# 7CS997 Independent Studies Progress Report

## Student/supervisor details

**Submission deadline for signed and completed progress report:**

* **Progress Report 1 — Scheduled for 2 June 2025 at 11:00 AM**

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 Samuel**

**Progress Report: 1**

## 1. Record of meeting

In our first supervisory meeting on 2nd June 2025, I presented a comprehensive overview of my MSc project titled "Early Prediction of Diabetic Complications Using Multi-Modal Deep Learning". The project focuses on predicting multiple diabetes-related complications, specifically retinopathy, nephropathy, neuropathy, cardiovascular disease, and peripheral vascular disease, by integrating structured EHR data, tabular clinical data, and time-series patterns. I shared early insights from my literature review, which revealed that most existing studies address complications in isolation, use unimodal data, and rarely capture temporal dynamics that are crucial for early prediction. These limitations helped me define a clearer contribution space for my work.

We discussed how integrating treatment history, clinical parameters, and longitudinal trends could form a novel methodological framework. My supervisor, Dr. Samuel, emphasised the need to articulate this contribution clearly and advised focusing on feasibility and alignment across data modalities. To support this direction, I presented three shortlisted datasets: (1) the UCI 130-US Hospitals dataset for real-world EHR encounters, (2) the Mendeley Micro & Macro Complications dataset with labelled complication outcomes, and (3) a Messidor-derived dataset containing retinopathy features. Although not linked at the patient level, these datasets offer complementary value for training and comparing models across modalities.

The meeting concluded with key action points: (i) finalise the literature review with emphasis on methodological gaps, (ii) complete acquisition and preprocessing of the selected datasets, and (iii) refine the problem statement to emphasise the early prediction goal. Our next meeting is scheduled for 16th June 2025.

## Evaluation of progress

Progress to date aligns well with the initial project timeline and reflects meaningful engagement with the core research objectives. The literature review has helped clarify key gaps in existing work: most studies focus on single diabetic complications and rely on unimodal data, overlooking both multi-complication prediction and true multi-modal integration. Feedback from the supervisor has helped sharpen the project’s unique value proposition, specifically, the use of longitudinal EHR data to combine treatment history, clinical indicators, and temporal progression markers for early risk stratification across multiple complications. While dataset acquisition is still underway, promising repositories have been identified that offer cross-modal features with patient-level alignment, such as harmonised lab results, medication timelines, and complication flags. Methodologically, the project has now shifted focus toward building a reproducible fusion framework capable of integrating heterogeneous data streams. This foundational work provides a strong basis for addressing real-world limitations in current diabetes risk prediction systems, with dataset finalisation and initial model development targeted for the upcoming phase.

## Actions/targets for next meeting

Ahead of the next supervisory meeting scheduled for 16 June 2025, my core focus will be on transitioning from planning to execution. The immediate actions are structured as follows:

**Literature Review Recap**: Finalise and submit a comprehensive summary of the 10–15 reviewed papers, with specific emphasis on how existing studies fall short in integrating multi-modal data or addressing multiple diabetic complications concurrently. The recap will help reaffirm the methodological novelty of this project.

**Dataset Validation**: I will complete acquisition, documentation, and preliminary quality checks on the shortlisted datasets (UCI 130-US Hospitals, Mendeley Micro/Macro Complications, and Messidor-derived features). This includes verifying their structure, assessing missingness, and confirming the presence of complication flags, clinical parameters, and timelines. In parallel, I will begin the application and credentialing process required to gain access to the PhysioNet-hosted eICU Collaborative Research Database and MIMIC-IV, which provide rich, patient-level EHR data with temporal and treatment history components. These datasets are potential candidates for enhancing the project’s multi-modal integration, particularly in linking time-series and structured clinical data for longitudinal modelling.

**Contribution Matrix**: I aim to map the scope of my contribution explicitly, contrasting prior models with the proposed framework. This includes highlighting how my study uniquely combines tabular, temporal, and possibly imaging features for multi-risk profiling, using explainable AI.

**Methodological Blueprint**: I will refine the modelling strategy, outlining how DL models (e.g., TabNet, LSTM, or Transformer for time-series) will be deployed. The blueprint will also include evaluation metrics, class balancing strategy, and interpretability approach using SHAP or LIME.

**Dataset Preparation and Exploration**: I will begin formal data wrangling, cleaning, and exploratory analysis of each dataset, documenting distributions, class imbalances, and feature interactions. This step also includes setting up pipelines for modality-specific preprocessing.

**Model Selection Decision:** Based on data characteristics, I will shortlist baseline models suitable for each modality and confirm the initial architecture for multi-modal fusion. If time permits, I will prototype the baseline model and prepare performance benchmarks for discussion.

## 4.Record of meetings held

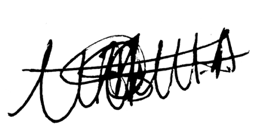
*The table below should contain the meeting schedule that you planned with your supervisor at the start of the project. In the “Confirmation” column on the right, you should confirm the meeting has been held by including a “Y” or “N” in the relevant box. If there has been a need to change the original meeting schedule planned, you should update this table.*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Project week | Meeting date | Meeting time | Venue | Confirmation |
| Week 1 | 2 June, 2025 | 11:00 AM | MS 217 | Y |
| Week 3 | 16 June 2025 |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

James Oluwafemi Adeshina Oluwarotimi W. Samuel

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

Student Name Supervisor Name



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

Student Signature Supervisor Signature

9th June, 2025

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

Date Date