Agricultura 12: No 1-2: 19-24 (2015)
DOI: 10.1515/agricultura-2016-0003

DE GRUYTER
OPEN

The accuracy of the germination rate of seeds based on image processing and artificial neural networks

Uroš ŠKRUBEJ[a], Črtomir ROZMAN[b] and Denis STAJNKO[b*]
[a]Lokovica 8a, SI-3325 Šoštanj, Slovenia
[b]University of Maribor, Faculty of Agriculture and Life Sciences, SI-2311 Hoče, Slovenia

ABSTRACT

This paper describes a computer vision system based on image processing and machine learning techniques which was implemented for automatic assessment of the tomato seed germination rate. The entire system was built using open source applications ImageJ, Weka and their public Java classes and linked by our specially developed code. After object detection, we applied artificial neural networks (ANN), which was able to correctly classify 95.44% of germinated seeds of tomato (*Solanum lycopersicum* L.).

Key words: image processing, artificial neural networks, seeds, tomato

INTRODUCTION

As one of the most important input in agriculture a quality seed is a basis for higher agricultural productivity and a key to economic growth.

A number of methods for seed quality evaluation and sorting have been developed so far, mainly based on the detection of various physical and chemical properties which correlate well with certain vigour and germination parameters (McDonald 1998).

Nowadays, seed testing is performed in accredited laboratories by trained human analysts. The tests are designed to evaluate the quality of the seed lot. Several tests are done. For instance, a germination test determines the maximum germination potential, or viability of the seed. The germination rate of a particular seed lot is a key indicator which shows the seed performances in the field and it is expressed as a percentage (for example a 90% germination rate means 90 out of 100 seeds are likely to germinate under proper growing conditions). This information is important for calculating optimal seeding rates and to determine whether a particular seed lot has the potential to produce a quality crop.

Since the manual counting is time-consuming and labour intensive process, we are looking at ways we can improve the process efficiency. We have been examining ways of automating a task by means of computer vision systems, based on image processing and machine learning. This can provide an alternative to manual counting and inspection of seed samples.

Image analysis was introduced in the field of seed technology already by Howarth and Stanwood (1994) who have developed a colour image database to characterize the phenotypic variation of genetic resources. Image processing also provided precise results in the field of seed identification or classification (Uchigasaki et al. 2000, Granitto et al. 2002) and germination assessment (McDonald et al. 1998). Dell'Aquila et al. (2000) used image analysis to characterize the imbibition of white cabbage seeds, while Geneve and Kester (2001) evaluated seeding size after germination by computer-aided analysis of digital images from a scanner (Ducournau et al. 2004).

Ureña et al. (2001) proposed a machine vision system which used automated data gathering process and a fuzzy logic-based system for automatic evaluation of germination quality.

Ducournau et al. (2004) presented a machine vision system

**Correspondence to:*
E-mail: denis.stajnko@um.si

designed to count the number of emergent radicle tips on seed lots under controlled lighting, temperature and hygrometric conditions. The automated acquisition system employed an algorithm that was able to count the germinated seeds and provided the mean germination time based on the difference between two consecutive pictures.

Modern computer vision mainly based on image processing procedures such as proprietary software MATLAB or other specialized expensive software. In our work a free image processing and analysis program named ImageJ was used, which is readily available, open source and public domain software developed at the National Institutes of Health (NIH), Bethesda, Maryland USA (Rasband 2012).

MATERIALS AND METHODS

Tomato seeds (*Solanum lycopersicum* L.) variety 'Marmande', were obtained from the seed company Semenarna Ljubljana d.d. Slovenia. Before the experiment, the uncovered seeds were stored for a month in an incubator at 4 C°, 50% relative humidity to equilibrate to an identical seed moisture condition. Then we randomly chose 700 seeds from 3 bags as the sample. Next, we placed a dark filter paper inside twenty-eight glass Petri dishes (90x98x18mm) and moistened each with 3ml distilled water. The dark filter paper was used to obtain optimal contrast between seed, radicle and filter paper. Twenty-five seeds were placed on top of the wet filter paper in each dish and spaced them evenly. We put covers on the dishes. The seeds were germinated under a controlled condition and maintained in the dark at 20 to 30 ºC (±1 ºC) and 75% relative humidity for seven days in a Jacobsen incubator. The seeds were illuminated for 8 hours in every 24 hour period. Light was provided by a cool white fluorescent source of 750 lux. Images were captured by a Nikon D80 digital SLR camera with Sigma 18 – 200mm zoom lens. The camera was mounted on a stand with an easy vertical movement, which provided rigid stable support. The camera was set at a distance 450 mm. The images were obtained by 3872x2592 pixels, horizontal resolution 300 dpi, vertical resolution 300 dpi and a bit depth 24. We placed a warm white, 22 W fluorescent tube with a 210 mm diameter circular lamp with a rated voltage of 220 V around the Petri dish with a seeds sample. A light diffuser, a semi-spherical steel bowl of 270 mm diameter, covered the light bulb, prevented external influences and provided diffused light (Figure 1). All images were transferred from the digital camera to a personal computer PC (dual-core microprocessor Intel Pentium B950 2.10 GHz, 4 GB installed memory RAM) via universal serial bus (USB) cable.

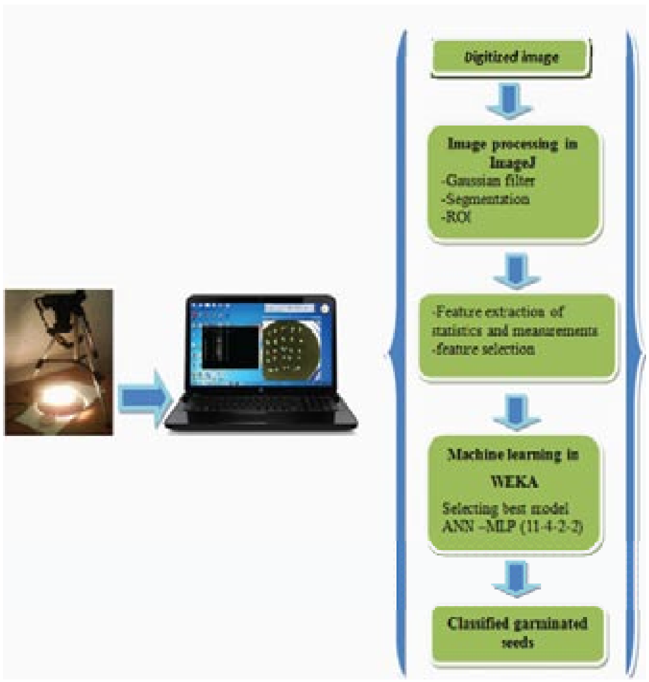

Fig. 1: Proposed computer vision system

Images processing

The ImageJ software was used for image processing and extracting features from original RGB images (Figure 2a). First, we cut off the frame from each RGB image to establish the correct region of interest (ROI) in the centre of each image by using the known radius, so the cropped image was received (Figure 2b).The cropping process reduced the size of the images so all the following manipulations were more efficient. In our study the original matrix of 3872x2592 pixels was reduced to 1854x1836 pixels. In the second step a Gaussian filter with sigma parameter σ set at 2 was used for smoothing the image. This filter used a convolution with a Gaussian function (Eq. 1) described by Rasband (2008):

$$\varphi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

where x is the intensity of pixel, μ is a mean,σ is the standard deviation, σ^2 is a variance, π=3.18 and e=2.718.

The image was then transformed from RGB color space into an 8-bit grayscale image (Figure 2c) which was finally converted by a threshold operation into a binary image (Figure 2d).For thresholding limits were defined by several pre-testing, so the lower limit was set to 85 and the upper limit to 255 grayscale intensity.

The minimum particle size was set to 1500 pixels and the maximum to 14.000 pixels. With this we additionally avoided the smaller and the bigger areas that could not be accounted as seeds. Finally, presented seeds were separated from the background and automatically labelled with the integer. The external perimeter of the seed was traced in yellow (Figure 2e).

The resulting set of extracted features together with statistics and measurements was saved in a separately table for each Petri dish, where each row represents a single seed and each column a single parameter of particular seed. Feature extraction was automated by using ImageJ's macro language facility, which enables any of ImageJ's GUI features to be invoked programmatically.

Simultaneously to image processing an expert technician

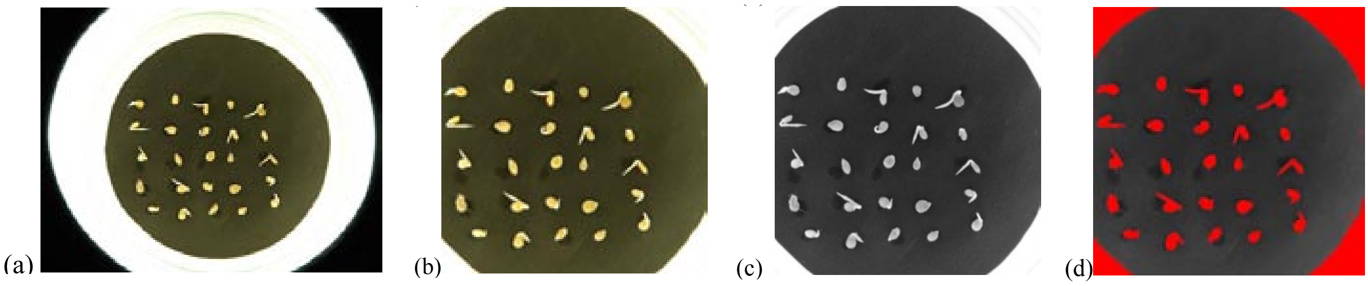

Fig. 2: Image processing. (a) Original image in RGB colour space. (b) Cropped image. (c) 8-bit grayscale image. (d) Binary image after thresholding. (e) Labelled seeds with numbers and yellow line. The next step in the processing was to extract the features from the image containing labelled seeds by Analyze Particles ImageJ command. The description of all measured and used image analysis parameters is reported in Table 1.

examined 28 images of each series of 25 seeds and provided a nominal class (germinated or not germinated) for each of the 700 seeds. Those results were also added to the previous mentioned tables.

Machine learning

Once the feature vectors had been generated and exported as *csv* (comma-separated values) formatted file, we used machine learning software WEKA (Waikato Environment for Knowledge Analysis) to perform further analysis. WEKA is a collection of machine learning algorithms for data mining tasks and contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. The algorithms can either be applied directly to a dataset or called from Java code. WEKA was developed at the University of Waikato, New Zealand for the purpose of identifying information from raw data obtained from agricultural domains and due to the usability and openness extended also to other fields. It is written in Java and well connected with an ImageJ. The combination of these two tools was first time described by Mayo et al. (2007) for classifying the moths by species after feature vectors were extracted from each of the moth images.

WEKA was first used for future selecting and ranking. For evaluator we used *InfoGAinAttributeEvaluator* which estimated the worth of an attribute by measuring the information gain with the respect to the class and *Ranker* which graded attributes by their individual evaluations. The biggest impact on classification had a parameter *Perimeter* which is the length of the outside boundary of the selection and lowest *Mean* which is the sum of the grey values of all the pixels in the selection divided by the number of pixels (Table 1).

For learning classifier models, the data were separated into 10 sets, each consisting of 70 instances. After that, the training was performed by using 9 of these sets, and testing was performed on the one remaining set (630 seeds as training and 70 seeds for testing for each run). This was repeated 10 times for each model.

Classification

For classification artificial neural networks (ANN) multilayer perceptron architecture (MLP) was implemented and compared with manual counting. For training we used a back propagation algorithm. The value of *learning rate* was set up

Table 1: Parameters measured by ImageJ and ranked by WEKA

Parameter	Description
Perim	Perimeter is the length of the outside boundary of the selection.
Kurtosis	Kurtosis is the degree of peakedness of a distribution.
Max	The maximum grey values within the selection.
Skewness	Skewness is a measure of the degree of asymmetry of a distribution.
StdDev	Standard deviation of the grey values used to generate the mean grey value.
Major	Major is the primary axis of the best fitting ellipse.
Area	Area of selection in square pixels
Mode	Modal gray value is a most frequently occurring gray value within the selection. Corresponds to the highest peak in the histogram.
Median	The median value of the pixels in the selection.
Minor	Minor is the secondary axis of the best fitting ellipse.
Mean	Average grey value within the selection. It is the sum of the grey values of all the pixels in the selection divided by the number of pixels.

at 0.3 and *momentum rate* at 0.2. The number of neurons in input and output layers was set to 11 and 2 respectively since the number of features was 11 and the number of possible classes was 2. In the next step several combinations of hidden layers and different number of hidden neurons were tested, so the training time varied from 100 to 2000 epochs.

Evaluation measures for classifier performance

To objectively evaluate the performance of ANN, we used a classification accuracy, precision, recall and the F-measure, which were derived from confusion matrix.

Classification accuracy refers to the percentage of correct predictions made by the ANN model when compared with the manual evaluation of the 25 test data. It is calculated as the number of correctly classified instances divided by the total number of instances:

$$accuracy=(TP+TN)/(TP+TN+FP+FN)*100\% \qquad (2)$$

where TP refers to true positive, TN refers to true negative, FP refers to false positive and FN refers to false negative. Then TP+TN+FP+FN is the total number of instances in the testing set and TP+TN is the number of correctly classified instances (Witten and Frank 2005).
In our case TP represented actually germinated seeds which were also predicted as germinated. TN represented actually not germinated seeds which were also predicted as not germinated. FP were actually not germinated seeds predicted as germinated and FN were germinated seeds predicted as not germinated.

Precision is the proportion of predicted positive instances which are actual positive among all those which were classified as positive. It is calculated as follows:

$$p=TP/(TP+FP) \qquad (3)$$

where TP refer to true positive and FP refer to false positive. A FP occurs when the class is incorrectly predicted as positive

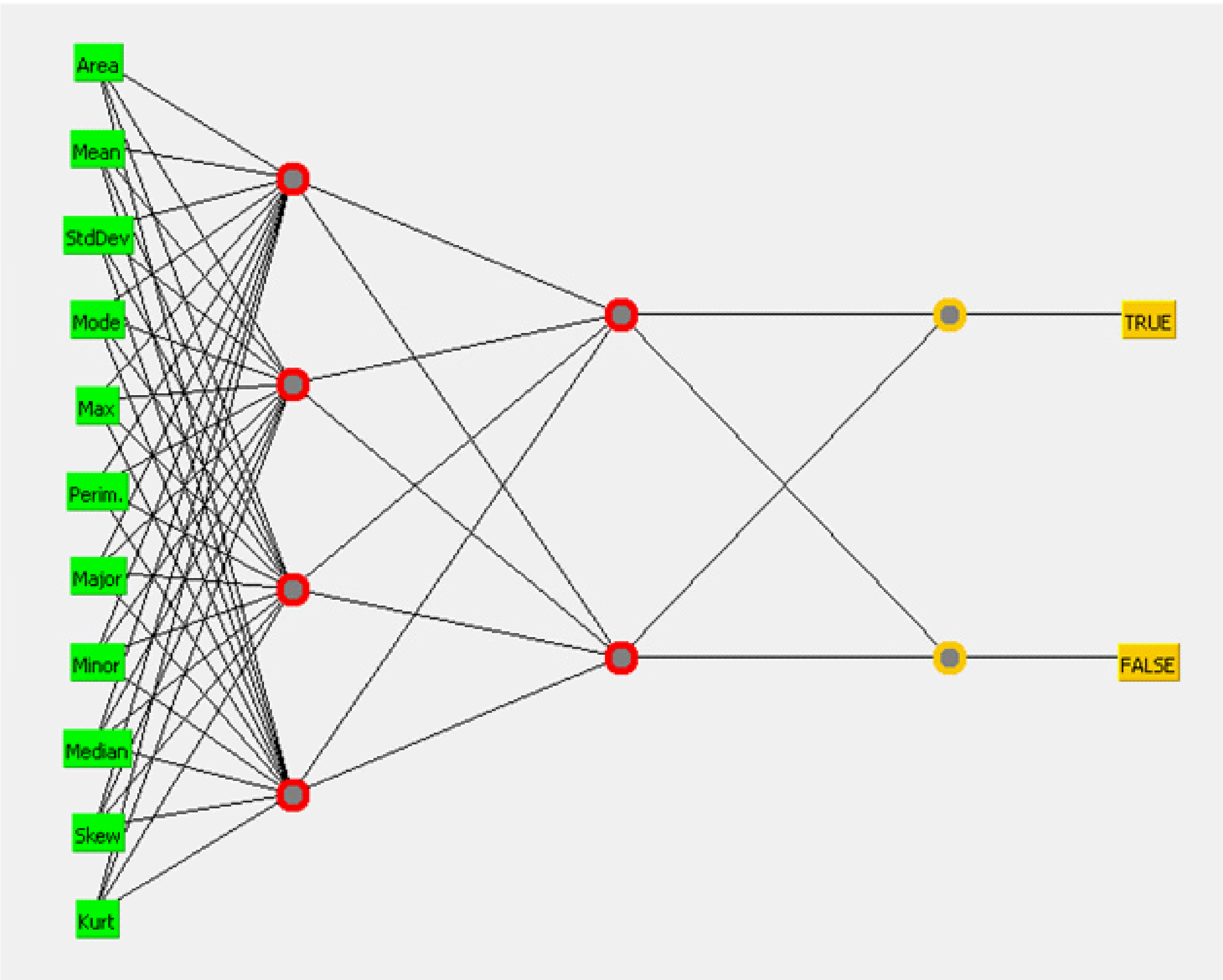

Fig. 3: ANN model with 11-4-2-2 topology

Table 2: Measures for classifier performance

model	Accuracy			Precision			Recall			F-measure		
	Mean	Std. dev	Sig.	Mean	Std. dev.	Sig.	Mean	Std. dev.	Sig.	Mean	Std. dev.	Sig.
ANN	95.44	2.14		0.9722	0.0233		0.9652	0.0218		0.9684	0.0148	

when it is actually negative (Witten and Frank 2005).

Recall is defined as the ratio of the true positive to the sum of the TP and FN. Recall is also known as sensitivity in some fields. It is calculated as follows:

$$recall=TP/(TP+FN) \tag{4}$$

FN occurs when the class is predicted as negative when is actually positive (Baeza and Riberio 1999, Witten and Frank 2005).

F-measure is defined as the harmonic mean of precision and recall. It is calculated as follows:

$$F\text{-}measure=(2\times Precision\times Recall)/(Precision+Recall) \tag{5}$$

It has a high value when both precision and recall have high values, and it is seen as way of finding the best compromise between these two measures (Baeza and Riberio 1999).

RESULTS AND DISCUSSION

The best accuracy of 95.44% (Table 2) was obtained with 2 hidden layers that contained 4 and 2 hidden neurons, respectively, and a training time set at 500. The algorithm realized a 97% TP rate, with a 3% FN rate and a 93% TN rate with a 7% FP rate. ANN achieved the precision (0.9722), which is also very high. The ANN reached a recall (0.9652) and F-measure (0.9684), respectively.

The main advantage of our procedure over the study of (Dell'Aquila et al. 2000), who detected the germanised seeds by comparing the difference of seed XY position or area size of seeds on two consecutive images (before and after germination), was that we used eleven features derived from a single picture instead, to assess the germination rate. The single acquisition is also favourable, because it decreased a possibility of fungal contamination caused by several opening of the Jacobsen incubator.

Additional significant difference between our study and the one of Jossen et al. (2010) is that in our case the image processing, feature extraction and classification were entirely automatic, without any manual interventions for transferring data between different software applications.

CONCLUSIONS

In the present study, a computer vision system based on image processing and machine learning techniques was developed which was implemented for automatic assessment of the seed's germination rate. The entire system was built by using open source applications ImageJ, Weka and their public Java classes and linked together by our specially developed code, which made it non-expensive and acceptable for many laboratories. The results show that ANN accuracy was very high (95.44%). The prototype system has classified one sample of germinated seeds in 4s and has out performed trained analysts, thus it shows a great opportunity for exchanging the time consuming manual counting in the laboratories for estimating the seed quality.

REFERENCES

1. Baeza R, Ribeiro B. Modern Information Retrieval. Addison-Wesley Longman Publishing Co., Boston, 1999: 544 p.
2. Dell'Aquila A, van Eck JW, van der Heidjen GWAM. The application of image analysis in monitoring the imbibition process of white cabbage (*Brassica oleracea* L.) seeds. Seed Sci. Res. 2000;10:163–169.
3. Ducournau S, Feutry A, Plainchault P, Revollon P, Vigouroux B, Wagner MH. An image acquisition system for automated monitoring of the germination rate of sunflower seeds. Comp. Electron. Agricult. 2004;44:189–202.
4. Geneve RL, Kester ST. Evaluation of seedling size following germination using computer-aided analysis of digital images from flat-bed scanner. Hortscience 2001;36(6):1117–1120.
5. Granitto PM, Navone HD, VerdesPF, Ceccato HA. Weed seeds identification by machine vision. Comp. Electron. Agricult. 2002;33:91–103.
6. Howarth MS, Stanwood PC. Extracting 3-D information using 2-D images of seeds. Comp. Electron. Agricult. 1994;10:175–188.
7. Jossen R, Kodde V, Willems L, Ligterink W, Van der Plas L, Hilhorst H. Germinator: a software package for high-throughput scoring and curve fitting of Arabidopsis seed germination. Plant J. 2010; 62(1):148–159.
8. Mayo M. Watson AT. Automatic species identification of live moths. Knowledge-Based Systems, 2007;20:195-202.
9. McDonald MB. Seed quality assessment. Seed Sci. Res. 1998;8:265–275.
10. Rasband WS. Image. US National Institutes of Health, Bethesda, Maryland, USA, 2008.
11. <http://rsb.info.nih.gov/ij/index.html> Accessed August, 2012.

12. Uchigasaki M, Serata K, Miyamoto S. An automated machine vision system for classification of seeds using color features. J. Agricult. Struct. 2000;30(4):325–332.
13. Ureña R, Rodriguez F in Berenguel M. A machine vision system for seeds germination quality evaluation using fuzzy logic. Comp. Electron. Agricult., 2001,32:1–20.
14. Witten, I., and Frank, E. (2005).Data Mining: Practical Machine Learning Toolsand Techniques (2nd Ed.). Morgan Kaufmann,

Natančnost določanja kalečih semen s pomočjo obdelave slik in nevronskih mrež

IZVLEČEK

Članek opisuje sistem računalniškega vida, ki temelji na tehnikah obdelave slik in strojnega učenja, ki je bil izdelan za avtomatsko oceno stopnje kaljenja semen paradižnika. Celoten sistem je bil zgrajen s pomočjo odprtokodnih aplikacij ImageJ, Weka in njihovih javno dostopnih javanskih kod, ki smo jih povezali v lastno originalno razvito kodo. Po odkrivanju predmetov na RGB slikah, smo uporabili umetne nevronske mreže (ANN), ki so bile sposobne pravilno razvrstiti 95,44% nakaljenih semen paradižnika (*Solanum lycopersicum* L.).