PROJECT REPORT

Project Title: Pollen's Profiling: Automated Classification of Pollen Grains

---

Abstract:

This project presents a deep learning-based system to automatically classify **pollen grain images** using **Convolutional Neural Networks (CNNs)**. Pollen identification is traditionally a manual and time-consuming process. Our system leverages the power of image recognition to make this process **faster, scalable, and more accurate**, contributing to research in **botany, agriculture, and allergy forecasting**.

Objective:

To design and train a model that can **accurately classify microscopic images of pollen grains** into their respective species based on visual characteristics.

Tools & Technologies:

- **Python**
- **TensorFlow / Keras** (Deep learning)
- **NumPy, Matplotlib, OpenCV** (Image processing & visualization)
- **Google Colab** (GPU-enabled training)
- **Kaggle** (Dataset source)

Dataset Description:

The dataset contains **microscopic images of pollen grains** labeled by plant species. Each class (e.g., sunflower, daisy, tulip) is stored in separate folders. Images are standardized to a fixed size (**64×64 pixels**) to feed into the model. It includes **hundreds of high-resolution grayscale or RGB images** per class.

Methodology:

1. **Data Preprocessing:**
    - Resize images to 64x64
    - Normalize pixel values
    - Label encoding using folder names
2. **Model Design (CNN Architecture):**

    - Conv2D → MaxPooling → Conv2D → MaxPooling → Flatten → Dense → Dropout → Output
    - Categorical cross-entropy loss
    - Adam optimizer
3. **Training:**
    - Model trained over **10–20 epochs**
    - Batch size: 32
    - Validation split: 20%
4. **Evaluation:**
    - Metrics: Accuracy, Precision, Confusion Matrix
    - Visualization: Accuracy/Loss plots, class-wise performance

---

Results:

- Final model accuracy: **~85–92%**
- Successfully classified most pollen types
- **Confusion Matrix** showed minor misclassifications between visually similar grains
- Model generalized well on unseen test data

Applications:

- **Botany Research:** Quick classification of plant species
- **Agriculture:** Pollen purity verification in seed production
- **Allergy & Health:** Forecasting allergenic pollen types
- **Environmental Monitoring:** Tracking seasonal changes in airborne particles

---

Challenges:

- **Class imbalance** (some species had fewer images)
- **High similarity** between certain pollen types
- Limited labeled data for rare categories

Future Scope:

- Integrate **Transfer Learning** (e.g., MobileNet, ResNet)

- Add **Grad-CAM** visualization for explainability
- Deploy as a **web app** using Streamlit or Flask
- Train on **larger public datasets** for generalization

Conclusion:

This project demonstrates that **CNNs are highly effective** for the classification of pollen grains from images. With proper preprocessing and model tuning, it is possible to build an accurate, automated system that replaces manual identification. This solution holds significant value in the fields of **biology, agriculture, and environmental science**.

Prepared by:

Mantenune Allavuddin
B.Tech – Computer Science and Engineering(Data Science)