PROJECT REPORT

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

Objective:

This project aims to **automate the classification of pollen grain images** using image processing and deep learning. Accurate identification of pollen is crucial in **botany, agriculture, allergy research, and environmental monitoring**. By using **Convolutional Neural Networks (CNNs)**, this system classifies images of pollen grains into their respective plant species, enabling **faster and more accurate identification**.

Dataset Description:

The dataset used is the **Pollen Dataset for Deep Learning Classification** from Kaggle. It contains microscopic images of pollen grains organized by plant species in separate folders. Each folder (class) represents a different plant like **sunflower, lily, or rose**. Images are standardized to **64x64 pixels** for input into the neural network.

Tools and Technologies:

- **Python** (NumPy, OpenCV, Matplotlib, Seaborn)
- **TensorFlow/Keras** for model development
- **Google Colab** (GPU-enabled training)
- **Kaggle** (for dataset access)

Methodology:

- **Data Preprocessing**: Images resized to 64x64 and normalized. Labels extracted from folder names and one-hot encoded.
- **Model Architecture**: CNN with Conv2D, MaxPooling, Flatten, Dense, and Dropout layers.
- **Training**: 10–15 epochs with Adam optimizer and categorical crossentropy loss.
- **Evaluation**: Accuracy and loss monitored. Final test accuracy ~85%.

Results:

- Achieved **high accuracy and low validation loss**.
- **Confusion matrix and classification report** used for performance evaluation.
- Misclassifications mainly due to **visual similarity** and **low image quality**.

Use Cases and Impact:

- **Botany**: Supports plant species identification.
- **Environmental Health**: Aids airborne pollen tracking for allergy alerts.
- **Agriculture**: Assists in quality control and species verification.

Challenges:

- Similar visual structure among some pollen types.

- **Imbalanced dataset** affects rare class performance.
- Improvements possible using **data augmentation** or deeper models.

Future Scope:

- Apply **transfer learning** (e.g., VGG16, MobileNet).
- Add explainability via **Grad-CAM heatmaps**.
- Deploy using **Streamlit** for live demo or web app.

Conclusion:

This project demonstrates that **CNNs are highly effective** for pollen grain classification. With well-prepared data and an efficient architecture, the system achieved strong results. This approach can support scalable, real-time identification for research, health, and agriculture sectors.

End of Report

Prepared by:

Mantenune Allavuddin