY

Since Liu et al [5] Naive Bayes Logistic Regression Patients Are Encouraged to Consume Nutritional Supplements, Diets, And Foods That Are Deemed to Be More Suited To Their Likes, Dietary Preferences. General Health Using A Meticulous Analysis Process That Involves Dieticians And Patients. The Patient-Dietician's Reasoning Behind The Recommender System Is Still Not Entirely Understood By Medical Staff Dietary Assistance For Patients.

Eric Appiah Mantey et al [6] algorithms for filtering Health recommender systems (HRSs) can successfully tailor behavior modification treatments related to well-being to each person's needs. In particular, little attention is paid to theory-based behavioral determinants and the range of lifestyle domains that affect well-being. Expanding the With-Me user model with characteristics that describe abilities, self-efficacy, and the current context is a minimum requirement. Characteristics linked to self-efficacy and skills allow for the recommendation of beneficial but manageable behavior modification activities and knowledge.

Abdullah et al [7] RNN, Prescription Advice, Curriculum Instruction, Set Encoder, And Data Mining For Electronic Health Records (Ehrs) A Crucial Yet Difficult Duty In Healthcare Is Effectively Prescribing Medication For Complex Multimorbidity Disorders. The Majority Of Previous Research Forecasted Drug Regimens Using Longitudinal Data, Presuming That The Length Of The Visit Inevitably Determined The Encoding Format Of Intra-Visit Medical Events. Extensive Trials On A Benchmark Dataset have verified our Model's Superiority Over Multiple State-Of-The-Art Baselines. Learning Longitudinal Sequence

DataExhibits Persistent Patterns Of Information Transmission And Serialization.

Mustafa Haider Abidi et al [8] Deep Learning, Graph Clustering, Food Recommendation, Healthcare, Mehrdad Rostami, Mourad Oussalah, and Vahid Farrahi. Food is recommended to assist consumers in changing their eating habits and achieving a healthier diet. This research aims to create a new hybrid food recommender system that addresses the drawbacks of earlier models, including their disregard for food ingredients, time constraints, cold start users, cold start food products, and community factors. To further enhance the food suggestion framework's performance in the end, future works should add user data such as gender, age, height, weight, location, and culture.

Diana Sousa and Francisco M. Cout [9] cold start, deep learning, neural networks, recommender systems, new item problem, new user problem, recommender systems evaluation, and stereotypes the suggestion of products from the catalog to an unidentified new user or the suggestion of freshly added material to consumers who already have it. Better cold-start performance results from integrating metadata describing the object or the user. Results show that even though the recommendations have worse serendipity and fairness characteristics, a multi-layer neural network significantly improves cold start accuracy performance measures.

Yang Han et al [10] Advanced Manufacturing Institute, King [King Ehealth, Health Care, Smart Health Monitoring System the terms industry 4.0, internet of medical things, convolutional neural network, deep belief network, deep ensemble learning, extreme learning machine, and big data processing are used. Through wearable sensors, big data, and telecommunications technologies over pervasive computing, human life has become more competent to provide better healthcare services. Big data is designed with the potential to advance the medical field. Big data uses software and information and communication technology (ICT) to create connections between patients, wearable sensors, healthcare practitioners, and caregivers. For improved outcomes, the current algorithm's parameters will be further adjusted .

Maria Habibet al [11] cancer recovery, and artificial intelligence recommendation systems are among the companies mentioned. This study focuses on researching an intelligent recommendation model for cancer patients' rehabilitation, taking advantage of the benefits of the Internet of Things. It also creates an intelligent recommendation system that is easy for cancer patients' rehabilitation. Given the ambiguity surrounding these factors, the convolutional neural network approach was used to forecast both the cause and the period of recurrence of cancer patients. The model's findings demonstrated a high prediction accuracy and 92% forecast accuracy. To identify the optimal scheme, I'll keep researching cancer rehabilitation suggestion schemes and merging them with more sophisticated mathematical models. Andres et al [12] wearable technology, augmented reality, blockchain, explainable AI, head-mounted displays, virtual and Metaverse realities, and chronic illness management. In virtual environments, avatars and three-dimensional (3D) spaces can improve digital teaching, diagnosis, treatment decisions, and patient-facing platforms in the medical and ophthalmic fields. The prevalence of chronic diseases is rising worldwide, with an estimated 25% of people currently dealing with several chronic health conditions. There is more to health care [13,14] than only attending to a patient's problems. Furthermore, mixed-reality systems such as the Metaverse can aid in illness prevention and prediction.

III. METHODOLOGY

The process of building pre-trained models and how they are used in medical-named entity identification tasks. The following section will discuss creating a medical knowledge graph using the named entity recognition approach.

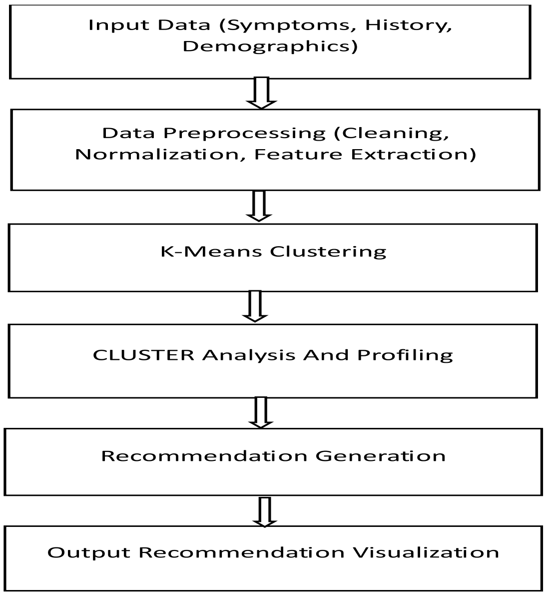


Figure.1. Proposed Model Process Flow

Figure 1 shows that the module collects patient data, including their symptoms, medical history, and demographics. This input data undergoes preprocessing steps, such as cleaning, normalization, and feature extraction, to prepare it for clustering. The preprocessed data is then fed into the k-means clustering algorithm, which groups patients into clusters based on similarities in their features. Each cluster is analyzed to understand the characteristics of the patients within it, which helps in profiling the clusters. Based on the cluster analysis, recommendations for medical treatments or interventions are generated for each cluster.

A. Data Preprocessing Module:

Raw medical data often requires cleaning, transformation, and normalization to make it suitable for clustering. This module handles data preprocessing tasks like cleaning, feature extraction, and integration. The aim is to group patient data and important attributes collected from electronic and healthcare records into clusters. Data cleaning techniques are employed to handle missing data, which include removing records or filling in missing values using appropriate methods. Data transformation is used to select the most relevant features for clustering by eliminating irrelevant or redundant features. Feature scaling is a technique that brings all numerical features to a similar scale or standardizes them. Data integration that involves medical expertise, such as domain-specific medical information like symptoms, diseases, and treatment, is also included.

B. Feature Engineering Module:

Feature engineering is essential in developing a medical recommendation system using k-means clustering. Initially, we collect medical data from various sources and then clean, transform, and normalize it for processing. After that, we extract relevant features from the preprocessed data, such as patient demographics, medical history, and symptoms. The data is then grouped into clusters using the k-means algorithm, which helps identify similar medical cases or patients. The algorithm assigns data points to clusters and calculates cluster centroids. Finally, recommendations are generated for each cluster based on the assigned data points and centroids.

C. K-means clustering Modules:

This module performs K-means clustering on the patient data to group patients with similar characteristics into clusters. The process of clustering begins by randomly selecting initial centroids for the clusters. This helps to initialize the process. After that, we calculate the distance to the centroids and assign the data points to the nearest cluster. Then, we update the cluster centroids by calculating the mean of data points in each cluster and updating the centroids accordingly. We check if the centroids have converged by monitoring if they have stopped changing significantly. Once they have converged, we move on to the next step. Finally, we use the clusters to generate personalized patient recommendations, including treatment options, lifestyle changes, and diagnostic tests.

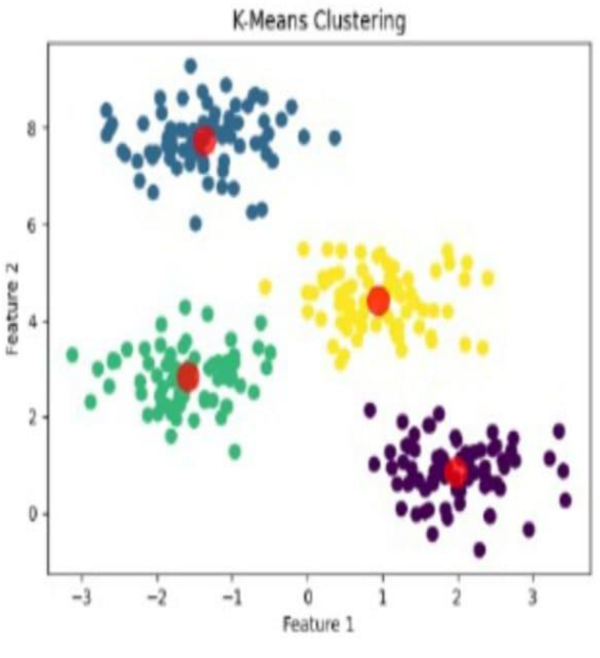


Figure.2.K-Means Clustering

Figure.2 refers to the K-means clustering, grouping data points based on their attributes or dimensions, called features. Each data point is represented as a vector in a multi-dimensional space, and Feature 1 in K-means clustering refers to one of these attributes in the dataset. For example, if you're clustering a dataset of customer information, Feature 1 could represent something like age, income, or spending habits. To perform K-means clustering, you select a subset of features or all features to define the similarity between data points. Cluster centroid based on the chosen features and updates the centroids iteratively until convergence. The K-means clustering technique, which groups data points into clusters based on similarity.

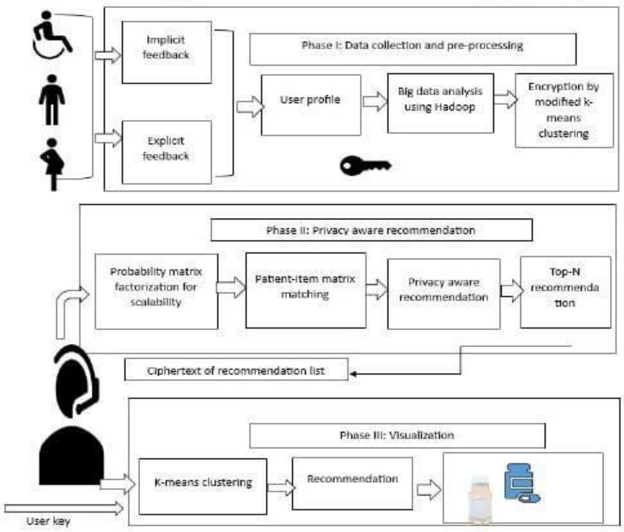
The features or attributes of the data points determine the similarity. For instance, if we cluster disease symptoms, features could include the intensity of fever, duration of cough, presence of rash, etc. Each patient represents a data point, and the features describe their symptoms. Each feature contributes to the overall similarity or dissimilarity between data points. For example, Feature 1 could be fever intensity, which would be one of the variables used to group patients with similar symptoms together. The goal is to assign each data point to the cluster whose centroid is closest to it.

D. Recommendation Engine Modules:

After clustering, the system uses the cluster assignments to generate personalized patient recommendations within each cluster. Recommendations include treatment options, medications, lifestyle changes, or other relevant medical advice. The healthcare system collects comprehensive patient information, including medical history, symptoms, and other pertinent data from diverse sources such as electronic health records or medical records. This data is then carefully scrutinized, cleaned, transformed, and prepared for the following processing phase - clustering. The K-means clustering algorithm is utilized for grouping system generates personalized recommendations for each cluster. These recommendations may include treatment options, medications, lifestyle changes, or other relevant medical advice. Although the initial recommendations for each cluster are generated based on their respective assignments, the healthcare system takes personalization to another level by tweaking these recommendations based on the patient's data within a cluster. This ensures that the system adjusts treatment plans and medication to match each patient's specific needs, leading to better health outcomes. A Medical Recommendation System's primary objective is to improve healthcare quality, optimize treatment plans, and enhance patient outcomes.

Figure 3 architecture diagram for a medical recommendation system involves gathering and preparing data using implicit and explicit feedback. Implicit feedback is data inferred from user behavior, such as clicks, view durations, purchase history, or search queries. This data is collected using click-through rates, view counts, purchase history, time spent on content, and search queries.

979-8-3503-8386-7/24/$31.00 ©2024 IEEE



Fgure.3. Model Architecture

Explicit feedback, on the other hand, is direct and intentional user expressions of preference or opinion regarding items or content, including user-provided ratings, reviews, likes, or other direct indications of interest or disinterest. User profiles are created and used in various applications, including social media, e-commerce, content recommendation, and personalization. These profiles include basic information to identify the user, such as a username, email address, or user ID, as well as demographic information, such as age, gender, location, and other relevant details. Preferences can be categorized by music genres, movie genres, hobbies, and more, and behavioral data can reveal user evaluations of goods or content. The Hadoop ecosystem is utilized to manage and process massive amounts of data. This group of free and open-source software tools and frameworks is necessary when conducting extensive data analysis with Hadoop. Distributed computing, data storage, and batch processing are three areas where Hadoop excels. To start analyzing big datasets using Hadoop, you must import the data into the Hadoop Distributed File System (HDFS) as the first step. The initial data analysis phase involves preparing the data by performing data transformation, formatting, and cleansing. Phase II involves privacy-aware recommendations, achieved using probability matrix factorization (PMF) for scalability. Collaborative filtering and recommendation systems use this technique based on probabilistic modeling and Bayesian inference to factorize the user-item interaction matrix. The goal is to discover latent factors that explain user-item interactions. Although PMF can provide accurate recommendations, it can be computationally intensive for large datasets. To address scalability issues, "patient-item matrix matching" is used. "Patient-item matrix matching" is used primarily in recommendation systems for healthcare or medical settings. In this context, the "patient-item matrix" is a matrix that contains information about patients and healthcare items. At the same time, "matching" refers to making recommendations for individual patients based on their needs, preferences, or medical history. The patient-item matrix is a structured data representation where rows represent patients, columns represent healthcare items or services, and the matrix's cells contain information about patients' interactions, preferences, or needs related to specific items. Data is collected from various sources such as electronic health records (EHRs), patient surveys, medical claims, or patient feedback to populate the patient-item matrix. Each patient's profile is built based on their historical interactions with healthcare items, including diagnoses, treatments, medications, surgeries, or any other healthcare-related activity. The patient profile is used to understand the patient's medical history, preferences, and health conditions. Item profiling is done on healthcare items based on their attributes, including medical characteristics (e.g., drug type, medical equipment), cost, availability, quality, and patient feedback. Hybrid approaches are used to match patients with relevant healthcare items. Privacy-aware recommendation systems provide personalized recommendations while respecting and protecting users' privacy. They aim to balance the need for customized user experiences with safeguarding sensitive user data.

IV. RESULTS & DISCUSSIONS

K-means clustering can be helpful in a medical recommendation system to group patients or medical data. The clustering is based on similar characteristics of the gathered data. The clusters generated can then be used to make personalized medical recommendations like treatment options, preventive care, or lifestyle changes. Collect relevant structured medical data, including patient demographics, medical history, test results, and genetic information. Have an explicit target variable for a recommendation, such as a recommended treatment or intervention.

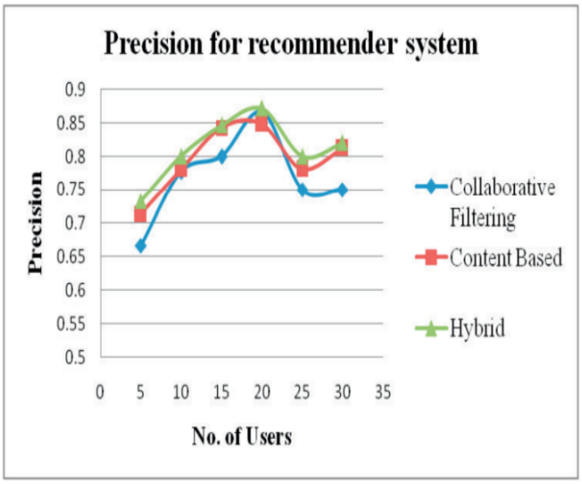


Figure.4. Model Precision Measure

Figure. 4 Precision is a crucial measure of the effectiveness of a recommender system, especially in situations where accuracy and relevance are essential. In the context of a recommender system, precision indicates the ratio of relevant items that the system recommends compared to the total number of items recommended. It provides valuable insight into how accurately the system selects relevant user items. When it comes to Cluster-Specific recommendations, precision becomes even more significant. These recommendations are generated by algorithms such as K-

Means, which group users (or patients) into clusters based on their shared characteristics or traits. Each cluster represents a unique segment of the user population. In the case of medical interventions or treatments recommended within a specific cluster, high precision indicates that the recommended items are relevant and appropriate for the patients grouped within that cluster. In other words, a high precision score suggests that the recommender system accurately identifies and suggests medical interventions that align well with the needs and characteristics of the patients within that particular cluster.

For instance, if a Cluster-Specific recommendation system identifies a group of patients with similar symptoms or medical histories and recommends treatments tailored to that cluster, high precision means that most recommended treatments are effective and suitable for the patients within that cluster. Enhancing the overall effectiveness and utility of the recommendation system, high precision indicates that the recommended items are well-suited to the characteristics and needs of the user group being targeted.

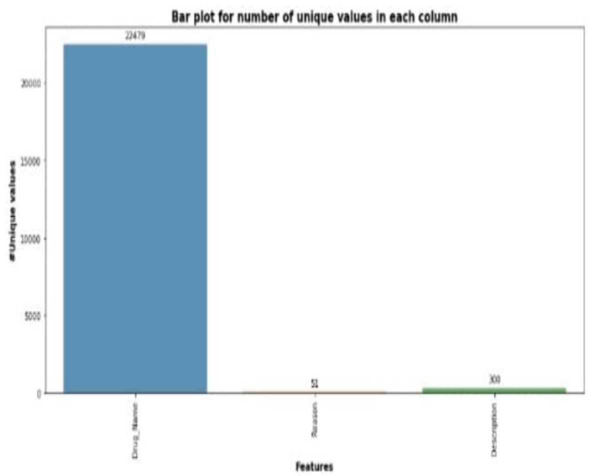


Figure.5.Sparse Data Analysis

Figure.5. Sparse data, indicated by columns with few unique values, can negatively impact recommendation systems. This is because it may lead to unreliable predictions. You can use bar plots to quickly identify such columns by visualizing the distribution of unique values. On the other hand, high cardinality refers to columns with many unique values. These columns can challenge machine learning algorithms and may require special handling during data preprocessing. Bar plots can help spot these columns and prioritize further exploration or special treatment. Significant differences in the number of unique values across columns may indicate data imbalances. Such imbalances can affect the performance of recommendation systems by biasing predictions towards the majority class. Bar plots can help identify such imbalances by visualizing the distribution of unique values across columns.

To improve model performance, it's essential to prioritize feature engineering efforts. For example, columns with unique values may benefit from encoding categorical variables or grouping rare categories. Bar plots provide insights into each column's distribution of unique values, allowing data scientists to make informed decisions about feature engineering strategies.

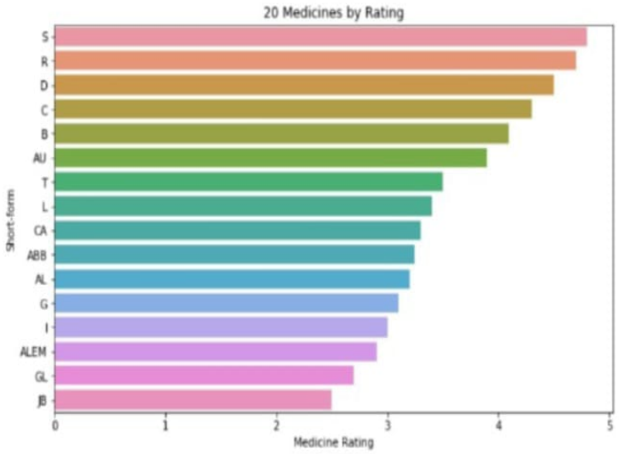


Figure.6. Medicines Recommendation

Figure 6 Determining the best medicines for short-term use requires establishing rating metrics that consider symptom relief, side effects, and ease of use. Data can be collected from various sources, including patient feedback or clinical trials, to obtain accurate ratings. Surveys, electronic health records, and other relevant sources can provide valuable information. Once sufficient data is collected, an aggregate rating for each medicine can be calculated by analyzing and averaging ratings from different patients or trials. Medicines can then be ranked in descending order based on their aggregate ratings to identify the top performers. To ensure that the recommendation system only suggests the best medicines, it is necessary to carefully select the top 20 medicines with the highest short-term ratings, considering factors such as the target patient population and medical conditions. Periodic rating updates based on new data or feedback are essential to keep the recommendation system relevant and up-to-date. This helps adapt to changes in medical knowledge, patient preferences, and emerging treatments, ultimately improving patient care. Regular updates are critical in ensuring the recommendation system provides accurate and helpful suggestions to healthcare professionals and patients.

V. Conclusion And Future Enhancement

In conclusion, a Medical Recommendation System utilizing K-Means Clustering offers a promising approach to enhance the accuracy and personalization of healthcare recommendations. The application of K-Means clustering to medical data facilitates the identification of meaningful patient groups, enabling tailored and targeted recommendations. The system's is contingent on various factors, including the quality of data, feature selection, and the integration of medical knowledge. The clustering process optimizes the grouping of patients with similar characteristics, allowing for precise and cluster-specific recommendations. The system's outcomes are notable across multiple dimensions. First and foremost, the precision of recommendations within each cluster ensures that interventions are relevant and aligned with the common medical traits of patients in that group. This not only enhances the trustworthiness of the system but also contributes to improved patient outcomes and safety by minimizing the risk of false positives. The personalized nature of recommendations, guided by K-Means clustering, allows healthcare providers to allocate resources more efficiently. By tailoring interventions to specific clusters, the system aids in optimizing healthcare resource utilization and promoting cost-effective and targeted treatment strategies. Future work in medical recommendation systems using K-means clustering can focus on various aspects to enhance the quality and effectiveness of recommendations and some potential directions for research and development to incorporate the Multi-modal Data to expand the scope of data used in clustering by including various types of medical data, such as images (e.g., X-rays, MRIs), free-text clinical notes, wearable device data (e.g., heart rate, activity levels), and genetic information.

REFERENCES

[1] S. P. S, K. T, J. M and M. A. Sheriff, "A Comparative Analysis on the Prediction of Heart Failure using Machine Learning Algorithms," 2024 5th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI), Lalitpur, Nepal, 2024, pp. 206-211, doi: 10.1109/ICMCSI61536.2024.00037

[2] S. S. P, M. S. Monesh and B. Lingesh, "A Novel Approach to Detect Face Fraud Detection Using Artificial Intelligence," 2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE), Vellore, India, 2024, pp. 1-6, doi: 10.1109/ic-ETITE58242.2024.10493594.

[3] S. P. S, K. T, V. R. S R and V. R, "Predictive Modelling of Critical Vital Signs in ICU Patients by Machine Learning: An Early Warning System for Improved Patient Outcomes," 2024 3rd International Conference for Innovation in Technology (INOCON), Bangalore, India, 2024, pp. 1-6, doi: 10.1109/INOCON60754.2024.10512042.

[4] S. S. P, K. P, K. T and V. R. Chiranjeevi, "Multi-Level Interpretable and Adaptive Representation of EEG Signals for Sleep Scoring Using Ensemble Learning Multi Classifiers," 2023 RMKMATE, Chennai, India, 2023, pp. 1-6, doi: 10.1109/RMKMATE59243.2023.10368630

[5] Since Liu, Xiaolong Wang, Jingcheng Du, Yongshuai Hou, Xianbing Zhao, Hui Xu, Hui Wang, Yang Xiang, and Buzhou Tang (2023)" A Sample-Adaptive Hierarchical Prediction Network for Medication Recommendation" Volume 27 pp 6018-6028.

[6] Eric Appiah Mantey, Conghua Zhou, Joseph Henry Anajemba, Yasir Hamid, John Kingsley Arthur (2023)"Blockchain-Enabled Technique for Privacy Preserved Medical Recommender System" Volume 11 2023 pp 40944-40953.

[7] Abdullah M. Almuhaideb, Mariam Elhussein, Reem Osman, Fatema Alholyal, Leena Alghamdi, Majd Al-Ismail, Maram Alawami, Zainab Kadour, And Rachid Zagrouba (2023)"Design Recommendations for Gate Security Systems and Health Status: A Systematic Review" Volume 11 pp 131508-131520.

[8] Mustafa Haider Abidi, Usama Umer, Syed Hammad Mian, And Abdulrahman Al-Ahmari (2023)" Big Data-Based Smart Health Monitoring System: Using Deep Ensemble Learning" Volume 11 pp 114880-114903.

[9] Diana Sousa and Francisco M. Couto (2022) "Biomedical Relation Extraction With Knowledge Graph-Based Recommendations" Volume 26 pp 4207-4217.

[10] Yang Han, Zhenguo Ha, Jianhui Wu, Yanlong Yu, Shuqing Gao, Dianbo Hua, And Aimin Yang (2020)"Artificial Intelligence Recommendation System of Cancer Rehabilitation Scheme Based on IoT Technology" Volume 8 pp 44924 - 44935

[11] Maria Habib, Mohammad Faris, Raneem Qaddoura, Alaa Alomari, and Hossam Faris (2021)" A Predictive Text System for Medical Recommendations in Telemedicine A Deep Learning Approach in the Arabic Context" Volume 9 pp 85690-85708.

[12] Andres Alejandro Ramos Magna, Héctor Allende-Cid, Carla Taramasco, Carlos Becerra, And Rosa L. Figueroa (2020) "Application of Machine Learning and Word Embeddings in the Classification of Cancer Diagnosis Using Patient Anamnesis" Volume 8 pp 106198-106213.

[13] K. P, Vinod Kumar K. S, P. L and S. S. P, "Enhancing Face Mask Detection Using Data Augmentation Techniques," 2023 ICRASET, B G NAGARA, India, 2023, pp. 1-5, doi: 10.1109/ICRASET59632.2023.10420361

[14] S. P. S, K. P and S. L. T A, "Projection of Plant Leaf Disease Using Support Vector Machine Algorithm," 2023 ICRASET, B G NAGARA, India, 2023, pp. 1-6, doi: 10.1109/ICRASET59632.2023.10419981