Introduction:

The purpose of this literature review is to analyze existing research on predictive analytics in healthcare. The findings will guide the development of a machine learning model aimed at predicting healthcare outcomes, improving patient care, and optimizing resource management. This report also defines the key stakeholders, user stories, and functional/non-functional requirements of the project.

Literature Review:

1. **Key words**
   - **Statistical Analysis in Healthcare**
   - **Diabetes Health Indicators**
   - **Classification of the patient's condition**
2. **Sources**
   - **Summary of Reviewed Sources**

Source	Summary	Relevance to the Project
Predictive Analytics in Healthcare: Trends & Challenges	Discusses common machine learning techniques (Logistic Regression, Random Forests, Neural Networks) and challenges such as data security and model interpretability.	Guides model selection and ensures data security best practices are followed.
Machine Learning for Disease Prediction Using Electronic Health Records (EHR)	Explores feature extraction techniques and handling imbalanced datasets in EHR-based predictive models.	Helps improve data preprocessing and feature selection to enhance model performance.
Bias and Fairness in Healthcare AI Models	Examines biases in predictive healthcare models and strategies to mitigate them, ensuring fair predictions across different population groups.	Ensures fairness in model training by applying techniques to balance datasets and reduce bias.

- **Google Scholar**

  https://academic.oup.com/jamia/article/26/12/1651/5542900

3. **Analysis methods from the source**

- **Predictive algorithms**
  1. **Machine learning algorithm (random forests, Logistic Regression)**
  2. **Deep learning**

4. Problems of used algorithms

- **"Black Box" : black box problem it happens when the algorithm be more complicated and can't explains from where it comes from with this results**

5. Integration of Findings into the Project

Based on the literature review, the following strategies will be incorporated into the project:

- **Improved Data Processing: Implementing techniques for handling imbalanced datasets to prevent model bias.**
- **Model Selection: Evaluating Logistic Regression, Random Forests, and Neural Networks to determine the most effective approach.**
- **Data Security Measures: Applying encryption and anonymization techniques to safeguard patient data.**
- **Model Transparency: Utilizing tools like SHAP and LIME to improve explainability and interpretability of model decisions.**
  Which solves the problem of "Black Box"

Stakeholder Analysis:

Key Stakeholders and Their Needs

Stakeholder	Role	Needs and Expectations
Healthcare Professionals	Doctors, Nurses, Medical Staff	Accurate predictive insights for early diagnosis and treatment recommendations.
Hospital Administrators	Management & Decision-makers	Data-driven insights for optimizing resource allocation and operational planning.
Patients	End Users	Personalized health predictions and preventive care recommendations.
Data Scientists & Developers	System Designers	High-quality datasets and well-defined evaluation metrics for model improvement.

User Stories and Use Cases:

User Stories

1. **As a doctor, I want to input a patient's health data and receive risk predictions so that I can take early intervention measures.**
2. **As a hospital administrator, I want to analyze patient risk trends to allocate resources efficiently.**
3. **As a patient, I want to receive predictive insights about my health to take proactive measures.**

Use Cases

- **Patient Risk Prediction: The system predicts disease likelihood based on health records.**
- **Trend Analysis: The system generates statistical insights into population health trends.**
- **Model Performance Monitoring: Continuous evaluation of model accuracy with periodic refinement.**

Functional and Non-Functional Requirements:

Functional Requirements

- The system must allow healthcare professionals to input patient data and receive risk predictions.
- The system must generate real-time visualizations of health trends.
- The model should support multiple machine learning algorithms for comparison.
- The system must provide an API for integrating predictions into hospital management software.

Non-Functional Requirements

- **The system should ensure data privacy through encryption and compliance with healthcare regulations.**

- **The model should be transparent and interpretable in its predictions.**
- **The system should be scalable to handle large datasets and increasing user demand.**
- **The API should provide responses within two seconds for optimal performance.**