Analysis of the statistical significance of 3D texture features in MRI images toward the detection of Tourette’s Syndrome

Murilo Costa de Barros *, Kaue TN Duarte†, Wang-Tso Lee‡, Chia-Jui Hsu§, Marco Antonio G de Carvalho*, murilo.barros.sn@gmail.com, kaue.duarte@ucalgary.ca, leeped@hotmail.com, jry730701@gmail.com, magic@unicamp.br

*Computing Visual Laboratory, School of Technology UNICAMP, University of Campinas, Limeira, R. Paschoal Marmo, 1888, Brazil
†Vascular Imaging Laboratory, University of Calgary, Calgary, 2500 University Dr NW, Calgary, AB T2N 1N4
‡Department of Pediatrics, National Taiwan University Children’s Hospital, Taipei, Taiwan
§Department of Pediatrics, National Taiwan University Hospital Hsinchu Branch, Taipei, Taiwan

Abstract—Tourette Syndrome (TS) is a genetically induced disorder that is believed to be caused by morphological alterations in brain structure, resulting in involuntary movements known as tics. The current clinical standard for diagnosing TS is by clinical assessments performed by physicians. Mild stages of TS, however, commonly go underdiagnosed as tics are infrequent or often suppressed. Brain imaging has been suggested to be a reliable tool to detect brain alterations and possible biomarkers that correspond to neurological disorders. In Magnetic Resonance Imaging (MRI), anatomical brain changes can be identified in the scan by variation in texture patterns of certain regions. The main goal of this work is to identify the statistical significance of texture features in specific brain regions to distinguish TS from control subjects. The proposed approach consists of four main steps: (i) image acquisition, where the data is also organized using demographic information; (ii) brain segmentation, where the structural MRI is parcellated into anatomical regions; (iii) registration, where functional MRI is aligned to structural MRI; (iv) obtaining texture features and statistical analysis, where texture features are extracted from the anatomical brain regions. We adopted 68 subjects aged between 6 to 14 years, divided equally into TS and Normal Control groups. We evaluated the texture features in a statistical manner, where our main findings are: (i) After False Discovery Rate (FDR) correction, only one texture feature was significant ($p$-value < 0.08) in structural MRI; (ii) Following FDR correction, eight texture features in functional MRI for three anatomical regions were considered significant; (iii) The right amygdala presented significance in distinct texture features, matching its importance in the literature. Texture features aligned with the literature can serve as a reliable tool to identify imaging changes, which can lead to future work applied to clinical studies.

I. INTRODUCTION

Tourette’s Syndrome (TS), also known as tic syndrome, is a hereditary disorder that comprises neurophysiological and neuroanatomical changes [1]. TS patients present repetitive and involuntary movements called tics. Motor and vocal actions are the most common types of tics and can vary in intensity across patients [2]. Tourette Syndrome is primarily developed between early childhood (age 2) to late adolescence (age 18) and is sparsely, if ever, diagnosed in adulthood.

The diagnosis of TS is challenging due to slight similarities with other syndromes, such as Attention Deficit Hyperactivity Disorder, Obsessive-Compulsive Disorder, and other possible comorbidities such as depression, sleep disorder, and anxiety [3].

According to the review detailed by Robertson et al. [2], there is no precise and direct method to diagnose TS other than clinical assessment. In addition, brain imaging has been extensively adopted in the medical field to identify TS and other pathologies [4].

The use of image processing, in specific texture description, allows the detection of patterns in the image, undetected by human sight. Magnetic Resonance Imaging (MRI) is by far the most common imaging in TS studies, but the importance of Positron Emission Tomography (PET) is also highlighted in the literature [5].

The subtle variation in brain images, either in shape or texture, represents distinct neurophysiological stimulation, which differs among brain regions. In fact, the stimulus varies according to the intensity of the TS. Studies in the literature often target regions such as the thalamus, basal ganglia, prefrontal and cingulate cortex [6], [7]. The main goal of this work is to evaluate the statistical significance of the texture features in anatomical brain regions to distinguish TS from control subjects. Our major contributions are: (i) provide an indication of impacted brain regions in subjects with TS; (ii) describe a texture-based workflow to discriminate significant regions to identify TS; (iii) present a statistical inference of radiomics features applied to TS subjects;

The remainder of this work is organized in the following sections: Section II presents the most relevant related work; The materials and methods are shown in Section III. The results and evaluations are presented in Section IV. The results are discussed in Section V. Section VI presents a summary and final considerations of this work.
II. RELATED WORK

The literature review performed in this work relied on related work that presented a computational approach toward TS detection. In addition, related works that presented statistical evaluation or similar approaches applied to other diseases or disorders were also considered for review. Greene et al. [4] proposed a multivariate approach to classify TS patients using the Support Vector Machine (SVM). Functional Magnetic Resonance Imaging (fMRI) scans of 84 children, distributed equally amongst TS and Normal Control (NC) group, are used to calculate correlation values in resting-state. The fMRI is pre-processed with inter-slice interpolation and intensity normalization, followed by registration to the TRIO_KY_NDC atlas. The registration process was performed using the bootstrap method, scaling the images in histograms and calculating the voxelwise average to align the images to the model atlas. After registration, 264 regions of interest (ROIs) were extracted, and a correlation matrix was computed using Pearson’s method, which is further used as input data for the SVM classification. The classification method demonstrated significant results with values of up to 70% accuracy.

A statistical analysis method, such as Spearman’s correlation, has been used in the work of Maneeka et al. [8]. The authors aimed to classify stereotypic motor corrections or Stereotyped Movement Disorder in adults with TS. The method analyzes 148 adult patients with a mean age of 30 years. The severity scored using the Yale Global Tic Severity Scale and the Stereotypic Severity Scale are obtained. The authors used Spearman’s correlation to analyze the relationship between tic severities and t tests to differentiate between analytic and control groups. They identified that there is a higher propensity of patients with Tourette Syndrome in males, as presented by [8] reaching 68.9%. In relation to the demographic data, no significant differences between patients were presented.

However, although texture features have been recently introduced to TS analysis, this topic has been extensively applied in dementia or other disorders. For instance, Shui et al. [9] adopted the use of texture descriptor towards Alzheimer’s disease (AD) to distinguish them from Normal Control patients. Each patient’s brain was parcellated into 106 regions, and texture features such as Gray-Level Size Zone Matrix and Gray level Difference Method were extracted. The authors reduced the dimensions of the features extracted with the K-best algorithm, followed by feature selection using the LASSO algorithm. Statistical analysis was performed using the Wilcoxon method. They concluded that the most significant regions were the hippocampus, the inferior parietal lobe, the precuneus, and the lateral occipital gyrus.

Addressing Parkinson’s disease, Tomer et al. [10] compared types of feature extraction techniques. The MRI scans were obtained from the Parkinson Progression Markers Initiative dataset with ages between 55 and 65 years. The scans were skull stripped, denoised, and normalized. The structures were parcellated by selecting only gray matter regions. Next, the feature extraction step is performed to identify the patterns in the image. The features are combined to perform a classification step via Random Forest (RF). The authors stated that Principal Component Analysis (PCA) requires high computational processing while Gray-Level Cooccurrence Matrix presents better results, reaching 90.5% compared to PCA, which reached 87.5%. Detecting brain tumors is also a field of interest for applied texture description. In Sadique et al. [11], it proposed an approach to differentiate recurrent brain tumors and radiation necrosis using texture extraction from MRI scans. Features such as structural heterogeneity, texture, intensity-based information, and histograms were used in their analysis. They grouped the features into matrices to perform a classification by means of RF. The results have shown better discrimination of the tumors using texture features.

Our work differ from the related work on the following aspects:

- (i) provides a statistical texture feature evaluation for anatomical regions;
- (ii) evaluates the statistical significance of the brain regions identified according to texture descriptors;

III. MATERIALS AND METHOD

The proposed workflow consists of four steps:

- (i) Image acquisition, which consists of obtaining MRI images, in the selection and demographic evaluation of the subjects;
- (ii) Volume segmentation, which parcellates the brain into anatomical regions;
- (iii) Image registration, align the sMRI and fMRI; and
- (iv) Feature extraction, which extracts texture features from anatomical regions.

The proposed workflow is illustrated in Figure 1.

![Fig. 1. Proposed workflow](image)

A. Image acquisition

A total of sixty-eight participants were enrolled for this work, thirty-four TS patients (TS group) and thirty-four presumed Normal Control (NC group), considering equal age- and gender-matched information. Their age ranged from 6 to 14 years old (Table I). All participants had no comorbidity after clinical evaluation. The mean functional (fMRI) volume and structural (sMRI) MRI were provided for each participant. No accidental findings were observed in either scan for all participants. The images were obtained using a TrioTim Siemens scanner using the following parameters: (i) sMRI - TE=2.98ms, TR=2000ms, TI=900ms; (ii) fMRI - TE=26ms, TR=2000ms, from a dataset obtained from National Taiwan University.
TABLE I
DEMOGRAPHIC INFORMATION

<table>
<thead>
<tr>
<th>Attributes</th>
<th>NC (N=34)</th>
<th>TS (N=34)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>6 to 14</td>
<td>6 to 14</td>
</tr>
<tr>
<td>Mean (years)</td>
<td>8.94</td>
<td>8.58</td>
</tr>
<tr>
<td>Gender (% Male)</td>
<td>71%</td>
<td>68%</td>
</tr>
</tbody>
</table>

B. Volume segmentation

The MR scans were pre-processed for bias field correction, noise removal, normalization, and skull stripping [12]. Freesurfer was used to parcellate the brain into anatomical regions based on the DTK atlas [12]. In essence, the tasks performed by Freesurfer can be divided into three main steps:

- Preprocessing, where it performs image normalization, skull stripping, bias field correction, and noise removal;
- smoothing and inflation of the surface, which is used to identify patterns and processing according to predefined volumes;
- Registration and cortical parcellation, which divides the brain into anatomical parts.

We restricted our analysis to 86 anatomical regions belonging to cortical gray matter, white matter, and deep gray matter. The anatomical region IDs are standardized, enabling comparison across distinct subjects. Each anatomical region was converted into a binary volume, where the intensity corresponding to the ROI was set to 1, and 0 otherwise.

C. Image registration

Intra-subject modality registration was done to align the structural and functional MR volumes. We adopted the Entropy Correlation Coefficient (ECC) to align the fMRI to the sMRI. The formulas presented below are references to the work of Craddock et al. [13]. Firstly, the singular entropy is calculated according to Equation 1:

\[
H(x) = \sum_{i=1}^{n} np(x_i) \log p(x_i),
\]

where \( p(\cdot) \) represents probability distribution, \( n \) and \( m \) is the limit of unique values that \( x \) can take.

The shared entropy is calculated according to Equation 2:

\[
H(x, y) = - \sum_{j=1}^{n} \sum_{i=1}^{m} p(x_j, y_k) \log p(x_j, y_k).
\]

The mutual information metric is calculated to utilize the singular and shared entropy, as formulated in Equation 3:

\[
I(x; y) = H(x) + H(y) - H(x, y).
\]

Finally, the ECC is calculated using Equation 4:

\[
ECC = \sqrt{\frac{I(x; y)}{\frac{1}{2}(H(x) + H(y))}}.
\]

After the ECC calculation, the sMRI and fMRI modalities are aligned. The registration step was manually checked for consistency.

D. Feature extraction

The feature extraction step involves obtaining volume information to capture different intrinsic structures in the images. Our work focused on texture features, which are reliable mechanisms to identify patterns in the image. We adopted Radiomics descriptors [14], which are widespread in the medicine field. Four types of texture features were used over the volume obtained in step B:

- **Gray-Level Co-Occurrence Matrix (GLCM)**, which calculates the co-occurrence of pixels in the image;
- **Gray Level Run Length Matrix (GLRLM)**, which detects length occurrence of pixels with the same intensity in an image;
- **Gray-level Size Zone Matrix (GLSZM)**, which evaluates homogeneous zone distribution of pixels in an image;
- **Neighboring Gray Tone Difference Matrix (NGTDM)**, which detects neighborhood information in the gray tones in the image.

In total, 61 texture descriptors were evaluated:

- **GLCM (N=24)**: Autocorrelation, ClusterProminence, ClusterShade, ClusterTendency, Contrast, Correlation, DifferenceAverage, DifferenceEntropy, DifferenceVariance, Id, Idm, Idmn, Idn, Imc1 Imc2 InverseVariance, JointAverage, JointEnergy, JointEntropy, MCC MaximumProbability, SumAverage, SumEntropy, SumSquares;
- **GLRLM (N=16)**: GrayLevelNonUniformity, GrayLevelNonUniformityNormalized, GrayLevelVariance, HighGrayLevelRunEmphasis, LongRunEmphasis, LongRunHighGrayLevelEmphasis, LowGrayLevelRunEmphasis, RunEntropy, RunLengthNonUniformity, RunLengthNonUniformityNormalized, RunPercentage, RunVariance, ShortRunEmphasis, ShortRunHighGrayLevelEmphasis, ShortRunLowGrayLevelEmphasis;
- **GLSZM (N=16)**: GrayLevelNonUniformity, GrayLevelNonUniformityNormalized, GrayLevelVariance, HighGrayLevelZoneEmphasis, LargeAreaEmphasis, LargeAreaHighGrayLevelEmphasis, LowGrayLevelZoneEmphasis, SizeZoneNonUniformity, SizeZoneNonUniformityNormalized, SmallAreaEmphasis, SmallAreaHighGrayLevelEmphasis, SmallAreaLowGrayLevelEmphasis, ZoneEntropy, ZonePercentage, ZoneVariance;
- **NGTDM (N=5)**: Busyness, Coarseness, Complexity, Contrast, Strength.

E. Implementation

Our approach was coded using Python 3.8 with the Spyder IDE platform in a macOS High Sierra version 10.13.6 configuration. The following techniques were used in this implementation:

- Demographic data was processed using Pandas and Numpy for grouping;
• Brain segmentation is performed using the Freesurfer tool [12] by the command `recon – all`.
• The binary volumes are generated using logical operations in Numpy;
• Feature extraction was coded using pyRadiomics, Numpy and Pandas;
• Pearson correlation, p-value calculation, and FDR correction are performed using pyStats.

Detailed information about our method, including the subjects used and processed, the processing specifications, the implementation of the patient selection software, and our result evaluation algorithm are available on GitHub1.

IV. RESULTS

In this section, the 3D texture features calculated in the last step of our approach are evaluated according to their significance criteria. Note that the segmentation and registration steps were manually quality checked. Our approach obtained a total of 61 texture features from four texture descriptors (GLCM, GLRLM, GLSZM, NGTDM), where each texture feature is applied to all examined regions.

Evaluation. The Pearson correlation ($r$) is adopted to identify the linear association strength between two groups (TS v CN) in a given region and texture feature. The $r$ value is calculated according to Equation 5:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{X})(y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{X})^2 \sum_{i=1}^{n}(y_i - \bar{Y})^2}}.$$  \hspace{1cm} (5)

where, $\bar{X}$ and $\bar{Y}$ stand for the mean of the groups $X$ and $Y$, respectively. In the sequence, we calculate the $t$-score to evaluate the significance of the correlation, and furthermore, the two tail $p$-value to identify the significance of each relation according to the null hypothesis. The $t$-score is calculated according to the Equation 6:

$$t = r \sqrt{\frac{n-2}{1-r^2}},$$  \hspace{1cm} (6)

where $n$ represents the sample size (i.e. 68). The $p$-value is calculated for a $t$-distribution with $n - 2$ degrees of freedom.

A $N \times M$ matrix is obtained after this significance test, where $N$ corresponds to the anatomical brain region ID, $M$ represents the texture features, and each cell is the $p$-value between the region $N_i$ to the texture feature $M_j$. Figure 2 illustrates the matrices from this step.

We adopted the False Discovery Rate (FDR) to identify if the statistical inferences were significant under a multiple hypothesis test perspective. It is often desired that multiple $p$-values calculated to the same sample groups have an additional correction step to account for variation in the simultaneous test. This practice avoids biases and reaffirms a possible discovery. Figure 3 shows the $p$-values uncorrected. Next, the FDR correction is applied to account for multi-test analysis, as shown in Figure 4.

The $p$-values less than 0.05 state that the null hypothesis can be rejected, since there is enough evidence that the hypothesis is true.

V. DISCUSSION

The $p$-values played a key role in understanding the significance of each texture feature associated with a brain region. This analysis is similar to Greene et al. [4], where a matrix is thresholded to extract only the significant values. Three major differences are notorious between their and our work:

• (i) they adopted RSFC matrices, which is represented as adjacency matrices, and
• (ii) no significance was found in their work after applying FDR correction, which differs from ours.
• (iii) a multivariate approach was adopted using a SVM to elucidate patterns in the brain.

However, the SVM can hide the features used to perform a prediction, which was avoided in this work since the features were analyzed independently.

In fMRI, eight descriptors were considered relevant for distinguishing TS patients (Figure 4.(a)). We grouped the texture descriptors by the anatomical regions, which enables extracting knowledge of that specific region.

$4^{th}$ ventricle. The Gray level variance (GLSZM) of the gray-intensity zones showed significant values. This value was

---

1https://github.com/muribarros/TS_Feature_Extractor

---

Fig. 2. Correlation matrix after significance tests and descriptors of textures and anatomical regions. (a) fMRI modality matrix; (b) sMRI modality matrix.

Fig. 3. Uncorrected correlation matrices to p-value < 0.05. (a) Analysis in fMRI; (b) Analysis in sMRI.
Right Amygdala. The mean inverse variance (GLCM) was higher in subjects with TS, and presented less coarse textures, since the long run emphasis (GLRLM) was smaller in patients with TS. In addition, the right amygdala in TS patients showed less variance in both per run lengths (GLRLM) and per zones (GLSZM), thus meaning that lower variance within the right amygdala is expected in TS subjects.

Right middle-temporal cortex. This region presented higher local homogeneity within the region in the TS subjects (measured by IDMN-GLCM), where the low gray-level emphasis values both in a small area (GLSZM) and in joint distribution for shorter run-lengths (GLRLM) were lower in TS subjects compared to NC subjects.

The Zone Entropy, a GLSZM-based feature, measures the randomness of the distribution in zone size. This feature was considered significant for distinguishing TS and Normal Control in sMRI. In essence, patients with TS showed higher Zone Entropy values, which indicates that the left medial orbitofrontal cortex has more heterogeneity in the texture patterns. These results highlight the importance of fMRI in TS analysis.

The biology of these regions has been addressed accordingly in the literature. The right amygdala, located in the limbic system, is responsible for regulating emotional aspects, thus being suggested as directly related to tics intensity in TS subjects [15].

The presence of the influence of the limbic system was also documented in the results of Barros et al [16].

VI. CONCLUSION

In this work, we propose to study the influence of texture features in brain regions for TS detection. Freesurfer carries out the segmentation of the brain regions. Three-dimensional texture features were extracted for each anatomical region. Later, the Pearson correlation was adopted to identify the similarity within a certain region between patients with and without TS. p-values were calculated to determine the significance of the relationship between a brain region and texture descriptors.

As suggested in the literature, fMRI was more significant for TS patients, although sMRI had one region with corrected p-value <0.08. The right amygdala had four texture descriptors that showed significance in the analysis. In essence, the right amygdala is responsible for regulating emotional aspects and has been indicated as correlated to TS subjects. In addition, the right middle-temporal cortex presented three texture descriptors with high significance, which surprisingly was found in the literature review as this region is commonly associated with learning, memory, and visual information.

As future works, the 4-dimensional information (time-series) of fMRI using texture descriptors needs to be explored to capture signal variation across time. In addition, the adoption of classification tasks (present in machine learning approaches) using texture features can be carried out to enlighten if the significant regions found in this work can be combined to perform predictions.

ACKNOWLEDGEMENT

The authors thank the group Department of Pediatric Neurology from National Taiwan University, located in Taipei-Taiwan, for providing T1-weighted MR scans used in this work. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001.

REFERENCES


