Consensus rule for wheat cultivar classification on VL, VNIR and SWIR imaging

Şahin Işık1 | Kemal Özkan1 | Duygu Zeynep Demirez2 | Erol Seke2

1 Department of Computer Engineering, Eskisehir Osmangazi University, Eskisehir, Turkey
2 Department of Electrical and Electronics Engineering, Eskisehir Osmangazi University, Eskisehir, Turkey

Correspondence
Şahin Işık, Department of Computer Engineering, Eskisehir Osmangazi University, Meselik Kampus, Eskisehir, Turkey
Email: sahini@ogu.edu.tr

Funding information
Turkey Scientific and Technological Research Projects Support Program, Grant/Award Number: 1160576; Eskisehir Osmangazi University Scientific Research Project Commissions, Grant/Award Number: 2016-1120

Abstract
To facilitate the quality assessment of wheat cultivars, diverse imaging tools and techniques have been applied in order to omit the expert decision, which can cause failure in the identification of the wheat cultivar's label and its quality simultaneously. To minimize the risks caused by the expert's decision, a promising framework for identification is greatly required to more effectively assess wheat type. Therefore, to be beneficial to this association, two methods have been developed by performing traditional and modern feature extraction algorithms on visible light (VL), visible near-infrared (VNIR) and short-wave infrared (SWIR) imaging as well as a fusion of these imaging systems. The proposed systems are called the bag of word (BoW) framework and convolutional neural networks (CNN) framework. With regard to wheat cultivar detection, the consensus rule has been established based on decisions predicted by CNN and BoW frameworks. The accuracy results obtained by consensus rule indicate that we have achieved 99.94% and 68.94% in case of CNN framework and BoW framework, respectively. Experimental results suggest that BoW features are not suitable to represent and match texture patterns such as repeated wheat kernels in an image, whereas CNN features always outperform handcrafted elements and properties for all datasets.

1 INTRODUCTION

Wheat is one of the oldest and most predominant cereal products around the world as well as the foremost produced and consumed variety of cereals. Considering wheat plants have all-embracing ability to adapt to various climate and soil conditions, it has precedence and is in the first place in terms of cultivation and production among the agriculturally cultivated plants required for human nutrition. The fertility of wheat is superior and its agricultural production is amenable and acquiescent therefore when compared to other cultivars it is more convenient. Wheat is primarily utilized in many food and industrial sectors, especially in baked goods. As a result, when wheat production is reduced, everyone is affected as a consequence. Accordingly, the quantities of the products would be decreased with the prices increasing correspondingly.

Moreover, one of the most important factors in the production of bread is the quality level of the wheat plant. The quality of wheat grains is influenced by genetic or environmental factors. In order to obtain adequate production, only the quality of the wheat plant is not sufficient. There is a direct connection between wheat seed ingredients and its distinctive calibre. Concurrently, factors such as foreign matter, inadequate cleaning, improper storage conditions, and contamination of the same or other kinds of seeds have a negative effect on the seed fertility or cause a decrease in the quality of the seed. The identification and classification of the varieties will ensure that the storage operations would be carried out correctly. Hence, the labour force and financial loss can be prevented by reducing the processing time wasted on determining wheat quality or specifying its cultivar. For these reasons, reliable methods should be developed for the identification of wheat cultivar.

Going through the literature, we can observe that a crowded set of methods have been established in this direction. An example among this set is the work of [1]. Canada West Red Spring (CWRS) wheat, Canada West Amber Durum (CWAD) wheat,
barley, oats and rye seeds were acquired by using an area scanning camera. From these images, 51 morphological features, 93 colour features, 56 textural features, and 135 wavelet features were extracted prior to the classification stage. They utilized the linear discriminant classifier (LDA) which combines all the morphological, colour, textural and wavelet features in CWRS wheat and achieved the best classification with 99.4% accuracy. In their system, rye, barley, oat and CWAD wheat have followed 99.3%, 98.6%, 98.5%, and 89.4% accuracy respectively. Following a similar attempt, the same authors of work [2], developed a classification system for Western Canada wheat types by using hyperspectral images of wavelet texture analysis obtained at 10 nm intervals in the 960–1700 nm wavelength range using the visible near infrared (VNIR) hyperspectral imaging method. Based on LDA, the top 100 features were selected and used for the classification of wheat varieties. Linear and quadratic statistical classifiers and a standard back-propagation neural network (BPNN) classifier were employed for classification with these features. When the best 90 features were used in the LDA classifier, the average highest classification rate for 8 classes had a 99.1% accuracy rate. Additionally, in instances of utilization of the BPNN classifier for the top 70 features, the success rate was reported as 92.1%. Conclusively, using wavelet texture features obtained from hyperspectral imaging resulted in a high success rate at 79.9%, respectively. Another study in which hyperspectral imaging was considered [3], the SWIR hyperspectral imaging system (1000–2500 nm) was proposed to estimate the a-amylase activity of individual wheat kernels. CWRS and CWAD wheat cultivars were carried out for classification purposes. A classification method was proposed for CWRS kernels to achieve a level of high and low a-amylase activity and accuracy of over 80% for aforementioned wheat cultivars. They were proposed a method to classify high and low-level a-amylase activity from CWRS kernels with an 80% accuracy rate. In their other work [4], CWRS wheat images were obtained using a VNIR camera. At the first stage, they used principal component analysis (PCA) to separate the VNIR data into four wavelengths that contribute to the revelation of the discriminative features of wheat types. Using a classification procedure that combines both spectral and spatial properties, 100% of the sound kernels, about 94% of the sprouting grains and 98% of the heavily sprouting grains were correctly classified. In a study [3], they utilized a hyperspectral imaging system in the VNIR (400–1000 nm) wavelength range for detection of Fusarium-damaged kernels (FDK) for wheat samples from Canada. They used PCA and LDA for classification. According to the observed performance, an overall accuracy of 92% with the LDA model was noted. Also, short wavelength near-infrared hyperspectral imaging systems in the 700–1100 nm wavelength range was utilized in another study [6]. They used healthy and fungal-damaged wheat grains. The 870 nm wavelength corresponding to the highest factor loading of the first major component was considered meaningful in terms of discrimination. The analytical and histogram properties obtained from the 870 nm wavelength image were selected and forwarded to the statistical discriminant classifiers (linear, quadratic and Mahalanobis). A total of 179 features (123 colours and 56 textures) were extracted from the colour images. They have classified healthy and fungal infected wheat kernels correctly with 97.3–100.0% accuracy using the hyperspectral imaging features along with the linear discriminant analysis classifier, the top ten features selected from colour images, and textural properties. Greyscale image processes were obtained for the identification of nine common varieties of Iranian wheat kernels by examining various textural feature groups of these kernel samples [7]. Overall, 1080 grey-scale images of bulk wheat seeds (120 images of each species) were obtained under a steady lighting condition (fluorescent ring light). A total of 131 textual features were extracted from grey level, grey level co-occurrence matrix (GLCM), grey level run length matrix (GLRLM), local binary patterns (LBP), local similarity models (LSP) and local similarity numbers (LSN). They have used the LDA classifier with best selected features. When using the best 50 of the features in the classifier, they achieved a 98.15% accuracy rate. Moreover, the investigation of hyperspectral imaging technique [8] was carried out for classifying the vitreous, yellow, and Fusarium-damaged types of wheat, which were examined by applying different chemometric techniques such as PCA and partial least squares discriminant analysis (PLS-DA) with NIR (1000–1700) camera. Based on comparison of the results, the best classification among the wheat species was reported with a low misclassification error (<1%) for the vitreous class and the yellow berry class observed slightly better results than the Fusarium damaged class, although in both cases the misclassification error is low. Near-infrared stethoscope was used in another study [9]. They have used NIR stethoscope with chemometric methods to estimate the geographic origin of wheat kernels and flour samples produced in different regions of Chile. Spectral data was used with discriminant partial least squares (DPLS). Implementing this method, the specimens were correctly classified, into 76% of wheat grain samples and 90–96% of the flour samples according to their geographical origins. With a different view, in the study of [10], the authors considered the SWIR imaging technique in their study. Contrary to given studies, they were able to establish aflatoxin contamination in corn kernels. The corn samples were inoculated with four concentrations of aflatoxin B1 (AFB1) (10, 100, 500 and 1000 mg/kg) and the control samples disinfected with a PBS solution. Both the infected and control sample were scanned with a SWIR hyperspectral system on a spectral range of 1100–1700 nm. A PLS-DA model was developed on a controlled classification procedure and infected cores with the highest overall classification accuracy attained as 96.9% with the developed model. Moreover, another study [11] was conducted on the hyperspectral imaging technique by proposing an algorithm for the automatic detection of Fusarium head blight in wheat, which was captured with the use of a particular NIR camera. They elaborated by formulating an algorithm to demonstrate its soundness and capacity regarding factors such as shape, orientation, ghosting and clustering of nuclei with a classification accuracy of over 91%. Additionally, in the study of [12], the authors used near-infrared (NIR) reflection spectroscopy to represent a fast and robust method for product authentication. The NIR spectra obtained from the grains and flours are used to make bread (n = 705), spelt (n = 673), durum
(n = 75), emmer (n = 75) and einkorn (n = 73). They used PLS-DA and the success rates was shown as 80–100%. In another study [13], researchers used the NIRS hyperspectral imaging system to detect ochratoxin (OTA) at five concentrations in contaminating wheat seeds. In order to determine the discriminative wavelengths that are more important in determining OTA contamination in wheat, they applied PCA to the wheat kernel samples. Classification models were derived from OTA contaminant wheat kernels and non-OTA producing *Penicillium verrucosum* inoculated wheat kernels with 100% accuracy. Concurrently, the classification system generated an accuracy rate, which is more than 98% for the five concentration levels of the OTA contaminant wheat grains and five infection levels of non-OTA *P. verrucosum* grafted wheat seeds.

Recently, the deep learning methodology has been applied for food, wheat and agricultural industry with a purpose of reducing the labour force and increasing the production growth. The motivation of such research works relies on three factors: (i) the deep CNN models are more robust than traditional classifiers for the generalization ability, (ii) CNN models are simple since the feature engineering is not required and (iii) the deployment of CNN model is not complex when it comes to be embedded into a real-time system. In a previous study [14], the rice blast disease recognition was carried out with a modest CNN architecture. The CovNet [15] architecture is employed for classification of 15 different wheat grain types consisting of 15,000 samples. To alleviate the necessity for manual categorization of maize seeds [16], a non-destructive approach that works beyond using a pretrained model was applied on a new dataset. Moreover, the 98.38% classifying accuracy rate was determined by using the transfer learning approach, which means that a pretrained model developed on a task is reused as starting point on wheat grain classification task [17]. In other words, the weights of pretrained model is updated through backpropagation process for wheat cultivar classification.

Additionally, in wheat transactions, the rates are usually defined over the expert’s judgment on the type and quality of the product which results in more dependence on the human factor. Specifically, different experts may offer different prices for the same wheat variety. Therefore, more objective quality determination methods are raised with the advanced technology. In line with this purpose, in this study, we have developed an intelligent framework that relies on the visible light (RGB, wavelength 400–800 nm), VNIR (wavelength 400–1100 nm), and shortwave infrared (SWIR, wavelength 900–1700 nm) imaging techniques. In order to accomplish this, the discriminatory information stated on the aforementioned imaging has been analyzed as a means to determine the identification or classification and label of wheat cultivar in an accurate way. With the use of these three different imaging techniques, a total of 8000 samples have been obtained from 40 varieties of wheat cultivar as there are 200 samples per each class. The objective of the proposed study is developing a prediction system for a wheat cultivar without knowing its label by performing the modern and traditional machine examination methods. This study resembles our other study, in which the information fusion over feature and image domain have been carried out for wheat kernel recognition [17]. Through the use of the developed system, the determination of the price and quality of the wheat variety and the specification of class, value, and designation of label would be facilitated. Thereby advanced and increased objective decision-making would be guaranteed once the human factor is reduced.

The rest of this study is as follows: Details about developed bag of word (BoW) [18] and CNN frameworks are available in Section 2. In Section 3, we have presented the objective results related to the absence and presence of the consensus rule for predicting the label of a wheat cultivar. Finally, a conclusion is given in the last section.

### 2 | CONSENSUS RULE BASED WHEAT CULTIVAR CLASSIFICATION

#### 2.1 | Motivation

From a technical point of view, the quality of end products like bread, biscuits, and macaroni, all of which are consumed and relied on by the majority of the population, depend on the derived wheat cultivar’s type. Also, the quality of wheat is directly connected to its graininess, growing, harvesting and storage conditions. However, the damages caused by microorganisms, insects, and pests negatively influence the quality standards of a wheat type. The identification of wheat types in the ministry of agriculture is usually carried out by an expert’s opinion, which is dependent on the experience, knowledge and skills of a person that may result in false labelling. Subsequently, some studies stated on addressing cultivars’ quality problems based on the small-scale datasets, but cannot be generalized into a whole system unless the count of cultivars increased by collecting them from different regions.

In order to correctly surmount the classification of wheat cultivars, a new machine learning technique capable of fulfilling identification processes with a good accuracy is greatly required to develop for the benefit of reducing the percentage of erroneous decisions made by an expert’s visual evaluation. Developed as an alternative for expert’s decision, each study in the literature of wheat classification relied on the policy of transferring the digital images from utilized imaging devices to the computer and extracting descriptive information by using machine learning algorithms, finally classifying them by using a machine learning approach. This study also aims to identify labels of wheat cultivars by creating and implementing an effective framework on samples returned from SWIR, VNIR and RGB cameras. The widely known feature detectors/extractors along with popular classifiers are planned to perform in order to recognize wheat type.

For this purpose, we have developed two frameworks: (i) using traditional BoW features and (ii) executing celebrated CNN features, respectively. Since using a single feature extraction algorithm with a predetermined classifier does not promise reasonable performance for the wheat representation, we have employed BoW [18] features on like scale invariant feature transform (SIFT) [19], speeded-up robust features (SURF) [20], dense-scale invariant feature transform (DSIFT) [18],...
LBP [21] and shape-based Fourier descriptor (SBFD) [22] for decision-making with consensus rule on SVM, ANN and KNN classifiers. Going through the literature on machine learning systems for wheat identification, one can observe that the single model [7, 23] has been usually employed in the case of identification process. On the contrary, we have proposed a consensus rule, consisting of majority voting (MV) based decision-making, in order to achieve favourable accuracy after combining prediction outputs returned from four classifiers (SVM (linear), SVM (polynomial), ANN, and K-NN). In the instance of the second framework, impacts of CNN features have been analyzed after conducting experiments on deep CNN architectures like VGG16, VGG19 and GoogleNet. In the second framework, we have also investigated the performance of consensus rule on returned features from CNN models. Regarding the developed CNN features based framework, the consistent and more objective decisions will be achieved when considering price estimation, gender identification and quality determination for wheat cultivar and thereby reducing the interference of varied and inaccurate human factors.

2.2 | BoW Framework: Traditional features and classifiers

The most significant objectives of BoW framework are (i) combining information from different imaging systems including RGB, VNIR and SWIR to improve the classification accuracy of wheat cultivars and (ii) evaluating the performance of well-known feature extractors and classifiers with a consensus rule.

To attain such objectives, we have pursued the system given in Figure 1. As a first step of classification system, the 40 different types of wheat cultivars have been stored by gathering from local offices, namely Eskişehir Commodity Exchanges and Eskişehir Directorate of Provincial Food Agriculture. Then, the three different imaging (RGB, VNIR, SWIR) for each sample of wheat cultivar have been captured with related cameras, which are supported by the Scientific and Technological Research Council of Turkey (TUBITAK). Additionally, the fourth dataset as a fusion of RGB, VNIR and SWIR images has been set up with the concept of common vector approach (CVA) method. The fusion procedure is explained in greater detail in Section 2.4. Thus four different datasets RGB, VNIR, SWIR and FUSION have been obtained to conduct experiments and to investigate performance of consensus rule for wheat type detection through BoW and CNN features.

The Figure 1 presents the first framework to distinguish wheat cultivars with the help of well-known traditional feature/descriptor extraction tools like SIFT [19], SURF [20], DSIFT [18], LBP [21] and SBFD [22]. Once the SIFT, SURF, DSIFT, LBP and SBFD features/descriptor sets are extracted from the aforementioned datasets, each feature set is encoded by using the BoW strategy. In BoW model [18], each local histogram, called descriptor set, is mapped to a visual word by seeking the closest centroid estimated from feature space with k-means clustering method. By considering the tradeoff between running time and accuracy, the size of a feature set is deduced as 600, which is a vector consisting of 600 scalar values returned from the BoW model. Since we have four datasets and five feature/descriptor extraction tools, there are 20 different varieties of feature set yielded in total. With respect to the classification stage, 80 different decisions have been estimated for each test sample by activating the four classifiers (SVM (linear), SVM (polynomial), ANN and K-NN). It means that there are 20 different feature sets and four classifiers, so that we have obtained the 80 decisions after separately generating models for selected feature sets and have executed associated classifiers. After running the majority voting procedure (‘consensus rule’), a single label has been determined for each processed wheat sample among the 80 decisions. The details and parameters of related utilized feature extractors and classifiers are given in Section 2.5.

2.3 | CNN Framework: deep learning features

As a second identification framework, we have employed robust features, determined from CNN architectures based on the deep learning strategy. To realize this framework, the VGG16, VGG19 and GoogleNet have performed to encapsulate discriminative information on weights used in convolution stages. In an effort to analyse the capability of CNN features on wheat cultivar classification, the aforementioned CNN architectures are trained through deep layers and massive weights.

Figure 2 shows the CNN-based consensus rule methodology on wheat cultivar classification. This second framework is composed of VGG16, VGG19 and GoogleNet, employed to improve the performance of wheat identification systems. The
objective of this second framework and structure is to compare the performance of deep neural network features against traditional features over consensus policy based on the objective evaluation metrics.

### 2.4 Fusion of SWIR, VNIR and RGB

Since system performance is likely suffering from using only RGB imaging, because there is low inter-class variation compared to inner-class variation, combining meaningful information from channels of three imaging systems provides a correct solution to determine the quality of wheat cultivars over their labels. Also, one can emphasize that the information through RGB, VNIR is in the form of 400–1700 nm will be encompassed within a single channel after the fusion process. In other words, the fusion refers to providing and obtaining a unique image having maximum information content after coupling different details stated on each channel together. In line with this motivation, a crowded set of fusion algorithms [24–26] have been recently applied along with different fusion methodologies to improve performance of machine learning systems.

As one of the novel points of this study, the CVA method [27] was performed for purposes of merging. Similarly, an attempt had been realized in works of CVA [28] with a purpose to fuse multispectral channels. Until now, the CVA has been performed for different tasks ranging from classification [29] to image processing [30, 31]. As a subspace based projection technique, the CVA method can be summarized as follows [32].

Assuming that a vector belonging to any class is represented with $\vec{a}_j$, it is possible to set the form of $\vec{a}_j = \vec{a}_{\text{com}} + \vec{a}_{\text{diff}}$. While the $\vec{a}_{\text{diff}}$ refers to the difference vector composed of details and unstable features over the processed class, $\vec{a}_{\text{com}}$ denotes the common vector including the shared information through vectors and summarizing the characteristic features of its class. Note that there are two cases for the CVA method as presence of sufficient data and insufficient data, respectively. In the event that the number of vectors of a class is less than the length of a vector, then it is considered as an insufficient data case, otherwise there is a sufficient data case for related class. Here, the insufficient data case occurring as five channels (Red, Green, Blue of RGB, VNIR and SWIR) were considered for integration and merging purposes. While RGB, VNIR is in the form of 256 × 320, also the SWIR consists of 175 × 200 size, it can be observed that there is insufficient data set for this study. In case of insufficient data, one can employ the Gram Schmidt decorrelation to determine orthonormal vectors as emphasized in the work of [27]. When replacing the formula of $\vec{a}_j = \vec{a}_{\text{com}} + \vec{a}_{\text{diff}}$ with $\vec{a}_{\text{com}} = \vec{a}_j + \vec{a}_{\text{diff}}$, then we can say that obtaining the difference vector $(\vec{a}_{\text{diff}})$ is enough to find the common vector $(\vec{a}_{\text{com}})$ for processed class, since the value of $\vec{a}_j$ is known.

$$\vec{b}_i = \vec{a}_{j=1} - \vec{a}_{j=i} \quad (i = 1, 2, \ldots, n - 1).$$

For this purpose, a reference vector, $\vec{a}_{i_1}$, is required to be selected to compute the difference space as shown in Equation (1). In this study, we have chosen the $\vec{a}_{i_1}$ as reference from n vectors. Thus a difference subspace {$\vec{b}_1, \vec{b}_2, \ldots, \vec{b}_{n-1}$} is obtained and forward into the Gram Schmidt decomposition procedure. Once the orthogonalization process is applied to the difference subspace, the orthonormal set {$\vec{z}_1, \vec{z}_2, \ldots, \vec{z}_{n-1}$} is obtained to further compute $\vec{a}_{i_1,\text{diff}}$ followed by projecting the reference vector onto the orthonormal vectors, which is expressed in Equation (2).

$$\vec{a}_{i_1,\text{diff}} = \sum_{k=1}^{n-1} \langle \vec{a}_{i_1}, \vec{z}_k \rangle \vec{z}_k$$

Finally, the common vector of the processed class is extracted as $\vec{a}_j - \vec{a}_{i_1,\text{diff}} = \vec{a}_{\text{com}}$. As a result, we can say that the common vector that contains rich information about shared details is obtained as $\vec{a}_{\text{com}}$.

![FIGURE 2](Image 47x609 to 287x734) Demonstration of CNN framework

![FIGURE 3](Image 330x602 to 527x735) Visualization for fusion of five channels including Red, Green, Blue, VNIR and SWIR for the #000012 sample of Tosunbey class
2.5 Tools and parameters

In this stage, we have introduced implementation details about utilized feature extractors and classifiers. As mentioned above, the experimental study consists of two frameworks. The first framework refers to the conventional feature extractors including SIFT, DSIFT, LBP, SURF and SBFD with different classifiers like SVM, ANN and K-NN, utilized to evaluate their performances on wheat cultivar identification. Moreover, as a second framework, the discriminative power of CNN features along with the Softmax classifier have been validated by conducting experiments over deep layers, namely VGG16, VGG19 and GoogleNet.

2.5.1 Feature extraction tools

As it is inferred by its name, the feature extraction refers to extracting true underlying traits of a category in order to facilitate the decision-making. When it comes to employing a feature extractor for classification purposes, then one must weigh all factors, that is the capability of increasing inner class variation as well as mitigating the inter class variations in the feature space. This study reveals the discriminative capacity of feature extractors in connection with wheat cultivar classification. To achieve this, in case of traditional features, we have randomly selected 6500 feature vectors from 20% of training samples per each class, prior to extracting vocabularies. Since we have considered 600 vocabularies, each image is represented with a feature vector having size of 600 × 1 after encoding features with obtained vocabularies, by using only one level of spatial pyramid.

- **SIFT** [19] is a popular feature extractor tool that relied on the histogram of descriptors.
- **DSIFT** [18] is another version of SIFT and extracts features by sliding the grids at the fixed scale and orientations instead of selecting the dominant key points as carried out in SIFT. In case of parameter settings, we have specified the bin size as 8.
- **LBP** [21] is a texture oriented feature extraction tool. In case of experiments, each image is divided into eight cells and descriptors are extracted from 3 × 3 neighbourhoods.
- **SURF** [20] detects the discriminative blob features and is widely known as a faster version of SIFT. We set the threshold to 10 selecting the strongest features.
- **SBFD** raises another novelty of this work in terms of applying a novel feature extractor tool for wheat cultivar classification. The SBFD is an improved version of Fourier Descriptors [22], which is employed for object classification. Similar to work of [33], we have extracted Fourier Descriptors on contour fragments to be forwarded into the classifier.
- **VGG16**: The Visual Geometry Group (VGG) [34] has developed a new standardized CNN architecture by improving the performance of AlexNet [35] features after utilizing small filters with constant size as 3 × 3. Since then, the attempts through reducing filters’ size in CNN architectures have been geared up to obtain minimum error rates for training large datasets over long deep layers. When conducting experiments with VGG16, the epoch size and batch size are fixed to 20 and 16, respectively. The optimization function is specified as Stochastic Gradient Descent with Momentum and Initial Learning rate is 0.0001.
  - **VGG19**: The VGG19 [34] closely resembles the VGG16 in terms of utilized structure, but the only difference is that the weight layers increased in VGG19. The same parameters of VGG16 are considered to carry out experiments.
  - **GoogleNet**: The underlying motivation behind the GoogleNet [36], is producing an efficient and effective deep learning tool by reducing the internal parameters of well-known CNN architectures like AlexNet [35], VGG16 and VGG19 as well as developing a fast network structure that gives high accuracy scores in machine learning tasks. Again, in case of extracting results, we have made experiments on GoogleNet with the same epoch size, batch size, initial learning rate and optimization function employed in VGG16.

### Table 1: List of the utilized 40 wheat cultivars

| Ahmetaga       | Altay2000 | Ayrin    | Dagdas    |
|----------------|-----------|----------|-----------|
| Harmankaya     | Izgi      | KateA1   | Energo    |
| Midas          | Mufitbey  | Nacibey  | Kirac     |
| Sonmez         | Soyer     | Sultan95 | Pehlivan  |
| Bayraktar2000  | Bezostaja | Cetiner  | Suzen     |
| Esperia        | Flamura85 | Gelibolu | Gerek79   |
| Kirgiz         | Kunduru   | Mesut    | Michelangelo |
| Porsuk         | Reis      | Renan    | Selmiye   |
| Tosunbey       | Yelken    | Yildiz   | Yunes     |

3 EXPERIMENTAL STUDY

3.1 Datasets

A total of 8000 images which are 640 × 512 in sizes are obtained from 40 wheat classes and each class has 200 images. 20% of the images are reserved for testing. The images are captured from visible light (RGB, wavelength: 400–800 nm), VNIR (wavelength: 400–1100 nm) and SWIR (wavelength: 900–1700 nm) cameras with the controlled environment. Table 1 illustrates a list of 40 wheat classes and Figure 4 shows some samples related to the RGB, SWIR, VNIR and FUSION dataset. The types of wheat samples are listed in Table 1.

Figure 5 shows that SWIR, NIR, RGB cameras are positioned at about 50, 30 and 20 cm above the samples, respectively. Their viewing is approximately 21 × 29, 9 × 7 and 9 × 7 cm² regions on the sample plane for SWIR, VNIR and RGB, respectively.
3.2 | Details about experiments

To give an overall insight into the performance of the proposed method, we have evaluated the obtained results with respect to the objective metrics—including Fscore and accuracy—estimated from confusion matrix. The accuracy refers to closeness of predicted results to actual values, hereafter called as ground truths. Also, we have visualized the Fscore of consensus rules for BoW and CNN frameworks. The Fscore is computed based on the true positive (TP), false positive (FP) and false negative (FN) rates.

In the case of the training stage, 80% of samples of each class are considered, while the rest of them are utilized for testing purposes. It means that 6400 samples have been employed in the training stage. To realize the consensus rule, the index associated with each training and testing sample is saved before the experimental stages and these saved samples have been considered at each experiment. The performance of each method has been investigated on 1600 test samples.

3.3 | Objective evaluation of BoW framework

To evaluate the performance of BoW and CNN frameworks, we have demonstrated some quantitative results based on the accuracy scores. For comparison, the accuracy result of each traditional method has been obtained and exhibited on Table 2. As shown in Table 1, the average accuracy rates obtained for different datasets are sorted as SWIR 12.90%, FUSION 16.33%, RGB 22.23% and VNIR 22.68%. From Table 2, we can deduce that the VNIR dataset promised the best performance against the other ones. For each category (FUSION, SWIR, RGB and VNIR), the best result is denoted as bold.

Moreover, the results suggest that the DSIFT is the most effective traditional feature extraction tool. Since the SBFT method is shape oriented, and is developed to replace the underlying gradients of a wheat sample with histograms of phase angles related to local Fourier transform (FT), its performance is considerably lower than the other ones. Interestingly, it means that shape descriptor extraction tools are convenient to match the texture patterns on similar wheat kernels. Upon inspecting the performance of classifiers, the SVM with polynomial kernel gets better performance than other ones.

Furthermore, the effectiveness of consensus rule has examined and commented to evaluate the BoW framework. Once the consensus rule applied on predicted labels is returned from traditional feature extraction tools, the best accuracy was achieved as 68.87%, which is a satisfactory accuracy rate in terms of wheat cultivar identification. The consensus rule was performed on 1600 x 80 decisions, which means that there are 1600 test samples and 80 decisions yielded per each sample. In case of assigning a class label of a wheat sample, the majority voting rule is considered for these 80 decisions.

3.4 | Objective evaluation of CNN framework

In this section, we enhance the robustness of CNN features of FUSION, RGB, SWIR and VNIR datasets. We have conducted three controlled experiments by forwarding each dataset to well-known and deep structures including GoogleNet, VGG16 and VGG19. By looking at the average accuracy results associated with each dataset, the following accuracy rates have been determined; FUSION 92.17%, RGB 98.58%, SWIR 85.73% and
TABLE 2  Accuracy (%) of BoW features for traditional methods

|                  | FUSION |          |          |          |          |
|------------------|--------|----------|----------|----------|----------|
|                  | SVM (Linear) | KNN | SVM (Polynomial) | ANN | Average |
| DSIFT | 36.19 | 25.00 | 39.00 | 28.06 | 16.33 |
| FFT | 3.19 | 2.94 | 2.88 | 3.31 |  |
| LBP | 20.94 | 18.38 | 22.51 | 20.31 |  |
| SIFT | 13.25 | 6.06 | 14.44 | 11.13 |  |
| SURF | 15.06 | 13.94 | 16.06 | 14.13 |  |

|                  | SWIR |          |          |          |          |
|------------------|------|----------|----------|----------|----------|
|                  | SVM (Linear) | KNN | SVM (Polynomial) | ANN | Average |
| DSIFT | 17.31 | 16.63 | 19.69 | 16.63 | 12.90 |
| FFT | 4.19 | 2.50 | 4.44 | 5.38 |  |
| LBP | 23.75 | 21.25 | 26.06 | 23.00 |  |
| SIFT | 10.75 | 5.31 | 10.63 | 7.94 |  |
| SURF | 11.81 | 7.81 | 12.69 | 10.31 |  |

|                  | RGB |          |          |          |          |
|------------------|-----|----------|----------|----------|----------|
|                  | SVM (Linear) | KNN | SVM (Polynomial) | ANN | Average |
| DSIFT | 36.94 | 23.44 | 41.81 | 34.81 | 22.23 |
| FFT | 4.13 | 2.69 | 4.63 | 3.88 |  |
| LBP | 30.63 | 23.56 | 33.50 | 32.00 |  |
| SIFT | 19.63 | 7.94 | 20.88 | 17.56 |  |
| SURF | 30.00 | 22.13 | 32.00 | 22.56 |  |

|                  | VNIR |          |          |          |          |
|------------------|-----|----------|----------|----------|----------|
|                  | SVM (Linear) | KNN | SVM (Polynomial) | ANN | Average |
| DSIFT | 37.69 | 25.63 | 42.06 | 43.13 | 22.68 |
| FFT | 4.63 | 2.50 | 4.50 | 4.44 |  |
| LBP | 28.88 | 22.38 | 30.38 | 29.00 |  |
| SIFT | 23.13 | 10.13 | 24.50 | 19.25 |  |
| SURF | 31.06 | 15.56 | 32.13 | 22.63 |  |

TABLE 3  Accuracy of each CNN method on FUSION, RGB, SWIR, VNIR

|                  | FUSION | RGB | SWIR | VNIR | Average |
|------------------|--------|-----|------|------|---------|
| GoogleNet | 86.44 | 98.13 | 87.94 | 94.88 | 91.84 |
| VGG16 | 95.50 | 98.13 | 84.44 | 98.38 | 94.11 |
| VGG19 | 94.56 | 99.50 | 84.81 | 98.81 | 94.42 |
| Average | 92.17 | 98.58 | 85.73 | 97.35 |  |

VNIR 97.35%. According to the overall results presented in Table 3, the best scores achieved on RGB dataset are compared to other ones. Besides, among the deep network models, the VGG19 provides the dominant results in almost all cases. It is interesting to find out that the favourable accuracies have only been obtained with the softmax classifier in case of executing experiments on convolutional models.

Likewise BoW framework, we have conducted the consensus rule on decisions estimated with CNN models and performance is reported with 99.94% accuracy rate. This obtained superior performance implies that only one sample was wrongly classified in case of performing the consensus rule on 1600 × 12 decision matrix, which includes 12 different predicted labels per each test sample. The empirical results provide that the deeper
network structure is, the better is the performance achieved, that is VGG19 is more robust than VGG16.

3.5 Trade-offs for CNN and BoW features

In this stage, we have compared the trade-off between CNN and BoW framework along with their advantages and disadvantages. For this purpose, the only performances estimated with consensus rule related to the both frameworks have been visualized in Figure 6(a) and 6(b). To evaluate the performance with all aspects, the accuracies and F-measures scores obtained for 40 classes have been summarized in Figure 6(a) and 6(b) for a fair comparison. The classes are labelled as from C1- to C40 by considering the letter order of names in Table 1.

Figure 6(a) shows the accuracy of each class obtained with CNN and BoW frameworks. As we can observe, CNN models pursued stable and successful results for 40 classes. On the other side, the BoW framework is not able to keep pace with the CNN framework, and it presents floating and low scores. When we take a look at the F-measure score, again, it is clear that the CNN framework attains top score, which is nearly always close to 1.

In order to comprehensively compare the performances of state-of-the-art methods together with the task to be processed, we have presented results of each study with utilized model, number of samples ($#N$), number of classes ($#C$) and imaging method. When zooming the results in Table 4, one can observe that for the rice blast disease classification [14], the performance was reported with 95% accuracy rate over 2906 positive samples and 2902 negative samples. With CovNet model [15], the recognition performance is around 97.33% when reserving 10,500 samples for training, 3000 samples for testing and 1500 samples for validation. In a comprehensive study [16], the seven popular pre-trained CNN architectures were utilized for two maize cultivars including haploid and diploid. The dataset includes 1230 haploids and 1770 diploids as 3000 maize seeds in total. The 2100 samples are used in training whereas the 900 samples are used for testing purposes. The highest accuracy rate was found with the VGG-19 model, which accounts for 94.22%. The performance of VNIR imaging based wheat cultivar classification is explored with a large dataset [17]. With a pretrained VGG16 framework, the 98.38% recognition result is achieved for 8000 wheat images. For the proposed study, we have obtained the superior accuracy results, reported to 99.94% classification score. This nice result indicates that only one sample is wrongly classified among 1600 test samples. This remarkable finding suggests that ensemble learning is important for further boosting the performance for predicting the wheat label. Moreover, with a single model, namely VGG19, we have achieved 99.50% accuracy rate. This nice achievement corroborates the results of earlier findings of a previous study [16].

These objective good scores indicate that in order to restrain the lower performance, one should consider the CNN models when it comes to develop a wheat cultivar identification system. Additionally, from overall evaluation, we can observe that deeper network models for wheat classification give more accurate results. Unfortunately, the effectiveness of traditional features for wheat identification is found to be suffering from weak performance when compared with CNN feature extraction tools. Since the BoW features are shape-oriented, they always collapse if applied on similar repeated texture patterns, like the so-called wheat kernels which strongly resemble patterned textures.

Although the quality of wheat cultivars cannot be simply restricted to a definition, we can categorize them in terms of their suitability for bread and macaroni production. In Table 1, the Cesit1252, Kandura, Yelken cultivars are used for macaroni production, whereas the rest are used for bread production.
### TABLE 4 Performance comparison with state-of-the-art methods

| Target grain | Imaging mode | Wavelength range (nm) | #N  | #C  | Accuracy (%) | Method     | Ref.  |
|--------------|--------------|------------------------|-----|-----|--------------|------------|-------|
| Rice         | RGB          | –                      | 5808| 2   | 95.00        | CNN        | [14]  |
| Rice         | RGB          | –                      | 5808| 2   | 95.82        | CNN+SVM    | [14]  |
| Wheat        | VNIR         | 400–1100               | 8000| 40  | 98.38        | VGG16      | [17]  |
| Wheat        | RGB          | –                      | 15000| 15  | 97.53        | CovNet     | [15]  |
| Maize        | RGB          | –                      | 3000| 2   | 92.67        | AlexNet    | [16]  |
| Maize        | RGB          | –                      | 3000| 2   | 92.78        | VGG16      | [16]  |
| Maize        | RGB          | –                      | 3000| 2   | 94.22        | VGG19      | [16]  |
| Maize        | RGB          | –                      | 3000| 2   | 90.89        | GoogleNet  | [16]  |
| Maize        | RGB          | –                      | 3000| 2   | 92.44        | Resnet18   | [16]  |
| Maize        | RGB          | –                      | 3000| 2   | 91.11        | Resnet101  | [16]  |
| Wheat        | RGB          | –                      | 8000| 40  | 99.50        | VGG19      | This work|
| Wheat        | RGB          | –                      | 8000| 40  | 99.94        | Ensemble   | This work|

The ground truth labels of wheat cultivars are determined by the Republic of Turkey, Ministry of Agriculture and Forestry. According to hectolitre weight [37, 38], the Yelken is a very suitable cultivar in terms of macaroni production, while the Bezoštaja or Konya-2002 is a good one for bread production.

## 4 CONCLUSION

This study investigates effectiveness of using different imaging techniques—including RGB, VNIR, SWIR—together with traditional and modern feature extraction tools with a purpose of wheat cultivar identification. For this purpose, we have made experiments over 40 different wheat cultivars as each cultivar consists of 200 samples, making a total of 8000 samples. Throughout the experimental stages, we have observed that the BoW features are not robust and they collapse when it comes to application on wheat cultivar detection. The performance degradation of the BoW framework is mostly due to characteristics of hand-crafted features like SIFT, SBFD, LBP, SURF and DIFT that are focused only on local variations. However, the local variations can resemble each other when considering within/between class variability and similar textures through the image domain. Unlike the BoW features, an integrated CNN framework, so-called deep learning structures, can incorporate the spatial information and cognize the discriminative details stated on wheat cultivars. The experimental works turn out that the CNN framework gives more robust performance compared with traditional feature extraction tools when looking at the achieved accuracy rates of 99.94% and 68.04% by consensus rule. As a system that relied on the VNIR camera and CNN features, it would be sufficient to generate more objective and accurate decisions without using expensive equipment in the wheat industry. The datasets are available at https://github.com/isahhin/wheat-classification.

## ACKNOWLEDGEMENTS

This work was part of research project number 1160576 supported by the Turkey Scientific and Technological Research Projects Support Program. This work was also partly supported by the Eskisehir Osmangazi University Scientific Research Project Commissions (grant no.: 2016–1120). The authors would like to thank the Mikro-Tasarım Company and the Eskisehir Commodity Exchange Laboratory for their help and support with their knowledge and experience.

## REFERENCES

1. Choudhary, R., et al.: Classification of cereal grains using wavelet, morphological, colour, and textural features of non-touching kernel images. Biosyst. Eng. 99(3), 330–337 (2008)
2. Choudhary, R., et al.: Identification of wheat classes using wavelet features from near infrared hyperspectral images of bulk samples. Biosyst. Eng. 102(2), 115–127 (2009)
3. Xing, J., et al.: Using a short wavelength infrared (SWIR) hyperspectral imaging system to predict alpha amylase activity in individual Canadian western wheat kernels. Sens. Instrum. Food Qual. Saf. 3(4), 211 (2009)
4. Xing, J., et al.: Detection of sprout damage in Canada western red spring wheat with multiple wavebands using visible/near-infrared hyperspectral imaging. Biosyst. Eng. 106(2), 188–194 (2010)
5. Shahin, M.A., Symons, S.J.: Detection of Fusarium damaged kernels in Canada western red spring wheat using visible/near-infrared hyperspectral imaging and principal component analysis. Comput. Electron. Agric. 75(1), 107–112 (2011)
6. Singh, C.B., et al.: Fungal damage detection in wheat using short-wave near-infrared hyperspectral and digital colour imaging. Int. J. Food Prop. 15(1), 11–24 (2012)
7. Pourreza, A., et al.: Identification of nine Iranian wheat seed varieties by textural analysis with image processing. Comput. Electron. Agric. 83, 102–108 (2012)
8. Serranti, S., et al.: The development of a hyperspectral imaging method for the detection of Fusarium-damaged, yellow-berry and vitreous Italian durum wheat kernels. Biosyst. Eng. 115(1), 20–30 (2013)

9. González, et al.: Chilean flour and wheat grain: Tracing their origin using near infrared spectroscopy and chemometrics. Food Chem. 145, 802–806 (2014)

10. Kandpal, L.M., et al.: Short wave infrared (Swir) hyperspectral imaging technique for examination of aflatoxin B1 (Afb1) on corn kernels. Food Control. 51, 171–176 (2015)

11. Barbedo, J.G., et al.: Detecting Fusarium head blight in wheat kernels using hyperspectral imaging. Biosyst. Eng. 131, 65–76 (2015)

12. Ziegler, J.U., et al.: Near-infrared reflectance spectroscopy for the rapid discrimination of kernels and flours of different wheat species. J. Food Compos. Anal. 51, 30–36 (2016)

13. Senthilkumar, T., et al.: Detection of ochratoxin a contamination in stored wheat using near-infrared hyperspectral imaging. Infrared Phys. Technol. 81, 228–235 (2017)

14. Liang, W.-J., et al.: Rice blast disease recognition using a deep convolutional neural network. Sci. Rep. 9(1), 1–10 (2019)

15. Lingwal, S., et al.: Image-based wheat grain classification using convolutional neural network. Multimed. Tools Appl. 1–25 (2021). https://link.springer.com/journal/11042/online-first?page=10

16. Altmuntas, Y., et al.: Identification of haploid and diploid maize seeds using convolutional neural networks and a transfer learning approach. Comput. Electron. Agric. 163, 104874 (2019)

17. Özkan, K., et al.: Identification of wheat kernels by fusion of RGB, SWIR, and VNIR Samples. J. Sci. Food Agric. 99(11), 4977–4984 (2019)

18. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms. In: Proceedings of the 18th ACM International Conference on Multimedia, Florence (2010)

19. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Internat. J. Comput. Vis. 60(2), 91–110 (2004)

20. Bay, H., et al.: Speeded-up robust features (SURF). Comput. Vis Image Underst. 110(3), 346–359 (2008)

21. Ojala, T., et al.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE PAMI., 24(7), 971–987 (2002)

22. Cevikalp, H., et al.: Return of the king: The Fourier transform based descriptor for visual object classification. In: 21st Signal Processing and Communications Applications Conference (SIU), Haspolat (2013)

23. Olgun, M., et al.: Wheat grain classification by using dense sift features with Svm classifier. Comput. Electron. Agric. 122, 185–190 (2016)

24. He, C., et al.: Exploiting score distribution for heterogenous feature fusion in image classification. Neurocomputing 253, 70–76 (2017)

25. Lin, G., et al.: Visual feature coding based on heterogeneous structure fusion for image classification. Inf. Fusion. 36, 275–283 (2017)

26. Wang, Y., et al.: Separable vocabulary and feature fusion for image retrieval based on sparse representation. Neurocomputing. 236, 14–22 (2017)

27. Gülmezoglu, M., et al.: The common vector approach and its relation to principal component analysis. IEEE Speech Audio Process. 9(6), 655–662 (2001)

28. Özkan, K., et al.: Multispectral image fusion based on the common vector approach. In: International Symposium on Innovations in Intelligent Systems and Applications, Sinaia (2016)

29. Cevikalp, H., et al.: Discriminative common vectors for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. 27(1), 4–13 (2005)

30. Işık, Ş., et al.: A new subspace based solution to background modelling and change detection. Int. J. Intell. Syst. Appl. Eng. 4(Special Issue-1), 82–86 (2016)

31. Özkan, K., Işık, Ş.: A novel multi-scale and multi-expert edge detector based on common vector approach. Int. J. Electron. Commun. 69(9), 1272–1281 (2015)

32. Ergin, S., et al.: A new implementation of common matrix approach using third-order tensors for face recognition. Expert. Syst. Appl. 38(4), 3246–3251 (2011)

33. Wang, X., et al.: Bag of contour fragments for robust shape classification. Pattern. Recognit. 47(6), 2116–2125 (2014)

34. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556 (2014)

35. Krizhevsky, A., et al.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, Tahoe (2012)

36. Szegedy, C., et al.: Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston (2015)

37. Aydoğan, S., et al.: Research of quality traits of some durum wheat and bread wheat varieties. Kahramanmaras Suçi University Tarım ve Doğa Dergisi 22, 264–271 (2019)

38. Aydoğan, S., Soylu, S.: Determination of yield, yield components and some quality properties of bread wheat varieties. Tarla Bitkileri Merkez Araştırma Enstitüsü Dergisi 26, (1), 24–30 (2017)

How to cite this article: Işık, Ş., et al.: Consensus rule for wheat cultivar classification on VL, VNIR and SWIR imaging. IET Image Process. 1–11 (2021).
https://doi.org/10.1049/ipr2.12206