Project Title

Class : _____
Project Participants : _____
Student ID : _____
Date : _____

NOTE: Delete the yellow highlighted regions below and ADD YOUR TEXT

1 Introduction

DEADLINE: SEPT 2nd, 2025 – 5pm

General Problem Description

- Briefly outline the main problem or challenge addressed by the project.

What Makes This Project Unique?

- **Databases (DBs):** Which databases are utilized?
- **Algorithms:** Are any algorithms sourced online? Which ones?
- **Solution/Approach:** Describe the methodology.
    - Did you implement code from scratch or adapt existing solutions (e.g., a notable paper like a CVPR Best Paper)?

Your Contributions

- **Databases:**
    - Did you collect your own data?
    - Are you using an existing or publicly available database?
    - Did you generate new data from online sources or pre-existing datasets?
    - Have you extended an existing database with additional data for testing or other purposes?
- **Code:**
    - Is the code entirely your own creation?
    - Are you leveraging toolboxes, such as MATLAB databases, OpenCV, or other resources (e.g., GitHub)? Be specific and provide citations.
    - Are you using someone else's code?
        - If so, are you utilizing it as-is, improving it, or combining it with other algorithms?
    - Describe your unique approach in integrating or modifying existing tools/algorithms.

2 Literature Review

DEADLINE: SEPT 20th, 2025 – 5pm

- **Find no less than 10 papers** most relevant to what you are trying to work on:

Example format expected for the review of your papers:

- Kong and Zhang [1] proposed several approaches to reduce eyelash and reflection noise on iris segmentation. The eyelash detection model was based on separable eyelash condition, non-informative condition and connective criterion. The reflection detection was based on strong and weak reflection tests. They concluded that the proposed noise reduction methods are effective in terms of detecting eyelash and reflections.
- Huang, Wang, Tan and Cui [2] explored an iris segmentation method using phase congruency. Iris is firstly localized and normalized to a rectangular block with fixed size. Then edge information is extracted based on phase congruency by a bank of Log-Gabor filters. Finally, the edge information of noise region is infused to iris segmentation, including eyelash, eyelid, reflection and pupil.
- Liu, Bowyer and Flynn [3] developed an improved segmentation algorithm based on Masek's algorithm. The detection order is reversed by segmenting the pupil boundary first. Then some edge points not from the iris boundary are eliminated in order to find correct iris boundary using Hough transform. Instead of each edge point voting in all directions, the proposed algorithm only let each edge point vote for 30 degrees on each side of local normal direction. Finally, hypothesize is performed to ensure the accuracy of the algorithm.

3 METHODOLOGY

Deadline: October 10th, 2025 – 5pm

3.1.1 Tools

- **Programming Languages and Tools:**
  Specify the programming languages (e.g., Python, MATLAB, R) and libraries/toolkits (e.g., scikit-learn, WEKA, NumPy, pandas) used.
- **Algorithms:**
  Provide a brief description of the machine learning algorithms applied.
  - Example: Explain their role in feature extraction, dimensionality reduction, classification, regression, clustering, or statistical modeling.
- **Overall Approach:**
  Outline the process flow clearly, from raw input to the final output.
  - Example: Input data → Pre-processing → Feature Engineering → Algorithm Training → Evaluation.
  - A **flowchart** is strongly recommended to visualize the workflow.
- **Significance of the Approach:**
  Explain why the chosen methodology and algorithms are effective for the problem at hand (e.g., material classification, predictive modeling).
  - Highlight advantages such as interpretability, efficiency, robustness, or better alignment with problem requirements compared to other methods.

4 EXPERIMENTS AND RESULTS

Deadline: October 31st, 2025 – updated in the follow-up report by November 30th, 2025

4.1.1 Database Description

- Provide details about the dataset, including:
  - Number of samples and distribution across categories.
  - Data diversity (e.g., conditions, resolutions, noise, class balance).
  - Preprocessing steps (e.g., normalization, feature scaling, handling missing values).

4.1.2 Experiment Details

- Clearly describe each experiment:
  - Objective (e.g., testing classification accuracy, evaluating clustering quality).
  - Setup (e.g., cross-validation, training/testing split, stratified sampling).
  - Parameters tested (e.g., number of neighbors in k-NN, kernel types in SVM, number of trees in Random Forest).

4.1.3 Error Analysis

- Analyze factors affecting performance, such as:
  - The influence of data preprocessing (scaling, normalization).
  - Impact of imbalanced datasets or noisy samples.
  - Differences in algorithm performance across categories.

4.1.4 Additional Aspects for Focus

1. **Feature Engineering and Selection:**
   - Describe how features were extracted, selected, or transformed.
   - Discuss dimensionality reduction techniques (e.g., PCA, LDA).
2. **Algorithm Tuning:**
   - Explain how parameters (e.g., C for SVM, depth of decision trees) were optimized.
   - Mention approaches like grid search, random search, or cross-validation.
3. **Comparative Analysis:**
   - Compare results across multiple algorithms or benchmarks.
   - Highlight trade-offs between accuracy, interpretability, and computational cost.
4. **Implementation Challenges:**
   - Document issues faced (e.g., overfitting, computational constraints).
   - Explain how these challenges were mitigated.
5. **Reproducibility:**
   - Ensure clear documentation of datasets, preprocessing steps, algorithm parameters, and code for reproducibility.
6. **Evaluation Metrics:**

- Describe evaluation metrics (e.g., accuracy, precision, recall, F1-score, ROC-AUC, clustering indices).
- Explain why these metrics are appropriate.

7. **Scalability and Application:**
   - Discuss how the solution scales to larger datasets or adapts to different applications.
   - Highlight real-world scenarios where the method could be deployed.
8. **Future Work:**
   - Suggest extensions such as testing with larger datasets, alternative algorithms, or hybrid approaches.
   - Identify limitations of the current experiments.

5 FINAL REPORT – CONCLUSIONS

Deadline: November 30th, 2025

5.1.1 Additional Experiments and Refinements

- Summarize any new experiments conducted to refine algorithm performance.
- Discuss parameter optimization and results from trying different algorithms or ensembles.

5.1.2 Additional Testing and Evaluation

- Describe extended testing efforts, such as:
  - Performance on unseen/real-world datasets.
  - Robustness under noisy or incomplete data conditions.

5.1.3 Insights from Experiments

- Present key findings on algorithm efficiency:
  - **Strengths:** Where performance was strong (e.g., high accuracy in balanced datasets).
  - **Weaknesses:** Where results fell short (e.g., difficulty handling imbalance).
- Compare results to initial expectations or baseline methods.

5.1.4 Additional Focus Areas for Conclusions

1. **Impact of the Work:**
   - Broader implications for the field or specific application area.
2. **Lessons Learned:**
   - Challenges faced, unexpected results, and key takeaways.
3. **Limitations:**
   - Clearly describe shortcomings of the current methodology.
4. **Future Directions:**
   - Suggest improvements (e.g., hybrid ML + DL approaches, better feature engineering).
5. **Closing Statement:**

- Concise summary of project contributions and significance.
- Reiterate how the project advances understanding or application of **machine learning algorithms**.

6 APPENDIX

Deadline: November 30th, 2025

- **Code** – Submit by the last week before finals (as a CD or GitHub repository).
- **Demo** – To be presented in the final week before exams.