Methodology

This project follows a structured methodology for developing a deep learning model to classify waste using the BDWaste Dataset. The process is divided into key stages, including data preprocessing, model selection, training, evaluation, optimization, and final testing.

---

1. Data Collection and Preprocessing

- The BDWaste Dataset is used, containing 2,625 images categorized into biodegradable and non-biodegradable waste.
- Data Cleaning: Removing duplicate, corrupted, or poor-quality images to ensure a high-quality dataset.
- Data Augmentation: Enhancing dataset diversity by applying transformations such as rotation, flipping, brightness adjustment, and zooming.
- Oversampling for Balance:
    - Since some waste categories contain fewer images, oversampling techniques are applied to balance the dataset.
    - Augmentation-based oversampling ensures equal representation across all categories.
- Dataset Splitting: The dataset is divided into:
    - Training Set (70%)
    - Validation Set (15%)
    - Testing Set (15%)
- Image Resizing and Normalization:
    - All images are resized to 224x224 pixels to maintain consistency across models.
    - Pixel values are normalized between 0 and 1 for better training efficiency.

---

2. Model Development

The project explores different Convolutional Neural Network (CNN) architectures to identify the most effective model:

1. Custom CNN Model: A model designed from scratch with convolutional layers, batch normalization, activation functions, and dropout layers.
2. Pretrained CNN Models for Transfer Learning:
    - VGG16
    - ResNet50
    - EfficientNet
    - MobileNetV2

    Each of these models is fine-tuned to improve classification accuracy.

---

3. Model Training and Evaluation

- Loss Function: The model is trained using categorical cross-entropy loss, which is suitable for multi-class classification tasks.
- Optimization Techniques: Different optimizers are tested, including:
    - Adam
    - Stochastic Gradient Descent (SGD)
    - RMSprop
- Performance Evaluation Metrics:
    - Accuracy – Measures overall classification performance.
    - Precision, Recall, and F1-score – Assesses model effectiveness in identifying each waste category.
    - Confusion Matrix – Analyzes correct and incorrect predictions.
- Early Stopping and Model Checkpointing:
    - Early stopping is implemented to prevent overfitting by monitoring validation loss.
    - Checkpointing saves the best-performing model for further improvements.

---

4. Hyperparameter Tuning

The best-performing CNN model is further optimized by fine-tuning key hyperparameters:

- Adjusting Learning Rate: Testing values such as 0.1, 0.01, 0.001, and 0.0001.
- Experimenting with Batch Size: Evaluating different batch sizes (16, 32, 64) for optimal training stability.
- Comparing Optimizers: Selecting the best optimizer from Adam, SGD, and RMSprop.
- Tuning Dropout Rates: Experimenting with dropout values (0.2, 0.3, and 0.5) to reduce overfitting.
- Modifying CNN Filters and Dense Layers:
    - Testing different filter sizes (32, 64, 128).
    - Experimenting with fully connected layer sizes (128, 256, 512).
- Automated Hyperparameter Search:
    - Techniques like Grid Search, Random Search, and Bayesian Optimization are used to identify the best configurations.

---

5. Model Optimization

Once the final model is selected, additional optimization techniques are applied:

- Parameter Pruning: Removing unnecessary parameters to make the model more efficient.
- Quantization: Reducing model size while maintaining accuracy.
- Regularization Strategies:
    - Dropout and Batch Normalization help stabilize training.
    - Weight Decay prevents overfitting.

---

6. Final Testing and Validation

- The optimized model is evaluated using the test dataset to confirm its generalization ability.
- Final performance is assessed using:
  - Accuracy
  - Precision
  - Recall
  - F1-score
- The best model is finalized for deployment and further research.

---

7. Ethical Data Handling and Project Management

- Data Privacy: Ensuring responsible dataset usage and compliance with ethical guidelines.
- Version Control:
  - All project files, including datasets and training scripts, are maintained using GitHub.
  - Regular commits document key project updates.
- Bias Reduction: Implementing strategies to ensure fair classification across all waste categories.

---

8. Final Report and Submission

- A detailed research report is prepared, summarizing:
  - Methodology
  - Experiments
  - Results and key findings
- The project deliverables include:
  - A structured research document
  - A well-documented code repository
  - Final presentation and viva examination
