Project Report

---

1. Student Information

- **Student Name: Anisha Raj**
- **Student ID: s222558616**
- **Date Submitted: 12/05/2025**

---

2. Project Introduction

- **Title of the Project:**

  AI-Powered Physical Activity Recognition (Health Behaviour - Use Case 8)

- **What is the project about?**
  This project focuses on recognizing physical activities using wearable sensor data from the PAMAP2 dataset. A series of deep learning models, including LSTM and GRU variants, were trained to classify human activities. The best-performing GRU model was then integrated into a Streamlit-based web app with all the other models trained for interactive evaluation and comparison.

- **Why is this project important or useful?**
  Human activity recognition is crucial for health monitoring, smart environments, and behaviour analysis. Using AI models like GRU enables accurate, real-time classification from raw sensor data, supporting practical use cases such as elderly care or fitness tracking.

---

3. API/Token Setup — *Step-by-Step*

Not applicable as no LLM/token-based API was used.

---

4. Environment Setup

- **Development Platform:**

  - Local Machine (macOS) — VS Code
  - GPU Available? [✗] Yes [✓] No (CPU-only training)
  - GPU Type (if applicable): NA

- **Python Version:** 3.12.2

- **Other Tools Used:**

  - Visual Studio Code (VS Code) for development
  - Jupyter Notebook inside VS Code for model building
  - Streamlit for interactive model exploration

---

5. LLM Setup

Not applicable — this project used GRU (deep learning, not LLMs).

---

6. Dataset Description

- **Dataset Name & Source:**
  PAMAP2 Physical Activity Monitoring Dataset (Provided by the UCI Machine Learning Repository)

- **Access Link:**
  https://archive.ics.uci.edu/dataset/231/pamap2+physical+activity+monitoring

- **Feature Dictionary / Variable Description:**
  The dataset contains readings from 3 inertial measurement units (IMUs) worn on the hand, chest, and ankle, along with heart rate data.
  Each row consists of:
  - Timestamp
  - Activity ID
  - IMU sensor features: accelerometer, gyroscope, magnetometer (x, y, z)
  - Heart rate
  - A total of 54 columns (including sensor channels and metadata)

- **Was preprocessing done? If yes, describe:**
  Yes, several preprocessing steps were applied:
  - Removed rows with missing heart rate values (NaNs)
  - Selected 25 most frequent activities
  - Applied normalization
  - Created fixed-length sequences using a sliding window
  - Encoded activity labels for model compatibility

- **Code: Load & Preprocess Dataset:**
  All loading, filtering, and preprocessing steps were implemented in **codeLogic.ipynb**.

---

7. Improving Model Performance

Not applicable for LLM – this section was adapted to deep learning training.

To identify the most effective architecture for physical activity recognition using the PAMAP2 dataset, a series of deep learning models were developed and evaluated. These included LSTM, Bidirectional LSTM, GRU, and Bidirectional GRU variants. Each model was iteratively improved using techniques such as L2 regularization, dropout, architecture simplification, and bidirectional layers. The results are summarized below.

LSTM-Based Models

Step	Method	Test Accuracy	Why It Improved or Declined
9	Baseline LSTM	39.86%	No regularization or tuning; overfitting and undertrained
10.1	LSTM + Dropout + Callbacks	41.89%	EarlyStopping helped improve stability slightly
10.2A	LSTM (64) + L2 + Dropout	50.34%	L2 regularization reduced overfitting; improved generalization

10.2B	Smaller LSTM (32) + L2 + Dropout	57.50%	Reduced model complexity + regularization = better generalization
10.2C	Stacked LSTM (64→32)	37.84%	Too deep; likely overfitting and unstable convergence
10.2D	Bidirectional LSTM (64)	48.00%	Better temporal context, but complexity didn't yield higher gains
10.2E	Smaller BiLSTM (32)	61.59%	Balanced context with lower complexity worked well

GRU-Based Models

Step	**Method**	**Test Accuracy**	**Why It Improved or Declined**
10.3A	GRU (32) + L2 + Dropout	**62.91%**	Light, regularized model with stable training; best overall
10.3B	GRU + Dense(64)	59.56%	Slightly increased complexity; small drop in generalization
10.3C	Bidirectional GRU (32)	60.00%	BiGRU added context; small benefit, but not better than simpler GRU
10.3D	Smaller BiGRU (16)	52.88%	Too small to capture complexity; underfitted slightly

The experiments clearly showed that simpler architectures with regularization, particularly the **GRU (32 units) with L2 and Dropout**, achieved the best balance of performance and generalization, with a **test accuracy of 62.91%**. Deeper or stacked architectures generally led to overfitting, while smaller, well-regularized models performed more consistently. Bidirectional models helped in some cases but didn't consistently outperform simpler ones.

Based on these results, a Streamlit app was designed to support **all trained models**, allowing users to explore, compare, and analyze each variant interactively.

---

8. Benchmarking & Evaluation

Metrics Used: Accuracy, Confusion Matrix, Classification Report (Precision, Recall, F1)
Why: Accuracy is intuitive for multi-class tasks; confusion matrix helps interpret class-wise performance.
Benchmark Dataset: PAMAP2 test split (~20% of cleaned sequences)
Visuals and evaluation results are displayed interactively through the Streamlit interface.

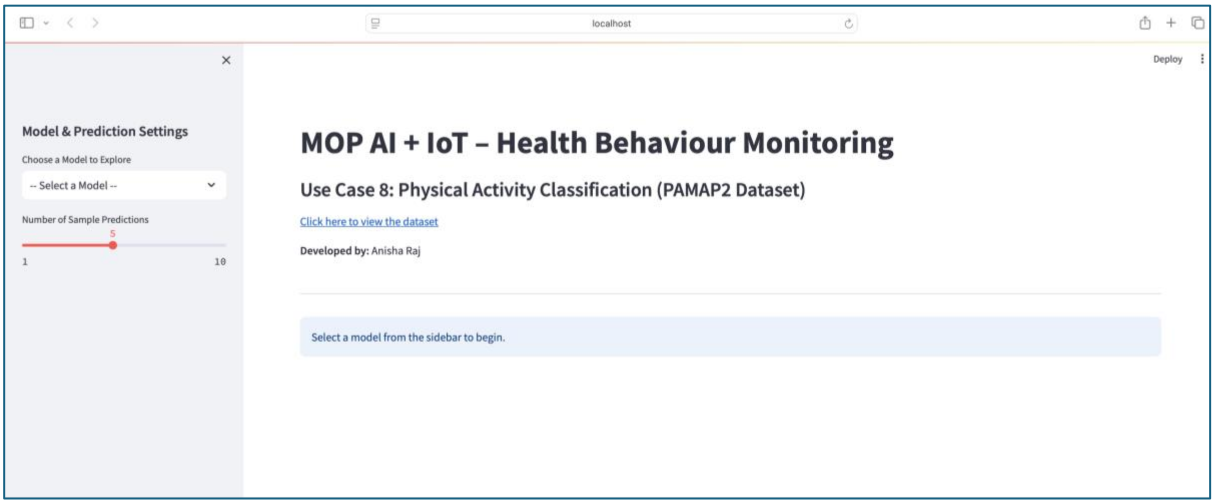
---

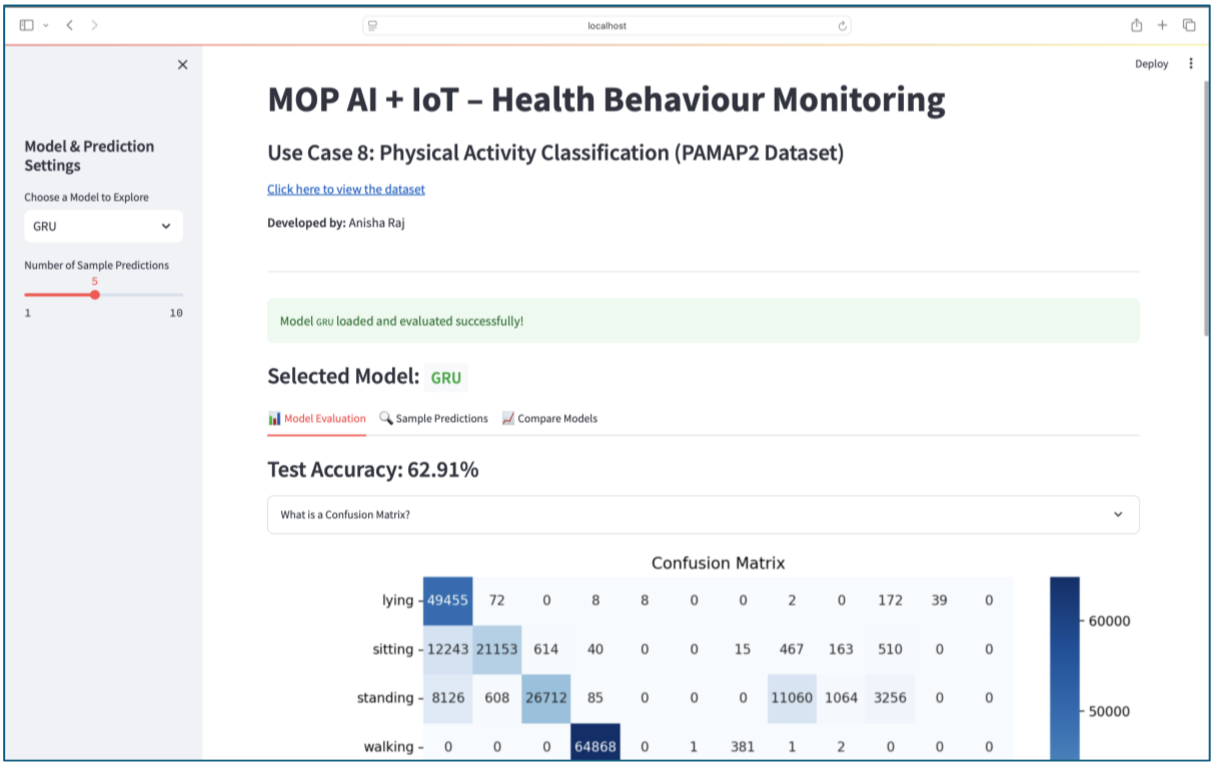
9. UI Integration

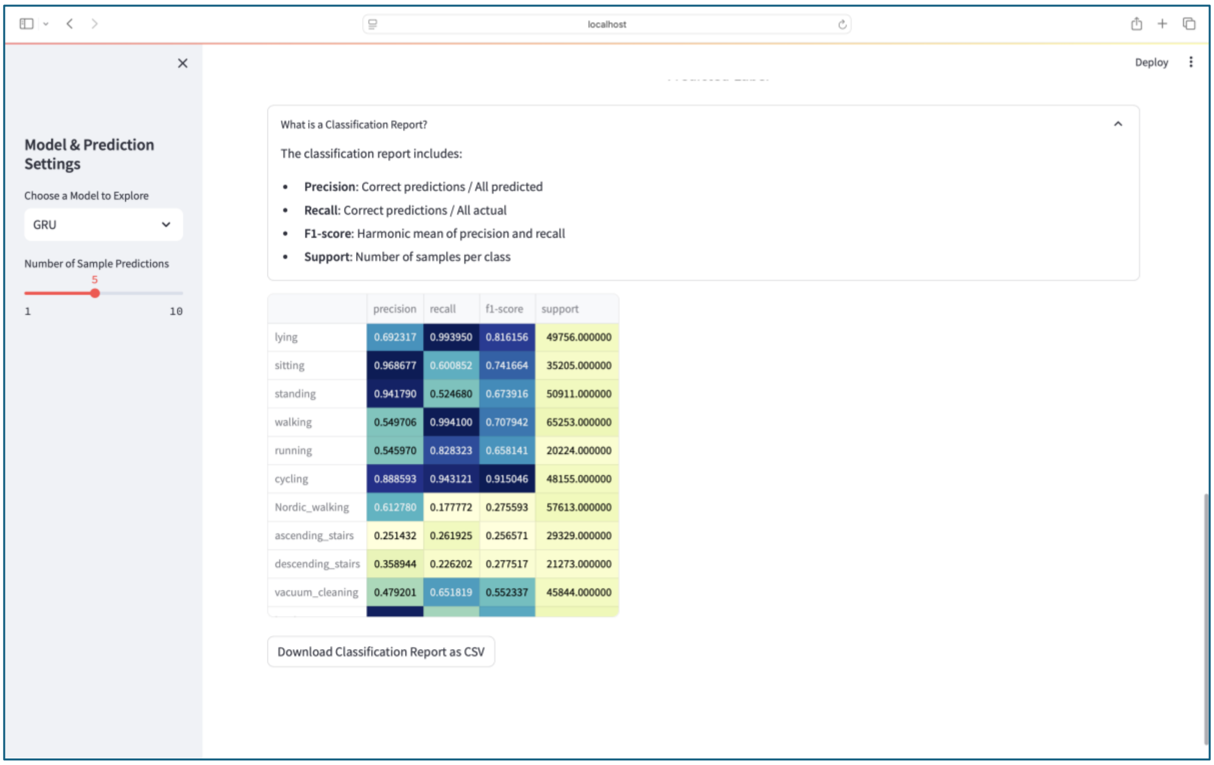
- **Tool Used:** Streamlit
- **Key Features of the Interface:**
  - Sidebar for selecting any of the trained models (LSTM, GRU, and their variants)
  - Dynamic display of test accuracy, confusion matrix, and classification report for the selected model
  - Random sample prediction display

- Explanatory tooltips and collapsible sections using st.expander to help users interpret metrics
- A dedicated *Model Comparison* tab showing a sortable table and horizontal bar chart of all model accuracies, allowing users to easily identify the best-performing model
- Optimized with @st.cache_resource and @st.cache_data to reduce latency when switching models

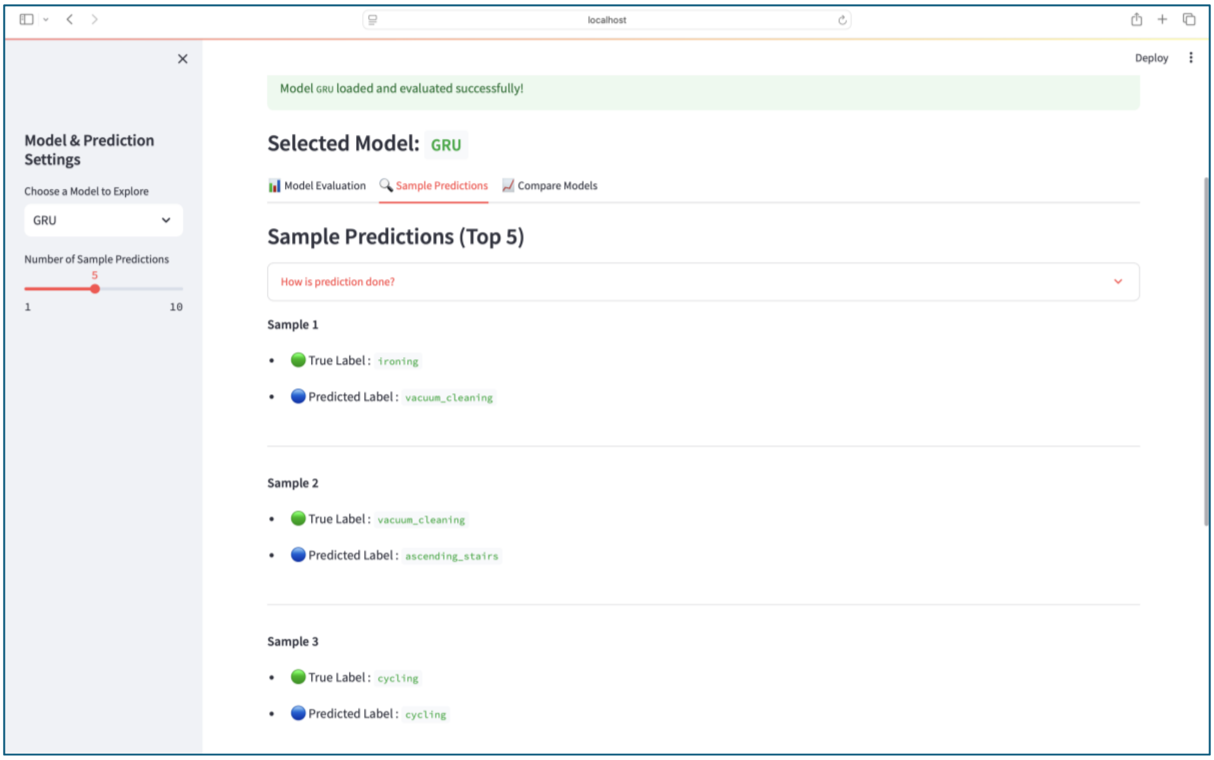
- **Screenshots of Working UI:**

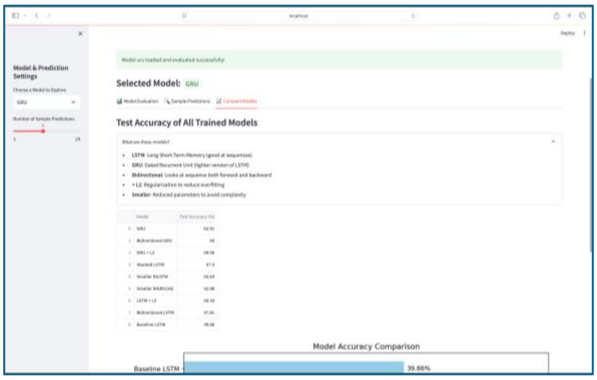




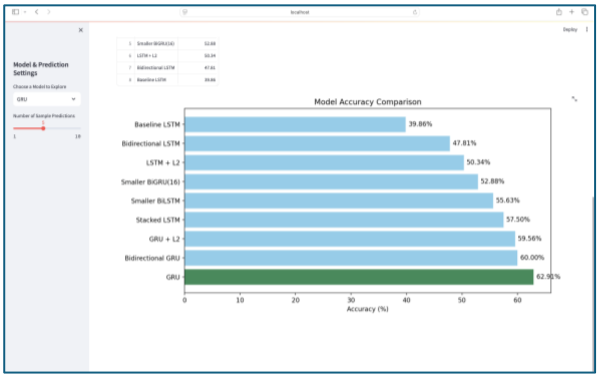


	precision	recall	f1-score	support
lying	0.692317	0.993950	0.816156	49756.000000
sitting	0.968677	0.600852	0.741664	35205.000000
standing	0.941790	0.524680	0.673916	50911.000000
walking	0.549706	0.994100	0.707942	65253.000000
running	0.545970	0.828323	0.658141	20224.000000
cycling	0.888593	0.943121	0.915046	48155.000000
Nordic_walking	0.612780	0.177772	0.275593	57613.000000
ascending_stairs	0.251432	0.261925	0.256571	29329.000000
descending_stairs	0.358944	0.226202	0.277517	21273.000000
vacuum_cleaning	0.479201	0.651819	0.552337	45844.000000





Model & Prediction Settings
Selected Model: GRU
Test Accuracy of All Trained Models
LSTM: Long Short-Term Memory (good at sequences)
Model Accuracy Comparison
Baseline LSTM
39.86%



Model & Prediction Settings
Model Accuracy Comparison
Baseline LSTM
39.86%
Bidirectional LSTM
47.81%
LSTM + L2
50.34%
Smaller BiGRU(16)
52.88%
Smaller BiLSTM
55.63%
Stacked LSTM
57.50%
GRU + L2
59.56%
Bidirectional GRU
60.00%
GRU
62.91%
0
10
20
30
40
50
60
Accuracy (%)