Student Name: Abdallah Abdulkarim Mohammad Sarhan

Student ID: 21110087

Instructor Name: DR. Elham Derbas

Course Name: Computer Research Project (CRP)

Introduction

In this research paper project, I will enroll in the healthcare field using the Ai system to predict the test result for the patient (Normal, Abnormal, Inconclusive). The data had been collected from several hospitals and patients and I have the access on this data via: <https://www.kaggle.com/datasets/prasad22/healthcare-dataset>

By clicking on the link you will have access to see the data and how it have been collected. I’ve worked on preprocessing the data before making the model learn to predict the test result, that was done by data analysis (EDA) and by cleaning the data, after that I’ve used multi-layer perceptron (MLP) to predict the test result. Also I’ve create a survey as a secondary primary data to get more data analytics, information’s, patterns and insights. Also in this research paper I will discuss the results and the findings.

**1 – Research Proposal**

**Literature Review**

Artificial Intelligence (AI) is becoming more popular in healthcare because it helps doctors diagnose and treat diseases better and faster. Many studies show that AI can predict patient outcomes, which helps in treating them more effectively.

One major use of AI is in detecting diseases early. For example, some computer programs are good at spotting heart disease early by looking at things like blood pressure and age. This early detection helps doctors treat patients sooner, which can lead to better health outcomes.

Also, some researchers have worked on making these AI programs better by tweaking their settings. This makes the programs more accurate, especially when they work with complex or large amounts of data.

Deep learning, a type of AI, has been used to analyze medical images and patient data over time. These methods are great because they can handle very detailed and large datasets, and they are better at understanding complex patterns in the data.

However, using AI in healthcare also comes with challenges. It's important to use AI fairly and safely, making sure that patient privacy is protected and that the AI doesn't become biased.

Overall, the studies show that AI has a lot of potential to improve healthcare by making diagnosis and treatment more accurate and efficient. This research will look more into how AI can be used to predict healthcare outcomes and help in treating patients better.

**Research Question:**

How can machine learning models effectively classify healthcare test outcomes (Normal, Abnormal, or Inconclusive) by leveraging detailed features from patient demographics (such as age, gender, and blood type), medical conditions (hypertension, diabetes, etc.), admission details (including room type and urgency), treatment specifics (like prescribed medications), financial data (billing amount and insurance provider), and healthcare providers (doctor and hospital information)? Additionally, can the classification model identify whether males or females are more susceptible to specific test outcome categories, and how do feature interactions contribute to these predictions while addressing challenges such as class imbalance and feature redundancy?

**2. List the research objectives.**

**Data Understanding and Preprocessing:**

* Analyze the dataset to understand the distribution of features, target classes, and potential biases.
* Clean and preprocess the data by handling missing values, outliers, and inconsistencies.

**Feature Engineering and Selection:**

* Transform categorical and numerical features into model-compatible formats.
* Identify the most significant features (age, gender, medical conditions, etc.) contributing to healthcare test outcomes.
* Investigate interactions between features to enhance predictive performance.

**Model Development and Classification:**

* Develop classification models (Decision Tree, Random Forest, Logistic Regression or MLP) to predict test outcomes (Normal, Abnormal, Inconclusive).
* Evaluate the models using metrics such as accuracy, precision, recall, F1 score.

**Gender-Specific Analysis:**

* Analyze whether males or females are more likely to have abnormal or inconclusive test outcomes.
* Incorporate gender as a key feature to assess its influence on predictions and outcomes.

**Optimization and Validation:**

* Apply hyperparameter tuning (e.g., grid search, random search) to improve model performance.
* Use techniques such as cross-validation to ensure the robustness and generalizability of the models.

**Insights and Interpretability:**

* Identify the key factors contributing to abnormal and inconclusive outcomes using interpretability tools (e.g., feature importance, SHAP values).
* Provide actionable insights into demographic and treatment patterns that can inform healthcare practices.

**Addressing Challenges:**

* Handle potential challenges such as class imbalance and multicollinearity in features.
* Ensure ethical and unbiased predictions to avoid misclassifications based on sensitive attributes like gender.

**Documentation and Recommendations:**

* Document findings on predictive accuracy and demographic-specific susceptibility to illnesses.
* Recommend how healthcare providers can leverage these insights to optimize patient care and resource allocation.

**Ethical and Practical Considerations:**

* Ensure that the model avoids bias, particularly in gender-specific predictions, by using fairness evaluation metrics.
* Assess the ethical implications of deploying predictive models in clinical practice, such as the risk of misclassification.

**Visualization and Reporting:**

* Create intuitive visualizations (e.g., confusion matrices, feature importance graphs) to present model performance and insights.
* Develop dashboards for healthcare providers to understand and act on the predictions.

**Methodology:**

In this project, I will use different ways to study and analyze data. There are three main methods: qualitative, quantitative, and mixed method.

**Qualitative methods** are about understanding ideas and opinions deeply but not through numbers. This might include interviews where you ask people about their experiences or thoughts.

**Quantitative methods** involve numbers and statistics. For example, in a survey, you might collect numerical data about how many people have a certain disease or how they respond to a treatment.

**Mixed methods** use both qualitative and quantitative approaches. This is helpful when you need a full picture that includes both statistical data and personal insights.

For my AI healthcare project, I will mainly use quantitative methods because my focus is on using data to predict patient test results. This will involve:

* **Surveys:** I’ll gather numerical data about patients’ health, like their age, blood type, and medical history.
* **Data Analysis:** Using machine learning and deep learning models to predict health outcomes based on the data from surveys.

The reason for choosing quantitative methods is that they are very effective for making predictions based on large amounts of data, which is essential for AI work. These methods allow us to see patterns in the data and make accurate predictions about patient outcomes.

I’ve decided not to use qualitative methods because this project doesn’t focus on personal opinions or experiences but rather on objective data that can be measured.

In conclusion, the quantitative approach is best for my project because it matches the goal of using AI to analyze health data and predict results. This method supports the project's aim to provide clear, data-driven insights that can help in healthcare settings.

**2 – Research Strategy Applied**