**Integrated Project - 1 (CS204)**



**PROJECT REPORT**

**On**

***Handwritten Digit Recognition System for Gurumukhi***

**BATCH-2022**

***“A Project on Handwritten Digit Recognition System for Gurumukhi Scripts”***

**Project Mentor Student Name & ID**

Dr. Pradeepta K Sarangi Shruti 2211981380

Sneha 2211981392

Shivani 2211981370

Rahul 2211981292

Introduction:

One of the many regional languages that are used to communicate with one another in India is known as Punjabi which is recognized as the official language of the Punjab (an Indian State), located in the northern region of India. Over fifty million people worldwide can communicate in the Gurumukhi language, which has a history that dates back more than a millennium and a half. Gurumukhi digits' complexity arises from each character's intricate design, which features numerous curves. Handwritten scripts exhibit increased complexity due to the unrestricted nature of freeform writing. Developing a machine learning model is crucial due to the vast number of historical documents in Gurumukhi scripts, some of which are over 1000 years old. Implementing a robust automation method facilitates the conversion of historical documents into a digital format, thereby ensuring the preservation of these valuable records.

**Objective(s):**

The main objectives of this project are:

* To understand the application areas of ML in real world.
* To implement ML for recognition of handwritten Gurumukhi digits.
* To create a sample data set of Gurumukhi numerals.

# Key Learnings:

To understand concepts associated with machine learning algorithms.

To recognize the practical benefits of machine learning.

To Design and implement various machine learning algorithms in a range of real-world applications.

# Driving Questions:

* What are the key features of the recognition system?
* How can machine learning and artificial intelligence be utilized for handwritten digit recognition in Gurumukhi scripts?

**Options to Execute:**

**Option 1*:* Web Scraping and Data Analytics**

Tech Stack: Python (Beautiful Soupier Scrappy for web scraping), Pandas for data manipulation.

**Option2*:* Machine Learning for Recommendation Systems**

Tech Stack: Python (scik it - learn, Tensor Flow, or Py Torch for ML), SQL for database management.

**Option3*:* Block chain for Transparency and Traceability**

Tech Stack: Ethereal or Hyper edge for block chain, Smart Contracts (Solidity),Web3.jsforintegration.

# Methodology:

Agile Methodology:

**Iterative Development:**

* Emphasize the iterative nature of Agile.
* Plan for incremental improvements in OCR accuracy.

**Flexible Adaptation:**

* Highlight the ability to adapt to changing requirements.
* Address challenges dynamically during development.

**Continuous Feedback:**

* Incorporate regular feedback loops.
* Engage stakeholders to refine OCR model performance.

**Collaborative Approach:**

* Stress collaboration among team members.
* Foster communication to enhance problem-solving.

**Prioritization of Features:**

* Use a backlog to prioritize OCR features.
* Focus on high-priority functionalities first.

# Individual Roles:

**Rahul -** Dataset Collection:

Source and curate diverse Gurumukhi Numerical datasets.

Ensure datasets are comprehensive and meet project requirements.

**Shivani -** Literature Research Specialist:

Conduct thorough literature research in Gurumukhi OCR.

Summarize existing studies and provide insights.

**Sneha -** Coding and Model Development:

Lead coding efforts for the Gurumukhi Numerical recognition model.

Optimize machine learning model using suitable frameworks.

Compile detailed reports on project progress and outcomes.

Compile detailed reports on project progress and outcomes.

**Shruti –** Data processing:

Ensure datasets are comprehensive and meet project requirements.

**Timeline:**

**1. \*Week 1-2: Problem Identification\***

- Define the scope and objectives of character detection.

- Conduct literature review on existing methods.

- Identify potential challenges and limitations.

**2. \*Week 3-4: Solution Framework\***

- Research and choose suitable algorithms for character detection.

- Develop a detailed solution framework.

- Plan for data collection and preprocessing.

**3. \*Week 5-10: Problem Implementation\***

- Implement the chosen solution framework.

- Test and iterate on the algorithm.

- Collect and preprocess data as needed.

- Address any unexpected challenges during implementation.

**4. \*Week 11-12: Report Writing\***

- Document the entire process, including methodology and results.

- Analyze the outcomes and discuss any limitations.

- Create clear visualizations to aid understanding.

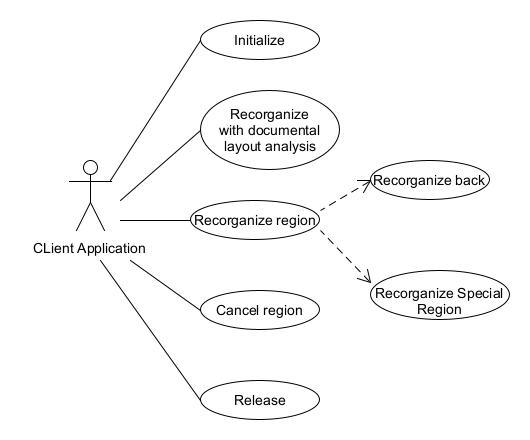
- Review and refine the report for clarity and completeness.

This timeline provides a structured approach to ensure a comprehensive and well-documented character detection project. Adjustments may be needed based on the project's complexity and any unexpected challenges that arise.

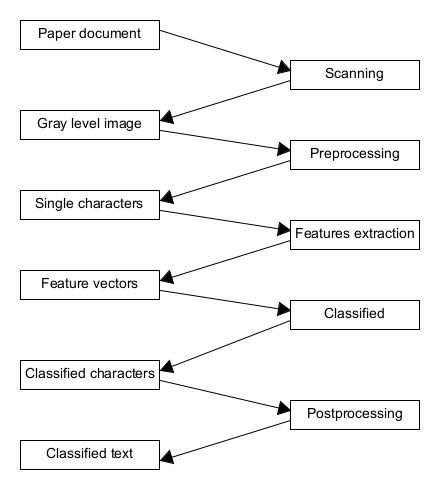
**Sustainable Development Goals:**

* **Gender Equality:** Ensure unbiased character detection, avoiding gender-based stereotypes.
* **Reduced Inequalities:** Develop inclusive algorithms to avoid biased related to race, ethnicity, or other characteristics.

**USE CASE DIAGRAM :**



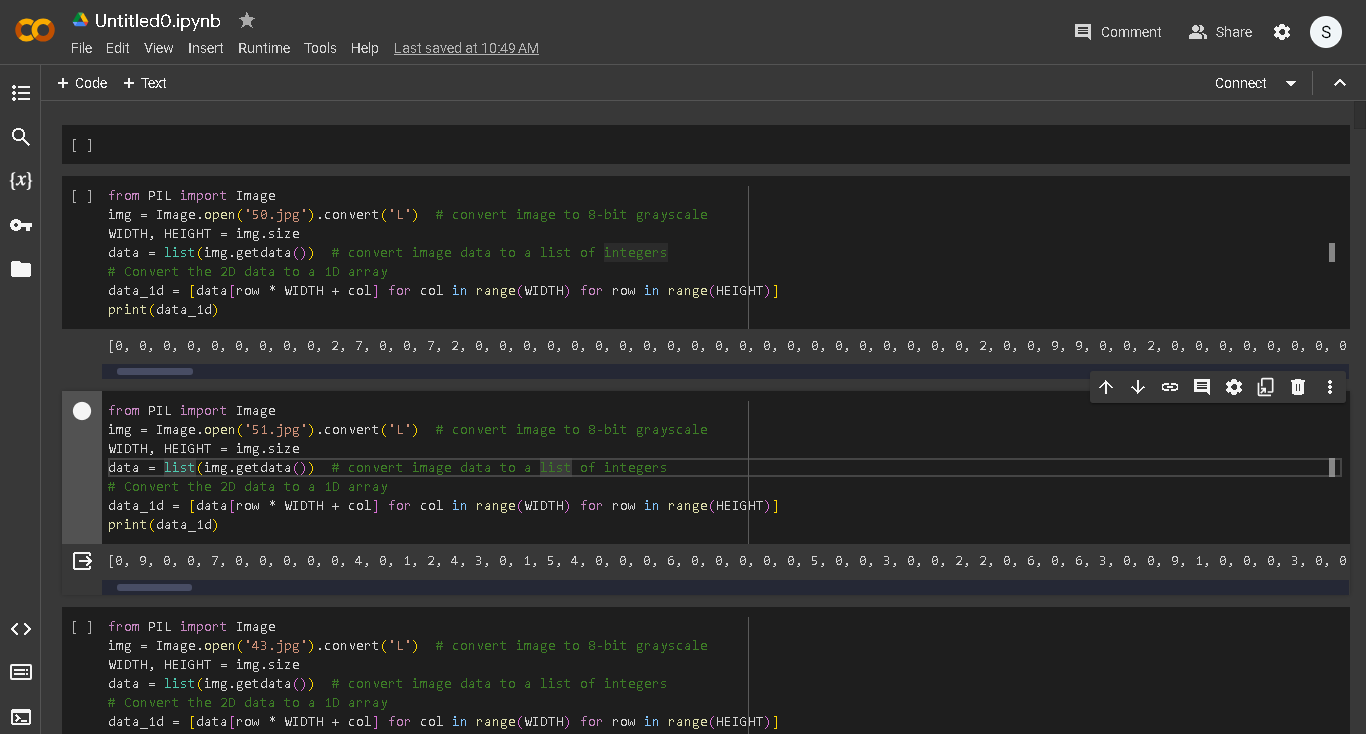
**CLASS DIAGRAM :**

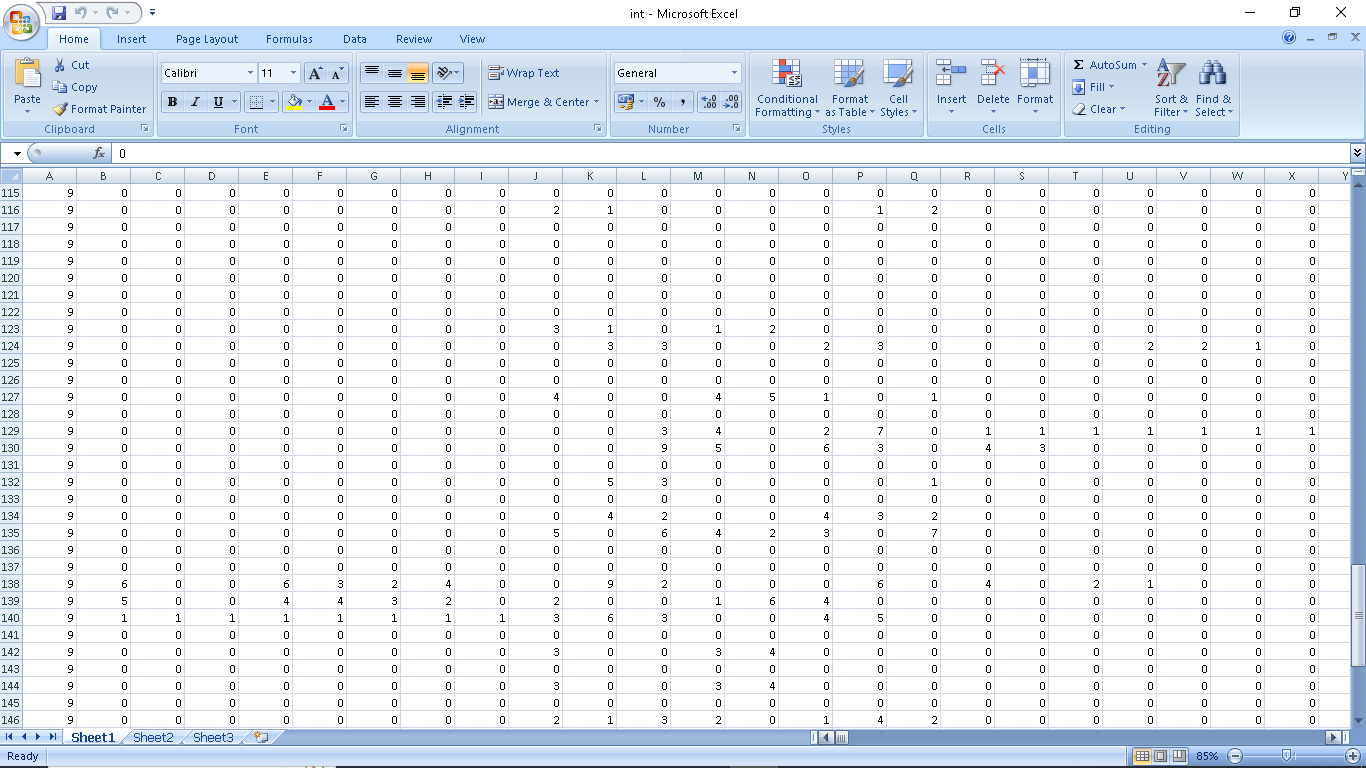


**SEQUENCE DIAGRAM :**

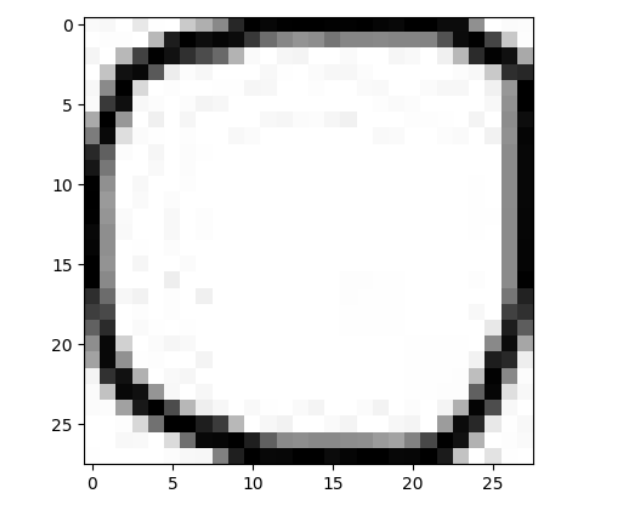


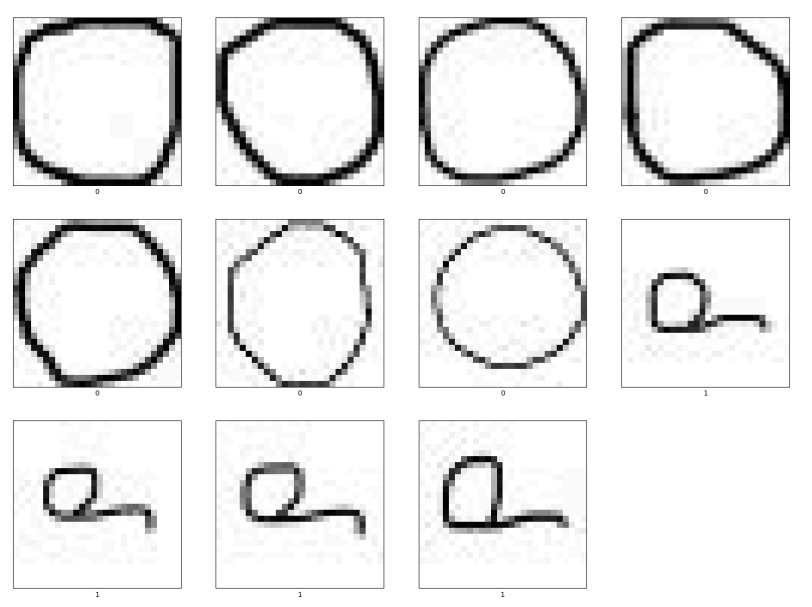
**PROTOTYPE PHOTOGRAPHS:**

****

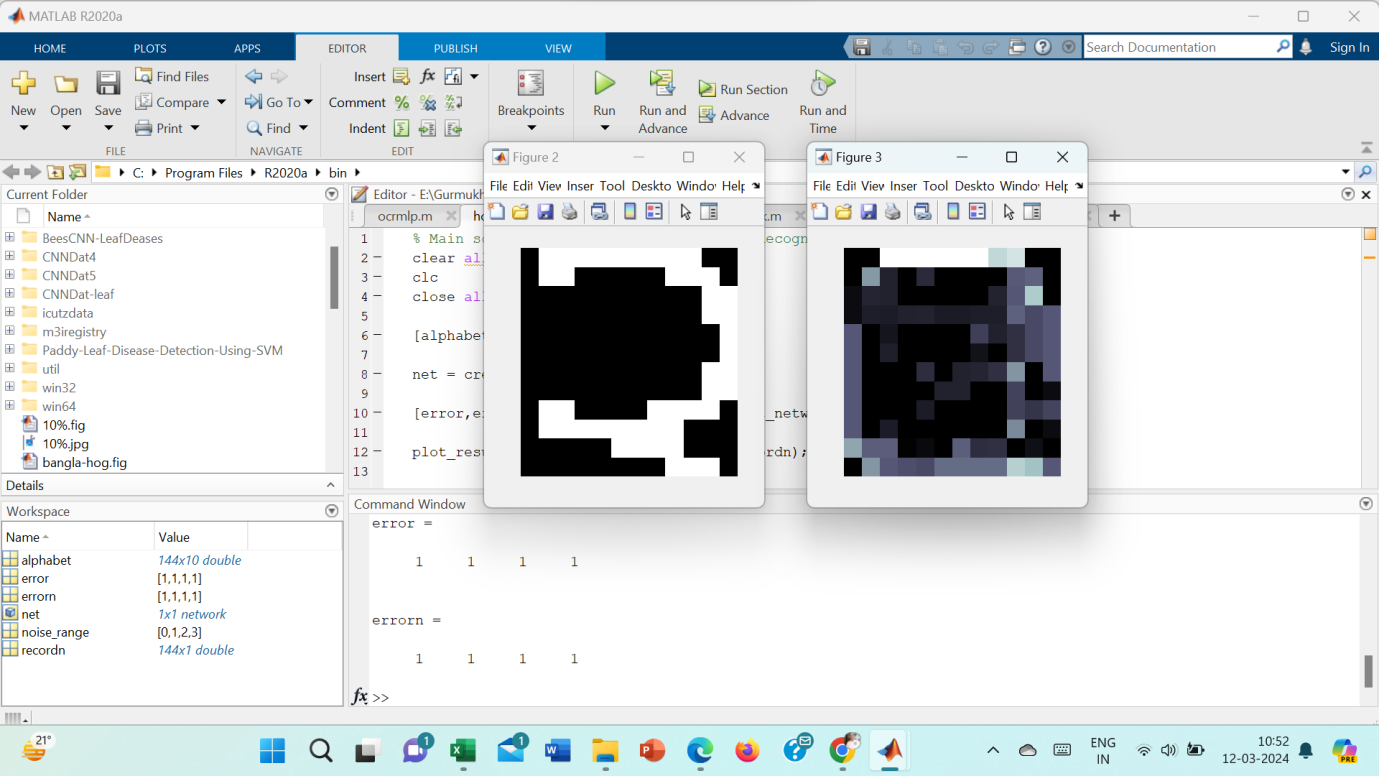
****

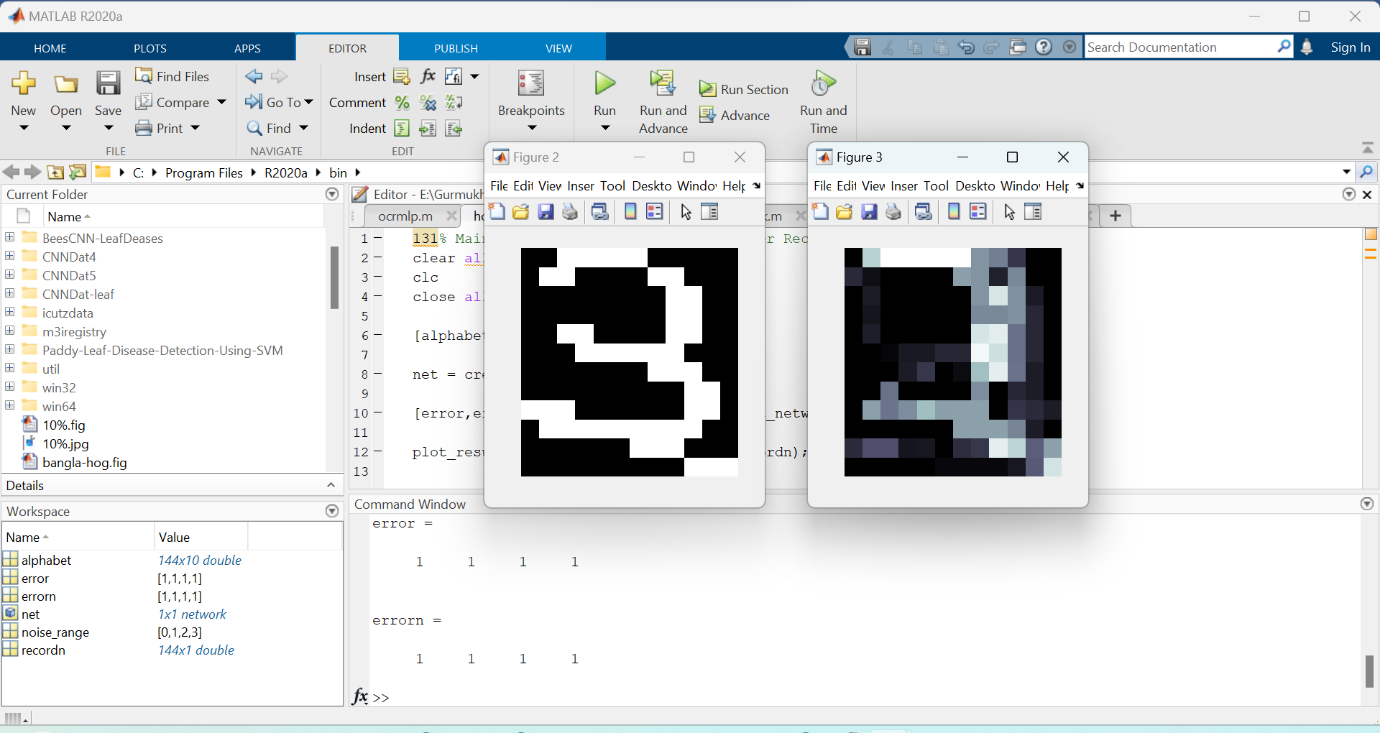
**These are the training shapes for the data:**

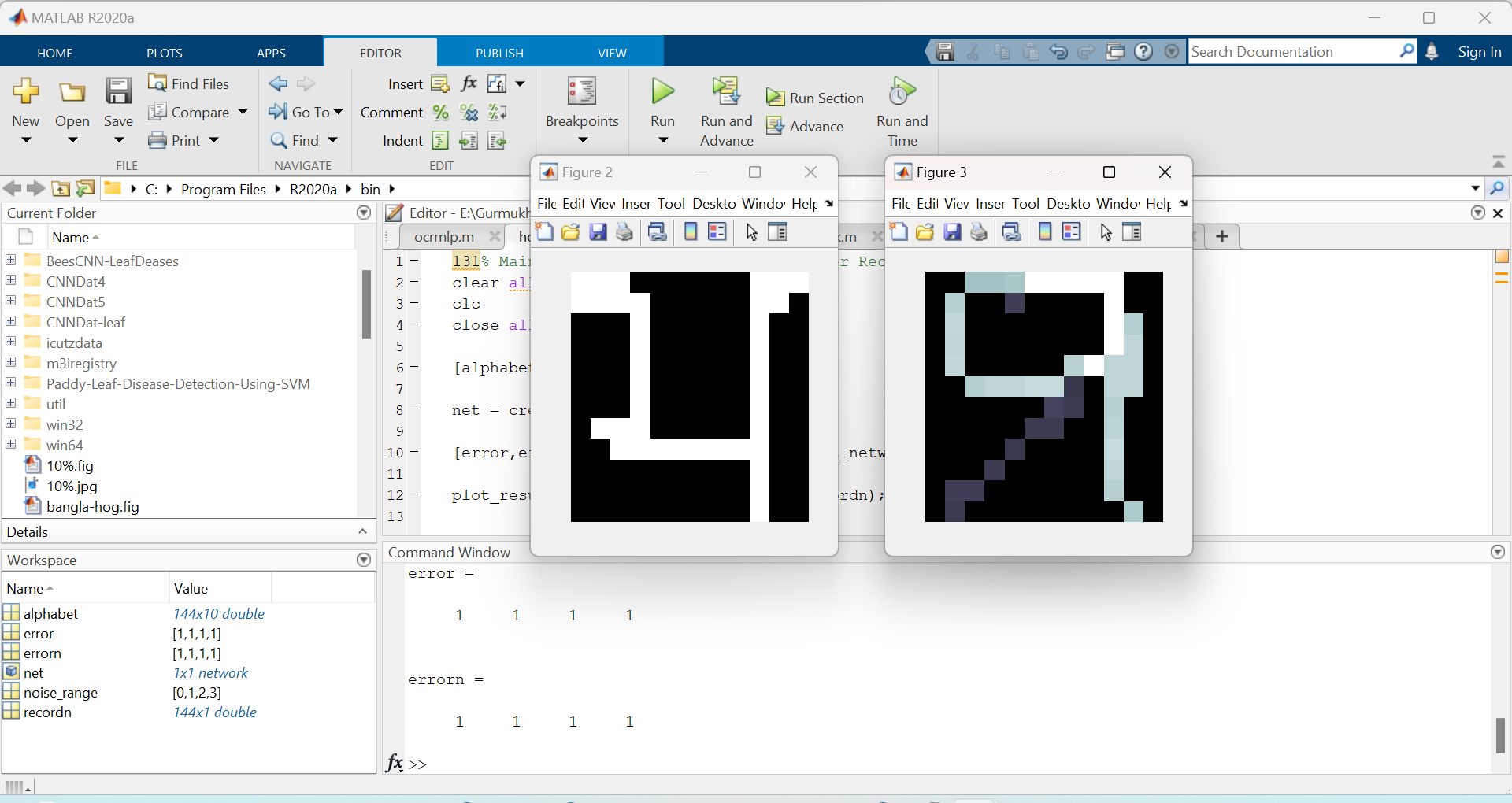
****

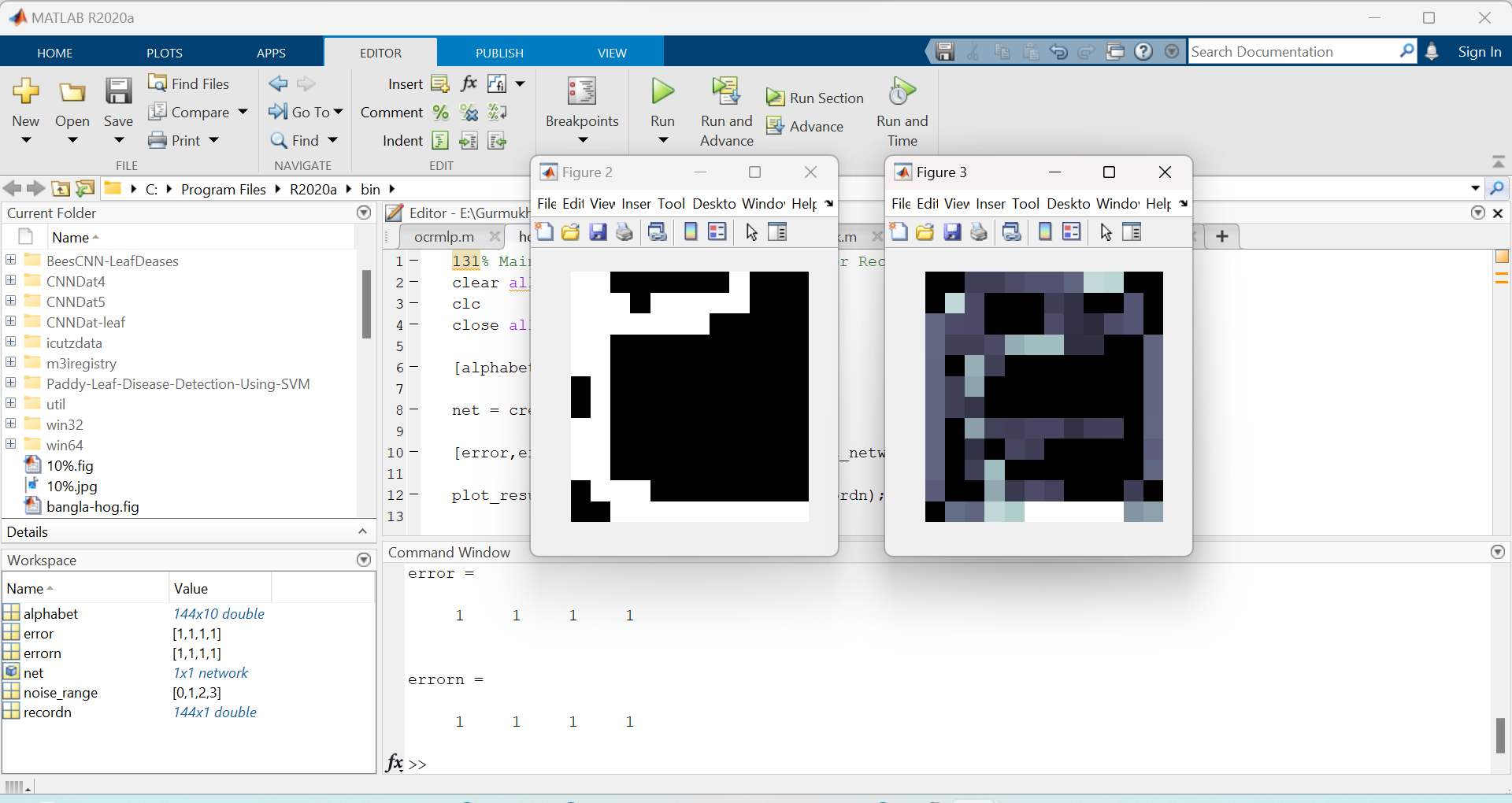
****

**The followings are the misclassifications of characters in Hopnet Model :**

****

****

****

****

**DIVISION OF WORK IN TEAM:-**

| **Name of student** | **ID** | **Role** |
| --- | --- | --- |
| **A Sneha** | **2211981392** | **Code execution and report writing** |
| **B Shivani** | **2211981370** | **Literature work** |
| **C Sneha** | **2211981392** | **Dataset processing** |
| **D Rahul** | **2211981292** | **Dataset collection** |

**Conclusion:-** In this project, CNN model is used for recognition of handwritten Punjabi character and the dataset is generated in ms paint . After the data pre-processing, the CNN model is implemented in Python. Total number of datasets are 1000, 100 for each digit. The overall accuracy is found to be 80%.