Leveraging Pretrained Ensemble Models for Fine-grained Cereal Quality Inspection

Duvvuri Prathamesh-*23154005*, Dept. of Mining Engineering
Abhishrey Mishra-*23154001*, Dept. Mining Engineering
Vyas Giri- 22034017, Dept. of Ceramic Engineering
Aditya singh parihar - 22155012, Dept. of Mining Engineering

Paper Link

October 12, 2025

Introduction

Problem Definition & Motivation

- Manual Grain Appearance Inspection (GAI) is time-consuming — inspecting 60 g of wheat (1600 kernels) takes 25–30 minutes for a skilled inspector.
- Grains of the same species show extremely similar texture and color, making automated differentiation a **Fine-Grained Visual Categorization (FGVC)** problem.
- Existing deep learning approaches often fail to generalize well across different devices and lighting conditions.
- Our work leverages **transfer learning** and **ensemble learning** to develop a robust, scalable system for fine-grained grain inspection.

Other Related Work

Comparison of Prior Works on Grain Appearance Inspection (GAI)

Study / Year	**Approach & Dataset**	**Key Limitations**
Anami & Savakar (2009) [3]	Machine vision using color & texture for wheat & impurities detection	Low accuracy; limited to few wheat types
Zapotoczny (2011) [48]	Texture-based neural network on 11 wheat varieties	Small dataset; no damaged grain detection
Golpour et al. (2014) [14]	Color-feature neural network for rice classification	Focused only on rice; lacks generalization
Guzman et al. (2008) [15]	ANN for 5 rice varieties (Philippines dataset)	Small regional data; low transferability
Shantaiya et al. (2010) [39]	Pattern-based rice seed classification (6 varieties)	No damaged grain categories; limited scope
Pearson (2009) [33]	Device-based grain inspection imaging setup	Closed-source; dataset not publicly available
Qiu et al. (2018) [34]	Hyperspectral imaging with CNN for rice seeds	High-cost sensors; limited sample diversity
GrainSpace (Fan et al., 2022)	Large-scale dataset (5.25M images, 3 devices: P600/G600/M600)	Models trained from scratch; limited domain adaptation

Research Gap: Prior GAI studies were constrained by small or species-specific datasets, costly acquisition setups, and poor generalization, while even large-scale efforts like GrainSpace trained models from scratch without leveraging pretrained or ensemble-based methods—limiting robustness across devices, grain types, and data imbalance.

Related Work: Grainspace

Prior Research on Automated Grain Inspection

- **GrainSpace Dataset (Fan et al., 2022)** introduced 5.25 million images of wheat, maize, and rice kernels captured using three imaging systems (P600, G600, M600) across 30+ regions.

Model	Description
ResNet50	Residual CNN backbone for robust texture and structure feature extraction.
DCL	Fine-grained recognition model improving class separability via part-based learning.
Swin Transformer	Vision transformer capturing global context using shifted windows.

Problem Formulation

Formal Definition of the Task

- Let x_i denote an image of a single grain kernel, and $y_i \in \{1, 2, \ldots, N\}$ the corresponding class label.
- The objective is to learn a classification function $f_\theta(x_i) \rightarrow y_i$ that accurately predicts the class of each grain while remaining robust to **class imbalance** and subtle intra-class variations.

Optimization Objective:

$$\max_\theta \frac{1}{N} \sum_{n=1}^{N} F1_n$$

where $F1_n$ represents the **class-wise F1-score**, ensuring equal importance to both major and minor classes.

Model Architecture

- The model integrates two complementary pretrained networks — **ConvNeXt-Tiny** and **ResNet50**
- **ConvNeXt-Tiny:** A hierarchical convolutional-transformer hybrid pretrained on ImageNet. Captures long-range dependencies with large 7×7 kernels, layer normalization, and GELU activation.
- **ResNet50:** A deep residual CNN using skip connections:

$$y = F(x, \{W_i\}) + x$$

  providing stable gradients and texture-level representation.
- Both backbones extract features independently and output logits:

$$z_{conv} = f_{ConvNeXt}(x_i), \quad z_{res} = f_{ResNet}(x_i)$$

- Logit-level ensemble fusion:

$$z_{ens} = \frac{1}{2}(z_{conv} + z_{res}), \quad \hat{y}_i = \mathrm{softmax}(z_{ens})$$

- The ensemble enhances robustness under class imbalance and improves Macro-F1 .

Training Objective and Evaluation Metric

- The model is trained using the **Cross-Entropy Loss (CE)**, a standard objective for multi-class classification.
- Since the dataset is imbalanced, performance is evaluated using the **Macro-F1 Score** instead of accuracy.

1. Cross-Entropy Loss:

$$\mathcal{L}_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{i,c}\log(\hat{y}_{i,c})$$

where:

- $y_{i,c}$ – ground truth label for class c
- $\hat{y}_{i,c}$ – predicted probability for class c

Experimental Setup

Dataset:

- Wheat_R19–22_G600 subset of **GrainSpace** [Fan et al., 2022].
- High-quality industrial captures of normal and damaged kernels.
- Data split: 80% training, 10% validation, 10% testing.

Training Details:

- Framework: PyTorch (GPU-enabled Jupyter env. on Windows 11).
- Optimizer: SGD (momentum 0.9, weight decay 1×10^{-4}).
- LR: $0.01 \rightarrow$ decayed $\times$ 0.1 every 10 epochs (StepLR).
- Batch size $= 32$, Epochs $= 25$.

Implementation:

- Hardware: Intel Core i7 @ 1.80 GHz, 8 GB RAM.
- Logit-level averaging performed before softmax for inference.

Results

Quantitative Performance

Model	Macro-F1 (%)
ConvNeXt-Tiny (pretrained)	73.32
ResNet50 (pretrained)	74.67
Ensemble (ConvNeXt + ResNet)	**76.42**

- The **ensemble model** achieved the highest Macro-F1 score of **76.42%**, consistently outperforming both individual pretrained backbones (ConvNeXt-Tiny and ResNet50).
- The improvement of approximately **0.32%** over the GrainSpace benchmark demonstrates measurable gains in stability and recognition consistency.
- The average Macro-F1 score over 5 independent runs was **75.11%**, indicating consistent performance and low variation across trials
- The ensemble produced more balanced classification outputs, minimizing the bias toward dominant classes compared to standalone models.

Conclusion

Summary of Contributions

- Developed a pretrained ensemble model integrating ConvNeXt-Tiny and ResNet50 for fine-grained cereal quality inspection.
- Improved Macro-F1 score (76.42%) on the Wheat_R19–22_G600 subset of GrainSpace.
- Demonstrated that transfer learning + ensemble averaging enhance recognition under imbalanced data.

Future Work

- Extend evaluation to other subsets of the GrainSpace dataset (P600, M600) to study cross-device and cross-condition generalization of the ensemble model.
- Apply the same framework to different crop types such as maize and rice to verify adaptability across grain categories.
- Explore ways to improve computational efficiency without compromising accuracy, such as pruning or model quantization.

References

- L. Fan, Y. Ding, D. Fan, D. Di, M. Pagnucco, and Y. Song. *GrainSpace: A Large-scale Dataset for Fine-grained and Domain-adaptive Recognition of Cereal Grains.* arXiv preprint arXiv:2203.05306, 2022.
- H. Zhou, X. Liu, Y. Zhao, et al. *Recognition of Rice Seed Varieties Using Machine Vision and Deep Learning. Sensors*, 19(22):5100, 2019.
- H. Ponce, D. Lopez, and M. Martinez. *Grain Classification of Mexican Maize Varieties Using Computer Vision and Deep Learning. Computers and Electronics in Agriculture*, 167:105097, 2019.
- K. He, X. Zhang, S. Ren, and J. Sun. *Deep Residual Learning for Image Recognition.* In *CVPR*, 2016.
- Z. Liu, H. Mao, C.-Y. Wu, C. Feichtenhofer, T. Darrell, and S. Xie. *A ConvNet for the 2020s (ConvNeXt).* In *CVPR*, 2022.
- T. Dietterich. *Ensemble Methods in Machine Learning.* In *International Workshop on Multiple Classifier Systems*, 2000.