Computers and Electronics in Agriculture 169 (2020) 105200

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

journal homepage: www.elsevier.com/locate/compag

A method for counting and classifying aphids using computer vision

Elison Alfeu Lins[a], João Pedro Mazuco Rodriguez[a], Sandy Ismael Scoloski[a], Juliana Pivato[b], Marília Balotin Lima[b], José Maurício Cunha Fernandes[a,b], Paulo Roberto Valle da Silva Pereira[b], Douglas Lau[b], Rafael Rieder[a,*]

[a] *University of Passo Fundo (UPF), Passo Fundo, Rio Grande do Sul, Brazil*[1]

[b] *The Brazilian Agricultural Research Corporation (Embrapa Wheat), Passo Fundo, Rio Grande do Sul, Brazil*[2]

ARTICLE INFO

Keywords:
Aphids
Classification
Computer vision
Counting
Measurement

2010 MSC:
00-01
99-00

ABSTRACT

Aphids are insects that attack crops and cause damage directly, by consuming the sap of plants, and indirectly, by vectoring microorganisms that can cause diseases. Cereal crops are hosts for many aphid species, including *Rhopalosiphum padi* (an economically important aphid species). Recording and classifying aphids are necessary for evaluating and predicting crop damage. Thus, serving as a basis for decision making on the utilization of control measures. It can also be useful to evaluate plant resistance to aphids. Traditionally, the recording process is manual and depends on magnification and well-trained staff. The manual counting is also a time-consuming process and susceptible to errors. With this in mind, this paper presents a method and software to automate the counting and classification of *Rhopalosiphum padi* using image processing, computer vision, and machine learning methods. The text also presents a comparison of manually counts from experts and values obtained with the software, considering 40 samples. The results showed strong positive correlation in counting and classification (r_s = 0.92579) and measurement (r = 0.9799). Concluding, the software proved to be reliable and useful to aphid population monitoring studies.

1. Introduction

Aphids can have an economic impact on agricultural production. The yield losses are due to direct damage (sieve drain and plant reaction) and indirect damage (often the most important, due to virus transmission) (Dedryver et al., 2010). Cereal crops are hosts for many aphid species, which are important pests like *Sitobion avenae*, *Rhopalosiphum padi*, *Schizaphis graminum*, *Metopolophium dirhodum* and others (Shavit et al., 2018). These aphids are accountable for transmitting *Barley/Cereal yellow dwarf virus*(B/CYDV) one of the most important cereal viruses in the world (Shavit et al., 2018).

Wheat is among the crops affected by aphid species. This crop is the second most produced cereal in the world, with significant importance in the world economy. Direct damage by aphids is responsible for mean annual losses of 700,000 tonnes of wheat in Europe (Dedryver et al., 2010). According to Yahya et al. (2017), aphids have fast multiplication rate and have potential to affect crop development within few days, and the yield losses can be up to 7.9 to 34.2%.

Aphid counts may be required for crop monitoring as well as population studies of these insects. In the first case, tools to monitor the size of aphid populations and their growth are useful for control decision making. In the second, studies aimed at determining how the aphid population reacts to biotic and abiotic factors require that both growth and population structure be estimated (Savaris et al., 2013).

Considering that small populations can transmit diseases and multiply rapidly, chemical control is recommended based on a population density threshold (calculated for each region and crop) (Hansen, 2000). Therefore, counting, classification, and measuring aphids is critical to estimate if there is a risk factor for crops (Shajahan et al., 2017). Weekly evaluations should take place for monitoring aphid density at random locations in the field (Martin et al., 2015).

Thus, several field samples are collected and taken to the laboratory for aphid identification and counting. One of the methods traditionally employed is the use of a stereoscopic microscope and a background grid

* Corresponding author.

E-mail addresses: 14787@upf.br (J.P.M. Rodriguez), 150662@upf.br (S.I. Scoloski), julianapivato15@gmail.com (J. Pivato), mariliabalotinlima@gmail.com (M.B. Lima), jmauricio@upf.br (J.M.C. Fernandes), paulo.pereira@embrapa.br (P.R.V. da Silva Pereira), douglas.lau@embrapa.br (D. Lau), rieder@upf.br (R. Rieder).

[1] http://www.upf.br.

[2] http://www.embrapa.br/trigo.

https://doi.org/10.1016/j.compag.2019.105200
Received 3 April 2019; Received in revised form 26 December 2019; Accepted 28 December 2019
Available online 10 January 2020
0168-1699/

of lines for visualization and counting (or sampling) of the insects. In addition to counting, for aphid population studies, insects are separated by developmental stage into nymphs, wingless adults and winged adults (Carter et al., 1982). In samples with a high density of insects, partial counting, and estimation of the total population of insects is required. It is a time-consuming method that tends to have a high error rate, since human beings are susceptible to physiological and psychological phenomena like fatigue, visual illusions and boredom (Barbedo, 2014).

In this context, using computer vision techniques to evaluate the dynamics of the aphid pest populations would have a value on integrated pest management and decision making in agriculture - in particular, the wheat crop (Barbedo, 2014). With an automatic counting system, it should be possible to shorten the examination time per sample. It also should contribute to a more reliable and accurate final result, indicating, for example, the infestation density of a given area (Gonzalez and Woods, 2006).

Besides, through different computational systems and image collections, different patterns of analysis can be configured (Nissimov et al., 2015). The algorithm should include a standard method of counting and classifying aphids defined by scientific methods (Parer and Hamilton, 2010; Arambula Cosio et al., 2003). Another feature is that the software is not susceptible to errors caused by human physiological and psychological phenomena (Lu and Ferrier, 2004).

A computer program with this approach can contribute to the reduction in the processing time of aphid samples allowing, for example, the fast and accurate collection of information (currently obtained manually by counting and identifying each aphid in the sample). In addition, a future application that enables automated monitoring of field aphid populations can change the way pest management is carried out, enabling near real-time decisions. Currently, the monitoring of aphids in crops is extremely time consuming and depends on specialized labor.

Hence, this paper presents details about the development of a computer vision method for counting, classifying, and measuring aphids using the OpenCV and TensorFlow libraries. This work aims to offer a tool for agriculture specialists to improve monitoring aphid population dynamics.

2. Related work

Several applications in agriculture involve or require the use of image processing. Among them, a portion relates to aid in pest detection and decision making. By means of break-even analysis, it is possible to know if it is needed to apply a particular pesticide, considering a pest population threshold at which the resulting damages are equal to the costs of control measures. The proposed solution can also reduce the use of pesticides and, thus, preserving the environment. Therefore, the use of image processing techniques has become a trend in recent years (Mande et al., 2018).

Some approaches for pest detection in plants have already been presented in other studies. Barbedo (2014) showed whitefly counts on soybean leaves, identifying the characteristics in each phase of the life cycle of these insects, based on the staining of individuals at each stage of growth. In the next step, he proposed an algorithm for counting units of whitefly.

In another study, Liu et al. (2016) reported monitoring of aphid populations and species identification, providing important data related to pest population dynamics and integrated pest management. Within this context, the authors developed software for counting and identifying aphids in the wheat crop, which does not require traps or an explicitly defined background. Images for processing are captured and processed directly on infested plants.

In addition to direct plant detection, Xia et al. (2015) proposed image processing techniques to identify three species of greenhouse pests using low-resolution images. The process begins with the capture of insects using adhesive traps, which are digitized and then processed. Another computational method was artificial neural networks.

Wadhai et al. (2015) introduced a pest detection system in greenhouses using image processing and neural networks exclusively applicable to greenhouse whitefly (*Trialeurodes vaporariorum*) and aphids.

These studies demonstrated that the use of digital image processing already brings satisfactory results, even without the use of recent machine vision techniques, like deep learning approach using convolution neural networks. According to Abdullahi et al. (2017), this combination is the most effective approach generating fast and excellent results for image classification.

The use of the two methods, therefore, should increase the accuracy and reliability of detection systems. Our computer vision proposal assumes this as a feature of a software solution to count, classify, and measure insects.

3. Materials

Embrapa Wheat supported the development of this work. The institution located in Passo Fundo is one of the 47 units of the Brazilian Agricultural Research Corporation (Embrapa) (Embrapa Trigo, 2019). It provided the materials needed to obtain the images used during the development of the system.

This work proposes the development of a software (**AphidCV**) that implements a method of aphids counting and classification using computer vision resources. This software is a tool to enable aphid population studies and factors affecting population growth. In general, controlled population studies require countings, classification (development stage count), and population morphometry. In developing the software, we used the programming language Java 7, libraries OpenCV 3.0, and TensorFlow r1.1.

Java is one of the most important programming languages in the world, based on the object-oriented paradigm. Maintained by Oracle Corporation, it features security, portability, robustness, and multithreaded capabilities for applications in a neutral, interpreted, distributed, and dynamic architecture (Schildt, 2017).

OpenCV (Open Source Computer Vision) is an open-source, cross-platform computer vision library with more than 2500 algorithms for video and image analysis (OpenCV Library, 2019). Originally developed by Intel, it is currently openly maintained by its community. It has interfaces for C++, C, Python, Java, and MATLAB.

TensorFlow is an open-source library for machine learning. Initially, it was developed by the Google Brain Team for research purposes in deep neural networks (Tensorflow, 2019), to detect and decipher patterns and correlations, analogous to human learning and reasoning.

4. Methods

4.1. Acquisition

To the classification method definition, the image acquisition step is essential, since it is necessary to establish a standard configuration for sample acquisition and reduction of bias. This step, with the exception of the digitization, is based on the manual counting method used at Embrapa Wheat.

The process begins with the collection of samples by field researchers, placing the aphids in test tubes containing aqueous solution. Then, the samples are transferred to circular and transparent Petri dishes, 14 cm in diameter. In the sequence, the dishes are scanned with the collected materials. The scanning is done on ordinary desktop scanner, at the resolution of 1200 dpi, dimensions of 7000×7000 pixels and color image. The development of the software is performed with samples containing colonies of the *Rhopalosiphum padi*. However, our method is projected to support the study of any aphid species, considering calibration by species.

Petri dish has 1 cm of height. Even closing the scanner cover, external light may input and affect the quality of the image. As the

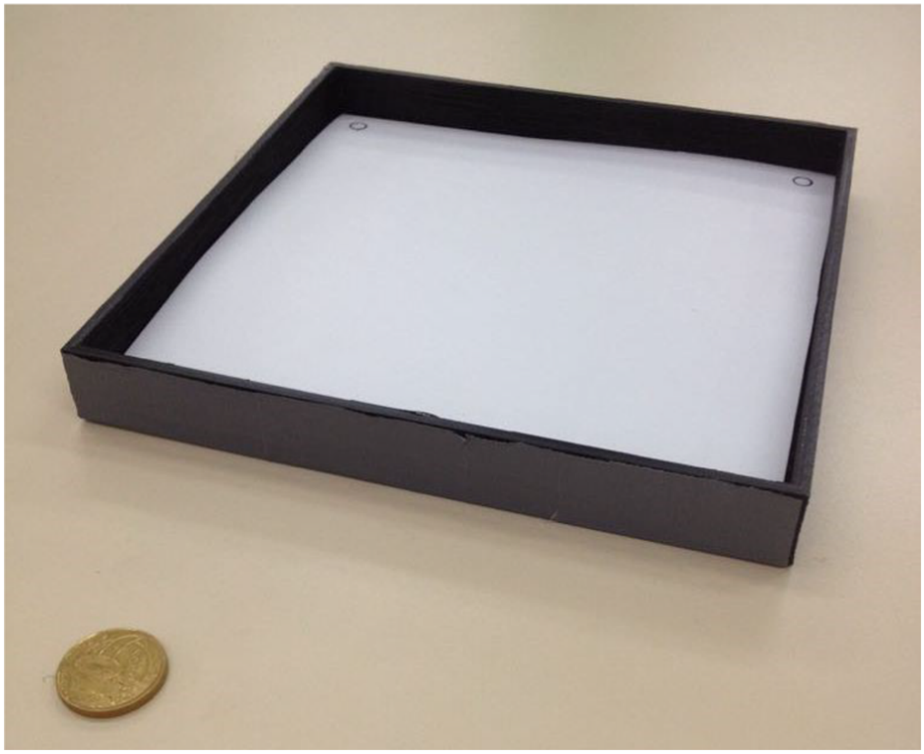

Fig. 1. Black-box used during image acquisition.

classification step depends on the small details present in the image it was necessary to solve this problem using a black-box. Fig. 1 demonstrates the device developed in a 3D printer in the dimensions of 15 cm × 15 cm that completely seals the external light input on the board.

Fig. 2 demonstrates a comparison, where (a) was scanned without the use of the carton, resulting in a much darker image than (b), scanned using the carton. Details of the wings of the aphids were more visible with the use of the equipment.

Circles of 0.5 cm in diameter drawn in the four corners of the box were used as reference objects (Fig. 3). These references can be easily identifiable in image (based on either location or appearances), and they have known dimensions (in terms of millimeters). So, we defined a ratio in order to convert pixel to millimeter unit, measuring the real size of aphids with computer vision.

To obtain the dataset of images for artificial neural network (ANN) training, we developed an auxiliary software, called CropAphid, developed in Java language. From this tool, the inputs were processed using the same digital image processing techniques described in the next section. Objects similar to aphids were selected out constituting the primary database. The sample with aphids was restrained in Petri dishes, as shown in Fig. 4.

After, an aphid specialist separated the images classifying as nymphs, adult wingless, winged adults, and false (plant tissue, soil, debris, etc.).

4.2. Processing

For the image processing, three methods were used to extract the regions of interest from the images, and to eliminate trash like exoskeletons, using OpenCV resources.

The process begins with macro segmentation. A scanned image contains the data outside the perimeter of the Petri dish. Therefore, it is necessary to apply a method of excluding these external areas.

Fig. 5 represents this situation, where the black part was considered unnecessary (right) in relation to the input image (left). For that, the Hough gradient method was applied to detect circles in an image, considering the gradient information of edges. The first stage involves edge detection and finding the possible circle centers and the second stage finds the best radius for each candidate center (Kaehler and Bradski, 2016). This implementation enables the algorithm to run much faster and helps to reduce the noise substantially.

The next step also performs segmentation but intending to exclude the exoskeletons discharged by aphids as they grow. Remains of plant

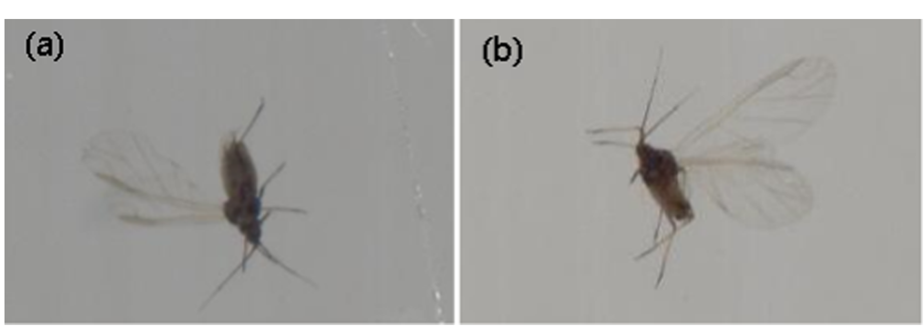

Fig. 2. Comparison of the image obtained in a scanner (a) without the box and (b) scanned with the box.

Fig. 3. Reference objects in image to calculate the size of aphids using a pixel per metric ratio.

parts, in most cases, have a lighter color hue than insects. This was also removed by a thresholding algorithm applied to the region of interest.

Thresholding is considered an essential and fundamental step during the segmentation process (Szeliski, 2010). Its principle is to separate areas of an image in two classes, such as the background and an object. The function sets a cut-off value (T) for an image that, when processed, will consider all pixel values higher than the threshold equal to T, otherwise as zero. By means of calibration, the value of T was set at 110. Fig. 6 shows the result of this step in an enlarged area of the image. Thus, potential false positives are eliminated, since exoskeletons have a high degree of resemblance with the insects.

In the third step, the contours of the objects found in the image are recognized by edge detection techniques. After recognizing each contour, the dimensions of areas of outstanding interest are examined to determine whether it can be considered an aphid. The contour measurements are the length, the width, and the perimeter.

If the contour is regarded as a likely insect, it will be extracted from the image – the extraction results from a rectangular contour cut.

The weight and body measurements were recorded in a sample of 328 wingless adults of *Rhopalosiphum padi*. The measures with a high precision ruler included the length, the perimeter, and the area of each aphid. The weight of each aphid was recorded on a high precision scale. A model was fitted to data to determine the relationship between weight and measures. The model equations are shown in Table 1. The estimated weight of each aphid is calculated by entering the area and perimeter in the above equations.

4.3. Classification

Each recognized area is evaluated through z learning techniques to determine the development stage. The characteristics of the current image are compared with positive and false-positive images in the database (generated by CropAphid).

After obtaining the potential contours, it is necessary to label the insects in three classes according to the stage of development (nymphs, wingless adults, and winged adults). Nymphs are aphids in the early stages of life that are smaller in size if compared to adults. Also, they have a less rounded morphology. Wingless adults are characterized by their larger body size and more rounded morphology. Winged adults have wings.

To make the classification, we used deep learning algorithms available in the TensorFlow library. From a Java Console module, input image clippings and a primary database were created, based on the previously presented image processing steps.

To generate this dataset, 120 samples similar to Fig. 4 were used. Out of these images, approximately 30,000 cutouts of 120 × 120 pixels were obtained. These cutouts were submitted to human classification, considering four classes: “nymph”, “wingless”, “winged” and “false”. For machine learning, we assigned 2475 images of the “winged” class (Fig. 7), 4173 of “wingless” class (Fig. 8), 15896 of “nymph” class (Fig. 9) and 4835 of “false” class (Fig. 10).

A portion of 90% of the dataset was used for training and 10% for validation. The ANN model used was Inception-v3 (Szegedy et al., 2016), a Google-structured model for the classification of general-purpose color images. As a result, machine learning with a precision of 98.10% was obtained in the validation dataset after 30 epochs.

However, the validation accuracy may not be compatible with reality since the results after each training cycle influence the back-propagation algorithm. Therefore, a new test dataset was collected with new 220 dishes never before analyzed by the trained model. The results are described below.

Fig. 4. Scanned dish example.

4.4. Presentation

Our proposed method and tools mentioned above resulted in the application called AphidCV. The system is for Windows platform, user-friendly, without the need for advanced computer skills or complicated configurations.

The user communicates with the software through a graphical user interface built-in JavaFX. Simply, the user has to select a source folder containing the images for processing. After that, each available image is processed. Fig. 11 shows the window with processing results.

Output images are also saved for visual analysis by users. Fig. 12 shows an enlarged cutout of a post-processing image, where the outlines in yellow were considered nymphs, the wingless adults were in red, and winged adults were in the purple. For the demonstration, the numbers on the side represent the measures in millimeters of the area, perimeter, length, and degree of certainty of the ANN classification. The estimated weight derived from equations presented in Table 1.

In addition to the processed images, AphidCV also generates JSON files for Embrapa Wheat third-party software, and CSV files with a list of all detected insects and their data, as shown in Fig. 13. JSON is a lightweight data-interchange formats used to transmit data objects consisting of attribute–value pairs and array data types, and CSV is a

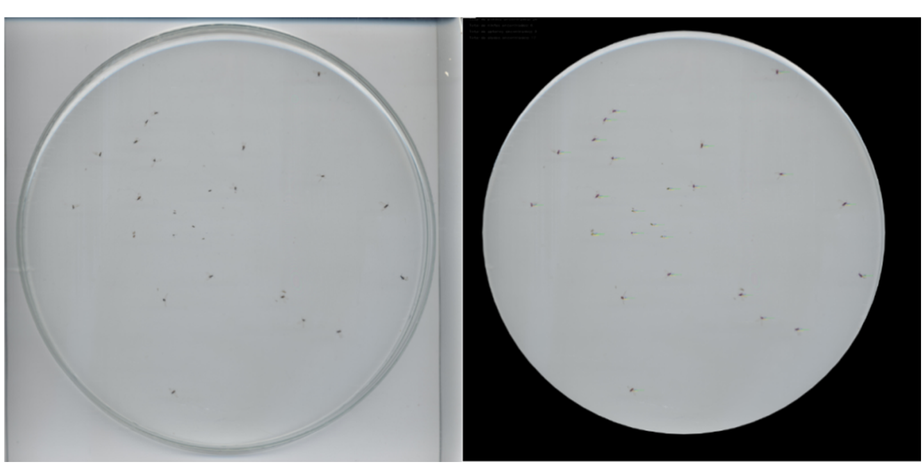

Fig. 5. Macro segmentation.

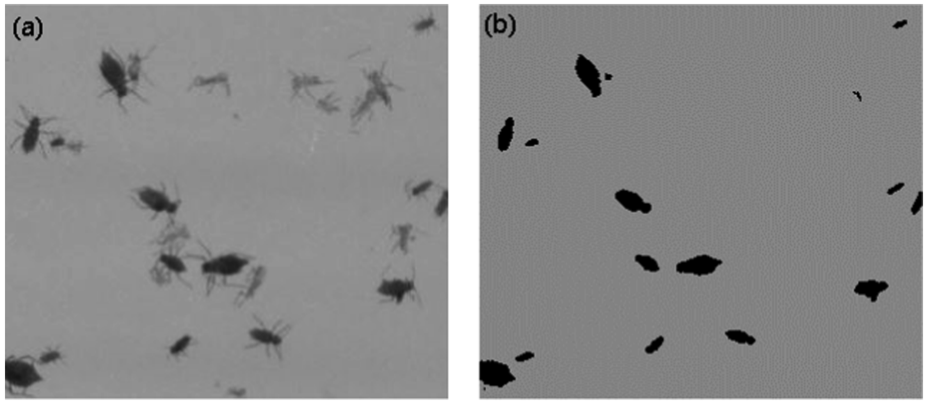

Fig. 6. Thresholding results.

Table 1
Transformation equations of area (mm^2) and length (mm) for weight (g).

Area (a)	Length (l)
$0.0002a^2 + 0.0002a + 0.00002$	$0.0006l^2 - 0.0009l + 0.0005$

common data exchange format that is widely supported by spreadsheets. The software also creates a "Doubtful" folder containing the insects in which the ANN has a degree of certainty less than 80%, conceding a human reassessment in these cases.

5. Results

At the end of the proposed software development, the validation step was conducted, comparing the data obtained by standard counting and AphidCV. We also implemented an examination using the CropAphid software to count and classify the cuts made by the AphidCV.

The standard is the method currently in use at Embrapa Wheat. The CropAphid cuts and detects contours, but the classification is done visually by experts. The AphidCV consisted of a fully computerized method, with counting, sorting, and printed output.

A set of 40 samples was used to evaluate the software. The recorded variables were the number of aphids, their classes, and time. First, the number of aphids was recorded and classified by experts at the Embrapa Wheat laboratory. Afterward, the samples were scanned and submitted to CropAphid, and the number of aphids determined. The classification was done by experts. Finally, the collection of images was processed in AphidCV.

For statistical analysis, the Shapiro-Wilk test was used first to check the data normality. All sets presented a non-normal distribution at $n = 40$ with a significance level of 5%. The Spearman Correlation (r_S) examined the relationship between methods.

The comparison pairs were:

- Standard × AphidCV;
- CropAphid × AphidCV.

All tests considered the total number of aphids to evaluate the counting method efficiency. The number of nymphs, wingless, and winged adults was used to compare the effectiveness of the classification method. Table 2 shows the number of aphids (total and stage of development) for each method.

5.1. Standard × AphidCV

Fig. 14 indicates a marked association between the two methods regarding the total number of aphids ($r_s = 0.92579$). The Spearman's correlation by stage was "nymph" ($r_s = 0.89601$), "wingless" ($r_s = 0.86138$) and "winged" ($r_s = 0.63149$).

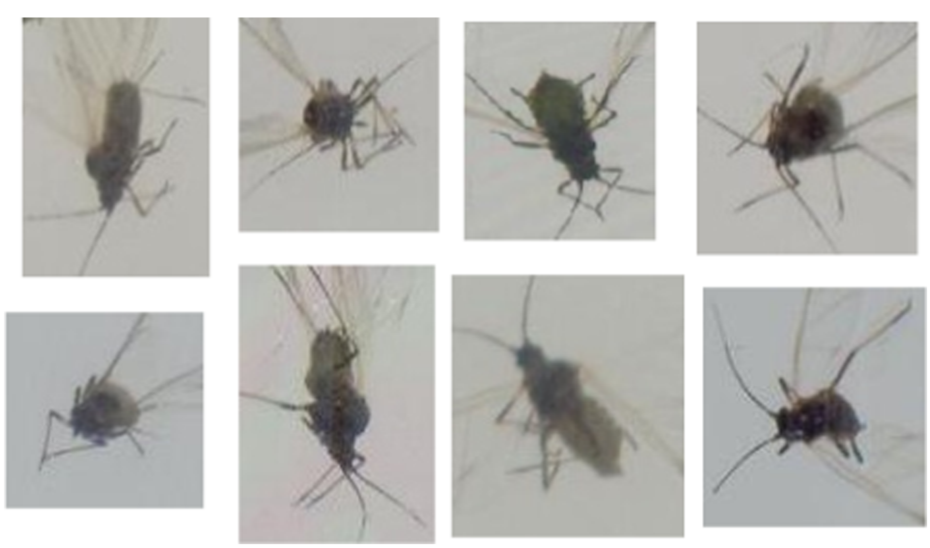

Fig. 7. Examples of "winged" class images.

Fig. 8. Examples of “wingless” class images.

5.2. CropAphid × AphidCV

Fig. 15 also shows a marked association between the two methods regarding the total number of aphids (r_s = 0.99249).The Spearman’s correlation by stage was “nymph” (r_s = 0.98555), “wingless” (r_s = 0.84755) and “winged” (r_s = 0.69146).

5.3. Measuring comparison

A sub-sample consisting of 40 randomly selected aphids was used. The aphid body length, in millimeters, was determined by Motic microscopy software (Motic - Moticam Software Series, 2019).

A Shapiro-Wilk test was used to check the data normality. Normal distribution was confirmed, and Pearson’s correlation was applied.

Aphid’s body length measures obtained under a microscope were highly correlated (r = 0.9799) with measures obtained in AphidCV (Fig. 16).

A comparison of the aphid’s weight determined on a scale and in AphidCV was also evaluated. First, we organized six samples representing aphid development stages: first-instar nymphs (N1 = 680), second instar nymphs (N2 = 174), third-instar nymphs (N3 = 181), fourth-instar nymphs (N4 = 227), winged adults (N5 = 135), and wingless adults (N = 50). Next, we estimated the mean weight per aphid.

Body length and area were determined in sub-samples of 30 aphids at each development stage. Again, Motic software was used to obtain length, perimeter, and total area. Based on the measurements, the mean weight was estimated.he medium-weight of these populations.

Using the same sub-samples, we used AphidCV to generate length, perimeter, area, and weight of each specimen.

The means of length, area, and weight of each subsample were used for comparison. Shapiro-Wilk test confirmed normal distributions, and Pearson’s correlation coefficient was applied.

The results showed a high correlation (r = 0.995) between weight values measured with the aid of a scale and with AphidCV (Fig. 17).

6. Discussion

The method implemented in AphidCV was efficient, especially when the total number of aphids (the sum of all stages and morphs) are compared. The sorting can be considered suitable for “nymph” and “wingless” classes. However, the “winged” class presented a weaker relationship in comparison to the others. The separation process for the class ”winged” needs more elements in the training database. Data from suction or tray traps could be a source for collecting more winged adults.

The best correlation was between CropAphid × AphidCV. CropAphid is a software that does not use machine learning (it depends on humans for the classification). Due to the almost perfect positive association, it is speculated a possible human error related to the total number of aphids.

A possible source of error could be the presence of tiny nymphs (less than 1 mm) that, in some cases, are not visualized in the standard method, but are detected by both software. For example, reviewing the

Fig. 9. Examples of “nymph” class images.

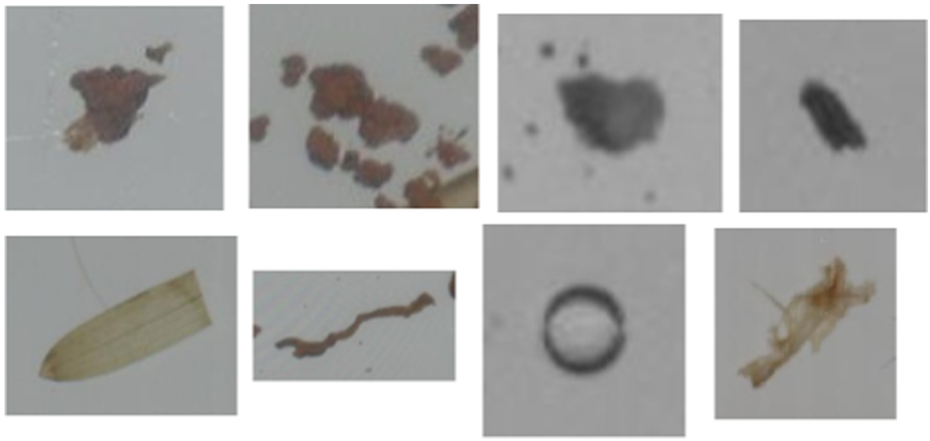

Fig. 10. Examples of “false” class images.

total number of nymphs, in all samples, the figures were: 12095 in the standard, 16329 in CropAphid, and 15784 in AphidCV. In the future, for a better assessment, comparisons should also include different experts.

The most prominent gain in the use of the software is undoubtedly the time need to evaluate a sample. Fig. 18 shows a comparison between the amount of time consumed by the three methods. AphidCV is four times faster than the standard method. The time presented for the AphidCV and CropAphid considers the sample preparation and scanning.

Besides, AphidCV no longer depends on a human after having the images scanned. The execution can be scheduled to start after working hours. Obviously, the software is not susceptible to tiredness or fatigue, allowing 24 × 7 evaluations with the same performance for all images.

Through this time saving, it will be possible to examine four times more samples than currently attained with the standard method. Hence, increasing data access and reliability.

There was a high correlation between the values obtained from a microscopic ruler and the AphidCV. Currently, aphids are not measured because of the amount of time required for such a task. Nevertheless, it is important data for many studies that can be obtained with the AphidCV. The acquisition of the added information is rapid, automated, and reliable.

An AphidCV fault is the discarding of objects of interest that have overlapping elements. Currently, our tool allows configuring size ranges for body length and area of aphids according to the characteristics of each species (Stroyan, 1984). AphidCV treats cases that extrapolate these measures as “Doubtful” for, human reassessment. In the future, modifications in the algorithm should include simple morphological details about antennas and legs present in doubtful objects of interest. In this case, we believe that it will be possible to use the same training images already available and adjust AphidCV to support this new feature. On the other hand, other morphological details, such as wing, should require extra care during capture to avoid damages in the wing structure. In this case, there will be a need for the acquisition of new images for training.

In terms of counting and detecting objects’ accuracy, AphidCV performed better than studies that realized pest detection in plants (Barbedo, 2014; Liu et al., 2016). Probably, this may have been due to the use of the ANN, which ignores false-positive counts, and the use of cleaner images, as it is in a controlled environment.

Moreover, in terms of ANN accuracy used for each stage, we obtained a lower correlation compared to Inception-v3 itself in the validation of ImageNET (Szegedy et al., 2016) or wild animals classifiers (Norouzzadeh et al., 2018). Almost certainly, this may have been due to the dataset difference used in model training, since these models were trained with datasets of 14 and 3 million images, respectively. In studies with smaller datasets, such as those that detect humans diseases, our accuracy correlation becomes similar (Shen et al., 2019; Couture et al., 2018; Coudray et al., 2018).

In spite of that, the results obtained in this first experiment were extremely encouraging. They demonstrated that the software presents a high degree of reliability when evaluating the total number of aphids. Concerning classification, one can also consider that the results are

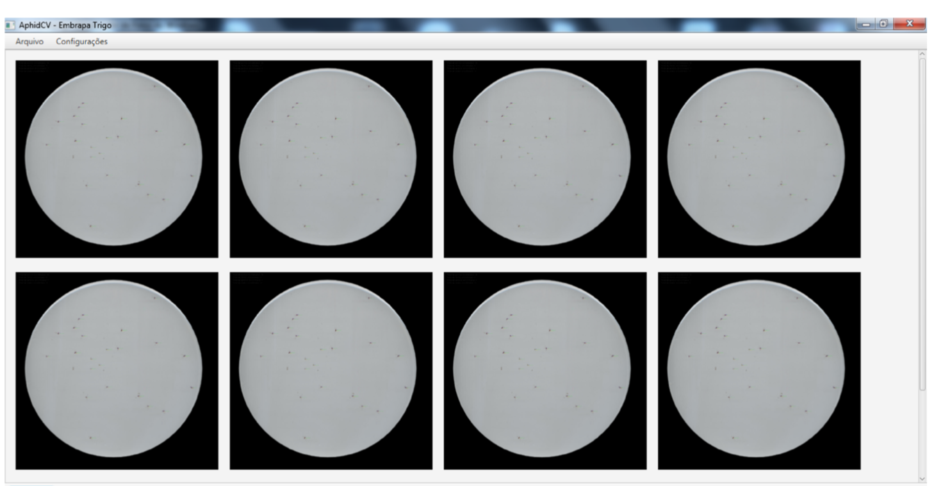

Fig. 11. Results presentation.

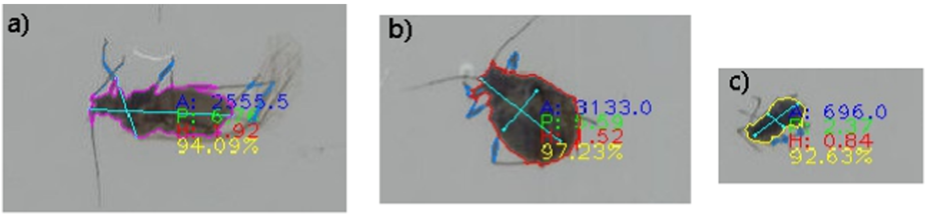

Fig. 12. Image details processed.

reasonably good, but with low precision with the "winged" class. We believe that with the improvement of the software in the classification of winged aphids, it could become extremely reliable. Hopefully, it will be conceivable to reach equivalent or higher levels of accuracy as compared to standard methods done by experts. Importantly, it always will be faster and will provide data regarding the size and the weight of each insect.

We propose two potential uses for this software. Firstly, to study the population dynamics of a given species under experimental conditions where it is necessary to evaluate progress and population structure according to factors under test. Secondly, to monitor field aphids for decision making in pesticide management, for example. The current version of Aphid CV is a first step that was born from the need of counting *Rhopalosiphum padi* populations. A similar approach is in development for other winter cereal species. One option for a wide variety of users working with other aphid species would be to provide this software with tools to facilitate machine learning, even for non-computer users such as biologists and agronomists. These experts could calibrate the program to their needs, creating a collaborative context to share new images and learning parameters for the public in general. For field monitoring, this solution can be useful as a background for counting trap samples containing winged aphids, a major agricultural challenge. Even for experts, identification still depends on accurate microscope analysis. Our approach is an initial effort to characterize the "body pattern" of aphids in order to develop computational tools helpful in monitoring their populations (regardless of species) under both laboratory and field conditions. These tools are intended to reduce dependence on specialized labor, reducing sample processing time and ensuring accurate identification and counting of aphids.

Table 2
Comparison of counting and classification of aphids between the Standard, CropAphid and AphidCV methods.

Class	Standard	CropAphid	AphidCV
Nymph	12095	16329	15784
Wingless	2080	2380	4296
Winged	149	170	206
Total	**14324**	**18879**	**20286**

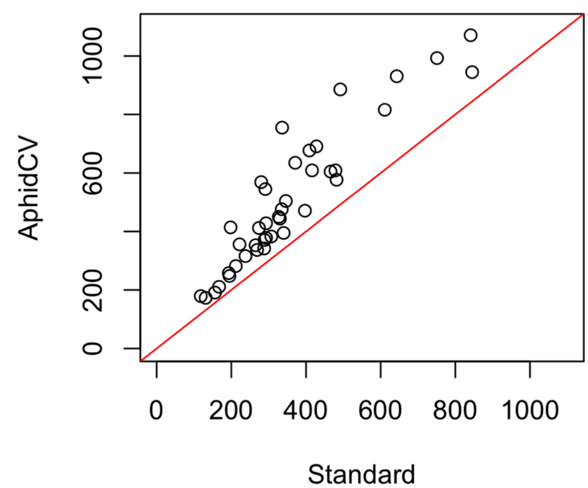

Fig. 14. Total number of aphids obtained by Standard and AphidCV methods.

7. Conclusion

This work presented the explanation of a method to automate the counting and classification of aphids of the *Rhopalosiphum padi* species using image processing, computer vision, and deep learning. A software named AphidCV was developed for implementing the proposed method.

Based on the validation tests with the software, we concluded that it is valuable for counting, sorting according to the development stage, and taking measures of the aphid species *Rhopalosiphum padi*.

If compared with the standard method used in the laboratory, we highlighted that the AphidCV saves time during the processes of counting and sorting. Besides, it added data that was previously not accounted for in samples, such as aphid morphometry. Also, AphidCV displays a complete report of the results in tabular format. The modularity feature enables easy integration with third-party systems.

	A	B	C	D	E	F	G
1	Species	Type	Probability(%)	Length(mm)	Weight(g)	Perimeter(mm)	Area(mm^2)
2	Rhopalosiphum padi	Nymph	64.31	1.19	0.00111989	3.69	0.66
3	Rhopalosiphum padi	Nymph	70.93	1.18	0.00110368	3.4	0.64
4	Rhopalosiphum padi	Nymph	56.61	1.17	0.00109906	3.48	0.67
5	Rhopalosiphum padi	Nymph	85.57	1.11	0.00103574	3.38	0.69
6	Rhopalosiphum padi	Wingless	82.83	1.43	0.00145886	3.79	0.83
7	Rhopalosiphum padi	Wingless	56.24	1.29	0.00125357	3.69	0.72
8	Rhopalosiphum padi	Wingless	81.53	1.26	0.00120517	3.41	0.67
9	Rhopalosiphum padi	Wingless	86.97	1.31	0.00127817	3.49	0.72
10	Rhopalosiphum padi	Wingless	81.26	1.26	0.00122204	3.56	0.74
11	Rhopalosiphum padi	Wingless	79.66	1.39	0.00140437	3.93	0.82
12	Rhopalosiphum padi	Wingless	35.02	1.42	0.00143792	3.95	0.8
13	Rhopalosiphum padi	Wingless	83.18	1.47	0.00150638	3.93	0.81
14	Rhopalosiphum padi	Wingless	87.19	1.42	0.00143792	3.95	0.8
15	Rhopalosiphum padi	Wingless	85.68	1.42	0.00143533	3.97	0.79
16	Rhopalosiphum padi	Wingless	86.1	1.27	0.00121961	3.49	0.68

Fig. 13. Example report per sample.

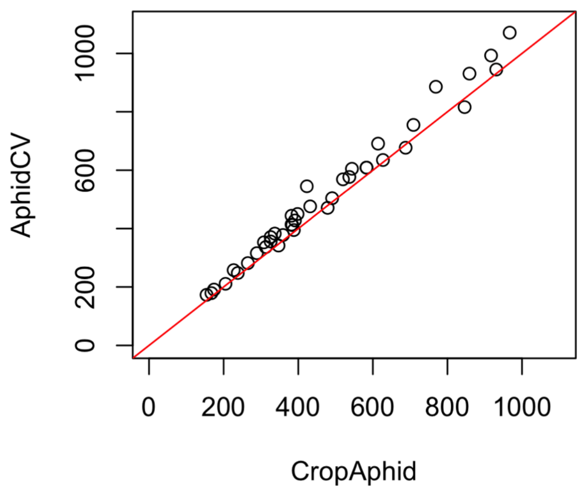

Fig. 15. Total number of aphids obtained by CropAphid and AphidCV methods.

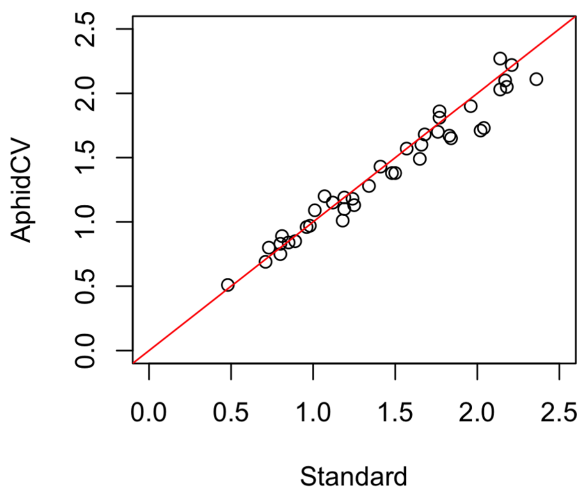

Fig. 16. Measurements' comparison considering insect body lengths (mm).

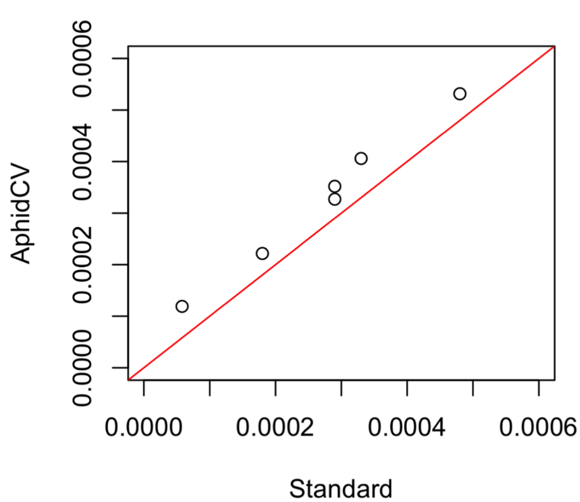

Fig. 17. Measurements' comparison considering insect weights (g).

The software still needs some fine-tuning to correct some limitations in sorting winged adult aphids, for example. To this end, it is recommended to balance and increase the training set, considering the application of techniques such as data augmentation for deep learning. Regular use of AphidCV may also feedback on the process.

Besides, we intend to test other ANN models to improve classification accuracy. We also plan to apply object detection techniques to classify more than one aphid in the same cropped image, since currently overlapping insects are classified as one, reducing the rate of success in the results.

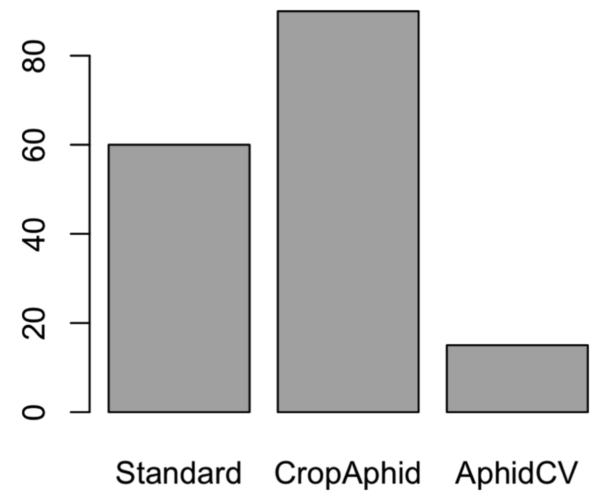

Fig. 18. Time comparison between methods (in minutes).

The proposed method and the software developed is an important move towards the application of computer vision in agriculture. AphidCV is a software tailored for use in the field of entomology where counting and sorting insects are common tasks. By using the software, it is possible to increase the number of samples in the same amount of time and with reproducible results.

Future work includes the use of a graphics processing unit (GPU) for fast calculations. We also have plans to train the software to recognize other cereal aphids species. Finally, a more challengeable goal is the use of software in samples collected directly from the field. An automated system for counting cereal aphids captured in traps installed in the field would be a valuable tool in decision-making on integrated pest management.

CRediT authorship contribution statement

Elison Alfeu Lins: Conceptualization, Methodology, Validation, Investigation, Software. **João Pedro Mazuco Rodriguez:** Validation, Writing - original draft, Writing - review & editing, Software. **Sandy Ismael Scoloski:** Validation, Software. **Juliana Pivato:** Investigation, Data Curation. **Marília Balotin Lima:** Investigation, Data Curation. **José Maurício Cunha Fernandes:** Investigation, Writing - review & editing. **Paulo Roberto Valle da Silva Pereira:** Investigation, Writing - review & editing, Data Curation. **Douglas Lau:** Methodology, Investigation, Writing - original draft, Writing - review & editing, Supervision. **Rafael Rieder:** Methodology, Writing - original draft, Writing - review & editing, Software, Supervision.

References

Dedryver, C.-A., Le Ralec, A., Fabre, F., 2010. The conflicting relationships between aphids and men: a review of aphid damage and control strategies. Comptes rendus Biol. 333 (6–7), 539–553. https://doi.org/10.1016/j.crvi.2010.03.009.

Shavit, R., Batyrshina, Z.S., Dotan, N., Tzin, V., 2018. Cereal aphids differently affect benzoxazinoid levels in durum wheat. PloS One 13 (12), e0208103. https://doi.org/10.1371/journal.pone.0208103.

Yahya, M., Saeed, N.A., Nadeem, S., Hamed, M., Shokat, S., 2017. Role of wheat varieties and insecticide applications against aphids for better wheat crop harvest. Pakistan J. Zool. 49 (6), 2217–2225. https://doi.org/10.17582/journal.pjz/2017.49.6.2217.2225.

Savaris, M., Lampert, S., Salvadori, J., Lau, D., Pereira, P.d.S., Smaniotto, M., 2013. Population growth and damage caused by *Rhopalosiphum padi* (l.) (Hemiptera, Aphididae) on different cultivars and phenological stages of wheat. Neotrop. Entomol. 42 (5), 539–543. https://doi.org/10.1007/s13744-013-0158-9.

Hansen, L.M., 2000. Establishing control threshold for bird cherry-oat aphid

(*Rhopalosiphum padi* L.) in spring barley (*Hordeum vulgare* L.) by aphid-days. Crop Prot. 19 (3), 191–194. https://doi.org/10.1016/S0261-2194(99)00094-0.

Shajahan, S., Sivarajan, S., Maharlooei, M., Bajwa, S.G., Harmon, J.P., Nowatzki, J.F., Cannayen, I., 2017. Identification and counting of soybean aphids from digital images using shape classification. Trans. ASABE 60 (5), 1467–1477. https://doi.org/10.13031/trans.12105.

Martin, A., Sathish, D., Balachander, C., Hariprasath, T., Krishnamoorthi, G., 2015. Identification and counting of pests using extended region grow algorithm. In: Electronics and Communication Systems (ICECS), 2015 2nd International Conference on. IEEE, pp. 1229–1234. https://doi.org/10.1109/ECS.2015.7124779.

Carter, N., Dixon, A.F.G., Rabbinge, R., 1982. Cereal Aphid Populations: Biology, Simulation and Prediction. Pudoc.

Barbedo, J.G.A., 2014. Using digital image processing for counting whiteflies on soybean leaves. J. Asia-Pacific Entomol. 17 (4), 685–694. https://doi.org/10.1016/j.aspen.2014.06.014.

Gonzalez, R.C., Woods, R.E., 2006. Digital Image Processing, third ed. Prentice-Hall Inc, Upper Saddle River, NJ, USA.

Nissimov, S., Goldberger, J., Alchanatis, V., 2015. Obstacle detection in a greenhouse environment using the kinect sensor. Comput. Electron. Agric. 113, 104–115. https://doi.org/10.1016/j.compag.2015.02.001.

Parer, J.T., Hamilton, E.F., 2010. Comparison of 5 experts and computer analysis in rule-based fetal heart rate interpretation. Am. J. Obstetr. Gynecol. 203 (5), 451–e1. https://doi.org/10.1016/j.ajog.2010.05.037.

Arambula Cosio, F., Marquez Flores, J.A., Padilla Castaneda, M.A., Solano, S., Tato, P., 2003. Automatic counting of immunocytochemically stained cells. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE Cat. No.03CH37439), vol. 1. pp. 790–793. https://doi.org/10.1109/IEMBS.2003.1279883.

Lu, C., Ferrier, N.J., 2004. Repetitive motion analysis: segmentation and event classification. IEEE Trans. Pattern Anal. Mach. Intell. 26 (2), 258–263. https://doi.org/10.1109/TPAMI.2004.1262196.

Mande, A., Gurav, G., Ajgaonkar, K., Ombase, P., Bagul, V., 2018. Detection of fruit ripeness using image processing. In: International Conference on Advances in Computing and Data Sciences. Springer, pp. 545–555. https://doi.org/10.1007/978-981-13-1813-9_54.

Liu, T., Chen, W., Wu, W., Sun, C., Guo, W., Zhu, X., 2016. Detection of aphids in wheat fields using a computer vision technique. Biosyst. Eng. 141, 82–93. https://doi.org/10.1016/j.biosystemseng.2015.11.005.

Xia, C., Chon, T.-S., Ren, Z., Lee, J.-M., 2015. Automatic identification and counting of small size pests in greenhouse conditions with low computational cost. Ecol. Informat. 29, 139–146. https://doi.org/10.1016/j.ecoinf.2014.09.006.

Wadhai, M., Gohokar, V., Khaparde, A., 2015. Agriculture pest control using computer vision technique. Int. J. Adv. Res. 3 (8), 309–314.

Abdullahi, H., Sheriff, R., Mahieddine, F., 2017. Convolution neural network in precision agriculture for plant image recognition and classification. In: 2017 Seventh International Conference on Innovative Computing Technology (INTECH). IEEE, pp. 1–7. https://doi.org/10.1109/INTECH.2017.8102436.

Embrapa Trigo. https://www.embrapa.br/trigo/apresentacao (accessed: 2019-04-01).

Schildt, H., 2017. Java: The Complete Reference, 10th ed. McGraw-Hill Education Group.

OpenCV Library. https://www.opencv.org (accessed: 2019-04-01).

Tensorflow - an open-source software library for machine intelligence. https://www.tensorflow.org (accessed: 2019-04-01).

Kaehler, A., Bradski, G., 2016. Learning OpenCV 3: Computer Vision in C++ with the OpenCV Library. O'Reilly Media, Inc.

Szeliski, R., 2010. Computer Vision: Algorithms and Applications. Springer Science & Business Media.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE, pp. 2818–2826. https://doi.org/10.1109/CVPR.2016.308.

Motic - Moticam Software Series, https://www.motic.com/As_Microscope_software/ (accessed: 2019-04-01).

Stroyan, H.L.G., 1984. Aphids-Pterocommatinae and Aphidinae (Aphidini), vol. 2 of 6. Royal Entomological Society of London, 1984, Handbooks for the Identification of British insects.

Norouzzadeh, M.S., Nguyen, A., Kosmala, M., Swanson, A., Palmer, M.S., Packer, C., Clune, J., 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc. Nat. Acad. Sci. 115 (25), E5716–E5725. https://doi.org/10.1073/pnas.1719367115.

Shen, L., Margolies, L.R., Rothstein, J.H., Fluder, E., McBride, R., Sieh, W., 2019. Deep learning to improve breast cancer detection on screening mammography. Sci. Rep. 9 (1), 1–12. https://doi.org/10.1038/s41598-019-48995-4.

Couture, H.D., Williams, L.A., Geradts, J., Nyante, S.J., Butler, E.N., Marron, J., Perou, C.M., Troester, M.A., Niethammer, M., 2018. Image analysis with deep learning to predict breast cancer grade, er status, histologic subtype, and intrinsic subtype. NPJ Breast Cancer 4 (1), 30. https://doi.org/10.1038/s41523-018-0079-1.

Coudray, N., Ocampo, P.S., Sakellaropoulos, T., Narula, N., Snuderl, M., Fenyö, D., Moreira, A.L., Razavian, N., Tsirigos, A., 2018. Classification and mutation prediction from non–small cell lung cancer histopathology images using deep learning. Nat. Med. 24 (10), 1559. https://doi.org/10.1038/s41591-018-0177-5.