# 7CS997 Independent Studies Progress Report

## Student/supervisor details

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

* **Progress report 2**

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**

Progress report: 2

## Record of meeting

The meeting commenced with a deliberation on the advancements achieved so far in relation to the project. Some of the progress made include:

1. **Literature Review Consolidation**: I presented insights from newly reviewed papers that allowed us to benchmark the proposed work against existing diabetic complication prediction models. We discussed how most existing models rely on unimodal or single-complication approaches, whereas this project uniquely aims to integrate multi-modal data for early prediction across five complication categories.
2. **Refinement of Label Anchoring Methodology**: I outlined how diabetic complication labels were re-anchored using true diabetes diagnosis dates in the MIMIC-IV dataset. We discussed how this correction addressed errors from earlier pipelines that had mistakenly tied complication timelines to first hospital admission. Label definitions were further aligned with NHS prevalence estimates to improve realism.
3. **Dataset Finalization and Filtering Strategy**: After exploring both the MIMIC-IV and eICU-CRD datasets, we determined that MIMIC-IV would serve as the primary dataset for both training and evaluation due to its superior longitudinal structure. The eICU dataset lacked reliable anchoring for complications and was excluded. Filtering approaches were revised to avoid excluding too many valid records, and we agreed to balance inclusion criteria with dataset richness.
4. **Planning for Modelling and Interpretability**: Dr. Samuel advised focusing first on a solid baseline using deep learning models suited to tabular data (e.g., TabNet), before layering in more complex architectures. The importance of model explainability using SHAP was reinforced as a key requirement for real-world relevance.

The meeting concluded with a shared understanding of current achievements and prioritised actions for the next phase, which will include finalising the analytical cohort and beginning formal model training.

## Evaluation of progress

Progress aligns closely with the planned timeline in the Gantt chart, especially regarding literature consolidation, dataset validation, and refinement of the modelling objective. As of this phase, the project has moved beyond foundational design and into implementation. I have completed a comprehensive prevalence and data quality audit of the diabetic cohort (N = 10,807) and validated longitudinal coverage across key lab biomarkers (e.g., A1C, creatinine, albumin). Variables with insufficient representation, such as BUN and CRP, were excluded after structural review.

I now possess a clearly defined modelling population and a confirmed prediction window for each complication. Efforts have been made to mitigate survivorship bias and improve ICD code specificity using diabetes-restricted definitions. The project is now ready to transition to the model prototyping phase, where architectures such as TabNet, LSTM, or Transformer-based models will be evaluated. Emphasis remains on interpretability through explainable AI techniques such as SHAP. Documentation of data wrangling steps and labelling decisions has also commenced to support reproducibility.

## Actions/targets for the next meeting

1. Completion Complete full cohort preparation pipeline and finalize train-validation-test splits.
2. Begin prototyping baseline deep learning models (e.g., MLP, TabNet) using structured clinical data.
3. Define multi-modal fusion strategy and determine feasibility of including temporal sequence models for lab tests.
4. Implement early-stage SHAP value integration for model interpretability
5. Start compiling a detailed Methods section for the dissertation, outlining cohort creation, feature selection, model types, and evaluation metrics.

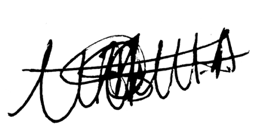
## Record of meetings held.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Project week | Meeting date | Meeting time | Venue | Confirmation |
| 1: 05/06/2023 | 05/06/2023 | 11:00-11:45 am | MS306 | Y |
| 3: 19/06/2023 | 19/06/2023 | 2:00-3:00 pm | MS306 | Y |
| 5: 03/07/2023 | 03/07/2023 | 2:00-3:00 pm | MS306 | Y |
| 7: 17/07/2023 | 17/07/2023 | 2:00-3:00 pm | MS306 |  |
| 9: 31/07/2023 | 31/07/2023 | 2:00-3:00 pm | MS306 |  |
| 11: 14/08/2023 | 14/08/2023 | 2:00-3:00 pm | MS306 |  |

James Oluwafemi Adeshina XXXXXXXXX

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

Student Name Supervisor Name



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

Student Signature Supervisor Signature

11th July, 2025 XXXXXXXXX

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

Date Date