Deep Learning – Assignment 1: Single-Word Audio Classification (20 marks)

🎯 Objective

Build a **Feedforward Neural Network (FNN)** to classify three spoken words from audio recordings.

🎙️ Audio Collection (4 marks)

- Record **10 individuals** (family/friends/neighbours)
- Each says **3 words**:
  → *"Subhanallah"*, *"Alhamdulillah"*, *"Allahuakbar"*
- **At least 3 recordings per word/person**
  → 3 words × 3 rounds × 10 people = **90 audio files**
- Preferred Format: WAV preferred, consistent sample rate (e.g., 16kHz)

📁 Dataset Organization & Splitting (2 marks)

- Organize files in folders based on word labels. Suggested dataset organization folder:

```
AUDIO_CLASSIFICATION/
├── TRAINING_DATASET/
│   ├── SUBHANALLAH/ | ALHAMDULILLAH/ | ALLAHUAKBAR/
└── TESTING_DATASET/
    ├── SUBHANALLAH/ | ALHAMDULILLAH/ | ALLAHUAKBAR/
```

- Manually split: **80% training**, **20% testing**

🛠️ Preprocessing & Data Handling (4 marks)

- Implement audio feature extraction (e.g., MFCC)
- Create a custom Pytorch DataLoader

🧠 Feedforward Neural Network Implementation (6 marks)

- Build an **FNN model**
- Train using appropriate loss function & optimizer
- Implement training loop and track accuracy/loss
- Use a suitable activation function and network depth

Evaluation & Results (4 marks)

- Evaluate model performance on test data
- Display:
  - **Overall accuracy**
  - **Confusion matrix**
  - **Training/validation loss/accuracy curves**
- Include **discussion of results** in Markdown cells

📦 Submission

- Submit a **PDF** version of your Jupyter Notebook (include code, outputs, and discussion)
- Include the **Google Drive link** to your audio dataset in the notebook
- **Deadline**: **2 May 2025**