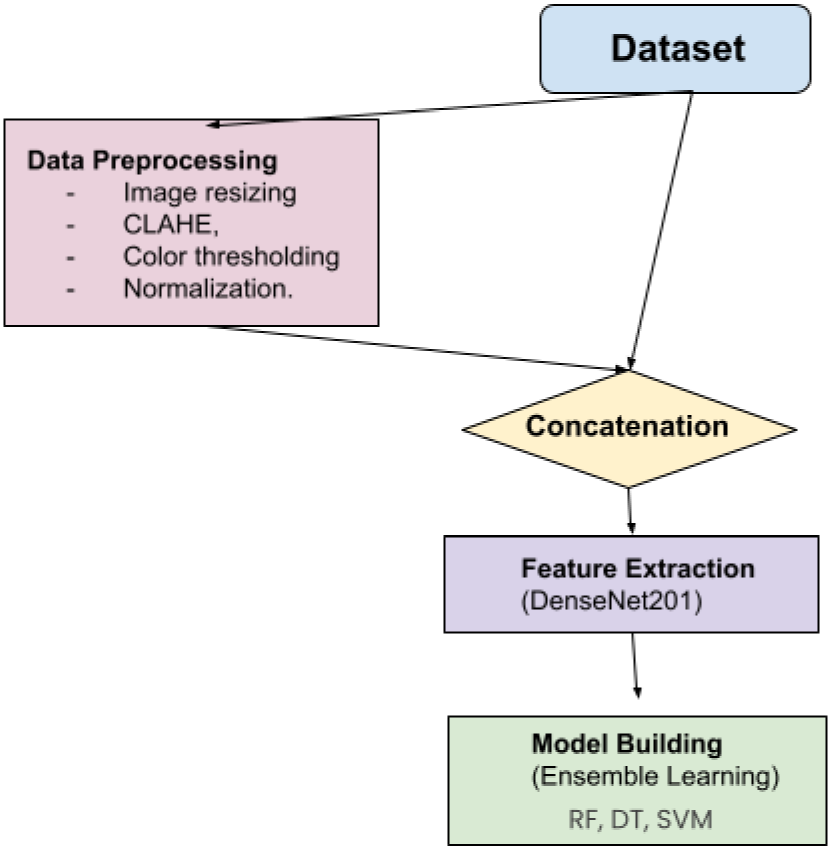
Dataset	Approch	CNN Architecture	Val Accuracy	Test Accuracy
Wheat Disease Detection Link	ADASYN+ CLAHE+ XGBOOST	ResNet50V2	0.95755....	0.90094....
		Xception	0.94275....	0.83310719....
		DenseNet121	0.986417.....	0.9565807.....
		Inception	0.92869.....	0.8778833...
		Inception_Res Net	0.93088.....	0.85074....
		DenseNet201	0.989813....	0.964721...
		MobileNetV2	0.9643463..	0.941655359...
		DenseNet169	0.988115....	0.9592944...
	SMOTETOMAKE + CLAHE + XGBOOST	DenseNet201	0.9864176.....	0.9667869...
	SMOTETOMAKE + XGBOOST (colsample_bytree =1.0, gamma=0.2, learning_rate=0.1, max_depth=5, n_estimators=200, subsample=0.8, random_state=42) + Contrast stretching	DenseNet201	0.972835...	0.948439...
	SMOTETOMAKE + XGBOOST + CLAHE + Color thresholding	DenseNet201	0.981324....	0.987788...
	SMOTETOMAKE	DenseNet201	0.9915110......	0.990502......

	+ XGBOOST + (CLAHE + Color thresholding)+ Origine data			
	SMOTETOMAKE + ADABOOST + (CLAHE + Color thresholding)+ Origine data	DenseNet201	0.9779286.....	0.97557666....
	Ensemble learning + (CLAHE + Color thresholding)+ Origine data	DenseNet201	0.9932088.....	0.99864314.....
Wheat Leaf Dataset Link	Ensemble learning + (CLAHE + Color thresholding)+ Origine data	DenseNet201	/	0.975609756097561



1. Data Preprocessing

- **Image Loading and Resizing**:
  - Images are loaded from the specified directories (train_dir and test_dir).
  - Images are resized to a fixed size (IMG_SIZE = (224, 224)) to ensure uniformity.
- **CLAHE (Contrast Limited Adaptive Histogram Equalization)**:
  - Applied to enhance the contrast of the images, particularly useful for improving the visibility of subtle features in the images.
- **Color Thresholding**:
  - Used to remove the green background from the images, isolating the wheat plants for better feature extraction.
- **Normalization**:
  - Pixel values are normalized to the range [0, 1] by dividing by 255.0.

2. Feature Extraction

- **DenseNet201 (Pre-trained CNN)**:
  - A pre-trained DenseNet201 model (trained on ImageNet) is used as a feature extractor.
  - The model is modified to exclude the top classification layers, and a GlobalAveragePooling2D layer is added to reduce the spatial dimensions to a fixed-size feature vector.
- **Feature Extraction for Original and Preprocessed Images**:
  - Features are extracted separately for both the original and preprocessed images.
- **Feature Concatenation**:
  - Features from the original and preprocessed images are concatenated to create a richer feature representation.

3. Model Building

- **Ensemble Learning**:
  - Three classifiers are used:
    1. **Random Forest (RF)**: A robust ensemble method that builds multiple decision trees and aggregates their predictions.
    2. **Decision Tree (DT)**: A simple tree-based classifier.
    3. **Support Vector Machine (SVM)**: A powerful classifier that finds the optimal hyperplane to separate classes.
  - These classifiers are combined using a **VotingClassifier** with hard voting, where the final prediction is based on the majority vote of the individual classifiers.

4. Training and Validation

- **Training**:
  - The ensemble model is trained on the concatenated features extracted from the training data.
- **Validation**:
  - The model is validated on a separate validation set to evaluate its performance and avoid overfitting.

    - Validation accuracy is calculated to assess the model's performance.

5. Testing and Evaluation

- **Testing**:
    - The trained ensemble model is tested on the test dataset.
- **Evaluation Metrics**:
    - **Accuracy**: Measures the proportion of correctly classified samples.
    - **Classification Report**: Provides precision, recall, and F1-score for each class.
    - **Confusion Matrix**: Visualizes the performance of the classifier by showing the true vs. predicted labels.

Strengths of the Methodology:

- **Preprocessing**: Enhances image quality and isolates relevant features.
- **Feature Extraction**: Leverages a powerful pre-trained CNN to extract meaningful features.
- **Ensemble Learning**: Combines multiple models to improve generalization and robustness.

Potential Improvements:

- **Hyperparameter Tuning**: Fine-tuning hyperparameters for the individual models (RF, DT, SVM) and the ensemble.
- **Cross-Validation**: Using cross-validation to ensure the model's stability and generalizability.
- **Advanced Ensembling**: Exploring stacking or boosting techniques for better performance.
- **Data Augmentation**: Adding more data augmentation techniques (e.g., rotation, flipping) to increase the diversity of the training data.