UNIVERSITA' DEGLI STUDI eCampus

Facoltà di Ingegneria Corso di Laurea Magistrale in

INGEGNERIA INFORMATICA E DELL'AUTOMAZIONE Classe delle lauree magistrali in Ingegneria informatica (LM-32)
Curriculum: Artificial Intelligence

TITOLO DELLA TESI

Artificial Intelligence and Machine Learning Techniques for the Characterization and Prediction of Diabetes

Relatrice:

Prof.ssa Ing. Patrizia Vizza, Ph.D

Candidato:

Dott.Ing. Leon Martial Doungala Nzoyem
Matricola :001447064

Anno Accademico 2023 / 2024

In Memory and Hope

This work is dedicated to all those affected by diabetes — those living with it and those who have lost their lives to the disease. It also honours their families, carers and loved ones who support them every day. May this work contribute to better care, better management and, one day, a cure for this disease.

Table of Contents

1. Introduction

1.1 Problem Statement

Diabetes is a chronic disease that affects millions of people worldwide. It is on the rise, causing serious health problems, economic challenges and reduced quality of life. According to recent reports, the number of adults with diabetes has quadrupled since 1980, mainly due to an increase in type 2 diabetes [1]. Early detection and good management of diabetes are essential to prevent serious complications such as heart disease, kidney failure and blindness [2]. However, traditional methods of diagnosing diabetes often miss the disease in its early stages, leading to delays in treatment and worse health outcomes [3].

Recently, advances in artificial intelligence (AI) and machine learning (ML) have shown great potential for improving healthcare. These technologies help in early diagnosis, personalised treatment and effective management of chronic diseases such as diabetes. AI and ML can analyse large amounts of data to find patterns that may not be obvious to doctors [4] . Using AI and ML to predict and manage diabetes could significantly improve patient care by providing more accurate, timely and personalised treatments [5].

1.2 Objectives of the Study

- The main objective of this thesis is to explore and develop AI and ML techniques to help predict and understand diabetes. Specifically, this study aims to:
- Build and evaluate machine learning models to predict diabetes based on clinical and demographic data [6].
- Explore the use of generative AI techniques to generate additional data for better model training and improved prediction accuracy [7].
- Use LangChain for natural language processing (NLP) to extract important information from unstructured medical records and improve diabetes management [8].
- Compare the effectiveness of different machine learning models and generative AI techniques to find the best approach for predicting and understanding diabetes [9].

- Add to the existing body of knowledge by showing how AI and ML can be practically applied in healthcare, with a focus on diabetes [10].

1.3 Structure of the Thesis

This thesis is divided into several chapters, each covering a specific part of the research:

- **Chapter 2**: Background and Literature Review - This chapter gives an overview of diabetes, AI, and ML. It also reviews the existing literature on how these technologies are used in healthcare, especially in predicting and managing diabetes [11].
- **Chapter 3**: Machine Learning Techniques for Diabetes Prediction - This chapter discusses the methods used to collect and process data, select and train machine learning models, and evaluate their performance [12].
- **Chapter 4**: Generative AI for Data Augmentation and Analysis - This chapter looks at how generative AI techniques can be used to improve data quality and the accuracy of predictive models [7].
- **Chapter 5**: LangChain for Natural Language Processing in Diabetes Management - This chapter explores how LangChain can be used for tasks like extracting information from medical records related to diabetes [8].
- **Chapter 6**: Experiments and Results - This chapter presents the experimental setup and the results of using machine learning and generative AI models, as well as LangChain implementations [13].
- **Chapter 7**: Conclusions - The final chapter summarises the main findings, discusses the contributions of the study, and suggests recommendations for future work .

5. Conclusions

5.1 Summary of Findings

This thesis examined the nexus of artificial intelligence and diabetes management, offering a comprehensive examination of prediction, characterisation, and management solutions. The study employed a range of methodologies, integrating machine learning , deep learning , generative AI, and natural language processing, which culminated in the development of Elyon, a Python Flask-based web application. Elyon provides a practical tool for healthcare professionals to facilitate effective analysis and management of diabetes.

Key findings include:

1. **Extensive Literature Review:**
    - The text provides a comprehensive and detailed understanding of diabetes mellitus, encompassing its various types, the global prevalence of the disease, and its impact on overall health and well-being.
    - The study identified the most significant challenges currently facing those managing diabetes, emphasising the necessity for the implementation of AI-driven solutions.

2. **Machine Learning in Diabetes Prediction:**
    - Supervised learning models such as logistic regression, decision trees and neural networks were used to predict incident diabetes.
    - Exploratory data analysis (EDA) identified key health measures such as HbA1c, BMI and cholesterol levels as significant predictors.
    - Advanced feature engineering and hyperparameter tuning significantly improved model accuracy and reliability.

3. **Generative AI for Data Augmentation:**
    - Techniques such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs) addressed data scarcity by generating synthetic data sets.
    - These augmented datasets improved the robustness of predictive models and facilitated better generalisation to unseen data.

4. **LangChain and Retrieval-Augmented Generation (RAG):**
    - LangChain's NLP capabilities enabled the interpretation of complex medical records and user queries.
    - RAG was used to combine data retrieval with language generation to provide accurate, contextually relevant insights.

5. **Web Application Development: Elyon :**
    - The Elyon app integrated machine learning models, generative AI and conversational agents.
    - It featured intuitive visualisations of patient metrics and interactive chat capabilities, making health data accessible and actionable for clinicians.

5.2 Contributions to the Field

This thesis makes several significant contributions to AI and healthcare research:

1. **Comprehensive AI Framework for Diabetes Management**:
    - Demonstrated the integration of machine learning, generative AI, and NLP into a unified system for disease prediction and patient characterization.
    - Showcased the practical application of AI methodologies to tackle real-world healthcare challenges.

2. **Enhancements in Data Augmentation:**
    - Developed frameworks for generating synthetic datasets using GANs and VAEs, addressing common issues such as data scarcity and class imbalance in medical datasets.

3. **Exploration of LangChain and RAG in Healthcare:**

    - Provided an implementation blueprint for using LangChain and RAG in medical decision support systems, highlighting their ability to interpret and respond to user queries in a data-driven manner.

4. **User-centred design:**
    - Developed **Elyon**, a web application that combines AI models with user-centric interfaces to improve accessibility and usability for clinicians.

5. Knowledge Contribution:

- Provided a comprehensive review of diabetes management challenges, current AI techniques and their applications.
- Encouraged further research by making partial implementations available via a GitHub repository, balancing transparency with privacy concerns.

5.3 Final Thoughts on AI in Diabetes Management

challenges of diabetes management. This thesis highlights how AI can contribute to:

- **Enhanced Predictive Accuracy**:
  - Advanced machine learning techniques enable early identification of at-risk individuals, facilitating timely medical interventions.
- **Personalized Healthcare**:
  - AI systems like Elyon analyze patient-specific data to deliver tailored recommendations, promoting precision medicine.
- **Streamlined Decision-Making**:
  - AI-powered tools such as LangChain and RAG reduce the cognitive load on healthcare providers by offering evidence-based insights and efficient information retrieval.
- **Scalable and Sustainable Solutions**:
  - Generative AI models address data limitations, ensuring scalability in predictive healthcare applications.

5.4 Closing Remarks

The development of Elyon clearly shows the strong potential of artificial intelligence in transforming healthcare. By combining supervised learning, generative AI, and natural language processing, this project offers a practical and scalable solution to manage diabetes effectively. As AI technology continues to advance, its role in personalized medicine and decision-making will grow even more important, helping doctors and healthcare providers deliver better care and improve patient outcomes. Elyon not only addresses key challenges in diabetes management but also opens the door to exciting possibilities for future research. This project provides a strong foundation for exploring how advanced AI techniques can further enhance global healthcare systems, potentially as the focus of a PhD study.