Research, Society and Development, v. 11, n. 14, e297111436211, 2022
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i14.36211

Wavelets in the analysis of seed image similarity: an approach using the Hurst directional exponent

Expoente direcional de Hurst na análise de similaridade de imagens de sementes

Exponente direccional de Hurst en el análisis de similitud de imágenes de semillas

Received: 10/06/2022 | Revised: 10/19/2022 | Accepted: 10/22/2022 | Published: 10/27/2022

Fernando Ribeiro Cassiano
ORCID: https://orcid.org/0000-0001-6147-5278
Universidade Estadual de Montes Claros, Brasil
E-mail: fernando.cassiano@unimotes.br
Thelma Sáfadi
ORCID: https://orcid.org/0000-0002-4918-300X
Universidade Federal de Lavras, Brasil
E-mail: safadi@ufla.brr
Paulo Henrique Sales Guimarães
ORCID: https://orcid.org/0000-0001-9158-1688
Universidade Federal de Lavras, Brasil
E-mail: paulo.guimaraes@ufla.br

Abstract
Modernization is present in all fields of knowledge. The wavelet transform and the Hurst exponent are tools that have fundamental importance in many of these advances. In the present study, the wavelet decomposition technique was combined with the Hurst exponent calculation to analyze X-ray images of seeds and thus classify them as full, slightly damaged or damaged. To calculate the Hurst exponent the mean and median were used as measurements of position. A support vector machine was used to validate the proposed method. For the full, damaged and slightly damaged seed groups, the average accuracy of the method, using the mean as measure position, was 74.5%, and using the median was 57.05%. For the full and damaged seed groups, the average accuracy using the mean was 99.76%, and using the median was 80.93%. For the slightly damaged and damaged seed groups, the average accuracy, using the mean as measure of position, was 99.26%, and the median was 76.22%. When analyzing seeds with slight damage, we observed a decrease in accuracy because the classification of the X-rays was subjective. Therefore, for the image database used in this study, the proposed methodology is efficient for automatic classification.
Keywords: Autossimilarity; Image classification; Wavelets.

Resumo
A modernização está presente em todos os campos do conhecimento. Técnicas mais sofisticadas e aparelhos mais modernos têm surgido com frequência. A decomposição em ondaletas é uma ferramenta que tem fundamental importância em muitos desses avanços. No que se refere à análise de imagens, essa ferramenta tem cooperado para criação de diversas novas técnicas, seja para reconstrução, compressão, eliminação de ruído, dentre outros. Outra ferramenta que auxilia na análise de imagens é o expoente de Hurst, que mede o quão autossimilar uma imagem é, de forma que se capte informação sobre características da imagem que a olho nu não seria possível. Com isso, o objetivo deste trabalho será combinar a técnica de decomposição em ondaletas com o cálculo do expoente de Hurst para analisar imagens de sementes e assim classificá-las em cheias, levemente danificadas ou danificadas. No cálculo do expoente de Hurst serão usadas como medidas de localização a média e a mediana. Um *suport vector machine* foi usado para validação do método proposto. Para o grupo de todas as sementes a acurácia média do método, utilizando a média, foi de 74,5% e, para a mediana foi de 57,05%. Utilizando o grupo de sementes cheias e danificadas a taxa média de acurácia, com a média como medida de posição, foi de 99,76% e, com a mediana foi de 80,93%. No grupo contendo sementes levemente danificadas e danificadas a taxa média de acurácia, usando a média como medida de posição, foi de 99,26% e, com a mediana foi de 76,22%.
Palavras-chave: Autossimilaridade; Classificação de imagens; Ondaletas.

Resumen
La modernización está presente en todos los campos del conocimiento. Técnicas más sofisticadas y aparatos más modernos están surgiendo constantemente. La descomposición en ondiculas es una herramienta que tiene una importancia fundamental en muchos de estos avances. En lo que se refiere al análisis de imágenes, esta herramienta ha contribuido para creación de nuevas técnicas diferentes, como para la reconstrucción, comprensión y eliminación de ruidos, entre otros. Otra herramienta que auxilia en el análisis de imágenes es el exponente de Hurst, que mide cuanto

tiene de auto-similitud una imagen, de forma que se capte información sobre las características de la imagen que a simple vista no seria posible. Con ello, el objetivo de este trabajo será combinar la técnica de descomposición en ondiculas con el cálculo del exponente de Hurst para analizar imágenes de semillas y así poder clasificarlas en llenas, levemente dañadas o dañadas. En el cálculo del exponente de Hurst serán usadas como medida de localización la media y la mediana. Un modelo de Máquinas de Vectores de Soporte será usado para la validación del método propuesto. Para el grupo de todas las semillas la precisión media del método, utilizando la media, fue de 74,5% y, con la mediana fue de 57,05%. Utilizando el grupo de semillas llenas y dañadas, la tasa media de precisión, con la media como medida de posición, fue de 99,76% y, con la mediana fue de 80,93%. En el grupo que contiene semillas levemente dañadas y dañadas la tasa media de precisión, usando la media como medida de posición, fue de 99,26% y, con la mediana fue de 76,22%.
Palabras clave: Auto-similitud; Clasificación de imágenes; Ondiculas.

1. Introduction

Technological advances are increasing worldwide in different areas of knowledge. In Brazil, agriculture is an area undergoing constant research and technological advancement. One important agricultural activity is the production of grains, especially sunflowers (*Helianthus annuus* L.). This crop has expanded and consolidates as an economically viable crop because of its resistance to drought, pests and diseases, in addition to other factors. Among annual crops, sunflowers are responsible for 16% of the world's edible oil production and are the third largest crop in the world production of edible oil (EMBRAPA, 2018). To improve the quality of final products that have sunflowers as a raw material, the best grains for planting are selected. Among the various tools used for this purpose, wavelets and machine learning algorithms have gained prominence. The use of these tools allows analyzing the images of the seeds, without destroy them in the process of evaluating the quality of the produced lots. This generates enormous speed in the production process of these grains. The objective of this study was to use the non-decimated discrete wavelet transform together with the Hurst directional exponent to extract characteristics of X-ray images of seeds and classify them (as full, slightly damaged or damaged) with the aid of a machine learning algorithm. Two statistics were used to calculate the Hurst exponent as location measures: the mean and median.

The accuracy of these two methods was compared. Because this method is noninvasive, it avoids the destruction of seed samples for the quantification of sowing quality and increases the range of analysis options. Thus, we expect to reduce the time necessary for seed classification and selection and improve the overall method.

2. Methodology

The analysis of wavelets has led to substantial increases in scientific knowledge in recent years. This technique allows the decomposition of a function, $f(t)$, in a linear combination of other functions, called wavelets and scale functions. Some references on the subject are Daubechies (1992); Mallat (1989); Burger and Burge (2016), Morettin (2014); Goswami (2011); Nason (2008), Souza, et al. (2022), among others.

The wavelet functions, ψ, are defined in the integrable square function space, i.e.,$\int_{-\infty}^{\infty}|\psi(t)|^2dt < \infty$. Another property that the wavelets share is $\int_{-\infty}^{\infty}\psi(t)\,dt = 0$. From the wavelet ψ, which is called the mother wavelet, other wavelets can be generated by expansion and translation. Then, for $a, b \in \mathbb{R}, a \neq 0$, there are

$$\psi_{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right).$$

For the discrete case, we have $j, k \in \mathbb{Z}$ and $\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k)$. Decomposition allows the analysis of the function $f(t)$ in various levels of detail, such that

$$f(t) = \sum_{k=-\infty}^{\infty} c_{j_0,k}\varphi_{j_0,k}(t) + \sum_{j=0}^{\infty}\sum_{k} d_{j,k}\psi_{j,k}(t). \quad (1)$$

In expression (1), the coefficients $c_{j_0,k}$ are called scaling coefficients, and the coefficients $d_{j,k}$ are called detail coefficients (Nason, 2008). This decomposition is the discrete wavelet transform of function $f(t)$. This technique has led to

major advances in the analysis of images and signal processing. Wavelet analysis has been an alternative to Fourier analysis because of its flexible nature, by preserving local, no periodic and multiscale phenomena (Rodrigues, 2015). A large number of studies have used discrete wavelet transforms to analyze images. Ramirez-Cobo and Vidakovic (2013) presented a wavelet transform for the two-dimensional case. According to the authors, the two-dimensional wavelet bases are constructed by the translation and dilation of products of univariate wavelets and scale functions. They are defined by

$$\varphi(t_1, t_2) = \varphi(t_1)\varphi(t_2),$$
$$\psi^h(t_1, t_2) = \varphi(t_1)\psi(t_2),$$
$$\psi^v(t_1, t_2) = \psi(t_1)\varphi(t_2),$$
$$\psi^d(t_1, t_2) = \psi(t_1)\psi(t_2),$$

in which h, v and d symbolize the horizontal, vertical and diagonal directions, respectively. Therefore, considering the elements of the wavelet bases, we have

$$\varphi_{j,\boldsymbol{k}}(\boldsymbol{t}) = 2^{2j}\varphi(2^j t_1 - k_1, 2^j t_2 - k_2),$$
$$\psi^i_{j,\boldsymbol{k}}(\boldsymbol{t}) = 2^{2j}\psi^i(2^j t_1 - k_1, 2^j t_2 - k_2),$$

for $i = h, v, d$, wherein $\boldsymbol{t} = (t_1, t_2) \in \mathbb{R}^2$, and $\boldsymbol{k} = (k_1, k_2) \in \mathbb{Z}^2$. Considering the image X as a function in $L^2(\mathbb{R}^2)$, X can be represented as

$$X(\boldsymbol{t}) = \sum_{\boldsymbol{k}} c_{j_{0,\boldsymbol{k}}}\varphi_{j_{0,\boldsymbol{k}}}(\boldsymbol{t}) + \sum_{j\geq j_0}\sum_{\boldsymbol{k}}\sum_i d_{j,\boldsymbol{k}}\psi^i_{j,\boldsymbol{k}}(\boldsymbol{t}) .$$

The nondecimated wavelet transform (NDWT) is a popular version of the wavelet transform because it has many advantages in its application, including that it maintains the same number of coefficients at all levels of detail. For the NDWT, there are

$$\varphi_{j,\boldsymbol{k}}(\boldsymbol{t}) = 2^{2j}\varphi(2^j(t_1 - k)_1, 2^j(t_2 - k_2)),$$
$$\psi^i_{j,\boldsymbol{k}}(\boldsymbol{t}) = 2^{2j}\psi^i(2^j(t_1 - k_1), 2^j(t_2 - k_2)).$$

In many of these studies, the Hurst exponent is used as a key measure for decision making. The Hurst exponent was created in 1951 by Harold Edwin Hurst in order to study temporal data series of rainfall, temperature, and river levels. For example, suppose that there is a river that drains into a reservoir of infinite capacity, and an annual amount must be released to meet the seasonal needs of the plantations. Next, consider a period of N years and that the amount released annually is the mean volume that the river flowed into the reservoir during that period. In this context, Hurst found a linear relationship between the $\log\left(R/\sigma\right)$ and $log(N)$, where R is the absolute sum between the maximum accumulated stock and the maximum deficit in the period N, and σ is the standard deviation of annual discharges. Hurst defined his coefficient as the slope of the linear regression line between these 2 variables (Hurst, 1956).

There are several ways to calculate the Hurst exponent (Feng & Vidakovic, 2017; Lyashenko, et al. 2016; Silva, et al., 2021), one of which is to use the energy spectrum of the wavelet transform. In this study, the spectrum was calculated in two ways. First, the logarithm of the mean of the squares of the detail coefficients was obtained by means of the discrete wavelet transform (Nicolis, et al., 2011). Considering a fractional Brownian field, the wavelet coefficients are given by

$$d^i_{j,\boldsymbol{k}} = 2^j \int B_H(\boldsymbol{u})\psi^i(2^j\boldsymbol{u} - \boldsymbol{k})d\boldsymbol{u},$$

in which the domain of the integral is $\mathbb{R}^2$. These coefficients are random variables with zero mean and variance given by

$$E\left[\left|d^i_{j,\boldsymbol{k}}\right|^2\right] = 2^{2j}\int\int\psi^i(2^j\boldsymbol{u} - \boldsymbol{k})\psi^i(2^j\boldsymbol{v} - \boldsymbol{k})E[B_H(\boldsymbol{u})B_H(\boldsymbol{v})]d\boldsymbol{u}d\boldsymbol{v}, \quad (2)$$

(Nicolis, et al., 2011). From (2), the following can be derived

$$E\left[\left|d^i_{j,\boldsymbol{k}}\right|^2\right] = \frac{\sigma_H^2}{2}V_{\psi^i}2^{-(2H+2)j}, \quad (3)$$

where $V_{\psi^i} = -\iint \psi^i(\boldsymbol{p}+\boldsymbol{q})\psi^i(\boldsymbol{q})|\boldsymbol{p}|^{2H} d\boldsymbol{p}d\boldsymbol{q}$. Applying the logarithm in base 2 on both sides of expression (3) results in the following equation:

$$log_2 E\left[\left|d^i_{j,\boldsymbol{k}}\right|^2\right] = -(2H+2)j + C_i, \quad (4)$$

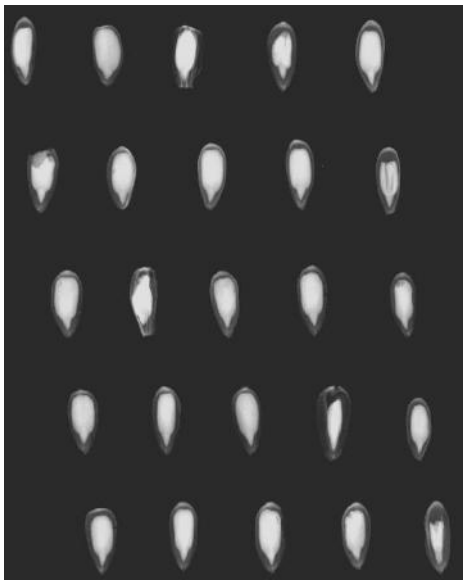
where $C_i = \frac{\sigma_H^2}{2} V_{\psi^i}(H)$(Nicolis, et al., 2011). Equation (4) was used to calculate the Hurst exponent (H) in each direction.

For the second method, the median logarithm of the square of the detail coefficients was used (Kang & Vidakovic, 2017). The median was a more robust alternative to possible outliers that could arise after the logarithmic transformation of the squares of the coefficients.

The Hurst exponent can be used as a measure of self-similarity along the scales of image decomposition. This decomposition was followed by a simulation of the data using the machine learning algorithm support vector machine to validate the proposed methods. Some references on machine learning techniques are Fernández-Delgado et al. (2014); Faceli, et al., (2017), Mello, et al., (2018), among others.

The data used in this study were X-ray images of sunflower seeds (*H.annuus* L.) captured at the Laboratory of Seed Analysis of the Department of Agriculture, Federal University of Lavras (Universidade Federal de Lavras). The samples of the sunflower seeds, cultivar Hélio-250, were produced in Uberlândia, state of Minas Gerais, from the harvest of 2010/2011. The images were analyzed by a seed specialist, and the seeds were separated into 3 categories: full, slightly damaged and damaged. A total of 175 images of full seeds, 130 images of slightly damaged seeds and 140 images of damaged seeds were considered. In Figure 1, an X-ray sheet is used for analysis and classification.

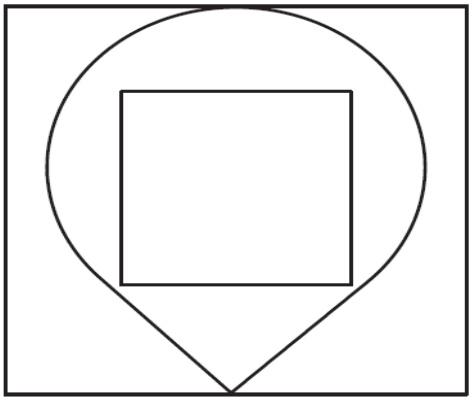
Figure 1 – Example of X-ray sheet used for classification of seeds as full, slightly damaged and damaged.



Source: Authors (2022).

Because the images had different sizes, we decided to work with subimages, which allows for the observation of the "same part" of all images, namely, a 64 x 64 pixel matrix. This cut was performed automatically using a function in the R software. Figure 2 shows the described procedure.

Figure 2 – Representation as the nucleus each image was extracted for decomposition.



Source: Authors (2022).

Wavelet decomposition was initially used in all images. The wavelet coefficients were obtained through the non-decimated discrete wavelet transform using the function proposed by Daubechies (1992). Nicolis, et al. (2011) studied several wavelet families, including Daubechies, and concluded that the differences between the parametric values were not significant, except for the Haar basis. We chose to work with Daubechies using 4 null moments and 4 levels of decomposition. Subsequently, the wavelet spectrum S(j), using the mean and the median as measures of location, which is defined as the ordinate presented, is

$$\left(j, log_2\overline{\left|d^i_{j,\boldsymbol{k}}\right|^2}\right),\ i = h, d, v, \qquad (2.1)$$

where $\overline{\left|d^i_{j,\boldsymbol{k}}\right|^2}$ is the mean of the squares of the detail coefficients in the scale j and

$$\left(j, median\left(log_2\left|d^i_{j,\boldsymbol{k}}\right|^2\right)\right),\ i = h, d, v, \qquad (2.2)$$

where h, v and d arethe horizontal, vertical and diagonal directions, respectively.

A linear regression was fitted to the ordered pairs using the ordinary least squares in (2.1) and (2.2). Considering α as the regression slope, the Hurst exponents are given by

$$H = -\frac{\alpha+2}{2} \text{ and } H = -\frac{\alpha}{2log2} - 1. \qquad (2.3)$$

The objective of the decomposition was to verify the Hurst directional exponent characteristics for the different types of seeds. These exponents then formed a set of data for training and set of data for validation with a machine learning algorithm. More specifically, we used support vector machine with the linear core function. Others core functions were tested, e.g., Gaussian, but the best result was found with the linear core. For both the training and validation sets, the decompositions were used in the horizontal, vertical and diagonal direction of each image. Decomposition was performed using the waveslim package (Whitcher, 2015) from the R software. Four groups were considered in this analysis: all seeds, full and slightly damaged seeds, full and damaged seeds, and slightly damaged and damaged seeds.

For each seed group, a random selection of 75% of the seeds was performed for the algorithm training. The remaining 25% were used to verify the learning, or accuracy, of the algorithm by constructing a confusion matrix. This procedure was repeated 1000 times in each seed group. Other ratios for training and validation were tested, but the ratio that provided the best

result was 75% and 25%, respectively. The R software packages that were used in this simulation were caTools (Tuszynski, 2014) e1071 (Meyer, et al., 2017) and the caret (Kuhn, 2017).

3. Results and Discussion

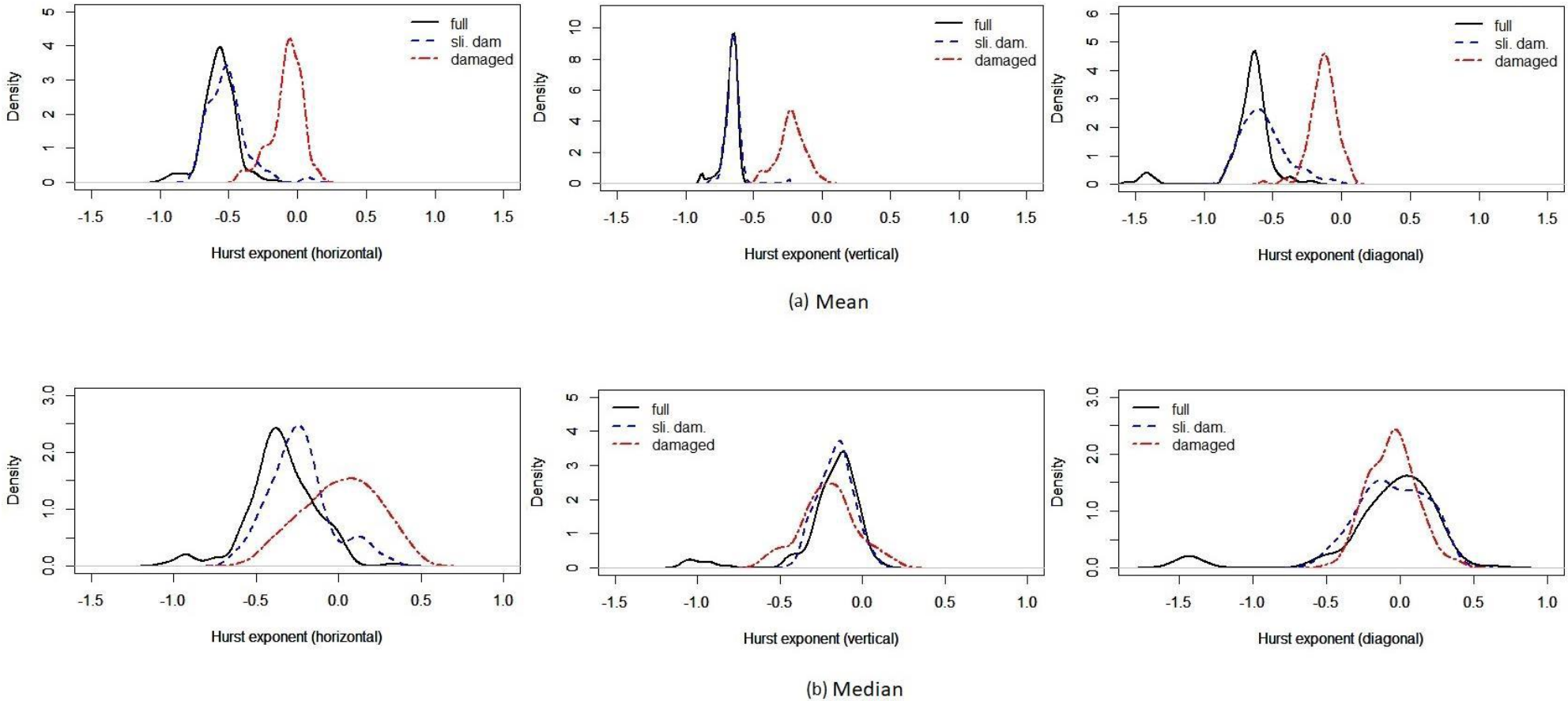
First, we present descriptive statistics of the Hurst exponents obtained with the wavelet decomposition for the mean and median as location measurements. Subsequently, the results from a thousand simulations are presented, with the objective of validating the proposed method. The results of the simulations are presented for each group, namely, all seeds, full and damaged seeds, slightly damaged and damaged seeds, and full and slightly damaged seeds.

Figure 3 shows the distribution of the Hurst coefficient, with the mean (a) and median (b) as measures of location in the horizontal, vertical and diagonal directions. Each subfigure contains the empirical distribution for the full, slightly damaged and damaged seeds. Regarding the mean value, the full and slightly damaged seed distributions overlapped in all cases. However, there was an increase in the variability of the exponent values in the diagonal decomposition of the slightly damaged seeds, resulting in a flattening of the distribution curve. We also observed a slight shift to the right in the mean value of the Hurst coefficients for the slightly damaged seeds group. The distribution of the damaged seed exponents was quite different from the distributions of the other seeds in the 3 directions. We found a considerable shift to the right in the mean of the Hurst coefficients of the damaged seeds. This technique is effective in extracting characteristics from each type of image. We also observed that most Hurst exponent values for all seeds were negative. This result may be related to the data from Jeon, et al. (2014), which indicated a correlation in the residuals.

The replacement of the mean with the median changed the results, albeit less than expected. Figure 3 (b) shows plots of the sample distribution of the Hurst coefficients in the 3 directions. The mean to median substitution caused an "overlap" in the empirical directional distribution of the coefficients for all seeds. In the horizontal direction, however, this overlap was smaller due to a gradual displacement to the right of the mean of the slightly damaged and damaged seed distributions, respectively. In general, there was a shift to the right in the distribution of the coefficients for all types of seeds, indicating a decrease in the correlation of residues, as shown by Jeon, et al. (2014).

The simulation showed that the mean rate of correct classification in the full and damaged seed groups was 99.76%, using the mean as a measure of location. This rate was 80.93% when the median was used. For this group of seeds, we found a clear increase in the classification power when the mean was used rather than the median. In addition, the correct classification rate when using the mean was much higher than the rates found in the methodology proposed by Sáfadi, et al. (2016). In the present study, the authors observed 64% accuracy using slopes s1 and s2 and 82% accuracy using slopes s1, s2, s3 and s4. We found that the rate was slightly lower than the results from Sáfadi, et al. (2016) for the 4 inclinations when the median was used as a measure of location. It is notable that for this group of seeds, 813 simulations of 1000 were made correctly.

Figure 3 - Empirical distributions of Hurst's exponent in horizontal, vertical and diagonal directions using the mean and the median as measures of localization.



Source: Authors (2022).

Table 1 presents the confusion averaging matrices for the 1000 simulations.

Table 1 – Mean confounding in 1000 simulations for full seeds (f) and damaged (d), using the mean and the median as measures of localization.

		Classified as			Classified as		
		f	d	Total	c	d	Total
Real classification	f	43.79	0.00	43.79	37.05	6.81	43.86
	d	0.19	35.02	35.21	8.26	26.89	35.15
	Total	43.98	35.02	79	45.31	33.70	79
		(a) Mean			(b) Median		

Source: Authors (2022).

In Table 1, the elements of the main diagonal represent the mean number of correct classifications. The other elements represent the mean number of classification errors (confusion). We observed almost no confusion in the classification of images using the mean as a measure of location. A mean of 0.19 damaged seeds were classified as full seeds by the proposed method. Using the median as a localization measure, there was confusion both for full seeds that were classified as damaged and for damaged seeds that were classified as full. Averages of 8.26 damaged seeds were classified as full and 6.81 full seeds were classified as damaged.

The results were similar for the slightly damaged and damaged seed groups. Considering the mean as a measure of location, the mean accuracy was 99.26%, and with the median the accuracy was 76.22%. Using the 4 slopes for this seed group, Sáfadi, et al. (2016) found an accuracy rate of 63%. In this study, both the mean and median provided better results, although using the mean resulted in a considerably better outcome than using the median.

Table 2 presents the confusion averaging matrices for the 1000 simulations.

Table 2 – Mean confounding in 1000 simulations for damaged(d) and slightly damaged seeds (sli. dam.), using the mean and the median as measures of localization.

		Classified as			Classified as		
		d	sli. dam.	Total	d	sli. dam.	Total
Real classification	d	35.10	0.03	35.13	26.38	8.90	35.28
	sli. dam.	0.47	32.40	32.87	7.27	25.45	32,72
	Total	35.57	32.43	68	33,65	34.35	68
		(a) Mean			(b) Median		

Source: Authors (2022).

The behavior of the group composed of the full and damaged seeds was maintained. There was little confusion in classification when the mean was used. An average of, 0.47 slightly damaged seeds were classified as full, and 0.03 damaged seeds were classified as slightly damaged. Using the median as a measure of location, these values were 7.27 and 8.90, respectively.

When the group of full and slightly damaged seeds was considered, there was a considerable decrease in the accuracy when either the mean or the median was used. When the mean was used as a measure of location, the mean rate of accuracy was 63.09%, and with the median the accuracy was 58.28%. This rate is much lower than that observed by Sáfadi, et al. (2016), who obtained a rate of 89%. It is important to note the subjectivity and difficulty in characterizing seeds as full and slightly damaged, as they are very similar groups (Leite, et al., 2013). Interestingly, the method used by Sáfadi, et al. (2016) was superior in detecting such differences.

Table 3 presents the matrices of confounding means for the 1000 simulations for the full and slightly damaged seed group.

Table 3 – Mean confounding in 1000 simulations for full (f) and slightly damaged (sli. dam.) seeds, using the mean and the median as measures of localization.

		Classified as			Classified as		
		f	sli. dam.	Total	f	sli. dam.	Total
Real classification	f	41.39	2.99	44.38	38.13	6.01	44.14
	sli. dam.	25.43	7.19	32.62	26.11	6.75	32.86
	Total	66.82	10.18	77	64.24	12.76	77
		(a) Mean			(b) Median		

Source: Authors (2022).

We observed (Table 3) considerable confusion in the classification of slightly damaged seeds that were considered full seeds. Using the mean as a measure of localization, we found a mean of 25.43 slightly damaged seeds that were classified incorrectly. When the median was used as a localization measure, the mean was 26.11.

For the group of all seeds, the mean rate of correct classification, using the mean as a measure of location, was 74.15%. When the median was used, this rate was 57.05%. We found increased accuracy of the classification algorithm when the mean was used instead of the median. The rate when using the median was similar to that observed by Sáfadi, et al. (2016), which was 57%. For the group of all seeds, we can compare our results with the work of Leite, et al. (2013). These authors, using independent components, achieved a rate of approximately 80% accuracy in seed classification. According to the variation in the number of independent components, this rate varied slightly. However, the authors did not use a simulation process to estimate the mean accuracy rate, which complicates the comparison between the methodology proposed in this study and that used in Leite, et al. (2013).

In Table 4, the mean confounding matrices are presented for the 1000 simulations for the group of all seeds. Again, we observed confusion in the classification of slightly damaged seeds. We found a mean of 25.58 slightly damaged seeds that were classified as full when the mean was used as the localization measurement. Using the median as a localization measure, this number was 24.36. In the same seed group, 80.1% were classified incorrectly using the mean. Using the median, the classification error percentage was 95%. This high confounding rate is a result of the overlapping distributions of the

Hurst coefficients for the group of slightly damaged and full seeds in the 3 directions of decomposition (Figure 3).

Table 4 – Mean confounding for all the seeds in 1000 simulations, by considering the mean and the median as measures of localization.

		Classified as			
		f	d	sli. dam.	Total
Real classification	f	41.56	0.00	2.68	44.24
	d	0.17	35.01	0.03	35.21
	sli. dam.	25.58	0.49	6.48	32.55
	Total	67.31	35.5	9.19	112

(a) Mean

		Classified as			
		f	d	sli. dam.	Total
Real classification	f	36.53	5.66	1.71	43.9
	d	7.29	25.82	2,07	35.18
	sli. dam.	24.36	7.01	1.55	32.92
	Total	68.18	38.49	5.33	112

(b) Median

f: full; d: damaged and sli. dam.: slightly damaged. Source: Authors (2022).

4. Conclusion

The use of the Hurst exponent in combination with the non-decimated wavelet transform was efficient for seed classification. The use of the mean as a localization measurement in the calculation of the Hurst exponent produced a superior result compared to the median for the considered set of images. The technique proposed in this study presented good results in most scenarios. The subjective difficulty in distinguishing between full and slightly damaged seeds was reflected in the proposed method, which exhibited reduced accuracy when this group of seeds was considered.

References

Burger, W., & Burge, M. J. (2016) *Digital image processing: An algorithmic introduction using java*. Springer.

Burrus, C. S., Gopinath, R. A., & Guo, H. (1998*) Introduction to wavelets and wavelets transforms: A primer*. Prentice Hall.

Daubechies, I. (1992) *Ten Lectures on Wavelets*. SIAM.

EMBRAPA. (2018) *Girassol*. https://www.embrapa.br/girassol

Faceli, K., Lorena, A. C., Gama, J., Almeida, T. A., & Carvalho, A, C. P. L. F. de (2017) *Inteligência Artificial: Uma abordagem de aprendizado de máquina*. LTC.

Feng, C., & Vidakovic, B. (2017) Estimation of the hurst exponent using trimean estimators on nondecimated wavelet coefficients. *Journal of latexclass files*, 14(8), 1-11. https://doi.org/10.48550/arXiv.1709.08775

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014) Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, 15(90), 3133–3181.

Goswami, J. C., & Chan, A. K. (2011) *Fundamental of wavelets: theory, algorithms and applications*. Wiley.

Hurst, H. E. (1956) The problem of long-term storage in reservoirs. *Hydrological Sciences Journal*, 1(3), 13–27. https://doi.org/10.1080/02626665609493644

Jeon, S.; Nicolis, O., & Vidakovic, B. (2014) Mammogram diagnostics via 2-d complex wavelet-based self-similarity measures. *Journal of Mathematical Sciences*, 8(2), 265–284. https://doi.org/10.11606/issn.2316-9028.v8i2p265-284

Kang, M., & Vidakovic, B. (2017) *MEDL and MEDLA: Methods for assessment of scaling by medians of log-squared nondecimated wavelet coefficients*. https://arxiv.org/pdf/1703.04180.pdf. https://doi.org/10.48550/arXiv.1703.04180

Kuhn, M. (2017) *caret: Classification and regression training*. **R**, Package Version 6.0-78, 2017. https://cran.r-project.org/web/packages/caret/caret.pdf.

Leite, I. C. C., Sáfadi, T., & Carvalho, M. L. M. de. (2013) Evaluation of seed radiographic images by independent component analysis and discriminant analysis. *Seed Science and Technology*, 41(2), 235–244. https://doi.org/10.15258/sst.2013.41.2.06

Lyashenko, V. V., Matarneh, R., Baranova, V., & Deineko, Z. V. (2016) Hurst expoent as a part of wavelet decomposition coefficients to measure long-term memory time series based on multiresolution analysis. *American jornal of systems and software*, 4(2), 51–56. 10.12691/ajss-4-2-4

Mallat, S. G. (1989) Multiresolution aproximations and wavelet orthonormal bases of L2(R). *Transactions of the American Mathematical Society*, 315(1), 69–87.

Mello, R. F., & Ponti, M. A. (2018) *Machine Learning: A practical approach on the statistical learning theory*. Springer.

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C-C., & Lin, C-C. (2017) e1071: *Misc functions of the department of statistics, probability theory group (formerly: E1071), TU Wien*. **R**, Package Version 1.6-8. https://cran.r-project.org/web/packages/e1071/e1071.pdf.

Morettin, P. A. (2014) *Ondas e ondaletas: da análise de fourier à análise de ondaletas de series temporais*. EdUSP.

Nason, G. P. (2008) *Wavelet methods in Statistics with R*. Springer.

Nicolis, O., Ramirez-Cobo, P., & Vidakovic, B. (2011) 2d wavelet-based spectra with applications. *Computational statistics and data analysis*, 55(1), 738–751. https://doi.org/10.1016/j.csda.2010.06.020

Ramirez-Cobo, P., & Vidakovic, B. A. (2013) 2d wavelet-based multiscale approach with applications to the analisys of digital mamograms. *Computational statistics and data analysis*, 58(2), 71-81. https://doi.org/10.1016/j.csda.2011.09.009

Rodrigues, L. L. M. (2015) *Análise de variância em séries temporais: uma abordagem usando ondaletas*. Tese (Doutorado) — Universidade Federal de Lavras, UFLA.

Sáfadi, T., Kang, M., Leite, I. C. C., & Vidakovic, B. (2016) Wavelet-based spectral for descriptors of detection of damage in sunflower seeds. *International Journal of Wavelets, Multiresolution and Information Processing*, 14(4), 235–244. https://doi.org/10.1142/S0219691316500272

Silva, A. S. A., Menezes, R. S. C., & Stosic, T. (2021) Análise multifractal do índice de precipitação padronizado. *Research, Society and Development*, 10(7), 1-10. https://doi.org/10.33448/rsd-v10i7.16535

Souza, P. M., Carneiro, P. C., Pereira, G. M., Oliveira, M. M., Costa Junior, C. A. da, Moura, L. V. de, Mattjie, C., Silva, A. M. M. da, Macedo, T. A. A., & Patrocínio, A. C. (2022) Towards a Classification Model using CNN and Wavelets applied to COVID-19 CT images. *Research, Society and Development*, 11(5), 1-17. http://dx.doi.org/10.33448/rsd-v11i5.27919

Tuszynski, J. (2014) *caTools: Tools: moving window statistics, gif, base64, ROC AUC, etc*. **R**, Package Version 1.17.1. https://cran.r-project.org/web/packages/caTools/caTools.pdf

Whitcher, B. (2015) *Waveslim: Basic wavelet routines for one-, two- and three-dimensional signal processing*. **R**, Package Version 1.7.5, https://cran.r-project.org/web/packages/waveslim/waveslim.pdf.