

A STATISTICAL ANALYSIS OF THE EFFECTS OF CT ACQUISITION PARAMETERS ON LOW-LEVEL FEATURES EXTRACTED FROM CT IMAGES OF THE LUNG

Joseph S. Wantroba, Daniela S. Raicu, Jacob D. Furst
Intelligent Multimedia Processing Laboratory
School of Computing, College of Computing and Digital Media
DePaul University, 243 S. Wabash Chicago, Illinois 60604, USA

1. ABSTRACT

We propose a solution for automatic classification of lung nodules in an environment with heterogeneous Computed Tomography (CT) acquisition parameters. Such a classification system needs to take into account the differences in CT acquisition parameters used when obtaining and processing each medical image. Using Analysis of Variance (ANOVA), our current research proposes to better understand the effects of CT acquisition parameters on predicting various semantic characteristics (such as spiculation, subtlety, and margin) used in the diagnosis interpretation process. All of the parameters were found to affect the low-level image features used in the classification models of these semantic characteristics. When this knowledge is used to normalize those parameters, the final semantic model will become unaffected by the CT acquisition parameters.

Index Terms – CT Acquisition Parameters ANOVA, CAD

2. INTRODUCTION

There has been a tremendous amount of imaging research done on computed tomography, and in the future, there will likely be more, as the need for and availability of CT imaging increases. However, to date, most of this research falls short in one significant aspect: since most of the research has been done at a single site, most of the imaging equipment from which CT images are produced is homogenous. The effects of differences in equipment remain ignored. With the publication of the Lung Image Database Consortium (LIDC) work, researchers now have a common dataset of CT images on which to investigate how the differences in equipment can affect the differences in images. This work analyzes how the different equipment's acquisition parameters affect extracted image features and the prediction of nodules' semantic characteristics.

3. BACKGROUND

CAD systems for lung nodules have been built using image features extracted from a set of CT images to predict malignancy [1, 2, 3]. Our previous research [4] on building

CAD systems has focused on using extracted images features to predict radiologist ratings of semantic characteristics from the LIDC [5]. In neither case do these CAD systems take into account the differences in CT scanners. However, the Lung Image Database Consortium (LIDC) provides a large public dataset in which CT images were taken from different CT scanners, creating a heterogeneous dataset. The dataset of 85 cases (patients) allows researchers to test whether or not the different CT scanners affect the results of a CAD system.

One possible solution for taking into account heterogeneous parameters is to split the data according to the parameters, for example, based on the values of Convolution Kernel. However, such a split would make eight datasets (one for each convolution kernel represented in the LIDC), each with too few cases to accurately predict semantic characteristics. Additionally, such a split would have difficulty predicting for new CT equipment.

ANOVA is a well-known statistical method that has been used in many domains, including medical imaging. Studies concerned with differences in the volumetric measurement of nodules over time using different CT scanner reconstruction parameters have used ANOVA to measure statistically significant changes in nodule's size [6]. Similarly, another study employed ANOVA to assess the image quality of CT-images when different scanner parameters were used [7]. Other studies used ANOVA to better understand differences in how radiologist rate images based on their level of expertise [8, 9, 10]. Another study likewise used ANOVA to understand which image features are most important for detecting high-level features for content-based image retrieval [11]. Our study is unique because ANOVA is used to understand how certain CT parameters affect the most important image features when making diagnostic predictions and provides insights on how to normalize these features to improve robustness.

4. METHODS

4.1 Dataset

The LIDC dataset includes 85 cases of which 60 cases have 149 distinct lung nodules with a diameter greater than 3mm. Up to four radiologists marked these images using ratings on a scale of 1 to 5 to assess nine semantic characteristics:

lobulation, spiculation, sphericity, calcification, texture, internal structure, malignancy, margin, subtlety. Internal structure and calcification are not considered because radiologists gave each image the same rating, ‘soft tissue’ and ‘no calcification’ respectively [4]. A set of 64 image features contained in 4 categories (shape, size, intensity, and texture) are extracted from each image (Table 1).

Table 1: Image features; SD stands for standard deviation, BG for background, and MRF for Markov Random Fields

Shape	Size Features	Intensity	Texture
Circularity, Roughness, Elongation, Compactness, Eccentricity, Solidity, Extent, Radial Distance SD	Area, Convex Area, Perimeter, Equivalent diameter, Major axis length, Minor axis length,	Minimum Intensity, Maximum Intensity, SD Intensity, Minimum Intensity BG, Maximum Intensity BG, Mean Intensity, Mean Intensity BG, SD Intensity BG, Intensity Difference	11 Haralick features, 24 Gabor features, 5 MRF features

4.2 Our Previous Work

Our initial predictors for LIDC semantic characteristics used decision trees for classification but did not consider CT parameters. Subsequent research used the same decision tree classifiers (DT’s) but this time added the CT acquisition parameters to the previously calculated image features [12]. The results have shown that CT acquisition parameters do appear within those DT’s and thus affect the classifier’s results. However the decision trees do not allow us to understand how to normalize those features in order to preprocess the images so that the classifiers are independent of the CT acquisition parameters. The decision trees approach did show that there are 7 CT acquisition parameters that affect semantic mappings: slice thickness, exposure, kVp, reconstruction diameter, convolution kernel, Z nodule location, and distance source to patient [12]. This research looks at specifically how the image features are affected by these 7 CT parameters through the use of ANOVA.

4.3 Analysis of Variance (ANOVA)

This research implements traditional techniques in new areas of research. Our research implements analysis of variance (ANOVA) in order to understand the relationship between CT acquisition parameters and extracted image features. ANOVA compares the mean square differences between groups and within groups in order to statistically prove whether or not one group is statistically different than another. A group is different when its mean square is not equal to the mean squares of all other groups [11]. To prove statistical significance of the mean differences between the two groups, the observed F- value, $\frac{MS_{BG}}{MS_{WG}}$, (where MS_{WG} =

within group variability and MS_{BG} = between groups’ variability) must be compared to the critical F-value. This can be done by computing a p-value, the probability that the observed F-value is different from the critical F-value [11]. If the p-value is smaller than a specified α (significance level), then it can be stated that the two groups are significantly different from one another.

Our null hypothesis H_0 states that CT parameters do not affect image features. Our alternative hypothesis H_a states that CT parameters do affect image features [11]. The null and alternative hypotheses are in the following form:

$$H_0(f, a) : \forall_{i,j;i \neq j} \mu_i(f, a) = \mu_j(f, a)$$

$$H_a(f, a) : \exists_{i,j;i \neq j} \mu_i(f, a) \neq \mu_j(f, a) \quad (2)$$

where a denotes any of the 7 acquisition parameters, i and j index the groups defined by the different values of acquisition parameter a , and f denotes any of the 64 low-level image features. In total, there will be 448 tests of significance to be performed (64 image features X 7 types of acquisition parameters). Note that the null hypothesis only tells us that the different values of an acquisition parameter don’t affect an image, while the alternative hypothesis tells us that at least one value of an acquisition parameter affects the image feature, but doesn’t tell us which one(s).

4.4 Least Significant Difference (LSD)

Using a post hoc test, like the Least Significant Difference (LSD) used in our research, further enhances the results of ANOVA by showing which group differs from the other groups, assuming the null hypothesis is rejected. LSD uses the following statistics similar to a T-test; the difference being that the mean squares within all groups is used instead of the mean squares of only the groups compared for each test:

$$LSD(f, a) = \frac{\mu_i(f, a) - \mu_j(f, a)}{\sqrt{MS_{WG}(f, a) \left(\frac{1}{n_i(f, a)} + \frac{1}{n_j(f, a)} \right)}} \quad (3)$$

where MS_{WG} = Mean Squares Within all groups (taken from ANOVA), $\mu_i(f, a)$ = Mean of grouping i in acquisition parameter a for image feature f , $\mu_j(f, a)$ = Mean of value j in acquisition parameter a for image feature f , $n_i(f, a)$ = total number of cases for value i in parameter a for image feature f , and $n_j(f, a)$ = total number of cases for value j in parameter a for acquisition parameter f .

The LSD value is compared to the critical t-value to assess specifically which value groupings of each parameter affect each image feature. This difference is shown when the LSD value is greater than the critical t-value [13].

5. RESULTS

ANOVA was run 448 times to see which of the 7 CT acquisition parameters affect which of the 64 image features. For every null hypothesis rejected, the LSD post-hoc test was performed to see which values of those CT acquisition parameters affect which image features differently. A p-value threshold of 0.0001 was used to differentiate the most important image features affected by the CT acquisition parameters: the standard threshold is typically much higher (e.g. 0.01) but in our case such a threshold is too high to accurately signify the most important image features affected by the CT acquisition parameters. Those CT acquisition parameters that contain continuous data, reconstruction diameter and exposure, were binned using equal-width bins (widths of 600 and 26 respectively) in order to properly run the LSD post-hoc test. The results of this function allow us to properly choose the most effective normalization technique to create the most robust prediction model for the semantic characteristics. To maintain the system's accuracy, the dataset should go through a preprocessing stage of normalization that makes the prediction model independent of the CT parameters.

Our research using ANOVA showed that CT acquisition parameters significantly affect image features (see Appendix for definitions). Convolution kernel affects all image features. It is the only CT parameter to do so, and this makes sense because the convolution kernel affects how the image is reconstructed (Figure 1). This shows the importance of this particular CT acquisition parameter, and this importance is reinforced by our decision tree classifier results [12] in which this CT acquisition parameter appeared the most (6 times), including multiple appearances within one decision tree. When the convolution kernel FC01 was used on images, all 6 of the Intensity features under the 0.0001 p-value threshold were significantly different from when the other convolution kernels were used to reconstruct images (Figure 2). All 15 Gabor texture image features under the 0.0001 p-value threshold were significantly different when a BONE convolution kernel was used as opposed to any to when the others were used. Convolution kernels are typically split up into two groups, edge enhancing or smoothing. When the kernels from these two groups were compared, there was no significant change in the features. This is additional evidence for what was already concluded in our previous research [12], mainly that these groups are too broad in scope to adequately distinguish the first group from the second group. The broadness stems from the fact that manufacturers each have proprietary convolution kernel algorithms. It is possible, though, to nullify the effects of convolution kernels FC01 and BONE, which have verifiably different square means from the other convolution kernel values for Intensity features and Gabor texture features respectively. A mean-shift normalization implemented during the data preprocessing stage would create identical means of all

convolution kernels, allowing us to ignore this CT acquisition parameter completely.

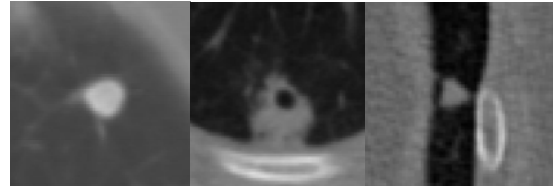


Figure 1: From left to right: Nodule images using three convolution kernels with different intensity ranges: FC01 (-998,1106), B31f (116,1864), and B30f (26,1833), respectively.

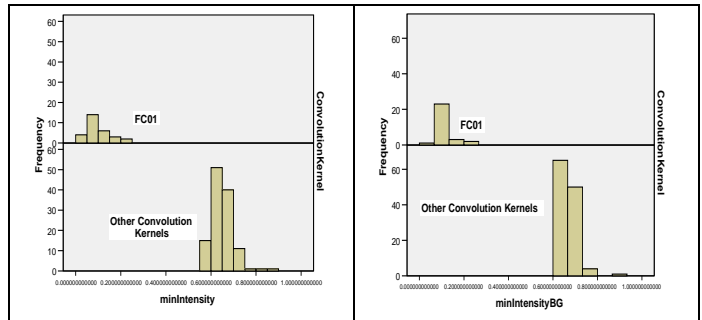


Figure 2: Histograms for two convolution kernels and two intensity features (min intensity –left, min intensity background –right)

Exposure also affects texture and intensity features. MinIntensityBG shows up consistently as the feature with the lowest p-value in every acquisition parameter except distance source to patient and Z nodule location. Area also appears for this CT parameter. ANOVA shows that the nodule size is important and should be taken into account when building robust CAD systems. MinIntensityBG and ClusterTendency appear in both our ANOVA results and our decision tree classifier results. Results show that exposure values ranging from 0-600 significantly affect the values for MinIntensity differently than all other exposure values. Values ranging from 1200-1800 change the values for Markov0 and Markov4 by a significant difference than other exposure values. All of the Gabor texture features under the p-value threshold change significantly when exposures of 1200-1800 are used as opposed to any other exposure value. Finally the post-hoc testing revealed that for 19 out of the 64 features the exposure value range of 0-600 changed those image features differently than a range of 1200-1800. Distance source to patient affected many texture features, although none of the features that were deemed affected by the decision trees appeared within the lowest p-values under the threshold (0.0001).

kVp is shown to affect Intensity and Texture features. The difference in penetration ability results in different intensity and texture values. A kVp value of 135 affects all 6 Intensity features below the 0.0001 threshold differently than any other value for kVp. These 6 features appear as most affected by kVp. All 7 of the Gabor features are affected by a kVp value of 140 in a different way than other

kVp values. Finally, elongation is shown to be unaffected by kVp (although affected by kVp in our previous [12]).

Reconstruction diameter only appeared once in our DTs and was shown to affect Markov4, a texture image feature. Markov4 appears with a low p-value when using ANOVA, proving its significance. Further post-hoc testing after binning revealed that a reconstruction diameter range of 286-312 changes Markov4 differently than any other reconstruction diameter value. 9 out of the 12 Gabor texture features were affected by their reconstruction diameter range of 386-312 differently than any other value of reconstruction diameter. Finally, all 6 of the Intensity features were found to be affected by the reconstruction diameter range of 364-390 differently than any other value of the parameter.

Slice thickness affects all image features except radialDistanceBG, a shape feature. This CT acquisition parameter is defined as the thickness of each CT image slice in mm. It appears in our decision trees, but does not affect any image features. ANOVA suggests, as previous studies have, that this parameter is very significant. All types of features are affected by this parameter when looking at those image features with p-values less than 0.0001. All 6 Intensity features under the p-value threshold of 0.0001 are affected differently by a slice thickness of 3mm in contrast to the other slice thickness values. 18 Gabor texture features appear under the threshold. All 18 are significantly different when a 1.25mm is used instead of other values.

Z nodule location appears in the decision trees and affects a shape and intensity feature. ANOVA results prove that these are also affected by Z nodule location, but no image feature falls below the p-value threshold (0.0001).

6. CONCLUSION

ANOVA statistically proves that CT acquisition parameters affect image features. Every parameter affects a number of image features, but Convolution Kernel is the only CT acquisition parameter that affects all image features. Intensity features prove to be affected the greatest by the acquisition parameters, while shape features prove to be affected the least. These results will assist in future work by giving insight into which normalization techniques for the CT acquisition parameters will be most effective. A normalization technique, unique to each CT parameter, can be used as a way to nullify the effects of those parameters, the ultimate goal being the creation of robust CT-independent classification systems for nodule interpretation.

7. APPENDIX

Table 2: Definition of all features that appeared in Results

Image & Param.	Definitions (based on [4] and [12])
<i>Area</i>	The actual number of pixels in the region
<i>ClusterTendency</i>	A Haralick co-occurrence texture feature
<i>Convol. Kernel</i>	An algorithm to reconstruct the image data
<i>Distance Source to Patient</i>	The distance from x-ray beam source to patient
<i>Exposure</i>	The number of x-ray photons in the beam

<i>GaborSD_45_03</i>	Standard Deviation of the Gabor image for direction of 45^0 and frequency of 0.3.
<i>kVp</i>	Beam quality; the ability of the beam to penetrate an object
<i>Markov0 and Markov4</i>	Markov features that captures the local contextual information (at 0^0 and 135^0)
<i>MinIntensity and MinIntensityBG</i>	Minimum gray-level intensity of the nodule and of the background pixels in the bounding box around the nodule
<i>RadialDistanceBG</i>	The standard deviation of the distances from every boundary pixel to the region's centroid
<i>Reconstruction Diameter</i>	The diameter of the region used to reconstruct the image
<i>Slice Thickness</i>	The thickness of each CT slice (in mm)

8. REFERENCES

- [1] Armato SG, III, Giger ML, MacMahon H. Automatic detection of lung nodules on CT scans: preliminary results. *Med Phys* 28 (2001):1552-1561.
- [2] Kanazawa K, Kawata Y, Niki N, et al. Computer-aided diagnosis for pulmonary nodules based on helical CT images. *Computerized Med Imaging Graph* 22(1998):157-167.
- [3] Lee Y, Hara T, Fujita H, et al. Automated detection of pulmonary nodules in helical CT images based on an improved template-matching technique. *IEEE Trans. MI* 20(2001):595-604.
- [4] Raicu DS, Varutbangkul E, Furst JD, Armato III SG. Modeling semantics from image data: opportunities from LIDC. *Int. Journal of Biomedical Engineering*, Vol. 2, (3), 2009.
- [5] NCI - Lung Imaging Database Consortium (LIDC). <http://imaging.cancer.gov/programsandresources/InformationSystems/LIDC>.
- [6] Goo, JM., T. Tongdee, R. Tongdee, Kwangjae Yeo, CF. Hildebolt, and KT. Bae. Volumetric Measurement of Synthetic Lung Nodules with Multi-Detector Row CT: Effect of Various Image Reconstruction Parameters and Segmentation Thresholds on Measurement Accuracy. *Radiology* 235 (2005): 850-856.
- [7] Shin JH, Lee HK, Choi CG, Suh DC, Lim TH, Kang W. The Quality of Recons. 3D Images in Multi-detector-Row Helical CT: Experim. Study Involving Scan Par., *Korean J Radiol* 3(1), 2002
- [8] Crowley RS, Naus GJ, Friedman CP. Development of Visual Diagnostic Expertise in Pathology. *JAMIA*, 2002.
- [9] Morita J, Miwa K, Kitasaka T, Mori K, Suenaga Y et al. Expertise in interactions of perceptual and conceptual processing. *Proceed. of the 27th Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Lawrence Erlbaum, 1541-1546 (2005).
- [10] Morita, J, Miwa K, Kitasaka T, Mori K, Suenaga Y et al. Interact. of Percept. and Concep. Proc.: Expertise in Med. Image Diagnosis. *Int J. Human-Computer Studies* 66 (2008): 370-390.
- [11] Shyu CR, Pavlopoulou C, Avinash CK, Brodley CE, Broderick LS. Using Human Perceptual Categories for CBIR from a Medical Image Database. *CV IU Journal*, 88 (2002): 119-151.
- [12] Yu S, Wantroba J, Raicu DS, Furst JD, Channin DS, Armato III SG. A study on the effect of CT imaging acquisition parameters on lung nodule image interpretation. *SPIE Medical Imaging* 2009.
- [13] Cohen BH, Lea B. *Essentials of Statistics for the Social and Behavioral Sciences*. John Wiley and Sons, 2004.

