**COE 691 – *Special Topics in Computer Engineering I: Explainable Artificial Intelligence***

**Assignment 1:**Dyslexia Detection with Classical Machine Learning Classifiers

**Dataset**:                         kagglehub.dataset\_download("luzrello/dyslexia")  : use [Dyt-desktop.csv]

**Link:**                             https://www.kaggle.com/datasets/luzrello/dyslexia

**Deadline:**                       4th October 2025

**Objective**

The goal of this assignment is to apply different classification methods to the Dyslexia dataset and evaluate their performance using cross-validation. You will implement, tune, and compare models while demonstrating good practices for explainable machine learning.

**Instructions**

1. **Classifiers to Implement**
   * k-Nearest Neighbors (KNN)
   * Perceptron Learning Algorithm (PLA)
   * Support Vector Machine (SVM)
   * Random Forest (RF)
   * Gradient Boosting (GB)
2. **Evaluation Strategy**
   * Perform **Stratified 10-Fold Cross-Validation**.
   * Report results **for each fold** (Accuracy, Precision, Recall, F1-score).
   * Provide **AUC-ROC curves (averaged)** across folds.
3. **Preprocessing**
   * Apply **any normalization technique** (e.g., MinMax, StandardScaler, RobustScaler).
   * Justify and report your chosen technique.
4. **Model-Specific Analysis**
   * **KNN**: Plot **F1-score vs. K**.
   * **SVM**: Identify and highlight **support vectors** (from any one fold).
   * **All models**: Report implementation details, such as:
     + KNN: number of neighbors (K), distance metric.
     + PLA: learning rate, iterations.
     + SVM: kernel type, C value, gamma.
     + Random Forest: number of trees, max depth, feature sampling strategy.
     + Gradient Boosting: number of estimators, learning rate, max depth.
5. **Comparison and Visualization**
   * Provide a **summary table** with **mean values** of Accuracy, Precision, Recall, F1-score across all folds.
   * Provide **bar plots of mean F1-scores** across classifiers.
6. **Report**
   * Write your findings in **ICASSP (IEEE Conference) LaTeX format**.
   * Structure should include:
     + **Abstract**
     + **Introduction** (short background on dyslexia & ML classifiers)
     + **Methodology** (dataset, preprocessing, classifiers, cross-validation)
     + **Results** (tables, plots, ROC curves)
     + **Discussion** (interpret findings, best model, challenges)
     + **Conclusion** (what you learned, potential improvements)
   * Figures and tables must be referenced in the text.

**Deliverables**

1. **Code notebook(s)** (well-documented, with comments and explanations).
2. **Final Report (PDF in ICASSP LaTeX template)**. (not more than 4 pages)

**Grading Rubric (Total: 20 Marks)**

| **Criterion** | **Marks** | **Notes** |
| --- | --- | --- |
| Implementation of 5 classifiers with correct setup | 5 | Proper cross-validation, parameter details reported |
| Evaluation metrics (per fold + mean summary) | 4 | Accuracy, Precision, Recall, F1, ROC |
| Model-specific analysis (KNN plot, SVM SVs, etc.) | 3 | Must show evidence |
| Visualization & tables (summary table + bar plots) | 2 | Clear, properly labeled |
| Report in ICASSP template | 4 | Structure, clarity, formatting |
| Code quality & documentation | 2 | Readable, reproducible |

📌 **Note**: Focus on **maximizing F1-score**, not just accuracy. Explain all choices clearly.