# 1. Project Overview

Project Title**:** Machine Learning-Based Medication Prescription Support System

## Summary:

This paper presents the development of a Machine Learning-Based Medication Prescription Support System designed to assist healthcare professionals in making accurate medication recommendations. The system will be trained on a Kaggle dataset, and various machine learning models will be evaluated based on their accuracy, precision, recall, and F1-score. Additionally, it will assess the performance of ensemble methods in comparison to individual classifiers and analyze model interpretability to ensure transparency and trust in decision-making.

Children with special health care needs, in most cases, can easily be treated with a combination of medication and parental care. However, the requirements involved can be pushed further when the parent caring for the child must also treat an illness while managing their child's care. In many instances, this leads to an inability of the parent to manage their own illness while also trying to manage their child's health (J. Rayle et al., 2022). Through various machine learning algorithms, it was found that there was a fair amount of accuracy when attempting to predict the overall intrusiveness of illness care for parents. However, the results show that the algorithms analyzed could accurately predict this parental intrusiveness of care within a small margin of error (Rayle et al., 2022). These findings highlight the potential of AI-driven predictive models in tackling complex healthcare challenges and improving patient outcomes through advanced data analysis.

Electronic Health Records (EHRs) provide valuable patient history data that can train predictive models to enhance patient care. A study on crisis prediction in individuals with depression found that combining structured and unstructured EHRs improved model accuracy. Among various approaches, a Long Short-Term Memory (LSTM) network trained on key features from a random forest model performed best, achieving a mean AUC of 0.901 on training data and 0.810 on a test set. These results suggest that predictive models could be integrated into clinical decision-support tools for mental healthcare (Msosa et al., 2023).

With the rapid development of artificial intelligence and big data analysis, the medical and healthcare field is undergoing a revolutionary change. Semi-supervised learning, as a machine learning method that effectively utilizes a large amount of un-labeled data, provides a new solution for health big data analysis. Traditional methods cannot effectively resolve issues related to inaccurate health data, making semi-supervised learning an effective approach to enhancing data quality and analysis (Ma, 2024).

Predictive analytics, a subfield of data analytics, enables future predictions about unknown events. It employs various machine learning algorithms, artificial intelligence, data mining, and statistical models to extract insights from historical data. In healthcare, past disease symptoms records can be used to predict diseases, and historical epidemic data can help forecast future pandemic trends. Machine learning techniques in predictive analytics can be categorized into classification and regression models. Classification algorithms classify discrete values, such as patient demographics, while regression algorithms predict continuous values, such as forecasting a person's income based on previous earnings. Studies have shown that machine learning algorithms for predictive analysis are effective in identifying patterns from large datasets and predicting underlying medical events (Bansal et al., 2021).

## Research Question:

* *Which machine learning model performs best in terms of accuracy, precision, recall, and F1-score for medication recommendations?*
* *How does the performance of ensemble methods compare to individual classifiers in predicting medication recommendations?*
* *How interpretable are the results of each model, and which model provides the most transparent decision-making process for healthcare professionals?*

## Objectives:

* Choose the best machine learning model that performs nice with standard evaluation metrics.
* Contrast the predictive performance of ensemble methods to that of the individual classifiers.
* Shapley Additive explanations, and LIME, for Local Interpretable Model-agnostic Explanations, to check the model interpretability.
* Select some amount of the dataset to test the system concerning its reliability and effectiveness.

# 2. Project Plan: Task List and Project Timeline

## Task List and Timeline

1. **Literature Review**(Feb 13 – Feb 25)

* Literature review on the applications of machine learning in medical prescription support systems.

1. **Methodology Development**( Feb 26 – Mar 11)

* Define and refine methodology for the choice and evaluation of machine learning models.

1. **Experimentation & Implementation**(Mar 12 – Apr 5)

* Train and test different types of machine learning models on the dataset.

1. **Preliminary Results Analysis**(Apr 6 – Apr 13)

* Compare the performance metrics of the models and choose the best among them..

1. **Final Report Writing**(Apr 13 – Apr 29)

* Document findings, methodology, and results for the final report submission.

1. **Viva Preparation**(May 2 – May 13)

* Prepare for the final viva with a review of key insights and justifications over model selection.

# 3. Data Management Plan

## Dataset overview:

* **Source:** Kaggle
* **Link:** [Medicine Recommendation System Dataset](https://www.kaggle.com/datasets/noorsaeed/medicine-recommendation-system-dataset)
* **License:** Apache 2.0
* **Usability Score:** 4.12
* **Description:** This dataset is designed for training ML models to predict diseases based on patient-reported symptoms. It includes various tables containing symptom weights, disease descriptions, medication recommendations, diet guidelines, precautions, and symptom-disease associations.

## Data Collection:

* **File Format: CSV**
* **Structure:**

**Symptom-Weight Table:** Contains symptom names and their significance scores.

(133, 2) represents 2 features and 133 data points.

**Disease-Description Table:** Provides disease names and descriptions.

(41, 2) represents 2 features and 41 data points.

**Disease-Diet Table:** Lists dietary recommendations for diseases.

(41, 2) represents 2 features and 41 data points.

**Disease-Medication Table:** Maps diseases to standard medications.

(41, 2) represents 2 features and 41 data points.

**Disease-Precautions Table:** Contains disease-specific precautionary measures.

(41, 6) represents 6 features and 41 data points.

**Disease-Symptoms Table:** Links diseases to common symptoms.

(4920, 6) represents 6 features and 4920 data points.

**Training & Testing Dataset:** Includes 132 binary symptom columns (1 = Present, 0 = Absent) and a prognosis column (target variable for disease classification).

(4920, 133) represents 133 features and 4920 data points.

# References

Msosa, Y. J. et al. (2023) 'Trustworthy Data and AI Environments for Clinical Prediction: Application to Crisis-Risk in People With Depression', *IEEE Journal of Biomedical and Health Informatics*, 27(11), pp. 5588-5598.

Link: <https://ieeexplore.ieee.org/document/10239317>

Rayle, J. et al. (2022) 'Using AI to Predict Caregiver Ability to Self-Manage Chronic Illness When Caring For Children With Special Health Care Needs', *2022 4th International Workshop on Artificial Intelligence and Education (WAIE)*, Xiamen, China, pp. 41-45.

Link: <https://ieeexplore.ieee.org/document/10035205>

Ma, Q. (2024) 'Research on the Application of Health Big Data based on Semi-Supervised Learning Algorithm', *2024 3rd International Conference for Innovation in Technology (INOCON)*, Bangalore, India, pp. 1-4.

Link: <https://ieeexplore.ieee.org/document/10512045>

Bansal, A. et al. (2021) 'Machine Learning Methods for Predictive Analytics in Health Care', *2021 10th International Conference on System Modeling & Advancement in Research Trends (SMART)*, MORADABAD, India, pp. 258-262.

Link: <https://ieeexplore.ieee.org/document/9676233>