# 7CS997 Independent Studies Project Plan

## Student/supervisor details

Project Title: **Early Prediction of Diabetic Complications Using Multi-Modal Deep Learning**

Name: **James Oluwafemi Adeshina**

Student Id: **100752659**

MSc: **Big Data Analytics**

Supervisor: **Dr. Oluwarotimi W. Samuel**

## Brief description

Diabetes is a major global health concern, contributing significantly to long-term complications such as neuropathy, nephropathy, retinopathy, cardiovascular disease, and peripheral vascular disorders (Rawshani et al., 2018; IDF, 2021). Effective prediction and diagnosis of these complications remain a clinical challenge. While many existing models rely on single-modality data and traditional machine learning approaches, these methods often struggle with transparency and generalizability in real-world settings (Dharmarathne et al., 2024; Tanim et al., 2023). This project aims to develop a multi-label deep learning model that can predict the presence of diabetic complications based on multi-modal tabular data. The dataset includes demographic, biometric, lifestyle, and treatment-related features alongside binary complication labels. Multiple algorithms, including Multilayer Perceptron (MLP), TabTransformer, Random Forest, and XGBoost, will be compared. The project emphasises model interpretability through the use of SHAP (SHapley Additive exPlanations) to identify key risk features influencing each predicted complication (Lundberg & Lee, 2017). By combining predictive performance with explainable AI techniques, this study contributes toward more trustworthy and actionable machine learning applications in clinical decision-making. The outcomes may assist in developing real-world diabetic risk stratification tools for use in hospitals and public health planning.

**References**

Rawshani, A., Rawshani, A., Franzén, S., Sattar, N., Eliasson, B., Svensson, A.-M., Zethelius, B., Miftaraj, M., McGuire, D.K., Rosengren, A. and Gudbjörnsdottir, S., 2018. Risk factors, mortality, and cardiovascular outcomes in patients with type 2 diabetes. New England Journal of Medicine, 379(7), pp.633–644. doi:10.1056/NEJMoa1800256

International Diabetes Federation (IDF), 2021. IDF Diabetes Atlas. 10th ed. Brussels: International Diabetes Federation. Available at: https://diabetesatlas.org/atlas/tenth-edition [Accessed 7 May 2025].

Tanim, S.A., Aurnob, A.R., Shrestha, T.E., Emon, M.R.I., Mridha, M.F. and Miah, M.S.U., 2023. Explainable deep learning for diabetes diagnosis with DeepNetX2. Journal of King Saud University - Computer and Information Sciences. doi:10.1016/j.jksuci.2023.101651

Dharmarathne, G., Jayasinghe, T.N., Bogahawaththa, M., Meddage, D.P.P. and Rathnayake, U., 2024. A novel machine learning approach for diagnosing diabetes with a self-explainable interface. Journal of Artificial Intelligence and Data Science, 1(1), p.100001. doi:10.1016/j.jaids.2023.100001

Lundberg, S.M. and Lee, S.-I., 2017. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30. Available at: https://arxiv.org/abs/1705.07874 [Accessed 7 May 2025].

## Project aims and objectives

1. To highlight the importance of diabetic complications, which contribute significantly to long-term patient morbidity and mortality.

2. To explore and implement classification algorithms capable of predicting multiple diabetic complications (e.g., neuropathy, nephropathy, retinopathy, cardiovascular disease, peripheral vascular disease) from structured clinical and behavioural data.

3. To conduct a comparative evaluation of deep learning approaches in the context of multi-label classification.

4. To integrate explainable AI techniques, such as SHAP, in order to identify the most influential features contributing to each predicted complication.

5. To develop a reproducible and interpretable framework that contributes to diabetic risk stratification and supports clinical decision-making.

## Plan of work

**Literature Review**

A comprehensive review of peer-reviewed papers will be conducted to establish a solid foundation for this study. The review will focus on the use of machine learning and deep learning in diabetes-related complication prediction, with particular emphasis on multi-label classification, multimodal data usage, and explainable AI. Databases such as IEEE Xplore, PubMed, Springer, and ScienceDirect will be explored. The review will also highlight limitations of existing models, such as a lack of interpretability, poor generalisation across datasets, or reliance on single-modality input.

**Methodology**

1. **Data Source and Preparation**

The project will utilise a structured clinical dataset containing demographic, behavioural, and clinical features relevant to diabetic complication prediction. It also should contain five binary complication labels: nephropathy, neuropathy, retinopathy, cardiovascular disease, and peripheral vascular disease. Preprocessing will involve missing value imputation, feature normalisation, and multi-label binarisation. No identifiable patient data is present.

1. **Feature Engineering and Input Structuring**

Derived features such as BMI categories or treatment adherence levels may be included. All features will be treated as structured tabular data to ensure compatibility across traditional ML and DL models.

**Algorithm Implementation & Evaluation**

The model implementation phase will involve both traditional machine learning and deep learning techniques. For benchmarking purposes, traditional models such as Logistic Regression, Random Forest, and XGBoost will be included. The deep learning models will focus on architectures suited to structured data, specifically Multilayer Perceptrons (MLP) and TabTransformer. Evaluation will be based on standard metrics for multi-label classification, including Macro F1-Score, ROC-AUC (both macro and per complication), Hamming Loss, and Subset Accuracy. Python will be used for development, along with libraries such as Scikit-learn, PyTorch, and SHAP for both modelling and interpretability.

**Analysis & Comparison**

This phase will compare the performance of all models using the metrics outlined above. SHAP analysis will be used to identify the most influential features for each predicted complication, helping assess both predictive strength and clinical plausibility. Results will be visualised through tables, graphs, and SHAP summary plots. Particular attention will be paid to:

* Model sensitivity across complications
* Feature overlap across conditions
* Interpretability vs performance trade-offs

Documentation & Reporting

The full project will be documented in a technical report structured as follows:

* Abstract, Introduction, and Related Work
* Methodology and Model Implementation
* Results, Interpretation, and Visualisations
* Limitations, Ethical Considerations, and Future Work

Documentation will include well-commented Python code developed in PyCharm, along with SHAP plots for interpretability, model comparison tables, and diagrams illustrating the model architectures used.

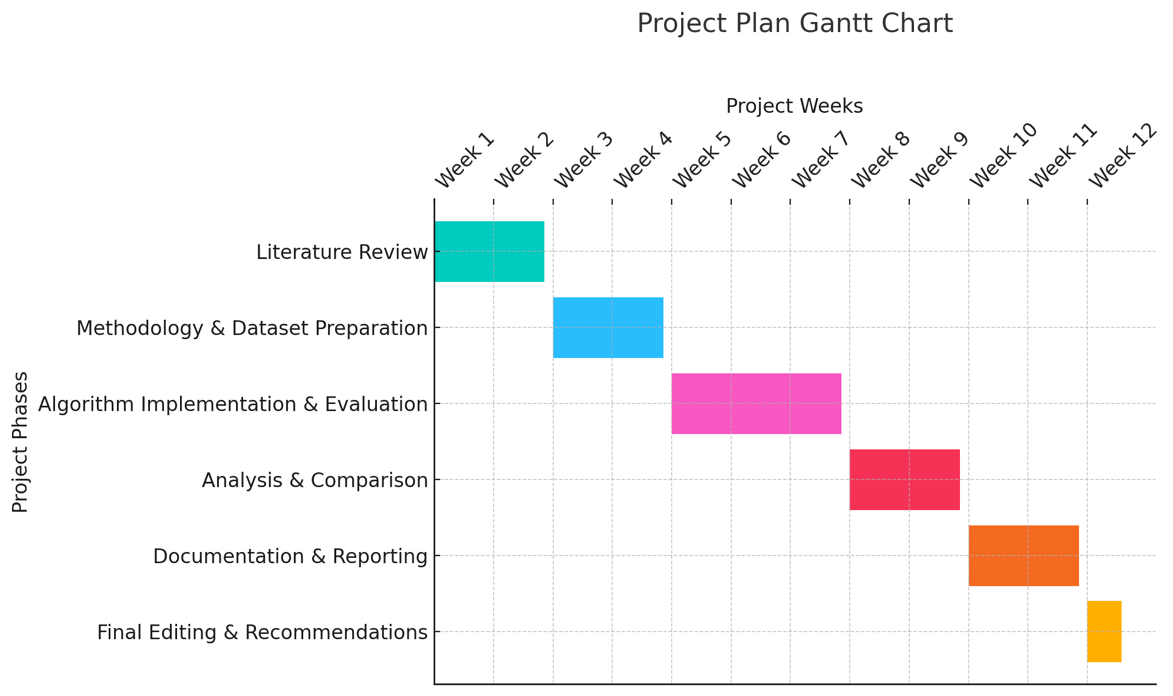
**Future Recommendations**

Based on the results, future work may include:

* Expanding to larger or longitudinal datasets for time-to-event modelling
* Exploring transformer-based models beyond tabular input (e.g., BERT for clinical notes)
* Developing a clinician-facing app using Streamlit or Dash
* Incorporating more real-world constraints like missing modalities and fairness metrics

## Gantt Chart

*The Gantt chart below outlines the key tasks, milestones, and deliverables for this project.*



## Meeting plan

*In the table below, the project week number and dates correspond to the week numbers that were presented in the introductory lecture that are designated meeting weeks. You should fill in the date, time, and venue for the meeting schedule that you have agreed with your supervisor as part of planning your project.*

|  |  |  |  |
| --- | --- | --- | --- |
| Project week | Meeting date | Meeting time | Venue |
| 1: 02/06/2025 |  |  |  |
| 3: 16/06/2025 |  |  |  |
| 6: 30/06/2025 |  |  |  |
| 8: 14/07/2025 |  |  |  |
| 10: 28/07/2025 |  |  |  |
| 11: 11/08/2025 |  |  |  |

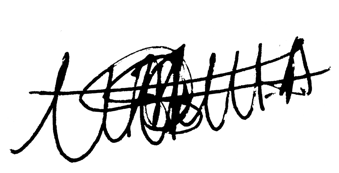
## Ethical approval

I confirm that I have discussed and agreed the project plan with my supervisor, and have completed the ethical approval process. The PDF confirming ethical approval has been completed and uploaded to the relevant submission point on course resources.

James Oluwafemi Adeshina

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Name



\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Student Signature

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Date