The impact of noise model on cancer diagnosis in MRI

تأثير نماذج التشويش على تشخيص مرض السرطان في صور الرنين المغناطيسي

Research proposed by

REEM MAJED ABWAN

Middle Technical University

Technical College of Management -Baghdad

Information Technology Department

**Abstract**

The presence of noise that can associate the MRI images, which was and still are problems that can affect the accuracy of the classification of the image and its contents which belongs to the infected or uninfected and the consequent increase in classification errors that affect the accuracy of the final decision

 For this purpose, the thesis presents to employ the Statistics features that are characterized by stability, the trend of the presence or absence of pictorial noise, as well as the difference in the type of this noise.

Some image transformations can be adopted such as (Fourier, Laplace, and Waft transform) and others, to observe the extent of the statistical distribution effect of this type of transfers, through the influence of the accuracy of the Taxonomic decision on the pictorial property of those dependent on this type

**الملخص**

ان وجود ضوضاء يمكن ان تصاحب صور الاستجابة المغناطيسية كان ولايزال من المشاكل التي يمكن ان توثر على دقة تصنيف الصورة ومحتوياتها لتعود الى مصاب او غير مصاب وما يتبع ذلك من تزايد في اخطاء التصنيف التي تؤثر على دقة القرار النهائي وفي سبيل ذلك جاءت الرسالة في محاولة لتوظيف الميزات الاحصائية التي تتميز بالثبات اتجاه وجود او عدم وجود ضوضاء(التشويش) صورية فضلا عن اختلاف نوع هذه الضوضاء

كما ويمكن اعتماد بعض التحويلات الصورية مثلل (تحويل فوريير، لابلاس ،تحويل ويفلت) وغيرها وذلك لملاحظة مدى تأثير التوزيع الاحصائي بهذا النوع من التحويلات من خلال مدى تأثر دقة القرار التصنيفي بالخاصية الصورية للمعتمدة على هذا النوع.

Keywords

noise model, statistical distribution, image digital, cancer disease

1-**Introduction**

Cancer currency in all countries of the world and different types of cancer and injury rates and mortality rates due to different countries and according to the conditions of each country, where different environmental conditions and the development of the medical system and health services in the diagnosis and treatment of disease and living conditions of the people of that country.(Aswathy, Devadhas et al. 2017)

The progress in the early discovery and diagnosis of cancer, especially brain cancer, requires the use of the latest medical devices and magnetic response images to reach the best diagnosis and to increase the accuracy of the diagnosis. The soft tissues of the body appear in high contrast and fine detail and can detect small changes within the body. This technique is used for accurate diagnosis. For strokes, cerebral tumor and also infection or spinal cord tumors.

To address noise problems and their impact on classification accuracy, then taking 100 images representing 72 injured and the rest uninfected, a small part of which was introduced due to adherence to research rules and then taking several statistical image features represented (entropy, medium, medium, mean, standard deviation, Variance, kurtosis) and study the extent of stability or change the statistical distribution as an indication of the extent of stability or change the image feature of medical images expressed

2-Previous work

Umbaugh as well as Gonzales methods and filters to remove noise from images, including mean filter and order filters, which include median, minimum, maximum, midpoint .... etc, which are all traditional mean to be without the use of a smart technology

In 2014 (Hari Babu Nandpuru and his colleagues) proposed a classification technique that classifies brain MRI images into natural images and abnormal images (images containing tumors). (28) properties are distributed among Gray level features, texture features, and symmetrical features, followed by the selection of effective features by applying Principal Components Analysis, and finally, the classification process using the classification algorithm. Automated Anchor Vector (SVM). 50 MRI images were used for this purpose. The results of the accuracy factor for the three functions used in the classification process were 74%, 84%, and 76%.

In 2017, researchers (Pranjal Jain and Harshita Didwania, Shivi Chaturvedi) showed a study addressing "brain tumor extraction from MRI using MATLAB" aimed at increasing the strategy of discovering and extracting brain tumors from MRI scans of a patient. Where at first he takes the person's name and age and then the brain MRI is used for the tumor detection process. It includes pre-treatment, segmentation, morphology, watershed segmentation, tumor area computation and tumor location.

 **3-Importance of Research**

The importance of this research is the need to avoid all varieties of problems associated with the process of diagnosis of the most serious diseases that harm humans, namely cancer, and before catching the biopsy.

**4- Objective of Research**

This study aims to increase the accuracy of the diagnosis of brain tumors through the formation of features and properties and methods of treatment by estimating parameters using different statistical distributions to find out the different estimation of these parameters to distinguish infected and non-infected

**5-cancer disease**

Cancer is a generic term that combines a wide range of diseases that can affect all parts of the body, also referred to as malignant tumors and cysts, and characteristic of the cancer is the rapid generation of abnormal cells that can grow beyond their known boundaries and then break into adjacent body parts and spread to other organs. This event is called metastasis, and metastases are the first cause of death from cancer.(Patil and Bhalchandra 2012)

Iraq suffers from the pollution of water, air, and soil resulting from the emissions of cars and generators in crowded areas, and the extreme use of chemical fertilizers, in addition to the remnants of war and shelling with depleted uranium, has exacerbated the difficulty of incurable diseases, especially cancer, where the incidence of cancer in Iraq, over the years The past few have risen significantly, compared to their rates in the early years following the US attack.(Avula, Lakkakula et al. 2014)

 **6-digital image processing**

Some can believe that digital image processing only means decorating, inserting or deleting some textures and drawings, and then looking in a strange appearance than the original. However, digital image processing goes beyond this and is almost non-existent. Importance is placed on appropriate digital encryption of images and ways to process these digital data so that the images or information displayed by the images can be used by a machine that can be a computer, a robot or other machine(Bhabatosh 2011)

1.6 **Image Enhancement**

There are many algorithms to optimize images as proposed. An important purpose is to purify the image from noise, which results in several reasons, such as a camera sensor or during image transfer and storage(Gonzales and Woods 2002)..

**2.6 image filtering**

Filtering is a method used to modify and improve the image. Using them we can, for example, highlight some of the features of the image such as edges or remove some defects such as noise or blur,(Gonzales and Woods 2002) such filters are using in this study

1.2.6 Median Filter

It is a non-linear digital filtering technology, often used to extract the noise from an image or signal. Noise abatement is a typical precursor to improving post-treatment results (for example, edge detection in an image). Medium filtering is generally used in digital image processing because it is under certain conditions (Melander, Bjorkman et al. 2000), with participants communicating with the region, the Asian region, the first floating Asian region N \* N in terms of interchangeability, or any, and after the risk of the legitimate project demand. Floor (Gonzales and Woods 2002) are the most common (worn, major, middle)

2.2.6 Sobel filtering

The Sobel worker performs a two-dimensional spatial scale on the image, through emphasizing the high spatial frequency domains corresponding to the edges. It is usually used for the analysis of each point in the grayscale image(Kaur and Gupta 2012), especially within edge discovery algorithms, to create an image that features the edges.

**7-Feature extraction**

A picture is worth a thousand words A phrase to the world (Gonzale 1992) The extraction of features and description is a critical step in the treatment of multimedia, where how to extract the ideal features that can display the basic content of the images as complete as possible is still a difficult problem in the process of computing and as is well known the most obvious features include It is common for color, texture, shape, etc. Among the features used in our research that are used for diagnostic and classification purposes are these features

 **1.7 Weighted Mean**:

This scale is one of the important measures of central tendency, which is in terms of the idea similar to the usual arithmetic mean, but the normal mean is the vocabulary of the sample under study has the same importance and effect in calculating any statistical indicator, but in some cases, some of the vocabularies are more powerful than others, which requires the use of another symbol to calculate The rate taking into consideration the importance of each of the sample items and this symbol is the weighted or weighted mean, and from here the value of this mean is more accurate than the usual mean, but it is less used than it was suggested in this research because of the similarity of values ​​when using the mean and to improve the accuracy of diagnosis and discover the injured Whether or not (Keller and Warrack 2004)

**2.7 Entropy**

It is an important and essential term for learning many of the events around us in the universe, and entropy is viewed through several sciences; Through the science of statistical mechanics, it is a measure of the amount of strength within any physical system in which this energy cannot be used to produce a work, which is the second law of the four laws of thermodynamics. Defined by check to interference in magnetic resonance images(Haralick, Shanmugam et al. 1973), and the entropy equation is shown

**3.7 Stander deviation**

It is the most used meaning among the distribution measures to measure the extent of statistical scattering, i.e. (Davies 2012)means the extent of the value fields extending within the statistical data set and symbolizes it σ The equation of number 3 represents the standard deviation equation

**8- NoiseModel**

Noise in images is generally defined as defective information that distorts and blurs the image. The first process that produces noise is the process of acquiring or recording a digital image, in which the visual image is transformed into a continuous electrical signal, which is later digitized. At each step of the process,(Patil and Bhalchandra 2012) variations due to natural events add random values to the brightness values of each element of the image, including.

1.8 Speckle Noise

This noise is a double noise. The appearance of lasers and radars due to continuous checking of waves, acoustics, etc., Speckle noise (point) can be similar to an image such as Gaussian noise. It follows the probability frequency function of the gamma distribution. When using any type of filter, the filter function filters dotted sums and reduces noise while keeping the sharpened edges in the image(Boyat and Joshi 2015) as in the equation below..

**2.8 Impulse Valued Noise (Salt and Pepper Noise)**

This is also described as low data noise because the default values ​​override the original data values. In this kind of noise, the image is not destroyed, but some pixel values ​​in the image change. Although the image is blurry, so chances are some pixels won't change (Chabay and Sherwood 2015)

This noise appears in the data transmission. If the number of bits to transmit is 8, the pixel value of the image will be replaced by a wicked pixel value with a value of minimum or minimum pixels, which is 255 or 0, respectively. (Joshi, Boyat et al. 2014) Its effect is evident in the photos taken quickly

# 9- Statistical Distributions

This means the shape that the data set takes. The shape of the data is very important in its analysis and representation, and as a step that precedes the decision to use any statistical method. The statistical distribution is usually connected with two types of related and separate data. The related type fits the nominal standards. There is a binary scale, meaning that it has only two values, and it is not called natural distributions, but moderately binary. One of the most important measures of distributions related to the binomial scale is due because the answer to the nominal scale is either yes or no.(Hosking, Wallis et al. 1985)

The study of random and discrete random variables and the probability distributions associated with each type of random variable to help us obtain outcomes that can be used in estimates of community features as well as tests of assumptions related to important decision making where such decisions are taken on a correct logical basis.(Munro 2005) The following are Some important probability distributions that have many important applications in practical life that were used in this research and that were used for the first time as the parameters of distributions as an indicator to help in deciding to contract the disease or not.

**1.9 Burr Distribution**

A Burr Type XII distribution or simply a Burr distribution is a continuous probability distribution of a non-negative random variable. Also known as the Singh-Maddala Distribution, it is one of several different distributions that are sometimes called "Generalized Logistics Distribution". (Evans and Ragab 1983)It is an important distribution in the analysis of data failure of equipment and channels and the study of reliability as well as applications of stress and durability and is one of the distinctions that were characterized to some extent by its resistance to the disturbance of the image magnetic resonance This distribution on several parameters as shown in the equations below (Wang and Rennolls 2005)

Parameters

- continuous shape parameter ()
α- continuous shape parameter (α >0)
β - continuous scale parameter (β >0)
 - continuous location parameter ( =0 yields the three-parameter Burr distribution)

Domain

Probability Density Function

.

Cumulative Distribution Function

Probability Density Function for three parameters

Cumulative Distribution Function for three parameters

2.9 Logistic Distribution

In probability theory and statistics, logistic distribution is a continuous probability distribution whose cumulative function is the same as the logistic function, which appears in logistic regression and anterior neural networks. (Shoukri, Mian et al. 1988)They resemble the normal distribution in shape but have disadvantages (high flattening). Logistics distribution is a special case of the distribution of Tukey lambda

Parameters

σ - continuous scale parameter (σ >0)
µ - continuous location parameter

Domain

In this distribution there are only two parameters so the probability density function and the cumulative function are in the equation below.

Probability Density Function

Cumulative Distribution Function

3.9 Generalized Pareto Distribution

The generalized Pareto distribution (GPD) is part of the probability distributions. It is often used to model another distribution. Determined by three parameters,(Hassan, Bakouch et al. 2013) sometimes only determined by scale and shape and sometimes only by its shape parameter, some references give the shape parameter κ = -ξ

Parameters

k - continuous shape parameter

σ - continuous scale parameter (σ >0)
µ - continuous location parameter

Domain

(The probability density function and the cumulative function of the double distribution are in the equations below)

Probability Density Function

Cumulative Distribution Function

**10- Data Research**

Neurosurgery Hospital was selected as a surgical hospital for tumor removal. The research sample included (100) images of brain tumors from MRI and different tumor location when imaging.

**Figure 1 Brain cancer MRI images of three infected & non-infected cases**

|  |  |
| --- | --- |
| Abnormal brain | Normal brain |
|  |  |
|  |  |
|  |  |

The suggested detection algorithm (SDA)

1/Start

2/Input MRI Image (Normal-Abnormal Image).

3/Adding Noise Model (Noise Image).

4/Filtering the Noise Image (Filtered Image).

5/Extract Statistical Features (Mean, Stander Division, Entropy) For (2,3,4) Steps.

6/Fitting A Suitable Statistical Distribution For (2,3,4) Steps.

7/End.

Table (1) the statistical features and distribution fitting for normal cases

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Image no. | wm | Entropy | Stander division | mod | statistical distribution |
| Normal images  | 1 | 0.00377  | 2.64118  | 0.04400  | 1.20128 | Gen.extream value |
| 2 | 0.00382,  | 2.52261  | 0.04015  | 3.10429 | Gen.extream value |
| 3 | 0.00386  | 2.45352  | 0.05100  | 3.574920 | Gen.extream value |
| Ab-Normal Images | 1 | 0.00389  | 2.47041  | 0.05812  | 0.0001114 | Burr(4p) |
| 2 | 0.00395  | 2.69872  | 0.06440  | 4.1512515 | Gen.extream value |
| 3 | 0.00383  | 2.55619  | 0.06082  | 3.33456 | Burr |

Table (2) the statistical features and distribution fitting for noisy(speckle) cases

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Noisy image | Image no. | Wm | Entropy | Stander division | mod | statistical distribution |
| Normal images  | 1 | 0.00377  | 2.51866  | 0.02659 | 0.00012 | Gen.extream value |
| 2 | 0.00377  | 2.55517  | 0.06062  | 1.04154 | Fréchet(3p) |
| 3 | 0.00391  | 2.53387  | 0.06885  | 1.54809 | Dagum |
| Ab-Normal Images | 1 | 0.00392  | 2.60240  | 0.06273  | 3.917497 | Burr(4p) |
| 2 | 0.00372  | 2.53063  | 0.05781  | 7.24083 | Burr(4p) |
| 3 | 0.004135  | 2.67772  | 0.04082  | 8.144460 | Gen.extream value |

Table (3) the statistical features and distribution fitting for filtered(median) cases

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Filtered image | Image no. | wm | entropy | Stander division | mod | statistical distribution |
| Normal images  | 1 | 0.00379  | 2.56217  | 0.01760  | 1.04863 | Gen.extream value |
| 2 | 0.00381  | 2.56113  | 0.03538  | 4.41711 | Fréchet(3p) |
| 3 | 0.00394  | 2.47432  | 0.06031  | 2.99011 | Gen.extream value |
| Ab-Normal Images | 1 | 0.00393  | 2.53029  | 0.02076  | 2.220340 | Burr(4p) |
| 2 | 0.00372  | 2.45909 | 0.05133  | 3.5874060 | Dagum |
| 3 | 0.00391  | 2.43245  | 0.02743  | 3.505694 | Burr(4p) |

Table (4) the statistical features and distribution fitting for filtered(soble) cases

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Filtered image | Image no. | Mean | entropy | Stander division | mod | statistical distribution |
| Normal images  | 1 | 0.00390 | 2.57376 | 0.01760 | 6.84178 | Gen.extream value |
| 2 | 0.00397 | 2.57657 | 0.03537 | 6.84178 | Dagum |
| 3 | 0.00372 | 2.48283 | 0.06030 | 2.21048 | Gen.extream value |
| Ab-Normal Images | 1 | 0.00388 | 2.54412 | 0.02076 | 2.25859 | Burr(4p)) |
| 2 | 0.00362 | 2.47172 | 0.05133 | 2.069078 | Gen.extream value |
| 3 | 0.00392 | 2.44142 | 0.02742 | 4.637635 | Gen.extream value |

From the above tables we can see that the statistical features and the statistical distribution effected with the kind of noise and the adopted filter so that the numerical values of each case deepened on the amount and kind of noise.

Figure (1) polynomial curve for statistical features for normal cases

Figure (2) polynomial curve for statistical features for filtered(speckle) cases

Figure (3) polynomial curve for statistical features for filtered(median) cases

Figure (4) polynomial curve for statistical features for filtered(soble) cases

From the above figure we can see that the polynomial line difference from each other, that’s because the effect of noise and filter on the numerical values of the statistical features.

Conclusions

The normal images have the values ​​of the features are similar, such that, the weighted mean range between (0.003 and 0.004), the entropy (2.4 and 3) and the standard deviation between (0.01 and 0.05), As for the images that are exposed to noise speckle, the values ​​of the features are also similar, such that, the mean range between (0.003 and 0.004), the entropy (2.5 and 3) and the standard deviation between (0.01 and 0.06), The images that used the filter medium & sobel have values ​​that are almost similar to the normal images, So if MRI images are exposed to a number of noise such as patient movement or when transmitted, a number of improvements leading to accurate diagnosis can be made by the following (SDA) ​​algorithm instead of re-checking.

Discussion

Stability of some of the image features towards the presence of noise accompanying medical images through the stability of the statistical distribution attributable to them, although there are a number of image transformations (instant, Laplace and Waffle) and others that can be adopted as well as increasing the accuracy of the classification by applying some filters that reduce or eliminate noise accompanying some Digital images with the possibility of adopting filters to isolate the injury area and applying the proposed classification models (noise filtering, static feature stability, network training) to determine the type of tumor as well as the injury or not with the use of other images scan, spiral secretion, x-rays and others to note the extent of the statistical distribution affected by the light accompanying this type of medical images(by researcher)

References

1. Aswathy, S., G. G. Devadhas and S. Kumar (2017). "MRI brain tumor segmentation using genetic algorithm with SVM classifier." Journal of Electronics and Communication Engineering,
2. Avula, M., N. P. Lakkakula and M. P. Raja (2014). Bone cancer detection from mri scan imagery using mean pixel intensity. 2014 8th Asia Modelling Symposium, IEEE.
3. Bhabatosh, C. (2011). Digital image processing and analysis, PHI Learning Pvt. Ltd.
4. Boyat, A. K. and B. K. Joshi (2015). "A review paper: noise models in digital image processing.
5. Chabay, R. W. and B. A. Sherwood (2015). Matter and interactions, John Wiley & Sons.
6. Davies, E. R. (2012). Computer and machine vision: theory, algorithms, practicalities, Academic Press.
7. Evans, I. and A. Ragab (1983). "Bayesian inferences given a type-2 censored sample from a Burr distribution." Communications in Statistics-Theory and Methods
8. Gonzale (1992).
9. Gonzales, R. C. and R. E. Woods (2002). Digital image processing, Prentice hall New Jersey.
10. Haralick, R. M., K. Shanmugam and I. H. Dinstein (1973). "Textural features for image classification." IEEE Transactions on systems, man, and cybernetics
11. Hassan, A. A., S. Bakouch and L. Esmaeili (2013). "Pareto Poisson-Lindley distribution and its application." Journal of Applied Statistics
12. Hosking, J. R. M., J. R. Wallis and E. F. Wood (1985). "Estimation of the generalized extreme-value distribution by the method of probability-weighted moments." Technometrics.
13. Joshi, A., A. K. Boyat and B. K. Joshi (2014). Impact of wavelet transform and median filtering on removal of salt and pepper noise in digital images. 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), IEEE.
14. Kaur, J. and P. Gupta (2012). "Fuzzy logic based adaptive noise filter for real time image processing applications." International Journal of Computer Science Issues
15. Keller, G. and B. Warrack (2004). Statistics for management and economics, Thomson/Brooks/Cole.
16. Melander, B., M. Bjorkman and P. Gunningberg (2000). A new end-to-end probing and analysis method for estimating bandwidth bottlenecks. Globecom'00-IEEE. Global Telecommunications Conference
17. Munro, B. H. (2005). Statistical methods for health care research, lippincott williams & wilkins.
18. Patil, R. C. and A. Bhalchandra (2012). "Brain tumour extraction from MRI images using MATLAB." International Journal of Electronics, Communication & Soft Computing Science and Engineering
19. Shoukri, M., I. Mian and D. Tracy (1988). "Sampling properties of estimators of the log‐logistic distribution with application to Canadian precipitation data." Canadian Journal of Statistics
20. Wang, M. and K. Rennolls (2005). "Tree diameter distribution modelling: introducing the logit logistic distribution." Canadian Journal of Forest Research