Wheat Grains Recognition System using Machine Learning

Wheat is one of the most planted grains in the world, almost all of which is produced for consumption, and is having high nutritional value. The Food and Agriculture Organize shows that the global wheat production is 756.51 million tons in 2018. Quality in wheat grains is estimated by checking for number of broken wheat grains among the samples of wheat grains. In practice, the number of wheat grains is often obtained by manual enumeration. This method is time-consuming, tedious and error-prone. It is thus necessary that a convenient and efficient grain counting method is developed. In the present scenario, as we all know, COVID19 is spreading and lacs of people died worldwide. Therefore, it is necessary to have a method of estimating the wheat grains quality without coming in contact of anything. Machine learning is best option in such conditions.

Publicly available datasets, including ImageNet, COCO and PASCAL VOC, do not contain the wheat grains of this study. Therefore, we manually collected the wheat grains dataset with white background of A4 paper is selected. Two images as sample are shown in Fig.1.

Fig.1. Sample images of Wheat Grains

Collection of wheat grains in a single image results in two challenges. The first one is to separate the object from the background, and the other is to resolve the difficulty in separating multiple grains that are adjacent. The typical approach for the first process is to select a single color plate as the background, and then easily separate the object from the background using a color extraction algorithm. The second process focuses on grain counting. Numerous existing segmentation algorithms for adjacent grains have been proposed, including expansion corrosion algorithms, watershed algorithms and feature point matching algorithms. Here, Watershed segmentation is performed to crop the wheat grains from the collection and these are classified using the classifier, Support Vector Machine (SVM).

Watershed Segmentation: The watershed is a classical algorithm used for segmentation, that is, for separating different objects in an image. Starting from user-defined markers, the watershed algorithm treats pixels values as a local topography (elevation). The algorithm floods basins from the markers until basins attributed to different markers meet on watershed lines. In many cases, markers are chosen as local minima of the image, from which basins are flooded. In the example below (Fig.2), two overlapping circles are to be separated. To do so, one computes an image that is the distance to the background. The maxima of this distance (i.e., the minima of the opposite of the distance) are chosen as markers and the flooding of basins from such markers separates the two circles along a watershed line.

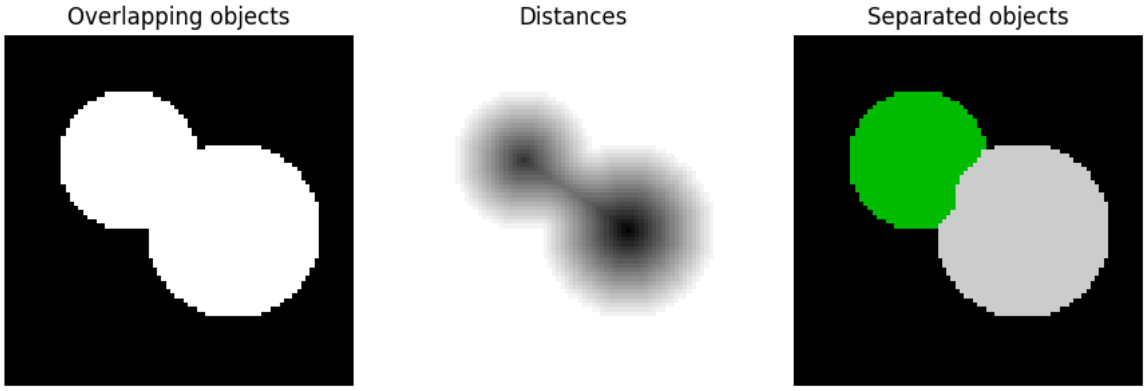

Fig.2. Watershed Segmentation Example

SVM: "Support Vector Machine" (SVM) is a supervised machine learning algorithm which can be used for both classification and regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well.

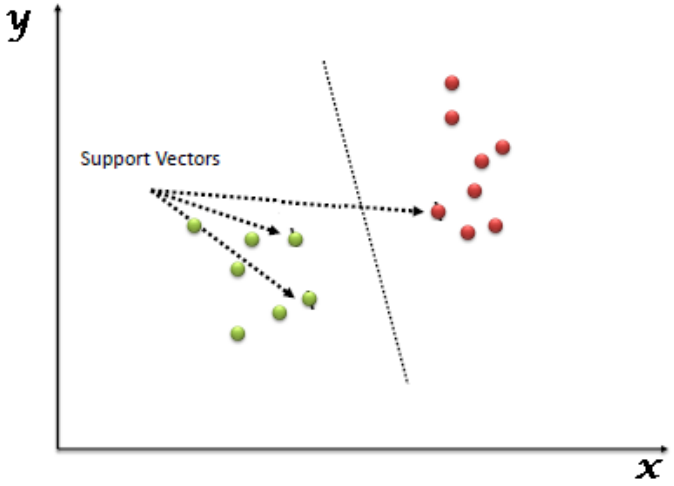

Fig.3. Basic diagram of SVM

All the images are resized to 750x1000 pixels. The input images were pre-processed, including the conversion of color space, image turnover, contrast change and so on. Stepwise results of preprocessing and segmentation are shown in Fig.4.

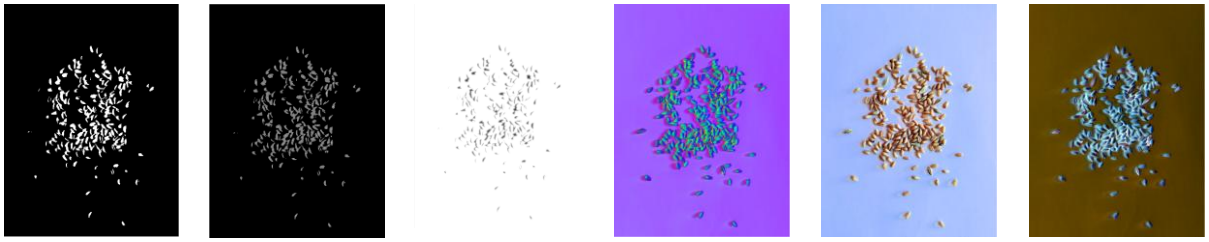

Fig.4. Steps in Wheat Grains Algorithm

All the images of wheat grains go through the pre-processing stage and then the method of segmentation. Images are resized to fix size of 750x1000 pixels. In pre-processing, image complement is performed to differentiate the background from the wheat grains in the image. Haziness is reduced in the image and then filtered to extract the wheat grains in the image. Saturation color scale is extracted from the filtered output as shown in Fig.4. After that, watershed segmentation is utilized to separate the wheat grains in the images. This segmentation technique is a region based method that utilizes image morphology. Then, the wheat grains are highlighted with the help of properties like centroid, major axis length, minor axis length and area covered as shown in Fig.5.

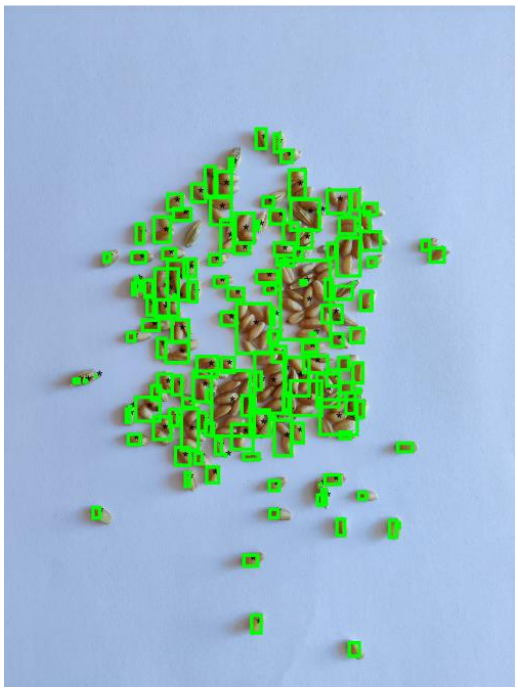

Fig.5. Segmented and labeled image of wheat grains

From the segmented and labeled image of wheat grains, the grains are cropped and stored in different classes. Classes are Complete wheat grains, broken wheat grains and infected wheat grains. Various random images, clicked manually, were utilized in this process. After that, these cropped images resized to 50x50 pixels are used to train the classifier (SVM).

Wheat Grains Recognition model is trained using the SVM classifier and then, this model is used to test the images. Output is observed as the number of Complete, broken and infected wheat grains. Also, percentage of complete wheat grains in the sample defines the quality of wheat grains sample is also the output.