Skin Cancer Classification Proposal

Proposal: Development of a Computer Vision and Machine Learning System for Skin Cancer Classification

Introduction

1. Skin cancer is one of the most common forms of cancer, and early detection plays a crucial role in successful treatment. Leveraging computer vision and machine learning technologies can greatly enhance diagnostic efficiency by providing accurate classification of skin cancer types, specifically distinguishing between benign and malignant cases. This project aims to develop a robust classification system using logistic regression and K-Nearest Neighbors (KNN) classifiers on a curated dataset of skin lesion images. The system will serve as a stepping stone toward more advanced AI-based diagnostic tools in dermatology.

Objectives

1. Develop a pipeline for loading, preprocessing, and splitting skin lesion image datasets.
2. Implement machine learning models, specifically Logistic Regression and KNN, to classify skin cancer types.
3. Evaluate model performance using standard metrics, including accuracy, classification reports, and confusion matrices.
4. Provide a user-friendly visualization of sample images and results for better interpretability.
5. Create a scalable framework for incorporating more advanced models in the future.

Scope

2. The project focuses on binary classification of skin cancer into benign and malignant categories. The dataset used will be processed to ensure balanced class representation and high-quality input images. Two machine learning algorithms will be explored:
3. **Logistic Regression** for its simplicity and interpretability.
4. **KNN Classifier** for its ability to capture complex decision boundaries with minimal assumptions.

Dataset

5. The dataset consists of high-quality labeled images stored in separate directories for training and testing. Images are preprocessed to normalize pixel values and resize to a uniform dimension of 64x64 pixels. Data augmentation techniques may be employed in future iterations to enhance model robustness.

Methodology

1. **Data Preprocessing:**
    - Normalize pixel values to a [0, 1] range for consistent input.
    - Resize all images to 64x64 dimensions.
    - Implement flow-from-directory loaders for efficient batch processing.

2. **Visualization:**
    - Display a set of sample images from each class to understand data distribution.

3. **Model Training:**
    - Train Logistic Regression with a maximum iteration limit of 1000 to ensure convergence.
    - Train KNN with k=3 to evaluate neighborhood-based decision-making.
4. **Evaluation Metrics:**
    - Accuracy
    - Confusion Matrix
    - Classification Report (precision, recall, F1-score)
5. **Results Interpretation:**
    - Generate confusion matrices with heatmaps to visualize misclassifications.
    - Provide detailed classification reports to understand model performance on individual classes.

Tools and Libraries

- **Programming Language:** Python
- **Libraries:**
    - TensorFlow/Keras for image preprocessing
    - Scikit-learn for model training and evaluation
    - Matplotlib and Seaborn for data visualization

Deliverables

1. A trained and evaluated skin cancer classification system.
2. Visualizations of sample images and results.
3. Comprehensive documentation of the project, including methodology, results, and future recommendations.

Timeline

6. The project is divided into the following phases:

1. Data Loading and Preprocessing: 2 weeks
2. Model Implementation and Training: 3 weeks
3. Model Evaluation and Visualization: 2 weeks
4. Documentation and Final Deliverables: 1 week

Budget

7. The budget includes resources for computing infrastructure, dataset acquisition, and tools for project execution. Details can be adjusted based on organizational requirements.

Conclusion

8. This project is a step forward in leveraging AI technologies for medical diagnosis, providing a scalable and interpretable framework for skin cancer classification. By focusing on binary classification initially, we aim to build a strong foundation for future advancements in this field.