FMDB *Transactions on Sustainable Health Science Letters*

Image Denoising and Feature Extraction Techniques Applied to X-Ray Seed Images for Purity Analysis

Suganthi M[1,*], J. G. R. Sathiaseelan[2]

[1,2,] Department of Computer Science, Bishop Heber College, Affiliated to Bharathidasan University, Trichy, Tamil Nadu, India.
sugan.swt09@gmail.com[1], jgrsathiaseelan@gmail.com[2]

Abstract: In this paper, image mining techniques are used for the purity test of various X-Ray seeds. Analyses of physical purity provide information on the percentage of pure intrinsic structure in a seed lot. An image mining technique has established a software application that can forecast seed pictures for seed lots. People may readily snap digital images anywhere, anytime, using a camera or mobile phone equipment, thanks to advances in camera technology. Additionally, using a computer system makes the transformation and processing simple. This research examines various image-mining methods that minimize the time and effort necessary to evaluate seedling growth performance while improving measurement accuracy. On the acquired X-ray image of the seed image, pre-processing techniques such as de-noising and feature extraction is performed to detect purity. Several denoise algorithms have been introduced, each with its benefits and controls. When choosing the right denoising algorithm, we need a good understanding of seed morphology. This paper presents a comparative analysis of four filter techniques. Feature extraction is associated with the seed's spatial, color, shape, texture, and statistical features. In order to classify various seeds using a feature extraction technique to produce the best results, this research developed a new texture feature extraction method.

Keywords: Guided Filter; Gaussian Filter; Median Filter; Feature Extraction; Reform Conventional Filter (RCF); Peak Signal to Noise Ratio (PSNR); Mean Square Error (MSE); Structural Similarity Index Metric (SSIM).

Received on: 05/11/2022, **Revised on:** 11/01/2023, **Accepted on:** 26/02/2023, **Published on:** 23/03/2023

Cited by M. Suganthi, and J. G. R. Sathiaseelan, "Image Denoising and Feature Extraction Techniques Applied to X-Ray Seed Images for Purity Analysis," *FMDB Transactions on Sustainable Health Science Letters*, vol. 1, no. 1, pp. 41–53, 2023.

1. Introduction

India is an agriculturally based country. But the farmers are facing many problems in growing better quality crops and diseases as they lack knowledge of diseases regarding the seeds. Once the quantity of illness has been determined, the seeds can be classed as infected, unaffected, or deadly. As a result, adopting image processing techniques might give farmers improved outcomes. The only way to meet strong market criteria is to have a weed-free seed lot. Purity tests may be improved by separating excellent quality seeds using suitable equipment and procedures. It can also assist in reducing the number of infected seeds in a seed lot and enhance the aesthetic, planting quality, and commercial. Seed quality is crucial for seeds to remain viable for a long time. The material structure of the seed may provide information about germination rates. The Association of Analytical Chemists (AOAC) designated X-ray imaging as an approved technique for detecting interior insect infestation in grains or seeds in 1961. X-rays have since been used to check the interior excellence of various agricultural items, including fruits, grains, and meals. X-ray-based quality monitoring systems have been used in several studies to investigate the internal component structures of seeds, which affect the seed's germination capacity.

[*]Corresponding author.

Noise reduction techniques are first used to detect the seeds' purity in this work. These methods of noise reduction are called image processing filters. The second element is featuring extraction, which is crucial to how any picture processing is presented due to its big impact on the outcomes of seed purity. The existing research study discusses many low-level structures, such as color, texture, and shape. Removing the colour feature of the image is not required because seed X-ray images, typically grey-level images, are employed in this area. To get the best results, this work employed a novel texture feature extraction technique for categorising seeds utilising a feature extraction technique.

2. Related Work

Using computer vision in agriculture has resulted in significant work on seed image filtering, seed texture, and form feature extraction via image mining.

Zin et al. [1] propose a fast Gaussian noise reduction method in this study. Noise suppression is accomplished via mass identical, 3D filtering, and biased nuclear standard drop. Even though these nonlocal pictures de-noising algorithms have a good quantitative presentation, the loss of high-frequency info outcomes in certain fine picture features is lost. To solve this problem, the revolutionary RAISR technology is improved and used immediately post-processing to the filtered image. With minimal processing overhead, it offers extreme results similar to cutting-edge methods while preserving essential picture structures. By improving mathematical translation and reducing the strength classes, they decrease the hash courses for the patch derived from the smoothed picture and the squares from the underlying data to 18 filters. Additionally, RAISR (Rapid and Accurate Image Super-Resolution) uses the demographic shift to combine the input images with the image that has undergone noise removal to produce results free of artifacts. Studies have shown this approach can provide more pleasing visual outcomes with much less recollection demand than other approaches.

Gajjar et al. [2] developed a complete system for detecting illnesses in a crop in real time. They suggested a novel CNN design for identifying and classifying 20 strong and sick greeneries from four distinct plants. It attained a precision of 96.88 percent, which is higher than existing designs.

SeedNet is a brand-new CNN architecture developed by Loddo et al. [3] for seed picture categorization and retrieval applications. The same dataset used in this study obtained 97.47 percent accuracy. However, the running time was rather lengthy, as is sometimes the case when using deep plans, particularly compared to standard machine learning techniques.

Although the authors started with a collection of seeds to address the issue of seeds relating to various phyla or classes, Gulzar et al. [4] used CNN and transfer learning to classify seeds. When selecting the pictures of the germs throughout information generation, his study uses symmetry. When scaling and labeling photos to abstract their features, regularity makes regularity. This resulted in a classification accuracy of 99 percent during the preparation set. For the examination usual of 234 photos, the suggested model gave findings that were 99 percent accurate. These results were significantly greater than those reported in previous studies.

Nadia Ansari et al. [5] developed a computer vision method to extract 20 crucial properties from 375 paddy germ images. The significance of the retrieved characteristics data amongst the seed types was then investigated using alteration examination. The separability of rice seed variants was also investigated using main component analysis. Throughout perfect creation, it was found that morphological picture topographies were more important than colour and textural picture components.

Laabassi et al. [6] offer a deep learning-based wheat varietal level categorization solution that produces reliable results (VLC). The wheat grain picture was classified into four varieties using the Convolutional Neural Network (CNN) (Simeto, Vitron, ARZ, and HD). Five conventional CNN designs were qualified using Transfer Learning to improve classification performance. We utilized data from 31,607 single-grain pictures taken after various Algerian locations plus acquired using various scanners to test the quality of the suggested models. According to the findings, the test accuracy for varietal level categorization varied from 85 percent to 95.68 percent. Consequently, the suggested approach's results are accurate and dependable, indicating that it might be used in practice.

Surekha et al. [7] introduced the multiple formulations technique, which uses connected component analysis to extract aspects such as spatial, color, texture, morphological, and statistical data from sample seed pictures. A Euclidian distance classifier is used to classify the seed pictures. The efficiency of the suggested approach is demonstrated using three grains: wheat, rice, and corn, with a 97 percent accuracy.

In order to distinguish between the eight canola seed variants, Salman Qadri et al. [8] developed a classification method built on a system idea method that combines multiple structures, such as top-level histogram, next-level arithmetical texture, binary, and spectral characteristics from majority pictures of canola seeds. Data was created using a 10-fold stratified cross-validation

technique, and these features were then applied to an ANN classifier. Colour digital photographs, an open sunlit setting, and the human collection of participation seed kinds to generate picture datasets are totally restrictions. All experiments were conducted in the open air with a modest computer and a digital camera. Very encouraging categorization results were obtained for the eight canola cultivars studied, with accuracy ranging from 95% to 98 percent.

3. Image Denoising

Unnecessary info, or noise, degrades the quality of images. Details on the kind of noise included in the original picture are crucial for eliminating background clutter. The noise taints typical photographs are described as a Gaussian, uniform, or salt-pepper pattern. Multiplicative speckle noise is another common type of noise. Following is a description of how each of these noises behaves.

3.1. Median Filter

The medium strainer is a statistically based nonlinear signal handling knowledge. The neighborhood's average assessment restores the numerical or the order's noisy value. The noisy value is recovered by saving the group's median value, which is determined by the grey levels of the pixels that make up the mask.

The result of MF is $g(x,y) = med\{f(x-i,y-j,,i,j \in W\}$ where $f(x,y), g(x,y)$are the normal picture and the filtered picture correspondingly, W is a 2d cover with dimensions of $n \times n$ (wherever n is usually odd), for example, $3 \times 3, 5 \times 5$, and so on; the mask form may be linear, rectangular, circle, fractious, or other.

3.2. The Median Filter's Ability to Reduce Noise

The medium filter's examination for a picture with random noise is rather sophisticated because it is a nonlinear filter. For a picture with no mean noise, the sound deviation of median filtering is approximately normal transmission.

$$\sigma^2 = \frac{1}{4nf^2(n)} \approx \frac{\sigma_i^2}{n+\frac{\pi}{2}-1}.\frac{\pi}{2} \quad (1)$$

The noise density function is where σ^2 input noise control (variance), n is the median filter mask level. And the average cleaning noise variance is $f(n)$

$$\sigma_0^2 = \frac{1}{n}\sigma_i^2 \quad (2)$$

The median filtering effects of (1) and (2) depend upon the mask level and the noise distribution. The average filtration performance is better than the average filtering performance of the random noise, but the pulsing noise is much smaller, and the pulse breadth is smaller than $n/2$ the median filter is extremely efficient.

3.3. Gaussian Filters

When filtering various varieties of surfaces, Gaussian filters are crucial [9]. This kind of purification is the primary option for filtering in many cases due to the ease of the method, ease of fabrication, and durability of the outcomes [10]. Academics frequently use the linear Gaussian filter, and it has become accepted as a reference for commercial purification in surface characterization. Gaussian filters can be used on the input ground by combining a Gaussian formula with the sample recorded. According to the formula, the Gaussian weighting factor is a chorus curve [11].

$$g(x) = \left(\frac{1}{\delta\lambda_c}\right) exp\left[-\pi\left(\frac{x}{\delta\lambda_c}\right)2\right]$$

The parameters of the predictive function for raising deconvolution could be determined by using the Gaussian function by

$$w_{i,j} = \left(\frac{1}{\delta\lambda_c}\right) exp\left[-\pi\left(\frac{x_{i,j}}{\delta\lambda_c}\right)2\right]$$

$$\delta = \sqrt{In\left(\frac{2}{\pi}\right)} \approx .4697$$

The output's roughness is controlled by the layering index, which in normal instances corresponds to the cut-off frequency λ_c. The constant value for $\lambda_c$is set to 0.01.

3.4. Guided filter

The guided filter's main presumption is a local linear relationship between instruction i and the filtration outcome q. Assumedly, q is a linear transform of I in the frame. w_k with pixel k as its centre.

$$q_k = a_k I_i + b_k \quad \forall i \in w_k \quad \ldots.. 1$$

Where other linear factors in w_k that are believed to be consistent are (a_k, b_k). We make use of an adjusted r - square frame. Because $\nabla q = a\nabla I$ in this local linear model, it is guaranteed that q has a margin only if I have one [12]. We require restrictions from the filtration element p to obtain the linear factors (a_k, b_k). We simulate the output q as the input p minus certain undesirable elements, such as noise or textures:

$$q_i = p_i - n_i \quad \ldots 2$$

We look for a solution that keeps the linear model but minimises the difference between q and p. (1). We specifically minimise the cost function shown below in the window w_k:

$$E(a_k, b_k) = \sum_{i \in wk}((a_k I_i + b_k - p_i)^2 + \in a_k^2) \ldots 3$$

The nonlinear regression-based model's formula and its answer are provided by:

$$a_k = \frac{\frac{1}{|w|}\sum_{i \in w_k} I_i p_i - \mu_k p_k}{\sigma_k^2 + \epsilon} \ldots.4$$

$$b_k = p_k - a_k \mu_k \ldots.5$$

3.5. Reform Conventional Filter

We proposed a previous concept to remove noise from the image in a new proposed filter, Reform Conventional Filter [18]. This proposed filter, an improved PSNR and SSIM, and a reduced MSE de-noise image are obtained.

4. Feature Extraction

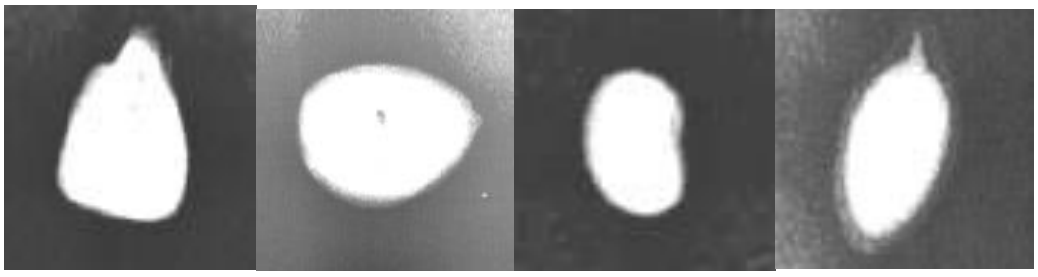
Shape, edge, and texture are the four seed characteristics discovered by feature extraction. The details for each feature are as follows.

4.1. Shape extraction

The form feature has two sub-features: the seed border and the mesocarp. The following information pertains to each sub-feature.

4.2. Seed boundary

The method used to determine seed boundaries involves cropping seed photos, as seen in Fig. 1. The machine then calculates the white pixels that make up the seed's boundaries by traveling in the appropriate boxes: Fig. 2 illustrates these directions: 1) from left to right, 2) from top to bottom, 3) from right to left, and 4) from bottom to top.



Corn Badam Bean Pumpkin

Figure 1: Seed cropping images

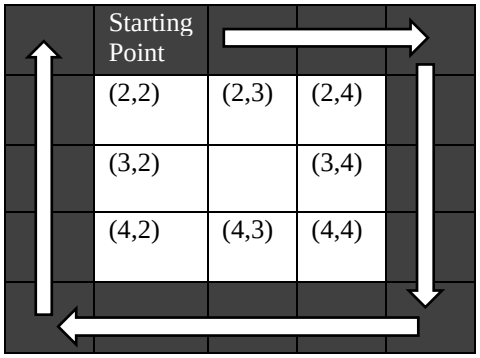


Figure 2: Boundary tracking direction

Algorithms were created to extract region descriptors to characterise the seeds with comprehensive and compact descriptors [13]. Area descriptions describe how the area's pixels are arranged. The area descriptors of an item can be obtained using various methods [14].

Eight geometric features, five form factors, ten standard moments, seven central moments, seven invariant moments(μ), and seven normalised essential instants (φ) are all extracted from the high-resolution photographs of the seed ounces for the feature analysis of the grains [15]. Area, perimeter, major and minor axis lengths, axis ratio, and compactness are some geometric descriptors. Shape factors 1–5 were computed from the values of axis length, area, and perimeter as follows:

$$Shape\ factor1 = \frac{1}{compactness},$$

$$Shape\ factor2 = \frac{Major\ axis\ length}{compactness}$$

$$Shape\ factor3 = \frac{Area}{Major\ axis\ length}$$

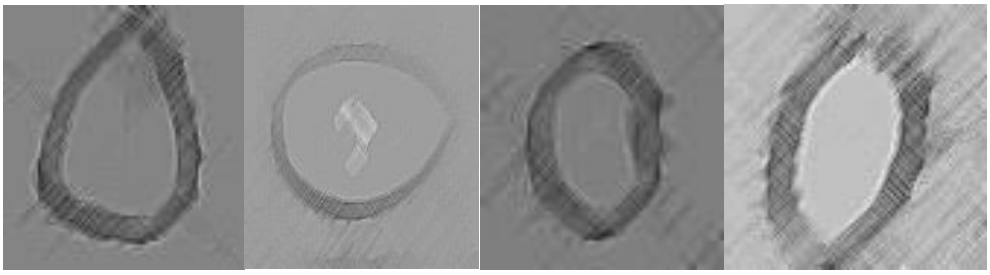
$$Shape\ factor4 = \frac{Area}{(Major\ axis\ length/2)(Major\ axis\ length/2)\pi}$$

$$Shape\ factor5 = \frac{Area}{(Major\ axis\ length/2)(Major\ axis\ length/2)\pi}$$

Instances include invariant, standard, central, and normalised central moments. According to Hu [16], moments were also utilised to determine the dispersion and thinness of grains [17]. However, moments include ten standard moments $m00$, $m01$, $m10$, $m02$, $m20$, $m03$, $m30$, $m11$, $m12$, and $m21$, seven central moments $\mu11$, $\mu02$, $\mu20$, $\mu03$, $\mu30$, $\mu12$, and $\mu21$, seven normalized central moments $\eta11$, $\eta02$, $\eta20$, $\eta03$, $\eta30$, $\eta12$, and $\eta21$, and 7 invariant instants $\varphi1$, $\varphi2$, $\varphi3$, $\varphi4$, $\varphi5$, $\varphi6$, and $\varphi7$. The moments remained computed using conventional formulas. Moments combine area, density, roughness, and higher-order descriptors to characterise a shape's design (the organization of its pixels).

4.3. Seed Mesocarp

Figure 3 shows the total oil content of loquat mesocarp and seed samples. According to the findings, the total oil content of loquat seeds (14.3%) was much greater than that of the mesocarp (1.21%). Because the seeds contain a high amount of oil, they can be considered high in energy.

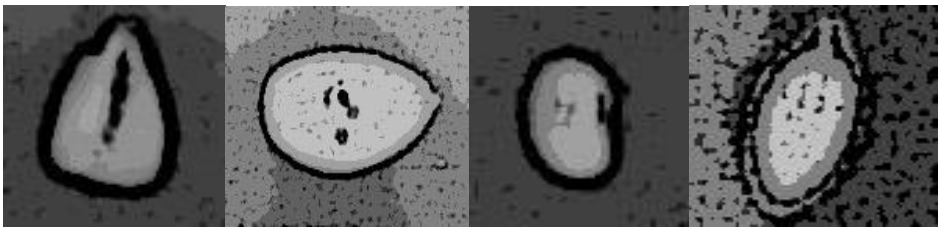


Corn Badam Bean Pumpkin

Figure 3: Seed mesocarp

4.4. Seed Edges

The system applies edge detection to X-ray seed images (Fig. 4). The technique then performs an AND operation with the destroyed seed picture to exclude the boundary. The remaining pixels are then multiplied by the area of the seed and divided by that quantity.



Corn Badam Bean Pumpkin

Figure 4: Seed Edge Detection

4.5. Texture Features Extraction

The texture comprises texture primitives called texels, which are adjoining groups of pixels with a tonal or geographical feature. The texture is defined by the texels' intensity (colour) attributes and the structural and spatial interactions between texels. The size of the textures is really important. Statistical/structural and syntactical texture characteristics are divided into two categories. In statistical texture features, texture characteristics are constructed using the statistical distribution of observed pairings of sensitivities at given places to one another in the image. Depending on how many intensity points (pixels) are present in each pair, statistics are categorised as first-order, second-order, and higher-order data.

4.6. Second-Order Statistics Features

The distributions of the image's grey levels can be determined using first-order statistical characteristics, but the comparative locations of the different gray values within the picture cannot be determined. When pixels are considered in pairs, second-order statistics elements do this. Using two or more parameters, such as distance away and pixel alignment, is common.

4.7. Gray Level Cooccurrence Matrices

The GLCM method, which excerpts second-order statistical texture structures, is a technique for doing so. The GLCM reveals the range of grey-level strengths. The number of rows and columns in a GLCM matrix matches the number of grey levels (G) in the picture. The regularity with which two pixels, divided by a pixel range ($\Delta x, \Delta y$), one with intensity I and the other with strength j, appear in a particular neighborhood is represented by the matrix element $P(i, j \mid \Delta x, \Delta y)$.

4.8. Contrast

The sum of squares variance is another name for this. This contrast or local intensity difference metric will be the inputs from $P(i, j)$ orthogonal to the diagonal, or $i \neq j$. It evaluates the overall intensity difference between a pixel and its surrounding pixels. If the picture is constant, the value is 0.

Contrast equation

$$CTR = \sum_{n=0}^{G-1} n^2 \left\{ \sum_{i=1}^{G} \sum_{j=1}^{G} P(i,j) \right\}, |i-j| = \quad (1)$$

The cell is on the diagonal, and $(i - j) = 0$ when I and j are equivalent. With a weight of 0, these numbers reflect pixels identical to their neighbours in every way. If there is a tiny contrast and the difference between I and j is 1, the weight is 1. The weight is four, and the contrast growing if I and j are different by two. As $(i - j)$ rises, the loads keep increasing exponentially.

4.9. Correlation

The grey-level linear dependency between the cells at the designated places in regard to one another is measured via correlation. It measures a pixel's correlation to its surrounding pixels throughout the image. The correlation's range is [1 1]. For a picture that is perfectly positively or negatively correlated, the correlation is either 1 or 1. For a static picture, the correlation is NaN.

Correlation equation:

$$COR = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i \sigma_j},$$

$$\mu_i = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} iP(i,j), \qquad \sigma_i = (i-\mu_i)^2 P(i,j)$$

$$\mu_j = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} jP(i,j), \qquad \sigma_j = \left(i-\mu_j\right)^2 P(i,j)$$

4.10. Energy

In the GLCM, energy returns the sum of the squared components. [0 1] is the region. Energy for a steady picture is 1.

$$Energy = \sum_{i=0}^{G-1} \sum_{j=0}^{G-1} \{P(i,j)\}^2$$

4.11. Homogeneity

A value that measures how closely the dispersion of GLCM components resembles the vertical is returned by uniformity. [0 1] is the area. For a vertical GLCM, uniformity is 1.

Homogeneity equation

$$\text{Homogeneity} = \sum_{i}^{G-1} \sum_{j}^{G-1} \frac{P(i,j)}{1+|i-j|}$$

4.12. Entropy

While non-homogeneous situations have low first-order entropy, homogeneous scenes have high entropy. When all probability is identical, total entropy is obtained.

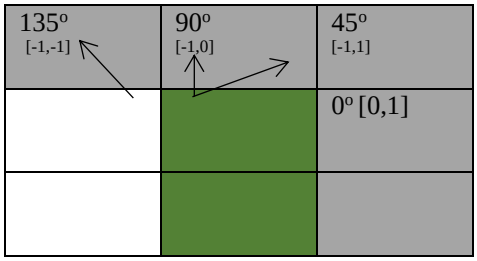


Figure 5: Offset values

Entropy equation:

$$Entropy = -\sum_{i=0}^{G-1}\sum_{j=0}^{G-1} P(i,j) \times \log\big(P(i,j)\big).$$

The offset (the connection of a pair of images) describes the separation between the target pixel and its neighbour. An angle is used to represent the offset. With respect to the pixel distance, which is here stated as 1, Figure 5 shows the distance between the sensor that defines typical angles.

Consider

Offset = {[0 1] for (0^o); [-1 1] for (45^o);

{[01] for (90^o); [-1 1] for (135^o);

5. Performance Evaluation Factors

5.1. Peak Signal Noise Ratio (PSNR) and Mean Square Error (MSE)

Related to these current approaches, the presentation of the developed method depends on qualitative and quantitative. The quantitative measurement uses Peak noise signal ratio (PSNR) and Mean Square Error (MSE). PSNR is listed

$$PSNR = 10\, log_{10} \frac{255^2}{MSE}$$

MSE is the Mean Square Error between the original seed picture (input seed image) and the resolution picture after noise removal (output seed image). MSE is defined as follows:

$$MSE = \frac{1}{NM}\sum_{x=l}^{M}\sum_{y=l}^{N}[g(x,y) - f(x,y)]^2$$

Where: M= Number of rows in the image, N: Number of columns in the image, g: Input image (Normal image), f: output image (Filtered image). Less value than MSN is the best result in the image.

5.2. Structural Similarity Index Metric (SSIM)

Depending on the presumption that the HVS has a structure closely related to the original to gather and process information about the structure from local features, and a high-quality image, the structural similarity index metric evaluates information about the overall. The image is divided into N × N-size frames, which are then used to execute the SSIM analysis on a local window. The three SSIM functions of light, contrast, and structure measurement are integrated to quantify picture quality.

The luminance comparison of images x and y is defined as:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}$$

Similarly, the contract comparison function is given by:

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}$$

Structure comparison is calculated using.

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}$$

where μ_x,μ_y are the mean values,

$\sigma_x\sigma_y$ are variance

σ_{xy} is covariance and C_1, C_2, C_3 are small constants

For a given local window, SSIM is calculated using

$$S(x,y) = \frac{2(\mu_x\mu_y + C_1)(2\sigma_x\sigma_y + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Finally, mean SSIM evaluates the image quality by combining local window SSIM values.

6. Results and Discussion

Based on the chosen image quality evaluation metrics, the experimental outcomes of the Four-filtering approach on noised seed X-Ray images are evaluated. The number of photos and filters was initialised before the studies began. The number of photos must be initialized to judge the excellence of the denoised picture. Therefore, tests are run to determine the ideal number of photos. The traditional mean square error (MSE), peak signal-to-noise ratio (PSNR), and structural similarity index are used to objectively assess denoising presentation (SSIM). In order to compare image quality more effectively, PSNR and MSE must be low. Figure 6 displays the outcomes of the picture implementation program, and Tables 1 to 3 display the MSE, SSIM, and PSNR values and the graph (Figure 7-9).

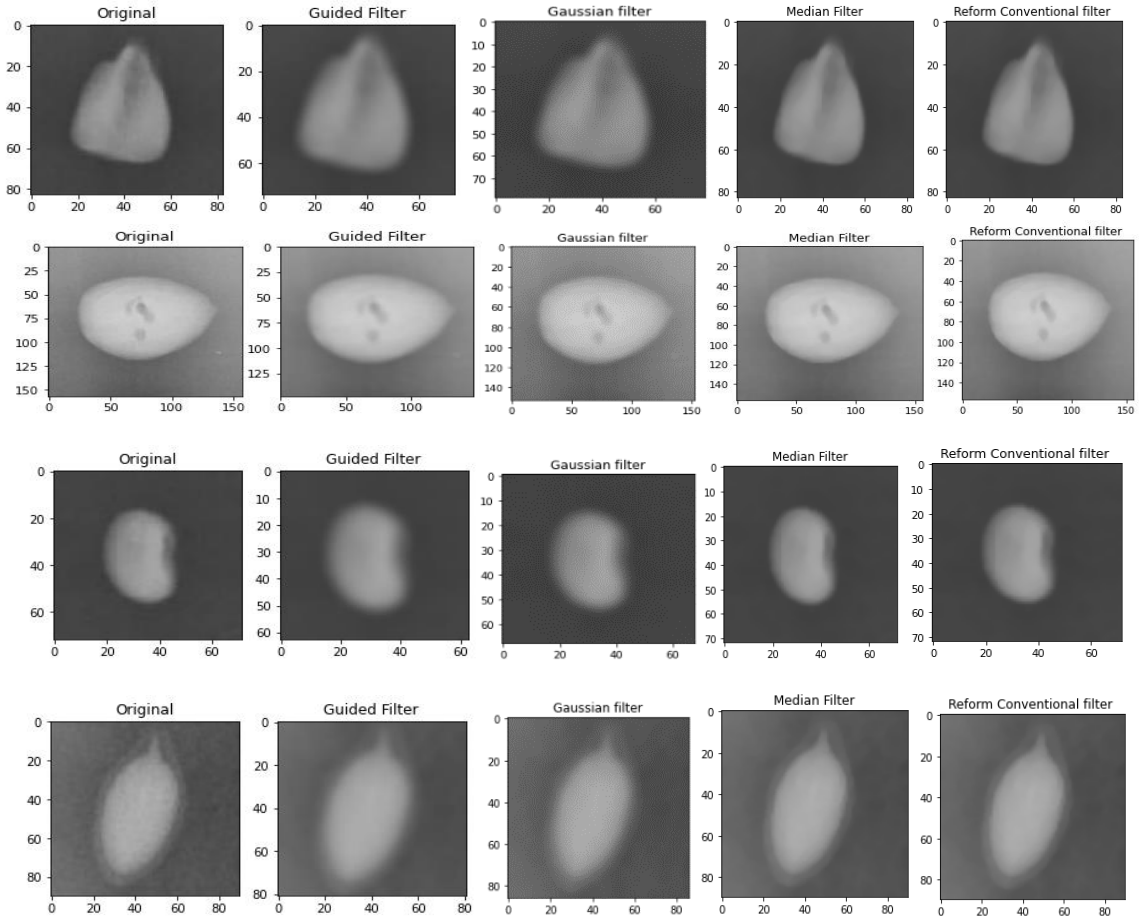


Figure 6: Original images (corn seed, Badam seed, bean seed, and pumpkin seed) and noise remove it's by using (Guided filter, Gaussian filter, Median filter, and Reform Conventional filter) methods

Table 1: Compare PSNR values to different filtering methods with the proposed RCF algorithm for various noises in the image

Filters / Seeds	Guided Filter	Gaussian Filter	Median Filter	Reform Conventional filter
Corn seed	33.9756	36.2119	46.6895	47.8361

Badam seed	35.1323	38.5609	49.5246	50.6482
Bean	34.7609	36.8680	47.3614	48.4721
Pumpkin seed	33.0614	36.4383	48.1539	49.2542

Table 2: Compare SSIM values to different filtering methods with the proposed RCF algorithm for various noises in the image

Filters **Seeds**	**Guided Filter**	**Gaussian Filter**	**Median Filter**	**Reform Conventional filter**
Corn seed	0.8403	0.9351	0.9997	0.9998
Badam seed	0.9152	0.9555	0.9981	0.9991
Bean	0.8837	0.9508	0.9980	0.9998
Pumpkin seed	0.8763	0.9280	0.9975	0.9985

Table 3: Compare MSE values to different filtering methods with the proposed RCF algorithm for various noises in the image

Filters Seeds	**Guided Filter**	**Gaussian Filter**	**Median Filter**	**Reform Conventional filter**
Corn seed	4.1220	0.9003	0.0682	0.0576
Badam seed	1.2246	0.3022	0.0445	0.0346
Bean	2.9184	0.6204	0.0530	0.0427
Pumpkin seed	1.9244	0.4714	0.0819	0.0751

From the result images and the data, the Proposed RCF algorithm is better than the other filter algorithm with a quality image.

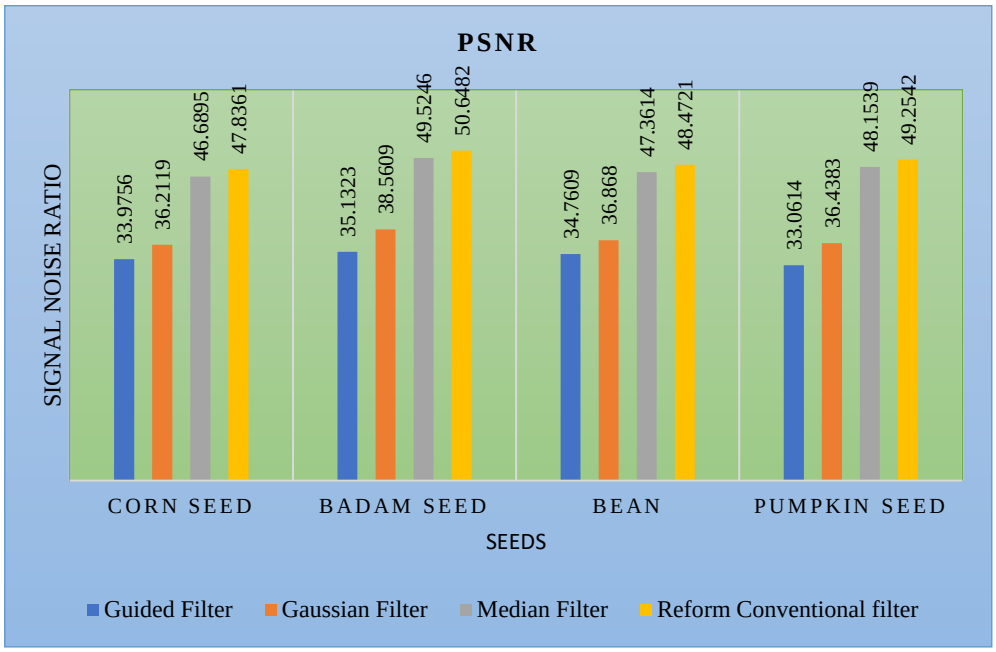


Figure 7: Graph of PSNR values for all seed images

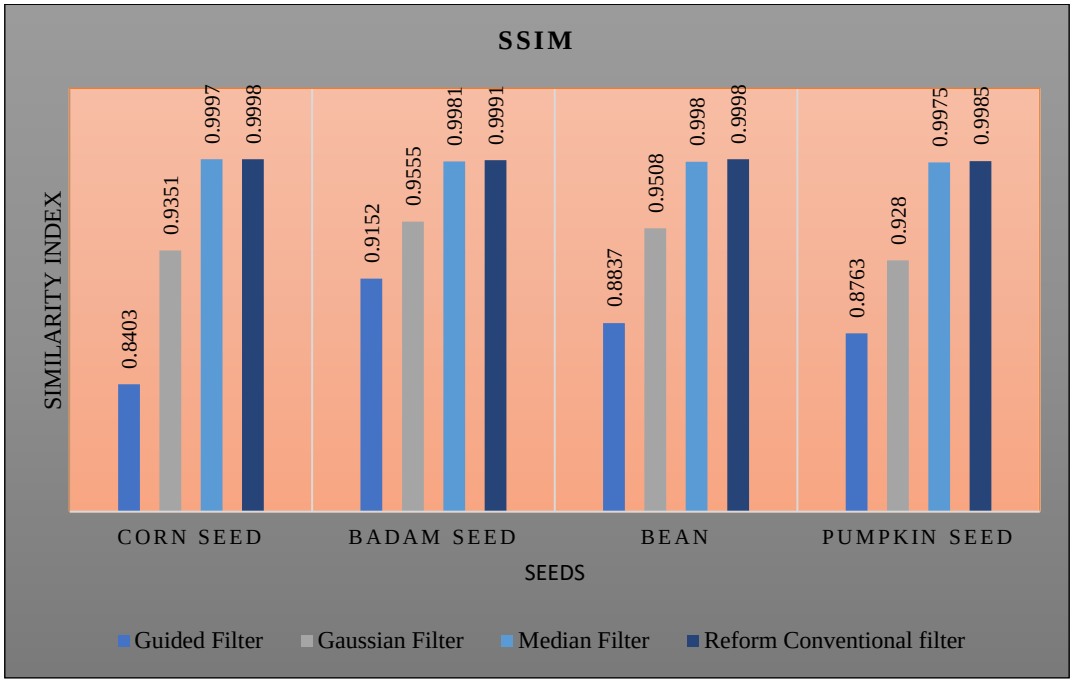


Figure 8: Graph of SSIM values for all seed images

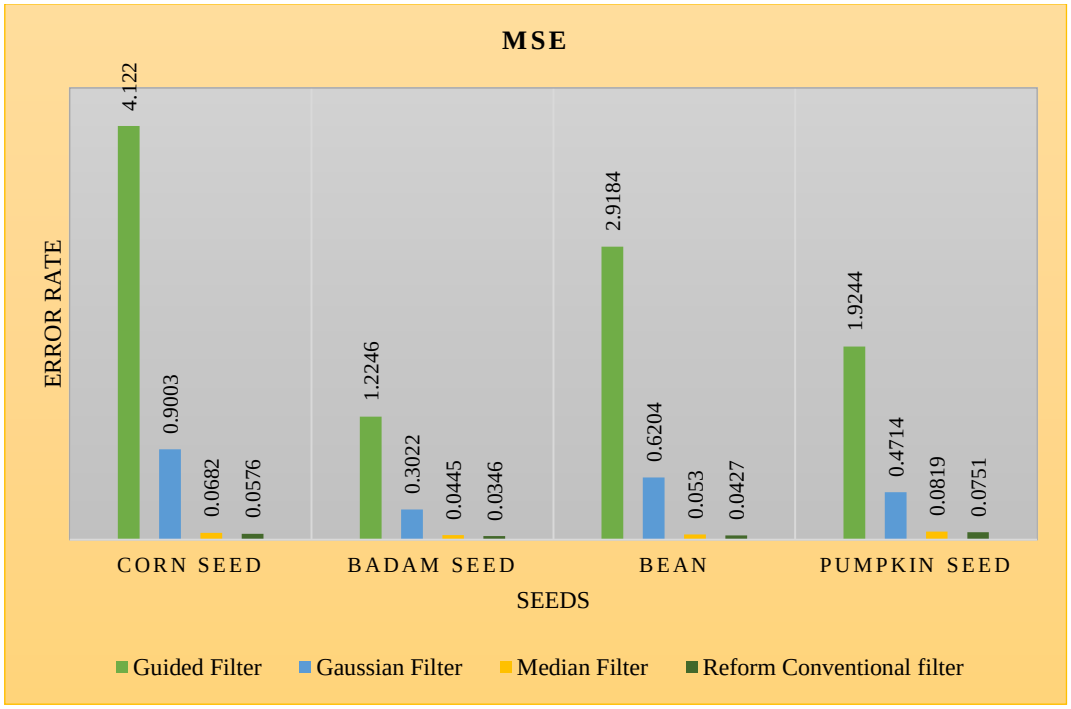


Figure 9: Graph of MSE values for all seed images

A training data set of ten species and a testing data set of five species, each of 15 different types of seeds, were used in the research. Six hundred photos of seeds from 10 different plant species can be found in the system database. The system was tested using 100 and 10 samples, accordingly, from each species in the training data set and testing data set. Match, mismatch, and unknown had 95%, 0.8%, and 4.2% precision rates in the training data set (Table 4). The precision rates for mismatch and unknown were 56.6% and 43.3% for the testing data set (as shown in Table 5). An image took 8.79 seconds on average to be accessed.

Table 4: Precision Rates for Training Data Set

Seed Name	Number of Testing data	Match	Mismatch	Unknown
Corn	100	85	0	15
Badam	100	96	0	4
Bean	100	99	0	1
Pumpkin	100	97	2	1
Sunflower	100	98	2	0
Sum	500	475	4	21

Table 5: Precision Rates for Testing Data Set

Seed Name	Number of Testing data	Match	Mismatch	Unknown
Cucumber	10	0	7	3
Grapes	10	0	0	10
Rice	10	0	10	0
Sum	30	0	17	13

7. Conclusions

This paper proposed an algorithm capable of noise reduction and feature extraction for X-ray seed images. The noise reduction part compared various noise-removal image filtering algorithms. According to the research findings, the effectiveness of the Median and Reform Conventional filters is superior to that of the Guided, Gaussian, and Median filters already in use. The evaluation of the image's quality is a very significant step in the process of processing X-Ray images. The feature extraction component of the plant seed recognition system extracts three critical traits to recognise plant seeds: 1) shape, 2) edge, and 3) texture. In this particular study, the method was evaluated using fifteen distinct kinds of plant seeds. The accuracy rates of the system for matching in a training dataset were 95 percent, whereas the accuracy rates for matching in an unknown testing data set were 56.6 percent. The time required to access each image was, on average, 8.79 seconds.

Acknowledgment: Support and guidance from my supervisor Dr J. G. R. Sathiaseelan.

Data Availability Statement: This study is conducted to predict seed quality. In order to classify various seeds using a feature extraction technique to produce the best results, this research developed a new texture feature extraction method. This is the new study conducted by the authors. The corresponding authors were notified to provide data from this work.

Funding Statement: No funding was received to conduct the research.

Conflicts of Interest Statement: Authors collectively produce this work where they all agree with the work done, issues faced during the project, and findings. The authors declared no conflict of interest. Citations and references are mentioned as per the used information.

Ethics and Consent Statement: This work is a draft from the corresponding author. Permission for data collection was obtained from the Institution, and Informed consent was obtained from the study participants. Authors of the work unanimously consent to make this publication available to all interested people for reading, teaching, and learning.

References

1. T. Zin, Y. Nakahara, T. Yamaguchi, and M. Ikehara, “Improved image denoising via RAISR with fewer filters,” Comput. Vis. Media (Beijing), vol. 7, no. 4, pp. 499–511, 2021.
2. R. Gajjar, N. Gajjar, V. J. Thakor, N. P. Patel, and S. Ruparelia, “Real-time detection and identification of plant leaf diseases using convolutional neural networks on an embedded platform,” Vis. Comput., vol. 38, no. 8, pp. 2923–2938, 2022.

3. A. Loddo, M. Loddo, and C. D. Ruberto, “A novel deep learning-based approach for seed image classification and retrieval,” Computers and Electronics in Agriculture, vol. 187, 2021.
4. Y. Gulzar, Y. Hamid, A. B. Soomro, A. A. Alwan, and L. Journaux, “A convolution neural network-based seed classification system,” Symmetry (Basel), vol. 12, no. 12, p. 2018, 2020.
5. N. Ansari, S. S. Ratri, A. Jahan, M. Ashik-E-Rabbani, and A. Rahman, “Inspection of paddy seed varietal purity using machine vision and multivariate analysis,” J. Agric. Food Res., vol. 3, no. 100109, p. 100109, 2021.
6. Laabassi, Karim, Mohammed Amin Belarbi, Saïd Mahmoudi, Sidi Ahmed Mahmoudi, and Kaci Ferhat. "Wheat varieties identification based on a deep learning approach." Journal of the Saudi Society of Agricultural Sciences, 2021.
7. R. Surekha, R. Shobarani, and G. Victo Sudha George, “Seed Classification using Multi-Feature Extraction,” International Journal of Innovative Technology and Exploring Engineering, Vol.8, no.8s, 2019.
8. S. Qadri et al., “Classification of canola seed varieties based on multi-feature analysis using computer vision approach,” Int. J. Food Prop., vol. 24, no. 1, pp. 493–504, 2021.
9. Y. Liu and W. Ma, “Fusing edge-information in image denoising based on CNN,” in 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), 2020.
10. D. Wu, H. Ren, and Q. Li, “Self-supervised dynamic CT perfusion image denoising with deep neural networks,” IEEE Trans. Radiat. Plasma Med. Sci., vol. 5, no. 3, pp. 350–361, 2021.
11. F. Jing, H. Shaohai, and M. Xiaole, “SAR image de-noising via grouping-based PCA and guided filter,” J. Syst. Eng. Electron., vol. 32, no. 1, pp. 81–91, 2021.
12. A. Limshuebchuey, R. Duangsoithong, and M. Saejia, “Comparison of image denoising using traditional filter and deep learning methods,” in 2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2020.
13. L. Lu, Y. Zhang, and K. Liu, “Non-iterative covariant feature extraction based on the shapes of local support regions,” IEEE Access, vol. 8, pp. 99354–99365, 2020.
14. Y. Adhitya, S. W. Prakosa, M. Köppen, and J.-S. Leu, “Feature extraction for cocoa bean digital image classification prediction for smart farming application,” Agronomy (Basel), vol. 10, no. 11, p. 1642, 2020.
15. J. Chaki and N. Dey, Texture Feature Extraction Techniques for Image Recognition. Singapore: Springer, 2020.
16. M.-K. Hu, “Visual pattern recognition by moment invariants,” IEEE Trans. Inf. Theory, vol. 8, no. 2, pp. 179–187, 1962.
17. A. Loddo, C. Di Ruberto, A. M. P. G. Vale, M. Ucchesu, J. M. Soares, and G. Bacchetta, “An effective and friendly tool for seed image analysis,” Vis. Comput., 2022.
18. M. Suganthi and J. G. R. Dr, “Seed Image Quality Analysis Using Proposed Filtering Technique for Image Denoising,” Design Engineering, pp. 3191–3204, 2021.