# Project Report : Medical Diagnosis Expert System - An expert system for medical diagnosis that assists healthcare professionals.

Md Tamim Ahamed:201426042
Email address: tamim.ahamed@nothsouth.edu

## Abstract

This project develops a Medical Diagnosis Expert System (MDES) to assist healthcare professionals in diagnosing diseases based on patient symptoms and medical history. Combining machine learning and rule-based inference, the system analyzes medical data to suggest potential diagnoses with confidence scores. It integrates symptom extraction and classification algorithms like decision trees for accurate predictions. Designed for seamless clinical use, MDES enhances diagnostic accuracy, reduces errors, and improves patient outcomes. Future improvements include expanding the knowledge base and integrating real-time patient data.

## Methodology

### 1. Problem Definition

- **Objective:** Accurate and timely medical diagnosis is critical but often hindered by human error, time constraints, and a shortage of medical experts. Misdiagnosis or delayed diagnosis can lead to severe health risks. Current diagnostic methods rely heavily on human expertise, which varies in accuracy.
  The Medical Diagnosis Expert System (MDES) addresses this by using machine learning and rule-based reasoning to assist healthcare professionals in identifying diseases based on symptoms and medical history. It enhances diagnostic accuracy, reduces errors, and supports decision-making, especially in resource-limited settings, ultimately improving patient outcomes.
- **Challenges:**
  - Data Quality and Privacy – Incomplete or biased medical data and strict privacy regulations limit access
  - Overlapping symptoms make accurate disease classification difficult.
  - System Integration – Compatibility with EHRs and hospital systems is challenging.
  - Regulatory and Ethical Concerns – Compliance with healthcare regulations is crucial for handling sensitive medical data securely.

### 2. Data Collection and Preprocessing

- Collect medical datasets from electronic health records (EHRs), public medical databases, and expert knowledge sources
- Perform data cleaning, feature extraction, and normalization to ensure high-quality inputs for the system.

### 3. Model Selection

To explore the performance of different machine learning models for segmentation, the following models will be implemented:

1) **Linear Regression:**
   - We create a baseline model to predict pixel intensities based on linear relationships.
   - Then converts the output into binary masks using a threshold.
2) **Decision Tree:**
   - First of all, we build a tree structure to classify pixels based on feature splits.
   - here Pruning techniques will be applied to prevent overfitting.
3) **Random Forest:**
   - Use an ensemble of decision trees to improve segmentation performance.
   - Perform pixel-wise classification by aggregating predictions from multiple trees.
4) **k-Nearest Neighbors (kNN):**
   - Classify pixels by considering the majority label of their nearest neighbors in the feature space.
   - Tune the number of neighbors ($k$) for optimal performance.
5) **Binary classification :**If the Medical Diagnosis Expert System (MDES) is designed to detect a specific disease (e.g., COVID-19: Yes/No, Diabetes: Present/Absent), then it is a binary classification model. However, if the system needs to classify multiple diseases based on symptoms (e.g., Pneumonia, Flu, Asthma, or No Disease), then it becomes a multi-class classification model will work.

### 4. Training the Models

- First of all we split the dataset into **training** and **testing** sets like 80-20 split.
- Train each model using pixel-level features from the training set.
- then Perform hyperparameter tuning for each model:
  - **Linear Regression:** Here we optimize regularization parameters if applicable.
  - **Decision Tree:** Here we tune maximum depth, minimum samples per leaf, and splitting criteria.
  - **Random Forest:** we optimize the number of trees and maximum features.

  - **kNN:** Here we determine the optimal value of $k$ and distance metric (e.g., Euclidean, Manhattan).
  - **binary classification :**If the Medical Diagnosis Expert System (MDES) is designed to detect a specific disease (e.g., COVID-19: Yes/No, Diabetes: Present/Absent), then it is a binary classification model. However, if the system needs to classify multiple diseases based on symptoms (e.g., Pneumonia, Flu, Asthma, or No Disease), then it becomes a multi-class classification problem.

## 5. Model Evaluation

- Evaluate the models on the testing set using segmentation-specific metrics:
  - **Accuracy:** Pixel-wise accuracy of segmentation.
  - **Dice Coefficient:** Measures overlap between predicted and true masks.
  - **Intersection over Union (IoU):** Evaluates the quality of segmentation.
  - **Precision, Recall, and F1-Score:** Assess class imbalance handling.

## 7 Deployment

- We package the best-performing model into a deployable pipeline.
- We have vision for scalability for processing large datasets in research or clinical environments.

## Conclusion

The Medical Diagnosis Expert System (MDES) leverages machine learning and rule-based inference to enhance diagnostic accuracy, reduce human errors, and support healthcare professionals in decision-making. By integrating AI-driven analysis with medical expertise, the system improves efficiency, especially in resource-limited settings. Despite challenges like data privacy, model interpretability, and system integration, MDES has the potential to revolutionize clinical diagnostics. Future improvements will focus on expanding the knowledge base, enhancing explainability, and ensuring regulatory compliance for real-world deployment.