**Explainable AI for Clinical Decision Support: Integrating explainable AI methods into clinical decision support systems to provide transparent and interpretable explanations for AI-driven medical recommendations, enhancing clinician trust and adoption.**

B.V.V. Mahesha Chandra, veera.maheshachandra2021@vitstudent.ac.in

K. Poorna Avinash, konakallapoorna.avinash2021@vitstdent.ac.in

# Abstract

The integration of Artificial Intelligence (AI) into Clinical Decision Support Systems (CDSS) is revolutionizing the healthcare sector by offering data-driven insights for medical decision making. However, a major barrier to the widespread application of AI in clinical practice is the opaqueness and lack of interpretability of AI-driven recommendations. By incorporating Explainable AI (XAI) approaches into a CDSS, this research seeks to address this challenge and increase the transparency of AI recommendations. The developed system combines medical domain knowledge with cosine similarity and Term Frequency-Inverse Document Frequency (TF-IDF) algorithms to make drug recommendations based on patient health status and preferences. The system offers interpretable explanations for each proposal, including side effects, appropriateness for certain conditions, and financial considerations, to guarantee that practitioners can trust and understand the AI's advice. The study shows that XAI techniques can increase clinician confidence and enable them to make better decisions by enhancing interpretability. According to the results, including XAI into CDSS improves decision-making and encourages healthcare professionals to embrace AI. This study shows how XAI could facilitate more open, evidence-based healthcare choices and ultimately lead to better patient care by bridging the gap between AI technology and practical clinical applications.

*Keywords*:

Explainable AI; Clinical Decision Support Systems; Medical Recommendations; Predicting Side Effects; Transparent AI Models; Trust in AI; Interpretable Decision Support; Cosine Similarity for Recommendations.

# 1. Introduction

The emergence of artificial intelligence (AI) has revolutionized a number of industries, with the healthcare industry being one of the most affected. AI is a vital tool for improving therapeutic outcomes because of its capacity to evaluate enormous volumes of data, identify complex patterns, and offer useful insights. Clinical Decision Support Systems (CDSS) with AI capabilities in particular have been a revolutionary development in medical practice. By utilizing AI's computational efficiency and accuracy, these systems support physicians by offering evidence-based suggestions for diagnosis, therapy, and patient management. The lack of openness and interpretability in their decision-making processes is one of the major obstacles to the adoption of AI-driven CDSS, despite its encouraging potential [1, 3, 4].

From medical imaging data to electronic health records (EHRs), AI-driven CDSS are made to handle complicated and frequently high-dimensional information. AI algorithms can produce predictive insights and individualized treatment suggestions by finding patterns and connections in these datasets. AI systems, for example, can evaluate radiological images to find anomalies like tumors or recommend treatment regimens based on patient-specific information [2, 7]. Significant advantages have been shown by these capabilities, such as increased diagnostic precision, less medical errors, and better use of healthcare resources [6, 8]. However, because of worries about trust, dependability, and ethical issues, physicians and healthcare institutions are still hesitant to use these technologies [4, 5, 10]. The use of "black-box" models is a significant obstacle to the broad adoption of AI-driven CDSS. Although these intricate algorithms frequently produce highly accurate predictions, they lack the transparency necessary for clinicians to comprehend how they make decisions. Clinicians are hesitant to depend on AI advice because of this opacity, which erodes system confidence, especially in situations where human lives are at stake [3, 4]. Additionally, explainability is emphasized by regulatory frameworks like the General Data Protection Regulation (GDPR), particularly in delicate fields like healthcare. AI systems run the danger of restricted adoption and encounter ethical and legal issues in the absence of unambiguous and interpretable outputs [6, 8].

To tackle these issues, Explainable AI (XAI) presents a strong argument. The goal of XAI is to create interpretable and accurate AI models so that medical professionals can comprehend how and why the system makes particular suggestions. Clinicians are encouraged to incorporate AI systems into their workflow by XAI's increased transparency, which increases trust in these systems [9, 10]. In general, there are two types of XAI methodologies: intrinsic and post-hoc. Interpretable models, such decision trees or linear regression, are created from the ground up using intrinsic approaches. Conversely, post-hoc approaches use methods such as SHapley Additive exPlanations (SHAP), Local Interpretable Model-Agnostic Explanations (LIME), or feature importance visualization to produce explanations for intricate, pre-trained models [11, 12]. Explainability is crucial in healthcare for reasons that go beyond technical difficulties; it has a significant impact on patient safety and clinical results. Medical decisions frequently have instantaneous and profound effects. An inaccurate medication advice, for instance, may result in serious side effects, hospitalization, or even death [3, 15]. Clinicians must be able to evaluate and verify the reasoning behind AI-generated recommendations in such crucial situations. In addition to giving clinicians more confidence in AI systems, explainable models enable them to spot and correct any potential biases in the algorithms. Inequities in healthcare delivery can be sustained if previous discrepancies are not addressed, as they are frequently reflected in medical databases. By identifying these biases, XAI approaches guarantee that suggestions are morally and equitably sound [9, 13].

Explainability not only addresses bias and trust, but it also improves clinician-AI system collaboration. XAI systems enhance human knowledge by offering concise and pertinent explanations, resulting in a synergistic approach to patient treatment. For example, clinicians' diagnostic thinking can be supported by visual explanations like heatmaps or feature significance graphs, which makes AI systems a natural extension of their expertise. Through this partnership, a mutual learning environment is created in which AI models adjust to user feedback and improve performance over time, while doctors gain insights from AI outputs [10, 13]. Even with XAI's advancements, integrating technology into clinical workflows is still difficult. The technical features of explainability have been the main emphasis of existing research, which frequently ignores the contextual needs of healthcare providers. Clinicians need outputs that adhere to ethical and regulatory norms, actionable insights that complement their workflow, and explanations that are specific to their degrees of knowledge [6, 9]. Standardized metrics to assess the interpretability and usability of XAI models in healthcare are also lacking. The majority of current assessments focus on accuracy and computational efficiency, which leaves out user-centered characteristics like simplicity of use and trustworthiness [5, 14]. A multidisciplinary strategy combining cooperation between AI researchers, doctors, ethicists, and legislators is required to address these issues.

By putting out a human-centered framework for a medication recommender system, this study seeks to close the gap between the practical requirements of healthcare practitioners and the technological developments in XAI. The suggested system uses XAI principles to give doctors contextual information, including possible side effects, appropriateness for specific conditions (such pregnancy, liver illness, or alcohol use), and economic considerations, in addition to medication recommendations. Key obstacles to the adoption of AI-driven CDSS are addressed by the system's integration of these aspects, which guarantees that recommendations are actionable, interpretable, and clinically meaningful [2, 9]. By emphasizing human-centered design concepts, the framework customizes the system to enhance rather than interfere with clinical workflows. Iterative feedback loops allow clinicians to interact with the model, refine predictions, and enhance accuracy over time, fostering a collaborative environment that empowers both clinicians and AI systems [13]. The ultimate objective of XAI in healthcare is to develop systems that are relatable to the human users who depend on them, rather than only enhancing the technical performance of AI models. XAI has the potential to revolutionize healthcare delivery by emphasizing interpretability, transparency, and human-centered design. It ensures that AI-driven CDSS are not only technically sound but also trusted and accepted by doctors, opening the door for safer, more moral, and patient-centered procedures [9, 11]. This work shows that incorporating explainability into clinical decision-making is a fundamental step toward the ethical and successful application of AI in medicine, not just a technical difficulty.

# 2. Related Work

Tarnowska K. A. et al. [1] suggested using a knowledge-based clinical decision support system (CDSS) to diagnose and treat hearing impairments such misophonia, hyperacusis, and tinnitus. In order to improve credibility and acceptance in clinical settings, the researchers created the eTRT system, which provides an explainable outcome. In order to handle data, their method combined machine learning approaches such as association- and action-rule discovery. The evaluation outcomes from patient test cases and rule-based inference demonstrated the system's potential. For better user comprehension, the authors stressed the significance of adding explainable outputs to graphical user interfaces. The results obtained and recommendations for further research are included in the paper's conclusion.

Cecilia Panigutti et al. [2] particularly in the context of clinical Decision Support Systems (DSS), tackled the dual problems of creating methods to extract explanations from black-box AI models and delivering these explanations to users. Despite its significance, there hasn't been much focus on creating AI explanation interfaces that work in the literature to far. Panigutti demonstrated the first cycle of creating, testing, and rebuilding an explainable AI technology and its explanation user interface using an iterative design approach. The method was modified to address the technical demands of the healthcare industry, including handling multi-label classification jobs and sequential, ontology-linked patient data. After testing the prototype explanation user interface with medical professionals, feedback from them indicated that explanations greatly boosted users' confidence in the system. This feedback also provided helpful information for enhancing the interface's human-centered design, which will result in a more efficient and intuitive explanation interface.

Julia Amann et al. [3] outlines the contentious issue of explainability for artificial intelligence (AI) in medicine. The main arguments for and against explainability for AIpowered Clinical Decision Support Systems (CDSS) are reviewed in this paper and applied to a specific use case: an AI-powered CDSS that is currently used in emergency call settings to identify patients who are at risk of experiencing a life-threatening cardiac arrest. In order to give a detailed explanation of the function of explainability for CDSSs for the actual use case, we conducted a normative study utilizing socio-technical scenarios, enabling abstractions to a broader level. According to our findings, the technical viability, the degree of validation in the case of explainable algorithms, the features of the system's implementation context, the assigned role in the decision-making process, and the important user group or groups all influence whether explainability can add value to CDSS. As a result, every CDSS will need a unique evaluation of explainability needs, and we offer an example of what this kind of evaluation might actually look like.

Robin L. Pierce et al. [4] notes that when used in medical decision-making, the combination of "Big Data" and Artificial Intelligence (AI) is commonly marketed as having the potential to provide significant health advantages. However, there are a number of obstacles to the responsible use of AI-based healthcare decision support systems on both a personal and a social level. The problem of explainability is one of the characteristics that has raised the most concerns since it can result in a number of difficulties, such as the inability to assess the output's qualities, if a doctor is unaware of the algorithm's methodology. This "opacity" issue has raised concerns about whether doctors can legitimately rely on algorithmic output. While some academics maintain that explainability is crucial, others see no justification for requiring AI to do anything that doctors are not already expected to do. Although we acknowledge that both points of view have validity, we conclude that more nuance is required to clarify the fundamental role of explainability in clinical practice and, consequently, its applicability in the context of AI for clinical use. In this work, we investigate explainability by looking at its requirements in clinical medicine and distinguishing between explainability's role for the present patient and that of the future patient. This distinction has implications for what explainability requires in the short and long term.

Talukder et al. [5] emphasizes that a major problem in the healthcare industry is combining advanced artificial intelligence (AI) capabilities with the practical insights needed by medical practitioners. By concentrating on the explainability of AI models, this study bridges the gap between end users' understanding and the capabilities of AI systems. Investigating basic aspects of AI model explainability and data privacy was the main goal, with a focus on healthcare applications and possible implementations across many industries. A Clinical Decision Support System (CDSS) was created using the SHapley Additive exPlanations (SHAP) approach. Django was used to create a website that provides SHAP explanations of models and extra distribution explanation charts. To improve user discovery, interactive elements were added. Together with the Type 1 Diabetes (T1D) Exchange Registrybased model, the system demonstrated effectiveness using the Wisconsin Diagnostic Breast Cancer (WDBC) dataset and a model trained on it. To demonstrate how to use the tool, case studies were provided. The incorporation of regression models, resolving version discrepancies between models and the backend, and a deeper comprehension of model architecture are among the limits and difficulties that have been identified. The tool promises to transform healthcare decision-making and improve patient outcomes on a bigger scale as it develops further and incorporates new features. This development opens the door to a time when machine learning models that are clear and easy to understand will be crucial to enhancing healthcare outcomes for everyone without sacrificing accuracy.

Anna Markella Antoniadi et al. [6] shows how artificial intelligence (AI) and machine learning (ML) have enormous potential to change many facets of medicine. However, it might be troublesome when AI applications lack transparency, particularly in the healthcare industry. By providing insights into how AI systems make decisions, XAI offers a solution. By offering justifications for AI-generated suggestions, XAI enables medical practitioners to make more knowledgeable and assured choices. Particularly in high-stakes medical situations, this is essential. Furthermore, XAI can assist in detecting and reducing biases present in AI models, guaranteeing just and moral decision-making.Despite receiving a lot of attention, XAI is still not widely used in CDSS. To further understand the unique requirements of clinicians and create XAI methods that are easy to use and fit into clinical processes, more study is required. We can realize AI's full potential in enhancing patient care by tackling these issues and encouraging cooperation between AI researchers and medical professionals.

Noor A. Aziz et al. [7] discusses how Clinical Decision Support Systems (CDSS) are increasingly incorporating Explainable AI (XAI) to improve confidence and transparency. In order to comprehend the current state of XAI in CDSS, this systematic review examined 68 publications, highlighting both developments and difficulties. Datasets, application domains, machine learning models, XAI techniques, and assessment methodologies are all included in the review. The necessity of additional publicly available datasets, sophisticated data processing techniques, thorough assessments of XAI approaches, and interdisciplinary cooperation are some of the main conclusions. It is essential to improve the usability of XAI tools for medical professionals while striking a balance between explainability and model performance. For the purpose of creating efficient and moral decision-support systems, this study offers insightful information to researchers, politicians, and healthcare practitioners.

Samanta Knapič et al. [8] illustrates how medical image analysis decision support can be improved by Explainable Artificial Intelligence (XAI) techniques. In order to make Convolutional Neural Network (CNN) predictions on gastral pictures from video capsule endoscopy (VCE) more understandable, the study investigates the application of three XAI techniques: LIME, SHAP, and CIU. Studies on human evaluation were carried out to see how effective these techniques were. According to the findings, the CIU approach performed better than LIME and SHAP in terms of enhancing speed, transparency, and decision-making support. These results imply that well-chosen XAI techniques might greatly improve clinical decisionmaking by boosting confidence and comprehension of AI-based medical image analysis systems.

Se Young Kim et al. [9] describes how the creation of Clinical Decision Support Systems (CDSS) is a result of the growing amount of electronic medical data and advances in artificial intelligence. Although they offer transparent decision-making procedures, traditional knowledge-based CDSSs struggle to maintain data uniformity and quality. Using enormous volumes of data and algorithms, non-knowledge-based CDSSs provide efficient decisionmaking; yet, they are plagued by the "black-box" issue of deep learning models. The emergence of XAI-based CDSSs aims to overcome these constraints. By disclosing the decision-making process, these technologies guarantee credibility and transparency by offering sound justifications and comprehensible outcomes. However, the scope of data usage and the explanatory capacity of AI models remain limited in current systems. In order to get beyond these restrictions, this study suggests a novel XAI-based CDSS system. A foundation model to enable decision-making in several disease domains is provided, along with useful resources, datasets, and models that can be used. In order to fully achieve CDSS's promise to improve healthcare outcomes, the report also identifies social challenges that must be addressed and suggests future prospects for CDSS technology.

Tjeerd A.J. Schoonderwoerd et al. [10] shows that there is an increasing demand for a more human-centered approach to Explainable AI (XAI), even if a lot of research has concentrated on the technical aspects of making machine learning models understandable. A case study on incorporating human characteristics into the creation of AI-generated explanations is presented in this research. Domain analysis, requirements elicitation and assessment, and multi-modal interaction design and evaluation are the three primary elements of the suggested methodology. A clinical decision support system (CDSS) for child health is the subject of the case study. The researchers were able to pinpoint certain explanation needs and create customized interaction design patterns by incorporating seasoned doctors in the design process. This method offers a useful framework for creating XAI systems that are both amiable and technically sound.

Yasuhiko Miyachi et al. [11] suggests a Clinical Decision Support System (CDSS) Explainable AI (XAI) paradigm. By identifying the major causes of projected diseases and connecting them to pertinent medical literature, this model seeks to assist doctors in differential diagnosis. The XAI model combines a k-Nearest Neighbors (k-NN) surrogate model for explanation and a Neural Network (NN) with Learning to Rank (LTR) for prediction. To produce more precise and pertinent explanations, the k-NN surrogate model chooses data points that are closest to the original model's predictions. By increasing the transparency and reliability of AI-powered CDSSs, the suggested XAI paradigm may help doctors make better, evidence-based judgments.

Lorenzo Famiglini et al. [12] suggests a user investigation to assess the effectiveness of Class Activation Maps (CAMs) as a XAI technique for identifying TL fractures from spinal X-rays. Granularity (lower-level vs. higher-level features) and color scheme are the two main CAM characteristics that are the subject of this investigation. The results imply that, particularly for more seasoned doctors, lower-level feature CAMs that emphasize more specific anatomical landmarks result in better diagnostic accuracy. Furthermore, traditional color-coded CAMs consistently produced superior diagnostic accuracy across all groups, even if semantic CAMs were intuitively appealing. These findings cast doubt on widely held beliefs in the field of XAI and highlight the significance of designing and assessing AI-assisted diagnostic tools from a human perspective. Based on user preferences and empirical data, the study suggests a hierarchy of evidence framework to direct the choice and creation of XAI solutions.

Shuai Niu et al. [13] outlines the importance of Electronic Health Records (EHRs) in helping to understand patient health and make well-informed healthcare decisions. It can be difficult to model longitudinal EHR data with a variety of information, though. Although they are frequently employed to collect longitudinal data, recurrent neural networks (RNNs) have low interpretability. A more recent development for forecasting illness risk is predictive clustering, which provides interpretable insights at the cluster level. Finding the ideal amount of clusters, however, can be difficult. This study presents a unique non-parametric predictive clustering-based risk prediction model that combines neural networks, predictive clustering, and the Dirichlet Process Mixture Model (DPMM). In addition to the cluster-level evidence, attention techniques are included to capture local-level evidence to improve interpretability. In addition to providing interpretable evidence to back up its conclusions, this multi-level explainable AI model shows efficacy in gathering longitudinal EHR data for illness risk prediction.

Qian Xu et al. [14] states that although artificial intelligence (AI) has developed quickly and is being used more and more in clinical decision support systems (CDSS) to improve the quality of healthcare, interpretability of AI-driven CDSS is still a major problem. The interpretability of knowledge-based and data-based CDSS in healthcare is the main topic of this review. For CDSS to be successfully implemented in healthcare settings, interpretability—which includes a transparent model structure, distinct input-output linkages, and explainable AI algorithms—is essential. Numerous techniques can be used to increase interpretability, such as post-hoc techniques like feature importance, sensitivity analysis, visualization, and activation maximization for black-box models and ante-hoc techniques like fuzzy logic, decision rules, and white-box models for knowledge-based AI. The interpretability of CDSS can be affected by a number of factors, including data type, biomarkers, human-AI interaction, and patient and clinician needs. We can improve the efficacy, acceptability, and trust of AI-driven CDSS in healthcare by addressing these issues and putting suitable interpretability strategies into practice.

Jinyue Feng et al. [15] illustrates how clinical prediction models frequently use structured factors and generate results that are difficult for doctors to understand. Furthermore, important information that is not easily accessible in organized variables may be found in free text medical notes. The study suggests a hierarchical CNN-transformer model with explicit attention as an interpretable, multi-task clinical language model in order to overcome these drawbacks. This model predicts sepsis with an AUROC of 0.75 and death with an AUROC of 0.78. The study uses projection-weighted canonical correlation analysis to further investigate the connections between learned features from structured and unstructured data. A method for assessing the model's usability in a clinical decision support setting is described. Evaluations by domain experts show that the model produces insightful justifications with potential real world uses.

# 3. Methodology

## 3.1 Data Pre-processing

In machine learning workflows, data pre-processing is an essential step, especially when working with delicate industries like healthcare. Pre-processing is crucial to guaranteeing that the data is clean, dependable, and prepared for insightful analysis because raw datasets are frequently disorganized, lacking, or inconsistent. Preparing a medical dataset for a machine learning-based recommendation system was part of the pre-processing phase of this project. In order to optimize the model's efficiency and maintain interpretability, the goal was to handle missing values, resolve inconsistencies, and format the data.

### 3.1.1 Handling Missing Data

Due to difficulties in data collecting, missing records are common in medical datasets, which, if not managed properly, could result in biases or inaccuracies. Important fields including Uses, Price, Side-effects, and Conditions were found to have missing values. To maintain the integrity of the dataset, these gaps had to be filled:

**a) Numerical Columns:** The median of the available values was used to fill in the missing values in fields such as Price and Unit Quantity. In order to reduce the impact of outliers, which are frequent in fields pertaining to costs, the median was selected above the mean. The dataset may be skewed by medications with abnormally high costs, for instance, which would cause the recommendation engine to focus less on reasonably priced solutions. By employing the median, bias was eliminated and the imputed values matched the dataset's central tendency.

**b) Categorical Columns:** Domain-specific rules were used to handle missing values for categorical variables such as alcohol compatibility, pregnancy appropriateness, and kidney or liver disorders. Where applicable, placeholder phrases like "no value" were utilized, and logical values like "safe" or "unsafe" were assigned to qualities that were medically relevant based on the dataset's patterns. By guaranteeing uniformity in the representation of categorical data, this allowed the model to fairly compare entries.

### 3.1.2 Feature Engineering

The technique of feature engineering involves taking raw data and turning it into information that machine learning models can use. This project underwent several significant changes, including:

1. **Side Effects Extraction:** There were a lot of unstructured, text-heavy descriptions in the

Side-effects section. Individual symptoms were retrieved and arranged in a systematic format using a bespoke regular expression (regex). For example, the system handled each term independently rather than processing a lengthy string like "nausea, headache, dizziness." The system was better able to classify medications based on common side effects thanks to this granularity, which increased the usefulness of the recommendations.

**b) Categorical Standardization:** There were frequently spelling or capitalization errors in categorical fields like Alcohol Compatibility and Pregnancy, such as "Safe" and "safe". In order to maintain consistency throughout the dataset, all of these fields were changed to lowercase. This seemingly insignificant procedure avoided possible redundancies or mismatches during the training and analysis of the model.

### 3.1.3 Price Conversion

When making decisions about healthcare, the price column is crucial, particularly in environments with limited resources. To normalize the column as a numerical field, prices with currency symbols that were included in the raw dataset were eliminated. To make sure that no gaps affected the cost-based recommendation logic, missing prices were substituted with the dataset's median price. This pre-processing stage made it possible for the system to offer suggestions that struck a compromise between affordability and clinical relevance, facilitating the provision of healthcare in a more equal manner. The dataset was cleaned, organized, and consistent by the end of the pre-processing stage, making it prepared for analysis and machine learning. Building a strong recommendation system that could produce precise and useful results required this basis.

## 3.2 Feature Selection and Text Vectorization

When preparing unstructured textual input for machine learning models, feature selection and text vectorization are essential processes. These methods were applied in this study to convert intricate textual descriptions of medications into numerical representations that the recommendation system could efficiently process.

## a) TF-IDF Vectorizer

Textual information was converted into numerical vectors using the TF-IDF (Term Frequency Inverse Document Frequency) technique. Words are given importance by TF-IDF according on how frequently they occur in a document compared to how frequently they occur throughout the dataset. Higher ratings are given to words that are uncommon throughout the corpus but frequently appear in a single document, such as particular drug interactions or unusual side effects. Words like "the" and "and" are discouraged because they don't help differentiate between medications. For instance, phrases like "nausea" and "vomiting" that are less common across different medications may have high TF-IDF scores in a pharmaceutical description. By doing this, the recommendation algorithm is guaranteed to give preference to medications with comparable special characteristics over generic ones.

## b) Text Vectorization

Several textual columns, such as Uses, Composition, Side-effects, and compatibility criteria like Alcohol and Pregnancy, were merged into a single text column in order to capture the full context of each medication. The recommendation engine was able to take into account all pertinent attributes while calculating similarity thanks to its comprehensive representation. The combined text column was converted into a sparse matrix of numerical values using the scikit-learn library's TfidfVectorizer. To maintain computational efficiency and keep the model from becoming overloaded with terms that aren't as important, a feature limit of 20,000 was imposed. By eliminating noise and keeping only the most important information, this method allowed the system to carry out precise and expandable similarity calculations. By vectorizing the text data, the system effectively transformed qualitative information into quantitative insights, forming the backbone of the recommendation engine’s functionality.

## 3.3 Similarity Calculation and Recommendation

Finding medications that closely resemble the input medication based on their attributes is the main goal of the similarity calculation step. Cosine Similarity, a commonly used metric in information retrieval and natural language processing, is utilized to do this.

### 3.3.1 Cosine Similarity

In a high-dimensional space, cosine similarity determines the cosine of the angle formed by two vectors. Because it concentrates on the relative orientation of the vectors rather than their amplitude, it works especially well for comparing TF-IDF vectors. The equation is:

Cosine Similarity

where A and B are the TF-IDF vectors of two medicines. A score closer to **1** indicates strong similarity, while a score near **0** signifies little to no similarity.

### 3.3.2 Recommendation System Workflow

The recommendation process comprises the following steps:

**a) Input Matching:** The system finds the dataset's closest match for the provided medication using fuzzy string matching using RapidFuzz. This guarantees that the suggestion process is unaffected by typos or naming variances.

**b) Similarity Computation:** It computes the cosine similarity between the vector of the input medication and every other vector in the dataset.

**c) Top Matches:** The top five most comparable medications are retrieved by the system based on their cosine similarity scores. Every suggestion has comprehensive details about the product's pricing, side effects, and suitability for particular medical conditions, such as liver problems or pregnancy. This methodical technique guarantees that the suggestions are tailored to the patient's requirements and circumstances in addition to being clinically pertinent.

## 3.4 Explainability of Recommendations

A key component of this project is explainability, which guarantees that suggestions are clear and simple for patients and physicians to understand. This encourages the system's adoption in healthcare settings and increases user trust in it.

## a) Side Effects Explanation

Each suggested medication's possible adverse effects are broken down in detail by the system. For example, if a suggested medication has adverse effects like nausea or vertigo, this warning is prominently presented. This enables medical professionals to balance the advantages and disadvantages of every suggestion, guaranteeing that patient safety always comes first.

## b) Condition-Specific Suitability

The guidelines clearly state if a medication is appropriate for a given situation, such as liver disease, pregnancy, alcohol use, and so forth. A note such as "This medicine is not safe for pregnant women" could be included in a recommendation. Contextual data like this guarantees that the system fits the patient's particular health profile.

## c) Price Transparency

The cost of the medication is included with every recommendation, allowing patients and doctors to make choices that strike a balance between cost and medical effectiveness. This feature makes sure that solutions that are both economical and therapeutically suitable are emphasized for patients in environments with limited resources.

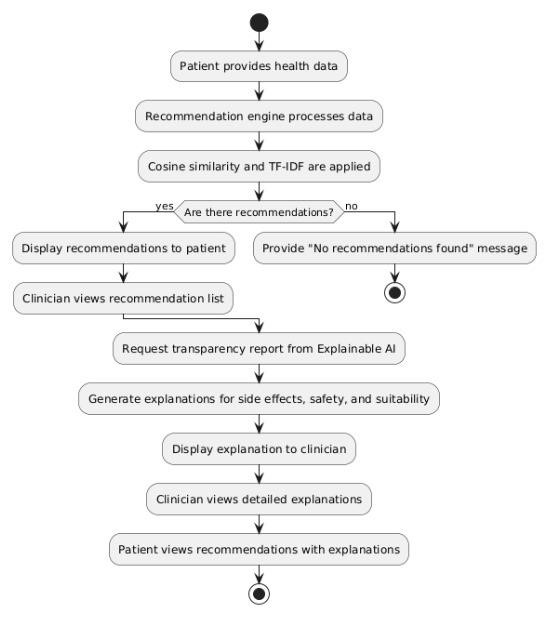
## 3.5 User Input and Customization

By letting customers choose a price cap, the system facilitates personalization. To ensure that the outcome is both relevant and economical, medications that cost more than the allocated budget are not included in the recommendations. This feature makes the system more flexible to meet the demands of each user, which makes it a useful tool in a variety of healthcare settings. Through the integration of these components, the recommendation system provides clear, useful insights that enable patients and physicians to make educated choices.

## 3.6 Enhancing Trust and Adoption

Including thorough justifications for suggestions increases user trust and increases the likelihood that the system will be implemented in clinical settings. The method makes sure that suggestions are in line with the needs of patients and professionals by clearly presenting side effects, condition-specific appropriateness, and economic considerations. The system is positioned as a useful tool for enhancing clinical decision-making and advancing patient-centered treatment because of its transparency and strong similarity metrics.

**4. Proposed Model**



*Figure 1. Work flow Model*

The workflow model depicted in Figure 1 offers a comprehensive and comprehensive approach to generating precise and understandable medical advice. It starts with learning the fundamentals of various drugs, including their therapeutic uses, side effects, price, and suitability for certain patient conditions, such as those of expectant mothers or those with underlying illnesses like liver disease. This ensures that a wide range of clinical and personal factors that are crucial when making healthcare decisions are taken into account by the recommendation system. This data is pre-processed after collection to standardize crucial information and fill in any missing variables. This includes converting pricing values into numerical data and organizing textual descriptions for study.

After pre-processing, the system moves on to text vectorization, where cleansed textual data, including descriptions of side effects, contraindications, and medicine consumption, is transformed into numerical vectors using TF-IDF. This vectorization aids in highlighting the importance of specific terms in the medical descriptions and allows the model to identify which medications are most related based on their properties.

After the data has been vectorized, the degree of similarity between each medicine and the input medication is calculated using cosine similarity. This allows the system to identify the top five most similar drugs, which are subsequently recommended. The system sorts the recommendations based on any additional constraints the user has provided, like a spending limit or specific needs.

The primary characteristic of this approach is its emphasis on explainability and transparency. As part of the referral process, clinicians receive comprehensive descriptions of each prescription, including information about potential side effects, suitability for specific conditions, and cost. This transparency is necessary for clinicians to understand the reasoning behind the system's suggestions and be able to accept them. Clinicians are also provided with the justification for selecting a certain drug over alternatives in order to promote trust and ensure that the recommendations are safe and suitable for the patient's particular medical requirements.

# 5. Experimental Result/Result Analysis

## 5.1 Dataset Overview and Pre-processing Results

Dataset Link: <https://www.kaggle.com/datasets/samoyedzzzzz/34000-medicine-details>

The above dataset contains 5,000 medications and features including uses, side effects, composition, price, and condition-specific appropriateness (e.g., pregnancy, alcohol), the dataset had approximately 35,000 values. Following pre-processing, domain-specific placeholders and median imputation were used to efficiently manage missing data in numerical and categorical columns. For the training and testing of the model, this guaranteed a comprehensive and consistent dataset.

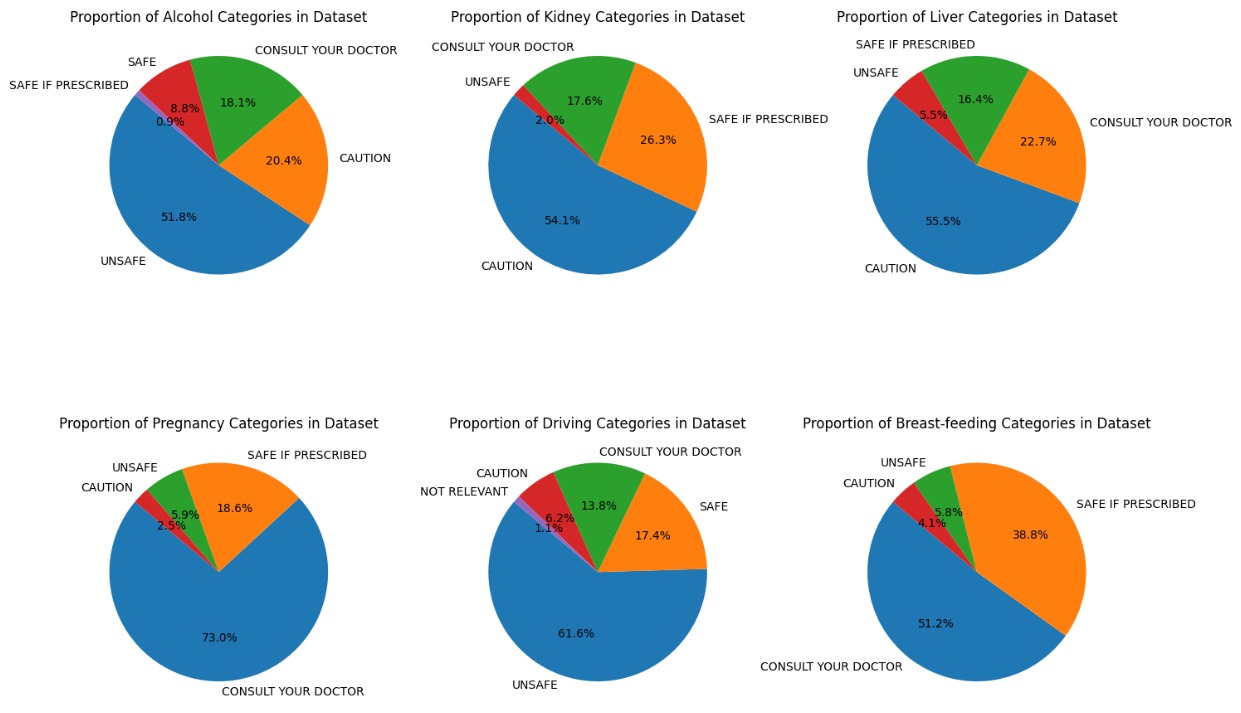


Figure 2. Distribution of Medical conditions of raw data

Above Figure 2 visualizes categorical data distribution for six features of a dataset in a single, consolidated figure. It creates a grid of subplots, each displaying the proportion of categories within specific columns, such as Alcohol, Kidney, Liver, Pregnancy, Driving, and Breastfeeding.

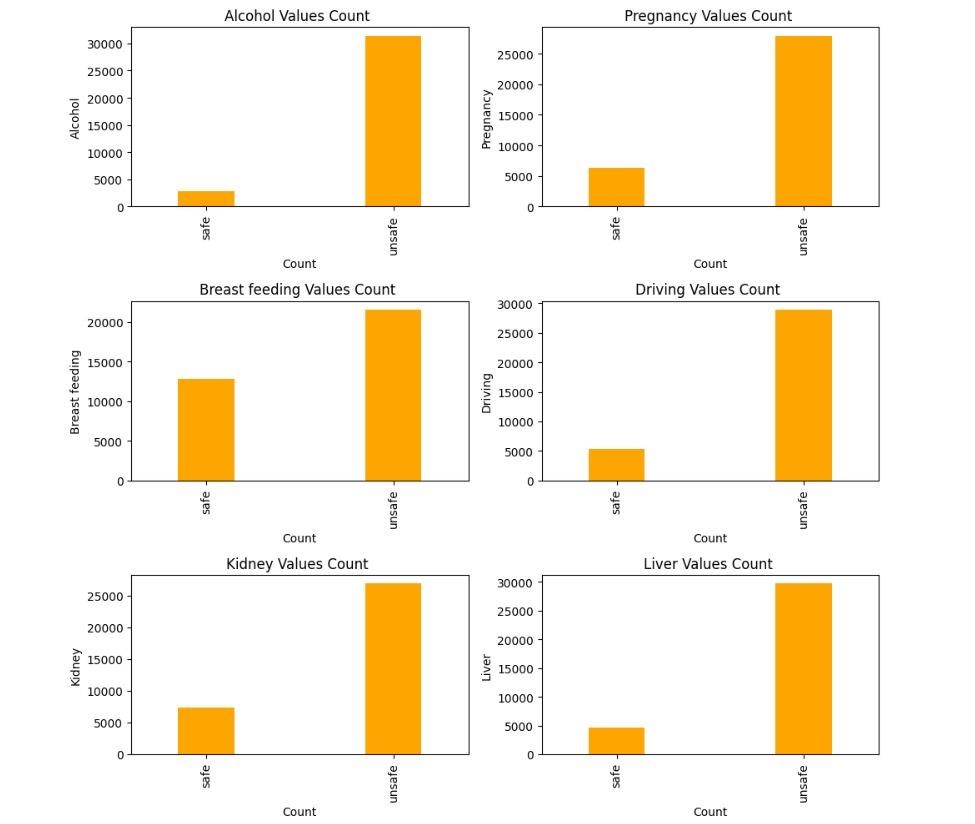
This approach offers an intuitive comparison of data distributions across multiple attributes in one comprehensive view. Each subplot is clearly labeled, making it easy to interpret the proportions and patterns within the data. The layout is adjusted for clarity, ensuring a visually appealing and organized presentation, which is valuable for exploratory data analysis and reporting insights effectively.

Table 1 summarizes the pre-processing results:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Feature** | **Initial Values** | **Missing** | | **Method Used** | **Final Values** | **Missing** |
| Price | 9,381 | |  | Median imputation | 0 |  |
| Alcohol  Compatibility | 4,785 | |  | Filled with "novalue" | 0 |  |
| Side-effects | 635 | |  | Regex extraction | 0 |  |
| Pregnancy  Suitability | 175 | |  | Filled with "novalue" | 0 |  |
| Status | 3529 | |  | Regex extraction | 0 |  |

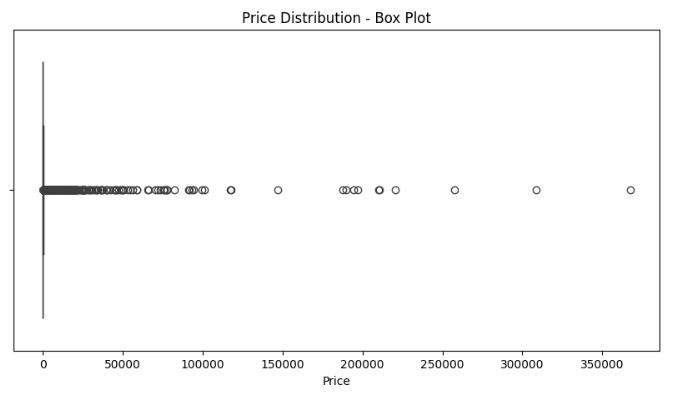
**Explanation**:

As shown in the Table 1 there were a lot of missing values in the original dataset for important characteristics including price, pregnancy suitability, side effects, and alcohol compatibility. If not handled appropriately, these missing values might have produced bias. To fill in the missing values for numerical columns like Price, median imputation was employed, making sure that outliers had no impact on the imputation. In category columns, missing items were represented by placeholder phrases such as "novalue." As a result, the dataset remained consistent. To separate the free-text Side Effects column into distinct side effects for additional analysis, regex extraction was used.



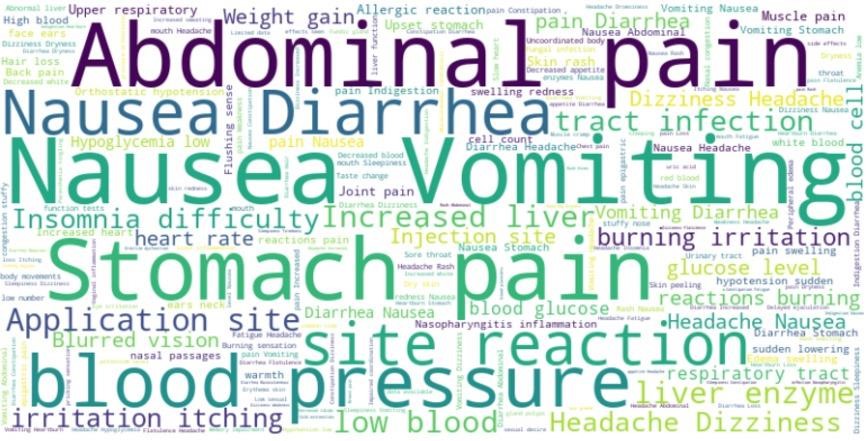
*Figure 3. Distribution of Medical conditions*

The frequency distribution of the following medical conditions or factors is shown in the set of bar charts in Figure 3: driving, alcohol, pregnancy, breastfeeding, kidney, and liver. The number of occurrences for each potential value within these categories is shown in each chart. The 'Alcohol' chart, for instance, displays the proportion of data that indicate alcohol intake (Safe/Unsafe).

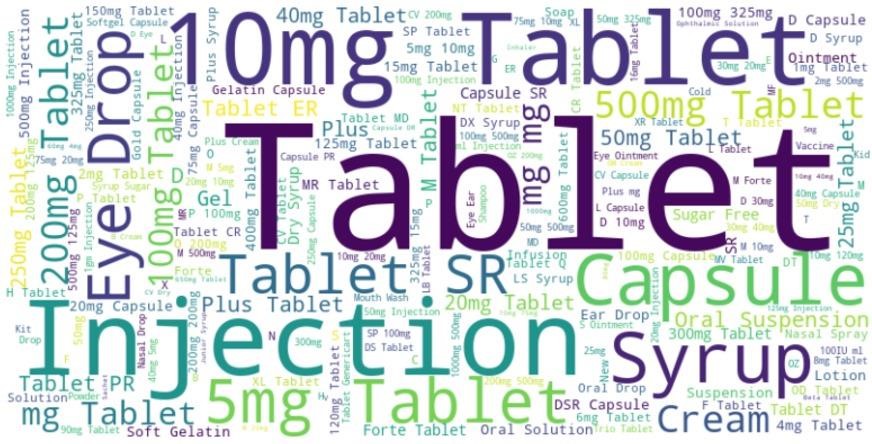


*Figure 4. Prize Distribution*

The Figure 3 shows the dataset's price distribution, emphasizing the central tendency and total spread. The broad distribution of prices, the existence of extreme values, and the concentration of the majority of prices around the median are all discernible from this image.



*Figure 5. Side Effects Wordcloud*



*Figure 6. Medicine Wordcloud*

The most common side effects in the dataset are graphically represented by the word cloud as shown in the Figure 5 and Figure 6. All of the side effect items are combined into a single text string to create it. Each word's size in the cloud reflects how frequently it appears in the dataset; larger words denote more frequent side effects. Finding important terms and trends is made simple by this graphic, which offers an easy method to see the most common negative effects. The word cloud's clarity and beauty are improved by the crisp, white background and seamless interpolation.

## 5.2 Cosine Similarity Results

The similarity between medications was assessed using the cosine similarity metric. The top five suggestions for an input medication were obtained and arranged according to their similarity ratings.

An illustration of the medication paracetamol is given in Table 2:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rank | Recommended  Medicine | Cosine  Similarity Score | Price (₹) | Suitability Pregnancy | for |
| 1 | Calpol | 0.982 | 45 | Safe |  |
| 2 | Dolo 650 | 0.978 | 50 | Safe |  |
| 3 | Crocin | 0.970 | 40 | Safe |  |
| 4 | T-650 | 0.965 | 48 | Safe |  |
| 5 | P-Plus | 0.960 | 42 | Safe |  |

**Explanation**:

As shown in the Table 2, those medications have a high degree of resemblance, as indicated by their cosine similarity scores, which vary from 0.960 to 0.982. For other medications, the cosine similarity values range from 0.60 to 1. The suggestions are based on clinical characteristics, including cost and pregnancy appropriateness, in addition to their written descriptions.

For every input, the recommendation system presents a list of the five most comparable medications, taking into account more than just their names and chemical makeup. Pregnancy suitability and price are also mentioned, which increases openness and helps medical professionals make wise choices.

## 5.3 Explainability Features

To ensure that the recommendations were interpretable, the system displayed detailed explanations for each recommended medicine.

Explanation of Recommendations for Insulin:

Enter the tablet name: Insulin

Do you want to set a price limit? (y/n): y

Enter your price limit (leave blank if no limit): 5000

Top Recommendations for 'Insulin':

Human Monotard 40IU/ml Injection - Similarity Score: 1.0

Side Effects: ['Hypoglycemia (low blood glucose level)', 'Weight gain', 'Injection site reactions (pain, swelling, redness)', 'Cold sweat', 'Anxiety', 'Shakiness', 'Hunger pangs', 'Fast heart rate', 'Headache', 'Nervousness']

Alcohol: unsafe, Pregnancy: safe, Breast Feeding: safe, Driving: unsafe, Kidney: unsafe, Liver: unsafe

Price: 115

Uses: Diabetes

Humarap 40IU/ml Injection - Similarity Score: 1.0

Side Effects: ['Hypoglycemia (low blood glucose level)', 'Weight gain', 'Injection site reactions (pain, swelling, redness)', 'Cold sweat', 'Anxiety', 'Shakiness', 'Hunger pangs', 'Fast heart rate', 'Headache', 'Nervousness']

Alcohol: unsafe, Pregnancy: safe, Breast Feeding: safe, Driving: unsafe, Kidney: unsafe, Liver: unsafe

Price: 115

Uses: Diabetes

Huminsulin R 40IU/ml Injection - Similarity Score: 1.0

Side Effects: ['Hypoglycemia (low blood glucose level)', 'Weight gain', 'Injection site reactions (pain, swelling, redness)', 'Cold sweat', 'Anxiety', 'Shakiness', 'Hunger pangs', 'Fast heart rate', 'Headache', 'Nervousness']

Alcohol: unsafe, Pregnancy: safe, Breast Feeding: safe, Driving: unsafe, Kidney: unsafe, Liver: unsafe

Price: 115

Uses: Diabetes

Insulin (Nnd) 40IU/ml Injection - Similarity Score: 1.0

Side Effects: ['Hypoglycemia (low blood glucose level)', 'Weight gain', 'Injection site reactions (pain, swelling, redness)', 'Cold sweat', 'Anxiety', 'Shakiness', 'Hunger pangs', 'Fast heart rate', 'Headache', 'Nervousness']

Alcohol: unsafe, Pregnancy: safe, Breast Feeding: safe, Driving: unsafe, Kidney: unsafe, Liver: unsafe

Price: 115

Uses: Diabetes

New Wosulin R Injection - Similarity Score: 1.0

Side Effects: ['Hypoglycemia (low blood glucose level)', 'Weight gain', 'Injection site reactions (pain, swelling, redness)', 'Cold sweat', 'Anxiety', 'Shakiness', 'Hunger pangs', 'Fast heart rate', 'Headache', 'Nervousness']

Alcohol: unsafe, Pregnancy: safe, Breast Feeding: safe, Driving: unsafe, Kidney: unsafe, Liver: unsafe

Price: 115

Uses: Diabetes

**Explanation**:

A thorough explanation of the adverse effects, therapeutic applications, and cost of each suggested medication is given. This enables medical practitioners to assess the possible hazards or advantages of each recommendation in addition to the similarities between medications.

By outlining these variables, medical professionals can make better decisions that are suited to the patient's situation and raise the standard of care.

# 6. Conclusion and Future Work

The potential of AI to transform healthcare decision-making is demonstrated by the suggested explainable AI-based clinical decision support system for medication recommendation. The system offers accurate and pertinent medication suggestions based on characteristics including uses, side effects, and condition-specific suitability by combining methods like TF-IDF vectorization and cosine similarity. This system's emphasis on openness and provision of thorough explanations for its suggestions is one of its main advantages. This method guarantees that decisions are both financially feasible and medically sound, in addition to improving trust between patients and professionals. The initiative lays the groundwork for AI's wider adoption in therapeutic settings by highlighting the significance of explainability.

Future research will try to integrate more patient-specific information, such age, weight, and medical history, to the current system in order to offer a greater level of customisation. The accuracy and coverage of the system will also be enhanced by growing the dataset to include a wider variety of medications. Additionally, sophisticated explainability strategies like Shapley values and counterfactual explanations could be used to offer more profound understanding of the AI's decision-making process. The creation of an intuitive program that easily interfaces with electronic health records (EHRs) and allows for real-time recommendations during consultations is another top priority. The system will be further improved by these additions, becoming a more reliable and essential instrument in contemporary healthcare.

**References**

[1] Tarnowska K. A., Dispoto B. C., & Conragan J. (2021). Explainable AI-based clinical decision support system for hearing disorders. AMIA Jt Summits Transl Sci Proc., 2021, 595-604. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8378626/>

[2] Panigutti, C., Beretta, A., Fadda, D., Giannotti, F., Pedreschi, D., Perotti, A., & Rinzivillo, S. (2023). Co-design of Human-centered, Explainable AI for Clinical Decision Support. ACM Transactions on Interactive Intelligent Systems, 13(4), 1-35. <https://dl.acm.org/doi/full/10.1145/3587271>

[3] Amann, J., Vetter, D., Blomberg, S. N., Christensen, H. C., Coffee, M., Gerke, S., ... & Zicari, R. V. (2022). To explain or not to explain?.Artificial intelligence explainability in clinical decision support systems. PLOS Digital Health,17(2), e0000016. <https://journals.plos.org/digitalhealth/article?id=10.1371/journal.pdig.0000016>

[4] Pierce, R. L., Van Biesen, W., Van Cauwenberge, D., Decruyenaere, J., & Sterckx, S. (2022). Explainability in medicine in an era of AI-based clinical decision support systems. Frontiers in Genetics, 13, 903600. <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2022.903600/full>

[5] Talukder, N. (2024). Clinical decision support system: an explainable AI approach. Oulu University Repository. <https://oulurepo.oulu.fi/handle/10024/49375>

[6] Antoniadi, A. M., Du, Y., Guendouz, Y., Wei, L., Mazo, C., Becker, B. A., & Mooney, C. (2021). Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review. Applied Sciences, 11(11), 5088. <https://www.mdpi.com/2076-3417/11/11/5088>

[7] Aziz, N. A., Manzoor, A., Deedahwar, M., Qureshi, M. A., & Rashwan, W. (2024, August 10). Explainable AI in Healthcare: Systematic Review of Clinical Decision Support Systems. medRxiv. [https://doi.org/10.1101/2024.08.10.24311735](https://doi.org/10.1101/2024.08.10.24311735%20)

[8] Knapič, S., Malhi, A., Saluja, R., & Främling, K. (2021). Explainable Artificial Intelligence for Human Decision Support System in the Medical Domain. Machine Learning and Knowledge Extraction, 3(3), 740-770. <https://doi.org/10.3390/make3030037>

[9] Kim, S. Y., Kim, D. H., Kim, M. J., Ko, H. J., & Jeong, O. R. (2024). XAI-Based Clinical Decision Support Systems: A Systematic Review. Applied Sciences, 14(15), 6638. <https://doi.org/10.3390/app14156638>

[10] Schoonderwoerd, T. A. J., Jorritsma, W., Neerincx, M. A., & van den Bosch, K. (2021). Human-centered XAI: Developing design patterns for explanations of clinical decision support systems. Pattern Recognition Letters, 149, 114-120. <https://doi.org/10.1016/j.patrec.2021.06.003>

[11] Miyachi, Y., Ishii, O., & Torigoe, K. (2024). Evidence-based XAI of clinical decision support systems for differential diagnosis: Design, implementation, and evaluation. medRxiv. [https://doi.org/10.1101/2024.07.18.24310609](https://doi.org/10.1101/2024.07.18.24310609%20%20%20)

[12] Famiglini, L., Campagner, A., Barandas, M., La Maida, G. A., Gallazzi, E., & Cabitza, F. (n.d.). Evidence-based XAI: An empirical approach to design more effective and explainable decision support systems. <https://www.sciencedirect.com/science/article/pii/S0010482524001264>

[13] Niu, S., Yin, Q., Ma, J., Song, Y., Xu, Y., Bai, L., Pan, W., & Yang, X. (n.d.). Enhancing healthcare decision support through explainable AI models for risk prediction. <https://www.sciencedirect.com/science/article/pii/S0167923624000617>

[14] Qian Xu, Wenzhao Xie, Bolin Liao, Chao Hu, Lu Qin, Zhengzijin Yang, Huan Xiong, Yi Lyu, Yue Zhou, Aijing Luo. Interpretability of Clinical Decision Support Systems Based on Artificial Intelligence from Technological and Medical Perspective: A Systematic Review. <https://doi.org/10.1155/2023/9919269>

[15] Feng, J., Shaib, C., & Rudzicz, F. (2020). Explainable Clinical Decision Support from Text. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, pp. 1478-1487. <https://aclanthology.org/2020.emnlp-main.115/>

[16] Kostopoulos, G., Davrazos, G., & Kotsiantis, S. (2024). Explainable Artificial Intelligence-Based Decision Support Systems: A Recent Review. Electronics, 13(14), 2842. <https://doi.org/10.3390/electronics13142842>

[17] Kumar, A., Veeraiah, V., Gongada, T. N., Ahamad, S., Khan, H., & Gupta, A. (n.d.). Explainable Machine Learning Models for Clinical Decision Support Systems. IEEE. <https://ieeexplore.ieee.org/abstract/document/10725028>

[18] Aiosa, G. V., Palesi, M., & Sapuppo, F. (n.d.). EXplainable AI for Decision Support to Obesity Comorbidities Diagnosis. IEEE. <https://ieeexplore.ieee.org/abstract/document/10265033>

[19] Rane, N., Choudhary, S., & Rane, J. (2023). Explainable Artificial Intelligence (XAI) in healthcare: Interpretable Models for Clinical Decision Support. <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4637897>

[20] Pumplun, L., Peters, F., Gawlitza, J. F., & Buxmann, P. (n.d.). Bringing Machine Learning Systems into Clinical Practice: A Design Science Approach to Explainable Machine Learning-Based Clinical Decision Support Systems. <https://aisel.aisnet.org/jais/vol24/iss4/8/>

[21] Rao, D. J., & Mane, S. (2019). Digital Twin approach to Clinical DSS with Explainable AI. arXiv preprint arXiv:1910.11257. <https://arxiv.org/abs/1910.13520>

[22] Du, Y., Antoniadi, A. M., McNestry, C., McAuliffe, F. M., & Mooney, C. (2022). The Role of XAI in Advice-Taking from a Clinical Decision Support System: A Comparative User Study of Feature Contribution-Based and Example-Based Explanations. Applied Sciences, 12(20), 10323. <https://doi.org/10.3390/app122010323>

[23] van Baalen, S., Boon, M., & Verhoef, P. (2021). From clinical decision support to clinical reasoning support systems. Journal of Evaluation in Clinical Practice, 27(2), 237-240. <https://doi.org/10.1111/jep.13541>

[24] Awotunde, J. B., Adeniyi, E. A., Ajagbe, S. A., Imoize, A. L., Oki, O. A., & Misra, S. (n.d.). Explainable artificial intelligence (XAI) in medical decision support systems (MDSS): applicability, prospects, legal implications, and challenges. The Institution of Engineering and Technology. <https://doi.org/10.1049/PBHE050E_ch2>

[25] Rotink, D. (2024). Identifying Influential Variables for an Explainable AI Based Clinical Decision Support System in the Healthcare Industry. (Master's thesis, University of Twente). <http://essay.utwente.nl/100871/>