An Ensemble CNN Transfer Learning Model for Wheat Grain Classification

Małgorzata Charytanowicz[1,3][0000-0002-1956-3941] and Urszula Kużelewska[2][0000-0003-4612-7640] and Rafał Stęgierski[1][0000-0001-7225-3275]

[1] Department of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland
[2] Department of Information Systems and Computer Networks, Bialystok University of Technology, Wiejska 45a, 15-351 Bialystok, Poland
[3] Systems Research Institute, Polish Academy of Sciences, Newelska 6, 01-447 Warsaw, Poland
u.kuzelewska@pb.edu.pl

Abstract. Automated image classification is a fundamental task in computer vision. One of the most powerful deep learning (DL) models employed in image classification is the Convolutional Neural Network (CNN). Its key strength is the automatic extraction of features during direct raw image processing. This work proposes an ensemble transfer learning model based on CNNs to classify wheat grains. The dataset consists of 288 images of the kernels of three wheat grain varieties. In order to build the ensemble model, VGG16, InceptionV3, DenseNet201-V1, DenseNet201-V2, and NASNetMobile were evaluated and compared on the basis of the performance metrics including accuracy, Cohen's Kappa, precision, recall and F1 score. The ensemble CNN model demonstrated a higher level of performance than its base models achieving an accuracy of 96.48% and F1 score of 97%. Despite the limited amount of data available, the study's findings indicate that the proposed approach is an effective means of enhancing wheat grain classification.

Keywords: Computer Vision, Convolutional Neural Networks, Ensemble Learning, Wheat Grain, Shapley Values.

1 Introduction

Wheat is the most extensively cultivated crop in the world, and for the world's population, it is a fundamental source of essential nutrients. The number of wheat varieties is constantly increasing. Therefore, the proper selection of the wheat variety is a key factor in determining yield, as well as in marketing the commodity. With the rapid advancement of computer-aided methods, object detection and recognition have achieved tremendous progress. These are widely used in many fields, including their application in agriculture [1,2].

With the advent of automatic image processing and artificial intelligence, classification of wheat grains can be carried out on an industrial scale. It should be noted that

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_20

this is more effective, cheaper and faster than expert judgement [3]. For these reasons, effective classification methods based on visual features of wheat kernels are still being developed [3,4].

Nowadays, convolutional neural networks (CNNs) and other forms of deep learning (DL) have been employed in digital image processing with the objective of enhancing prediction performance. In comparison to the outcomes yielded by traditional machine learning, these DL methods have demonstrated notable improvements. One of the first experiments using neural networks for classifying texture images was carried out in 1999 [5]. The results obtained for pollen images showed that the neural network was superior to some previously published statistical classifiers. Indeed, 100% accuracy was evidenced. Daood et al. [6] also developed a CNN model for pollen grain classification. In this case, achieving 94% classification rate on a dataset of 30 pollen types with a database containing 1000 images.

Recently, Agarwal and Bachan [3] proposed a computer vision based system for automatic quality grading of wheat grains. The authors obtained accuracy of 93%. In related work, Koklu et al. [7] developed a method for classifying rice varieties. The features dataset and the image dataset were used as inputs. The classification success rate was found to be 99.95% for Deep Neural Networks (DNNs) and 100% for CNN.

Brahimi et al. [8] reached 99.18% of accuracy and demonstrated that CNNs outperformed SVM and RF in tomato diseases classification. Yang et al. proposed a novel method for peanut variety identification and classification by improving VGG16 [9]. The average accuracy of the improved model was 96.7%, which was 1.6–12.3% higher than that of other classical models. The identification of grain diseases was successfully achieved using one of the transfer learning methods, DenseNet201, which resulted in the highest accuracy of 96.8% [10]. The recent article [11] is devoted to the diagnosis of tomato leaf diseases. The authors performed experiments with YOLOv5, MobileNetV2, ResNet18, and a custom CNN model. The dedicated CNN model demonstrated superior performance, attaining an accuracy of 95.2%.

In summary, despite great progress in the field of wheat grain classification, there are still challenges in achieving more efficient classifiers. Therefore, the aim of our paper is to propose an ensemble classifier based on CNNs for multiclass classification of wheat grain. A further challenge is posed by the limited amount of data available, which frequently proves to be a significant problem for CNN-based solutions. Analysis was performed in order to provide a model interpretation based on Shapley values.

2 Materials and Methods

2.1 Characteristics of the Study Material

The experimental material comprised cleaned wheat grain of three varieties: Canadian, Kama and Rosa. The study covered X-ray photograms containing 10 to 12 grain kernels groove down positioned. The photograms were then scanned by an Epson Perfection V700 table photo-scanner with 600 dpi scanning resolution and 8 bit gray scale image digitization [12]. This X-raying and photo-scanning procedure provided resolution images of sufficient quality for extracting distinct features important for

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_20

accurate kernel characterization. Finally a randomly selected sample of 288 kernels was studied. Figure 1 presents an exemplary images of wheat kernels.

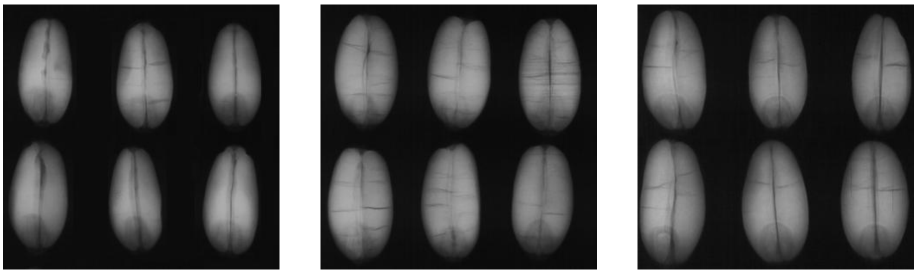

Fig. 1. X-ray image of kernels of three wheat varieties: Canadian, Kama and Rosa.

2.2 Deep Learning Models

In our study, a transfer learning approach was used that is rooted in pre-trained CNN models. This approach enables the application of pre-existing models to successfully adjust and tune to new domains and tasks. The growing number of studies on transfer learning models confirms their high adaptability and versatility [13]. The following CNN models were investigated: VGG16 [14], InceptionV3 [15], DenseNet201 [16] and NASNetMobile [17]. Combining a diverse set of individual CNN models can improve the stability of the overall model, resulting in more accurate predictions. Thus, ensemble learning was applied to produce one optimal and more robust predictive model [18, 19].

2.3 Model Evaluation and Interpretation

The evaluation of the proposed models was carried out using a multi-class confusion matrix and the performance metrics: accuracy, Cohen's Kappa, precision, recall and F1 score [20]. Five-fold cross-validation was applied to estimate the effectiveness of a classifier model in predicting the class of an unseen data. At the end, the error in cross-validating the data is determined as an average of all computed errors.

A game theory-based framework known as SHapley Additive exPlanations (SHAP) is applied to explain the CNN models [21]. The idea used for explanations of model predictions is to treat features that explain the prediction models as players and the prediction as the total payout [22]. A CNN model for image classification is to be trained and the final model is used to predict classifications of images in the test set. In the context of an image, each pixel is treated as a feature, therefore, Shapley value can be used to determine the pixel level importance in classifying images. The interpretation of visualization of the images shows highlighted parts in shades of red and blue. Red pixels represent positive SHAP values that increase the probability of the class, while blue pixels represent negative SHAP values that reduce the probability of the class.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_20

2.4 Proposed Ensemble Approach

In our work, five machine learning classification models, namely, VGG16, InceptionV3, DenseNet201-V1, DenseNet201-V2, NASNetMobile were built. All classifiers were trained from a total of 288 kernel images of three classes (Canadian – 108, Kama – 72, Rosa – 108). All input images were augmented with the following operations: rotation with angle 0-5, width shift 0-2%, height shift 0-0.5% [23].

The architectures and their optimal parameter values were selected based on prior experiments, which were then hypertuned. The final parameter specification of the employed models are presented in Table 1.

Table 1. Model parameters.

Parameter	Specification
Image size	200×150×3
Optimizer	Adam
Learning rate	0.001
Epochs	30
Dense Layer activation function	ReLU
Dropout	0.5
Output Layer activation function	Softmax
Loss function	Categorical Cross-entropy

The models were executed apportioning 80% of the data to train and 20% to test. Three best models were combined to build the soft voting ensemble CNN classifier.

3 Experimental Results and Discussion

3.1 Classification Results

This section presents the results of the study. Tables 2 and 3 display the outcomes of our study. The three best results for the Kappa, accuracy and weighted-average indicators of the individual CNN models are in bold.

As shown in Table 2, the VGG16 and InceptionV3 models achieved the lowest global accuracy of 91.97% and 88.89%. The DenseNet201-V1 and DenseNet201-V2 models demonstrated a 94.78% and 95.47% success rate in classification, respectively. The DenseNet201-V2 model displayed the highest level of efficiency in identifying the Canadian variety (98.2%) while DenseNet201-V1 achieved the highest result for the Kama variety (91.7%). The NASNetMobile model achieves an accuracy of 94.78% and a Cohen's Kappa score of 87%. However, this model was particularly effective in the recognition of the Rosa variety (99.1%). The three models with the highest Cohen’s Kappa and global accuracy were included in the ensemble model. In addition, the base models differed in their ability to identify specific wheat varieties.

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_20

Table 2. Classification results with CNN models.

VGG16				InceptionV3			
Accuracy	91.97%			Accuracy	88.89%		
Kappa	86.00%			Kappa	82.00%		
	Precision	Recall	F1 score	Class	Precision	Recall	F1 score
Canadian	0.88	0.94	0.91	Canadian	0.89	0.92	0.90
Kama	0.94	0.89	0.91	Kama	0.86	0.82	0.84
Rosa	0.95	0.92	0.93	Rosa	0.91	0.91	0.91
Average	0.92	0.92	0.92	Average	0.89	0.89	0.89
DenseNet201-V1				DenseNet201-V2			
Accuracy	**94.78%**			Accuracy	**95.47%**		
Kappa	**91.00%**			Kappa	**93.00%**		
	Precision	Recall	F1 score	Class	Precision	Recall	F1 score
Canadian	0.95	0.97	0.96	Canadian	0.97	0.98	0.98
Kama	0.96	0.92	0.94	Kama	0.97	0.90	0.94
Rosa	0.93	0.94	0.93	Rosa	0.93	0.96	0.95
Average	**0.95**	**0.95**	**0.94**	Average	**0.96**	**0.95**	**0.96**
NASNetMobile				**Ensemble model**			
Accuracy	**94.78%**			Accuracy	**96.48%**		
Kappa	**87.00%**			Kappa	**94.00%**		
	Precision	Recall	F1 score	Class	Precision	Recall	F1 score
Canadian	0.96	0.92	0.94	Canadian	0.96	0.98	0.97
Kama	0.89	0.89	0.89	Kama	0.97	0.94	0.96
Rosa	0.95	0.99	0.97	Rosa	0.96	0.97	0.97
Average	**0.94**	**0.94**	**0.94**	Average	**0.96**	**0.97**	**0.97**

The ensemble model used DenseNet201-V1, DenseNet201-V2 and NASNetMobile as base learners. The experimental settings remained constant. The ensembles were trained separately, and an average of all predictions was combined as the final result.

Table 3. The confusion matrix obtained from the ensemble model (5-fold cross validation).

		Predicted Class		
		Canadian	Kama	Rosa
Actual Class	Canadian	**104 (96.2%)**	2 (1.9%)	2 (1.9%)
	Kama	1 (1.4%)	**68 (94.4%)**	3 (4.2%)
	Rosa	3 (2.8%)	0 (0.0%)	**105 (97.2%)**

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_20

The confusion matrix can be found in Table 3. There was a significant improvement for the Kama variety (94.4%), and a deterioration for the Canadian (96.2%) and Rosa varieties (97.2%). However, weighted-average precision value, as well as the efficiency of this model and F1 score had the highest rate of 0.96, 0.97 and 0.97. The ensemble classifier gave the highest global accuracy (96.48%) and Kappa (94%).

3.2 Features Importance Analysis

For the interpretation of the CNN models, the Deep SHAP algorithm was applied to generate an explaining module. Figure 2 presents the results of DenseNet201-V2 model's SHAP generated for three inputs on the leftmost column.

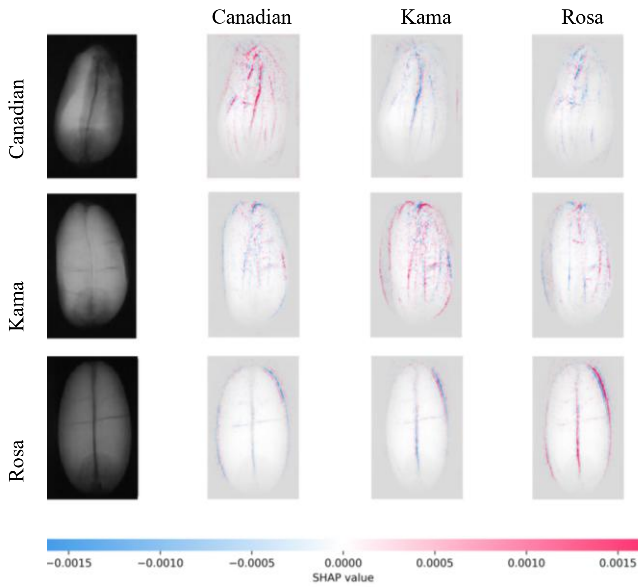

Fig. 2. DenseNet201-V2 model's SHAP values for wheat variety kernel images.

In the case of the image with the Rosa wheat variety, red pixels are presented mainly in the kernel groove. Here, the kernel groove is the longest. This region is also presented for the Kama and Canadian variety kernels. The average kernel groove length was significantly higher for Rosa, in comparison with the two other varieties [24]. Thus, the most discriminatory feature for the Rosa variety is kernel groove length.

In addition, red pixels emphasized regions of interest at the edges of the Kama and Rosa kernels, which contributed to the correct classification. This is consistent with the fact that according to the SHAP values, kernel perimeter is the second most discriminatory feature. The results for the Canadian and Kama variety kernels show red pixels concentrated in specific large internal kernel parts. It is evident that there is a substantial disparity between the average kernel area and the average kernel perimeter of the Rosa variety and those of the other varieties [24].

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_20

3.3 Discussion

The classification of small datasets is a challenging task, especially in the context of computer vision approaches. However, it is an area of interest for many researchers [25]. The issue of having a limited number of grain images by an ensemble of CNN models for image classification is addressed in our study.

The outcomes of almost all transfer learning methods were excellent, with accuracy, Kappa, precision, recall and F1 score exceeding 90%. In comparison, only InceptionV3 demonstrated slightly lower performance, achieving an accuracy of 88.89%, a Kappa of 82%, a precision of 0.89,a recall of 0.89, and an F1 score of 0.89. The DenseNet201-V2 model demonstrated a very high accuracy of 95.47%. The remaining outcomes, i.e. Kappa, precision, recall and F1 score, confirmed the accuracy. The Kappa score was 93%, the precision rate was 0.96, the recall rate was 0.95 and the F1 score was 0.96. The proposed soft-voting ensemble CNN model outperformed single classifiers achieving an accuracy of 96.48%, a Cohen Kappa of 94% and an F1 score of 97%. It also outperformed traditional ML models based on the main shape features as extracted from the same X-ray wheat kernel images, achieving an improvement in F1 score of 3% [24]. Furthermore, the feature importance analysis of the CNN models was consistent with the SHAP values of traditional models.

4 Limitations and Future Works

This study presented a soft-voting ensemble CNN approach for classifying X-ray images of wheat grains. A key limitation of this study is the relatively small size of the dataset, which may restrict the model's ability to generalize. However, the dataset is unique and challenging to obtain, which enhances the value of this work despite its limited scale. Future work will focus on more advanced ensemble techniques to improve the performance and robustness of the model using larger and more diverse datasets. The results demonstrate that the proposed approach effectively enhances wheat grain classification even when the size of the dataset is limited.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Acknowledgments. The research was partly carried out within project no. WZ/WI-IIT/3/2023 at Bialystok University of Technology and financed from the research subsidy of Bialystok University of Technology.

References

1. Utku, H.: Application of the feature selection method to discriminate digitized wheat varieties. J Food Eng (46), 211216 (2016)
2. Zhang, Q., Liu, Y., Gong, C., Chen, Y., Yu, H.: Applications of deep learning for dense scenes analysis in agriculture: A review. Sensors **20**(5), 1520 (2020)

3. Agarwal, D., Bachan, S.P.: Machine learning approach for the classification of wheat grains. Smart Agricultural Technology 3, 100136 (2023)
4. Mahony, N.O., et al..: Deep Learning vs. Traditional Computer Vision. Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC) (2020)
5. Li, P., Flenley, J.R..: Pollen texture identification using neural networks. Grana **38**(1), 59-64 (1999)
6. Daood, A., Ribeiro, E., Bush, M.: Pollen Grain Recognition Using Deep Learning, Advances in Visual Computing. Lecture Notes in Computer Science, 10072. Springer (2016)
7. Koklu, M., Cinar, I., Taspinar, Y.S.: Classification of Rice Varieties With Deep Learning Methods. Comput Electron Agric **187** (3), 106285 (2021)
8. Brahimi, M., Boukhalfa, K., Moussaoui, A.: Deep Learning for Tomato Diseases: Classification and Symptoms Visualization, Appl Artif Intell **31**(4), 299-315 (2017)
9. Yang, H., Ni, J., Gao, J., Han, Z., Luan, T.: A novel method for peanut variety identification and classification by Improved VGG16. Scientific Reports **11**(1), 1-17 (2021)
10. Adimas, A. K., et al.: Soybean leaf disease detection and classification using deep learning approach. Bulletin of Electrical Engineering and Informatics, **14**(4), 2697-2704 (2025)
11. Prama, T.T., Oni, M.K.: Optimized Custom CNN for Real-Time Tomato Leaf Disease Detection, ICCS 2025,, Lecture Notes in Computer Science, vol 15909. Springer (2025)
12. Charytanowicz, M., et al..: An Evaluation of Utilizing Geometric Features for Wheat Grain Classification using X-ray Images. Comput Electron Agric, 144, 260-268 (2018)
13. Rostami, M.A. et al.: Efficient pollen grain classification using pre-trained Convolutional Neural Networks: a comprehensive study. Journal of Big Data 10.1, 151 (2023)
14. Simonyan, K., Zisserman, A.: Very Deep Convolutional Networks for Large-Scale Image Recognition. Computer Vision and Pattern Recognition (2014)
15. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the Inception Architecture for Computer Vision (2015)
16. Zoph, B., Vasudevan, V., Shlens, J., Le, Q.V.: Learning Transferable Architectures for Scalable Image Recognition. Computer Vision and Pattern Recognition (2017)
17. Ganaie, M.A., Hu, M., Malik, A.K., Tanveer, M., Suganthan, P.N..: Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence 115, 105151 (2022)
18. Pudumalar, S., Muthuramalingam, S.: An ensemble deep learning recognition model for plant diseases. Journal of Engineering Research (2023)
19. Iqbal, M.S. et al.:An adaptive ensemble deep learning framework for reliable detection of pandemic patients. Computers in Biology and Medicine 168, 107836 (2024)
20. Yilmaz, A., Demircali, A.A., Kocaman, S., Uvet, H.: Comparison of Deep Learning and Traditional Machine Learning Techniques for Classification of Pap Smear Images. Image and Video Processing (2020)
21. Lundberg, S.M., Lee, S.-I. : A Unified Approach to Interpreting Model Predictions. Advances in Neural Information Processing Systems 30, 4765-4774 (2017)
22. Pitts, O.et al.: SHAP-Prioritised Machine Learning for Diagnostic-Grade Prediction of Lung Function. In: Computational Science – ICCS 2025 Workshops. ICCS 2025. Lecture Notes in Computer Science, vol 15910. Springer (2025)
23. Arora, J., Agrawal, U.: Systems. Classification of Maize leaf diseases from healthy leaves using Deep Forest. J. Artif. Intell. Syst. 2, 14-26 (2020)
24. Charytanowicz, M.: Explainable Ensemble Machine Learning for Wheat Grain Identification. In: The 23rd IEEE International Conference on Data Mining (ICDM). IEEE, 1-9 (2023)
25. Dong, L., Shu, H., Tang, Z., Yan, X.: Microseismic event waveform classification using CNN-based transfer learning models. Int J Min Sci Technol 33(10), 1203-1216 (2023)

ICCS Camera Ready Version 2026
To cite this paper please use the final published version:
DOI: 10.1007/978-3-032-29915-4_20