Computers and Electronics in Agriculture 69 (2009) 46–50

Contents lists available at ScienceDirect

Computers and Electronics in Agriculture

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

# Maize root complexity analysis using a Support Vector Machine method

D. Zhong[a], J. Novais[b], T.E. Grift[c,*], M. Bohn[b], J. Han[a]

[a] Institute of Control & Automation, Xi'an Jiaotong University, Xi'an, Shaanxi 710049, China
[b] Department of Crop Sciences, University of Illinois, Urbana, IL 61801, USA
[c] Department of Agricultural & Biological Engineering, University of Illinois, 1304 West Pennsylvania Avenue, Urbana, IL 61801, USA

ARTICLE INFO

*Article history:*
Received 12 January 2009
Received in revised form 13 May 2009
Accepted 18 June 2009

*Keywords:*
SVM
Feature extraction
Corn

ABSTRACT

Root complexity is an important factor in the growth and survivability of maize plants under biotic and abiotic stress conditions. To genetically improve root structure in the future, there is a need to identify the genes that govern root complexity. Root complexity itself is ill defined, but indicators derived from images of the root system such as Fractal Dimension can be used as proxies. A disadvantage of using Fractal Dimension as a complexity indicator is that the complexity of the root as seen in the images is captured into a single parameter.

This paper describes an alternative method, which translates a root image into a set of parameters. The method consists of computing the intercepts of circles drawn around the centre of the root image with the root branches. This led to characteristic curves from which parameters can be extracted using curve fitting. In addition to the parameters obtained by curve fitting, the density of the root images was included. All parameters were evaluated on their ability to classify the roots among their original genotypes using a method from the realm of Artificial Intelligence, the Support Vector Machine (SVM).

The results showed that whilst using merely three parameters originating from the characteristic curves, the SVM algorithm was capable of correctly classifying 99.95% of roots among 235 original genotypes.

Published by Elsevier B.V.

## 1. Introduction

The ability of plants to grow and produce seeds is directly related to a healthy, functional and efficient root system. Generally, root complexity and root development depend on genetic and environmental factors and their interactions (O'Toole and Bland, 1987). To assess the genetic basis of root complexity, earlier research determined the Fractal Dimension (FD) of thousands of maize roots recovered from specifically designed field trials using images of the roots (Bohn et al., 2006). A combined analysis of molecular linkage information and FD results led to the identification of Quantitative Trait Loci (QTL) for FD on most of the ten maize chromosomes. QTL are regions in the genome that carry genes involved in the inheritance of a quantitative trait, in this case root complexity. The FD has been shown suitable to describe the complexity of natural objects (Mandelbrot, 1983). A considerable amount of work has been done to capture biological complexity using FD, including studies on root systems (Tatsumi et al., 1989; Lynch et al., 1993; Shibusawa, 1994; Nielsen et al., 1997; Masi and Maranville, 1998; Oppelt et al., 2000; Eghball et al., 2003; Walk et al., 2004; Lontoc-Roy et al., 2006; Soethe et al., 2007), soil clod formation (Shibusawa, 1992), shoot systems and canopies of young trees (Morse et al., 1985; Foroutan-pour et al., 1999), seaweeds (Kubler and Dugeon, 1996), plant foliage (Da Silva et al., 2006), sponges (Abraham, 2001), neurons (Fernandez et al., 1994), and fungal mycelia (Mihail et al., 1995).

A disadvantage of the use of FD is that the complexity of the whole root as contained in gray scale images is captured in a single indicator. Therefore as an alternative, a method was devised which transforms the two-dimensional gray scale image into a set of parameters. This was accomplished by drawing circles around the known centre location of a root image, and to accumulate the intercepting pixels of these circles with the root branches. This method yielded a characteristic function where the accumulated number of intercepting pixels was plotted against the radius of the circles. This characteristic function was approximated by fit curves and the parameters of these curves were used to classify the roots among their original genotypes using the Support Vector Machine (SVM) algorithm (Vapnik, 1995). The SVM method is essentially a binary classifier based on finding the maximal margin hyperplane between two or more classes (Burges, 1998; Suykens and Vandewalle, 1999). The SVM method has been applied in a variety of applications such as in weed and nitrogen stress detection (Karimi et al., 2005), tissue classification (Furey et al., 2000; Pavlidis et al.,

* Corresponding author. Tel.: +1 217 333 2854; fax: +1 217 244 0323.
*E-mail address:* grift@uiuc.edu (T.E. Grift).

0168-1699/$ – see front matter. Published by Elsevier B.V.
doi:10.1016/j.compag.2009.06.013

2004), face detection (Osuna et al., 1997), gene selection for cancer (Guyon et al., 2002), as well as shape extraction and classification (Cai et al., 2001).

The objective of this study was to develop an alternative method of root image analysis, based on the Support Vector Machine method, enabling classification of maize roots among their original genotypes.

## 2. Materials and methods

Maize plants were grown in Urbana, IL, USA, using an incomplete block design with 235 entries (genotypes), 2 replications, and 47 incomplete blocks at 5 entries per block. Each plot was a single row measuring 4.6 m in length at a distance of 0.76 m separating the rows. Plots were composed of 25 plants/row or 71,525 plants/ha. The roots were harvested at the R1 (silking) stage. The first plant per row was discarded and the next five consecutive plants were trimmed at the third node and uprooted making sure that for each plant a 0.30 $m^3$ root core was recovered. Roots were individually labelled with a barcode tag. Root cores were soaked in water to remove a large part of the soil before they were thoroughly cleaned using high-pressure water hoses.

Images of the roots were obtained using a semi-automated image acquisition system containing two monochrome cameras (Unibrain, Fire-iTM 701b) with a resolution of $1280 \times 960$ pixels. The procedure was as follows: firstly, the operator scanned a barcode attached to the root to identify its genotype. Subsequently, the system acquired images of the background underneath and behind the root position. Secondly, the user placed a root on a spike in the centre of the imaging plane. The root then automatically rotated such that four lateral images and two (redundant) top view images resulted. All images of the maize root system were stored in a gray-scale TIFF format.

The research as described in this paper only uses a single top view image per root. The first processing step was to subtract the background image under and behind the root to obtain better contrast and detail. Sample images of maize roots from various genotypic origins are shown in Fig. 1. These images emerged after background subtraction and noise reduction.

The images contain noise which is caused by the shadow that the roots cast on the background (Fig. 2).

Since the spectrum of the noise was distinctly different from the root pixel spectra, the noise was removed using standard thresholding techniques. Subsequently the image was thresholded to obtain a binary image such as shown in Fig. 3.

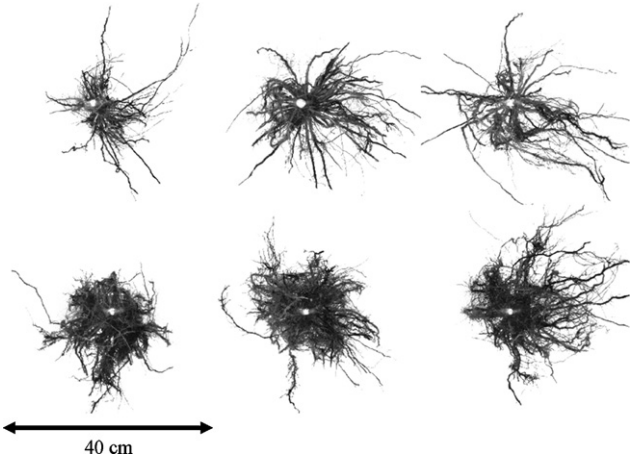


**Fig. 1.** Images of typical maize roots representing various genotypes after background subtraction and noise suppression.

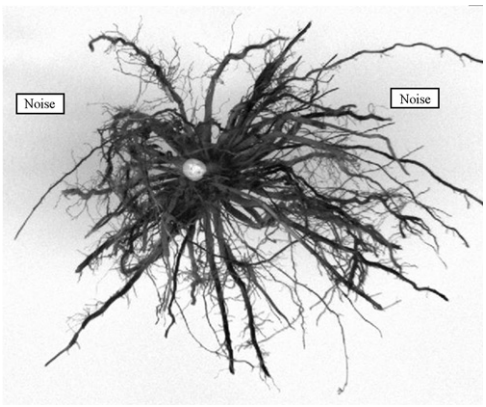


**Fig. 2.** Image of a maize root showing noise caused by the shadow it casts on the background.

Among the root images, there were 235 genotypes where each root was considered as one sample. A total of 2215 roots were analyzed where each genotype (class) had 10 roots (samples) or fewer.

### 2.1. Feature extraction

Every root image being analyzed has its own background image which contains a distinct spike. This spike location was used to obtain the centre of the root since the root was placed on the spike during imaging. Subsequently, circles were drawn around the centre of the root, with radii ranging from 0 to 500 pixels with an increment of 1 pixel. Fig. 4 shows this process where a limited number of circles is shown for clarity.

The pixels that intercepted the circle-related pixels as shown in Fig. 4 were accumulated for every circle yielding a characteristic function that relates the total number of interception pixels to the radius of the circle. Two distinct points characterize this function. At the centre of the root, the circle diameter is zero, and no circle intercepting pixels (CIPs) exist. Similarly, when the radius of the circle becomes large enough, there are no more pixels associated with the root, and again the number of CIPs is zero.

The hypothesis was that this process will produce characteristic functions that contain a monotonically increasing section and a monotonically decreasing section for increasing circle radii, with a distinct maximum number of CIPs separating the curves. Fig. 5 shows the characteristic function of the root shown in Fig. 4, where the number of CIPs is plotted as a function of the circle radius in pixels.

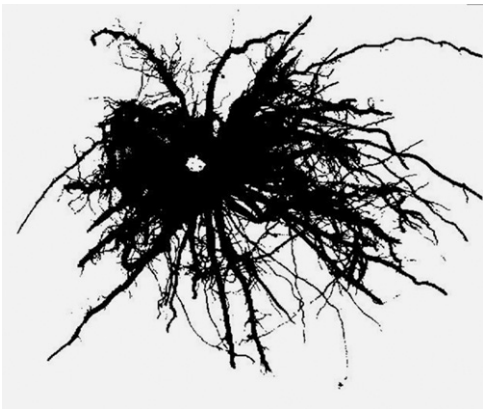

**Fig. 3.** Binary image of the maize root after thresholding. Although noise is no longer present, smaller root details have been lost in the process.

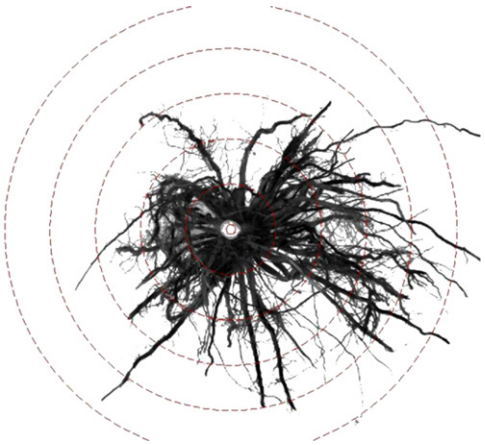

**Fig. 4.** Here the process of root characterization is shown. Circles were drawn around the known centre of the root, allowing the pixels representing intersections of the root and circle pixels (CIPs) to be accumulated as a function of the circle radius. The smallest circle represents the centre of the root; the radii of other five circles are 100, 200, 300, 400, and 500 pixels respectively.

Fig. 5 shows that the characteristic function can be approximated by two curves, intersecting at a maximum (*MaxPoint*) and tied to zero at the extreme radii. The two curves were fit in an Ordinary Least Squares sense by dividing the characteristic function into a left and right side using the maximum value *MaxPoint* as the separation point. As expected, the left side of the characteristic function is monotonically increasing and it was approximated by a polynomial curve, whereas the right side of the characteristic function was indeed monotonically decreasing and approximated by a negative valued exponential curve with an offset. Eqs. (1)–(3) show the functions $y_1, y_2$ that were fitted to the characteristic function.

$$y_1 = ax^3 + bx^2 + cx + d \tag{1}$$

$$y_2 = A + B\exp\left(\frac{-(x-C)}{D}\right) \tag{2}$$

$$f(x) = \begin{cases} y_1, x \le MaxPoint \\ y_2, x > MaxPoint \end{cases} \tag{3}$$

where $x$ is the radius of the circle in pixels.

In addition to the parameters derived from the CIP curves, the 'density' of the roots was evaluated. The reason for adding the density was that this value can be obtained quite easily as the ratio of the pixels belonging to the root divided by the total number of pixels in a circular plane encompassing the root as follows:

$$Density = \frac{\sum RootPixels}{\pi R_L^2} \tag{4}$$

where $R_L = 500$, the largest circle evaluated.

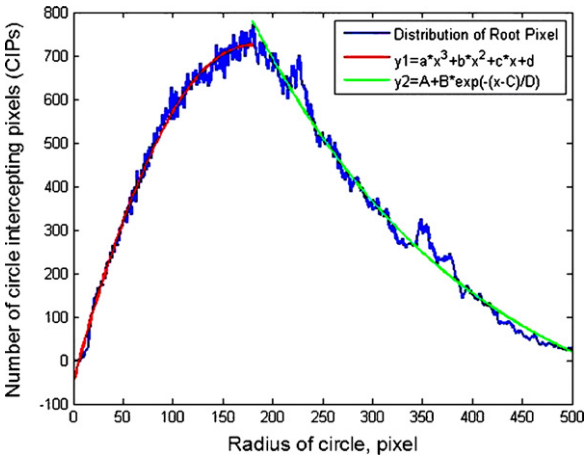


**Fig. 5.** Number of circle intercepting pixels (CIPs) as a function of the circle radius in pixels. Two curves were fitted: a third order parabola on the rising part, and a negative valued exponential function on the falling part. The coefficients of each function were considered root features and used for classification in the Support Vector Machine algorithm.

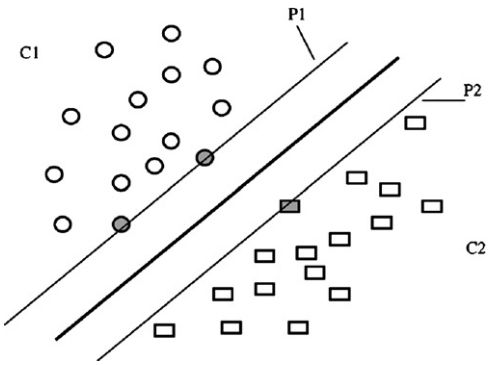


**Fig. 6.** Optimal separating plane (hyper-plane) between two hypothetical classes of data (Burges, 1998).

### 2.2. Support Vector Machine method

The Support Vector Machine is a "maximal margin classifier" (Karimi et al., 2005). The working of the SVM method is as follows: given two training sets containing a number of data points, the shortest distance between two points from each class is called the 'support vector'. Fig. 6 shows a simple example of SVM dealing with two classes of data. The solid circles and rectangle are support vectors; two thin lines P1 and P2 are hyper-planes according to the support vectors and the bold line is the optimal separating

**Table 1**
Classification accuracies using combinations of 2 features among 10 randomly selected classes (genotypes) containing 95 samples.

| | *b* | *c* | *d* |
|---|---|---|---|
| *A* | 86.32% (82/95) | 92.63% (88/95) | 100% (95/95) |
| *B* | 75.79% (72/95) | 82.11% (78/95) | 97.89% (93/95) |
| *C* | 77.89% (74/95) | 86.32% (82/95) | 100% (95/95) |
| *D* | 78.95% (75/95) | 83.16% (79/95) | 100% (95/95) |

**Table 2**
Classification accuracies using combinations of 2 features among 100 randomly selected classes (genotypes) containing 945 samples.

| | *b* | *c* | *d* |
|---|---|---|---|
| *A* | 45.50% (430/945) | 56.83% (537/945) | 97.35% (920/945) |
| *B* | 25.50% (241/945) | 36.30% (343/945) | 94.92% (897/945) |
| *C* | 27.94% (264/945) | 38.32% (362/945) | 95.34% (901/945) |
| *D* | 26.98% (255/945) | 36.19% (342/945) | 93.76% (886/945) |

**Table 3**
Classification accuracies using 2 minor features (*b* and *D*), 2 major features (*d* and *A*) and 3 features (*d*, *A* and *MaxPoint*) for various class sizes.

| Class size | 2 Minor features | 2 Major features | 3 Features |
|---|---|---|---|
| 2 | 95% (19/20) | 100% (20/20) | 100% (20/20) |
| 3 | 96.43% (27/28) | 100% (28/28) | 100% (28/28) |
| 10 | 78.95% (75/95) | 100% (95/95) | 100%(95/95) |
| 50 | 37.93% (176/464) | 99.14% (460/464) | 100% (464/464) |
| 100 | 26.91% (254/944) | 97.35% (920/945) | 100%(945/945) |
| 150 | 21.26% (301/1416) | 95.91% (1359/1417) | 99.93% (1416/1417) |
| 200 | 18.31% (344/1879) | 94.25% (1772/1880) | 99.95% (1879/1880) |
| 235 | 16.57% (367/2215) | 93.32% (2068/2216) | 99.95% (2215/2216) |

**Table 4**
Classification accuracies using a combination of "*Density*" and individual features derived from CIP curves (*b*, *c*, *d*, *A*, *B*, *C*, *D*, *MaxPoint*) among 10, 100 randomly selected classes (genotypes) and the complete dataset of 235 classes.

| | | *b* | *c* | *d* | *A* | *B* | *C* | *D* | *MaxPoint* |
|---|---|---|---|---|---|---|---|---|---|
| *Density* | 10 Classes | 15.79% (15/95) | 7.37% (7/95) | 41.05% (39/95) | 86.32% (82/95) | 75.79% (72/95) | 77.89% (74/95) | 78.95% (75/95) | 47.37% (45/95) |
| | 100 Classes | 3.39% (32/945) | 3.60% (34/945) | 11.96% (113/945) | 45.40% (429/945) | 25.50% (241/945) | 27.94% (264/945) | 27.20% (257/945) | 12.59% (119/945) |
| | 235 Classes | 1.63% (36/2214) | 1.72% (38/2214) | 6.68% (148/2214) | 27.24% (603/2214) | 14.86% (329/2214) | 16.21% (359/2214) | 16.62% (368/2214) | 6.79% (150/2214) |

line which can be calculated using the SVM algorithm. This idea can be extended to deal with multi-class data (Bredensteiner and Bennett, 1999).

Based on this support vector, a hyper-plane perpendicular to the support vector can be computed that separates the classes in an optimal way. The Support Vector Machine algorithm as used in this study is captured in a library called LIBSVM (Chang and Lin, 2008), which supports multi-class classification (Bredensteiner and Bennett, 1999).

## 3. Results and discussion

Although the order of the chosen polynomial shown in Eq. (1) was three, the curve fitting process showed that the value of parameter '*a*' was consistently close to zero, and therefore the value was ignored. This left nine potentially useful features from the curves, being "*b*, *c*, *d*", "*A*, *B*, *C*, *D*" and *MaxPoint* as well as *Density*.

To determine the most influential parameters, among the major features from parameter group 1 (*b*, *c*, *d*), group 2 (*A*, *B*, *C*, *D*) and *MaxPoint*, experiments were carried out for 10 and 100 classes by testing various combinations of the parameters. The classification results are presented in Tables 1 and 2. From the tables it is evident that the parameters "*d*", and "*A*" represent the major features among each group. *MaxPoint* was added as the third feature to increase the accuracy of the SVM.

To obtain classification results in a structured fashion, three groups of features were created being "2 Minor Features" (*b* and *D*), "2 Major Features" (*d* and *A*) and "3 Features" (*d*, *A* and *MaxPoint*). The number of classes to be classified varied from 2 to 235. The classification results are presented in Table 3.

When only 2 minor features were applied (*b* and *D*) the accuracy ranged from 95% for two classes to 16.57% for 235 classes. By applying the main features (*d* and *A*) the accuracy fell from 100% for 2 classes to 93.13% for 235 classes. Adding the third parameter *MaxPoint* caused the accuracy to start at 100% for 2 features and drop to 99.95% for 235 classes. It is clear that the combination of "*d*", "*A*" and *MaxPoint* represent the optimal combination in the classification process.

The classification accuracies based on a combination of *Density* and individual features derived from the CIP curves (*b*, *c*, *d*, *A*, *B*, *C*, *D*, *MaxPoint*) among 10, 100 randomly selected classes (genotypes) and the complete dataset are shown in Table 4.

From Table 4 it is clear that *Density* is not a major feature by comparing the highest scoring combination in Table 4 (*Density*, *A)* with the combination (*d*, *A*) in Table 3. For the class sizes 10, 100, and 235, the combination (*Density*, *A)* scored 83.32%, 45.4% and 27.2% respectively. In contrast, for the same class sizes the combination (*d*, *A*) scored 100%, 97.35%, and 93.32% respectively.

## 4. Conclusions

The complexity of maize roots was captured in images from which root characteristic functions were derived by drawing circles around the centre of the roots and accumulating the number of circle intercepting pixels as a function of the radius of the circle. The characteristic functions were approximated by fitting two curves and a characteristic *MaxPoint* where these two curves intersected. In addition to the fit curve related parameters, a *Density* parameter was evaluated. The curve related parameters as well as the density were regarded as features and used in a Support Vector Machine algorithm that classified the maize roots among their known original genotypes.

Combinations of features were tested to determine the major features in the curves. Out of a total of ten parameters, there were three that dominated the classification process. Using these three

parameters, the SVM algorithm was capable of correctly classifying 99.95% of the maize roots among 235 genotypes.

The most important finding of this research is not that the SVM algorithm works well; it is rather that the genotypical information is preserved over the growing season expressed in the phenotypical data such as root architecture and complexity.

## 5. Further research

The SVM algorithm was trained using 235 classes (genotypes). When a new unknown root is evaluated the network will classify this root in a class that closely resembles the phenotype of the root, represented in its architecture and complexity. An interesting extension of the method would be to evaluate crosses among the genotypes and to evaluate if the offspring classifies closely (and equally) to its parents.

The fact that the three main parameter were *d*, *A* and *MidPoint* indicates that the information from the CIP curves is mainly contained in the offsets, and the separation point between the left and right hand regression curves. This indicates that a simpler functional approximation may be feasible with the same classification accuracy.

## References


Abraham, E.R., 2001. The fractal branching of an arborescent sponge. Marine Biology 138, 503–510.

Bohn, M., Novais, J., Fonseca, R., Tuberosa, R., Grift, T.E., 2006. Genetic evaluation of root complexity in maize. Acta Agronomica Hungarica 54 (3), 291–303.

Bredensteiner, E.J., Bennett, K.P., 1999. Multicategory classification by support vector machines. Computational Optimization and Applications 12, 53–79.

Burges, C., 1998. A tutorial on support vector machines for pattern recognition. Data Mining Knowledge Discovery 2 (2), 121–167.

Cai, Y.D., Liu, X.J., Xu, X., Zhou, G.P., 2001. Support vector machines for predicting protein structural class. Bioinformatics 2 (3), doi:10.1186/1471-2105-2-3.

Chang, C., Lin, C., 2008. LIBSVM—A Library for Support Vector Machines. Available at http://www.csie.ntu.edu.tw/~cjlin/libsvm/. (visited December 22nd, 2008).

Da Silva, D., Boudon, F., Godin, C., Puech, O., Smith, C., Sinoquet, H., 2006. A critical appraisal of the box counting method to assess the fractal dimension of tree crowns. In: Lecture Notes in Computer Science, vol. 4291. Springer, Berlin/Heidelberg, doi:10.1007/11919476, pp. 751–760.

Eghball, B., Schepers, J.S., Negahban, M., Schlemmer, M.R., 2003. Spatial and temporal variability of soil nitrate and corn yield: multifractal analysis. Agronomy Journal 95, 339–346.

Fernandez, E., Eldred, W.D., Ammermuller, J., Block, A., Von Bloh, W., Kolb, H., 1994. Complexity and scaling properties of amacrine, ganglion, horizontal, and bipolar cells in the turtle retina. Journal of Computational Neurology 347, 397–408, doi:10.1002/cne.903470306.

Foroutan-pour, K., Dutilleul, P., Smith, D.L., 1999. Soybean canopy development as affected by population density and intercropping with corn: fractal analysis in comparison with other quantitative approaches. Crop Science 39, 1784–1791.

Furey, T.S., Cristianini, N., Duffy, N., Bednarski, D.W., Schummer, M., Haussler, D., 2000. Support vector machine classification and validation of cancer tissue samples using microarray expression data. Bioinformatics 16 (10), 906–914.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Machine Learning 46, 389–422.

Karimi, Y., Prasher, S.O., McNairn, H., Bonnell, R.B., Dutilleul, P., Goel, P.K., 2005. Classification accuracy of discriminant analysis, artificial neural networks and decision trees for weed and nitrogen stress detection in corn. Transactions of the ASAE 48 (3), 1261–1268.

Kubler, J.E., Dugeon, S.R., 1996. Temperature dependent changes in the complexity of form of *Chondrus crispus* fronds. Journal of Experimental Marine Biological Ecology 207, 15–24.

Lontoc-Roy, M., Dutilleul, P., Prasher, S.O., Han, L., Brouillet, T., Smith, D.L., 2006. Advances in the acquisition and analysis of CT scan data to isolate a crop root system from the soil medium and quantify root system complexity in 3-D space. Geoderma 137, 231–241.

Lynch, J., Johannes, J., Beem, V., 1993. Crop physiology and metabolism: growth and architecture of seedling roots of common bean genotypes. Crop Science 33, 1253–1257.

Mandelbrot, B.B., 1983. The Fractal Geometry of Nature. Freeman, New York.

Masi, C.E.A., Maranville, J.W., 1998. Evaluation of sorghum root branching using fractals. Journal of Agricultural Science 131, 259–265.

Mihail, J.D., Obert, M., Bruhn, J.N., Taylor, S.J., 1995. Fractal geometry of diffuse mycelia and rhizomorphs of Armillaria species. Mycological Research 99, 81–88.

Morse, D.R., Lawton, J.H., Dodson, M.M., 1985. Fractal dimension of vegetation and the distribution of arthropod body lengths. Nature 314, 731–733.

Nielsen, K.L., Lynch, J.P., Weiss, H.N., 1997. Fractal geometry of bean root systems: correlation between spatial and fractal dimension. American Journal of Botany 84, 26–33.

Oppelt, A., Kurth, W., Dzierzon, H., Jentschke, G., Godbold, D., 2000. Structure and fractal dimensions of root systems of four co-occurring fruit tree species from Botswana. Annals of Forest Science 57, 463–475.

Osuna, E., Freund, R., Girosi, F., 1997. Training support vector machines: an application to face detection. In: 1997 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 130–136.

O'Toole, J.C., Bland, W.L., 1987. Genotypic variation in crop plant root systems. Advances in Agronomy 41, 91–145.

Pavlidis, P., Wapinski, I., Noble, W.S., 2004. Support vector machine classification on the web. Bioinformatics 20 (4), 586–587.

Shibusawa, S., 1992. Fractals in clods formed with rotary tiller. Journal of Terramechanics 29 (1), 107–115.

Shibusawa, S., 1994. Modelling the branching growth fractal pattern of the maize root system. Plant and Soil 165, 339–347.

Soethe, N., Lehmann, J., Engels, C., 2007. Root tapering between branching points should be included in fractal root system analysis. Ecological Modelling 207, 363–366, doi:10.1016/j.ecolmodel.2007.05.007.

Suykens, J.A.K., Vandewalle, J., 1999. Least squares support vector machine classifiers. Neural Processing Letters 9, 293–300.

Tatsumi, J., Yamauchi, A., Kono, Y., 1989. Fractal analysis of plant root systems. Annals of Botany 64, 499–503.

Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer Verlag.

Walk, T.C., Van Erp, E., Lynch, J.P., 2004. Modelling applicability of fractal analysis to efficiency of soil exploration by roots. Annals of Botany 94, 119–128.