AI: BASED SCRATCH DETECTION

1. Problem Statement

The objective is to develop a prototype system for generic surface scratch detection using public datasets.

DATASET USED:

NEU Surface Defect Database (KAGGLE)

The dataset contains grayscale industrial surface images categorized into six defect classes:

1. Crazing

2. Inclusion

3. Patches

4. Pitted Surface

5. Rolled-in Scale

6. Scratches

Preprocessing:

- Images resized to 224 × 224

- Converted to RGB

- Normalized using ImageNet mean and standard deviation

- Dataset split into 70% training, 15% validation, 15% testing

Methodology Overview

Basically similarity matching aka pattern matching using cnn for embeddings

The approach consists of:

1. CNN-based feature extraction

2. Vector database indexing

3. k-Nearest Neighbour (kNN) classification in embedding space

Feature Extraction

A pretrained ResNet18 model was used as a feature extractor.

- The final fully connected classification layer was removed.

- The network outputs 512-dimensional embeddings representing surface texture patterns.

- Transfer learning enables strong representations with minimal training time.

This allows the system to learn visual similarity between surface defects rather than rigid class boundaries.

Vector Database & Pattern Matching

To enable efficient similarity-based inference:

- All training embeddings were L2-normalized

- A FAISS vector index using cosine similarity (inner product) was constructed

- During inference, test images are embedded and compared against the indexed vectors

Prediction Strategy:

- k-Nearest Neighbour search (k = 1)

- The label of the nearest embedding is assigned as the predicted class

This approach avoids overconfidence issues common in softmax classifiers and allows better generalization to unseen surface patterns.