

# Misure statistiche di texture analysis su immagini di risonanza magnetica

Miriam SANTORO

29 aprile 2019

## Indice

<b>1 Scopo</b>	<b>2</b>
<b>2 Introduzione</b>	<b>2</b>
2.1 QSM 2016 Reconstruction Challenge . . . . .	2
2.2 Texture analysis . . . . .	2
<b>3 Dati sperimentali</b>	<b>3</b>
<b>4 Immagini 2D</b>	<b>4</b>
4.1 Teoria . . . . .	5
4.1.1 ImageJ . . . . .	5
4.1.2 MATLAB . . . . .	5
4.2 Risultati . . . . .	8
4.2.1 ImageJ . . . . .	10
4.2.2 MATLAB . . . . .	11
<b>5 Immagini 3D</b>	<b>13</b>
5.1 Teoria . . . . .	13
5.2 Risultati . . . . .	21
5.3 Confronto tra le ROI . . . . .	26
<b>6 Valutazione dei risultati</b>	<b>35</b>
6.1 PCA . . . . .	37
<b>7 Conclusioni e lavori futuri</b>	<b>41</b>
<b>A Appendice</b>	<b>42</b>
<b>B Appendice</b>	<b>44</b>
<b>C Appendice</b>	<b>46</b>
<b>D Appendice</b>	<b>47</b>
<b>E Appendice: PCA per diversi ordini di grandezza</b>	<b>75</b>
<b>F Appendice: PCA features migliori in comune</b>	<b>83</b>

# 1 Scopo

Lo scopo di questo progetto è estrarre misure statistiche di texture analysis da immagini di risonanza magnetica della QSM 2016 challenge. Queste misure statistiche sono state realizzate sia su immagini 2D (in particolare su ROI definite al loro interno e su riferimenti biologici) che su immagini 3D (in particolare su VOI).

## 2 Introduzione

### 2.1 QSM 2016 Reconstruction Challenge

La QSM 2016 Reconstruction Challenge è una sfida realizzata con lo scopo di testare l'abilità di vari algoritmi QSM per recuperare la suscettività da dati di fase in maniera veritiera e realizzare una mappa di suscettività magnetica di riferimento. A questo fine sono state acquisite immagini gradient-echo di una volontaria sana a 3T con una risoluzione isotropica di 1.06 mm per 12 orientazioni della testa.

Durante un'acquisizione RM vengono misurate sia la componente di magnitudo che quella di fase; quindi, entrambe sono in grado di fornire un'immagine. Nella pratica clinica, però, le immagini di fase non sono di uso comune anche se portano con sè tutte le informazioni inerenti alle differenze di suscettività magnetica tra i tessuti.

L'estrazione di mappe di suscettività viene effettuata a partire dal segnale e dalla ricostruzione delle immagini di fase. Queste sono elaborate in modo da isolare i contributi di variazioni di fase riconducibili alle proprietà magnetiche dei tessuti rispetto ai contributi di variazioni di fase su larga scala dovuti alle disomogeneità del campo elettromagnetico statico.

### 2.2 Texture analysis

La texture è una proprietà che rappresenta la superficie o la struttura di un'immagine. In generale può essere definita come una ripetizione regolare di un elemento o di un pattern su una superficie. A questo proposito, una regione in un'immagine ha una texture costante se un set delle sue caratteristiche è costante, cambia lentamente o è approssimativamente periodico. [1]

La texture può essere intesa come feature ed essere usata per dividere immagini in regioni di interesse (roi) e classificarle, in quanto dà informazioni sulla disposizione spaziale delle intensità di un'immagine che può essere relativa ai singoli pixel o alla loro relazione con pixel vicini. Infatti essa descrive una varietà di tecniche di analisi dell'immagine che quantificano le variazioni in intensità della superficie o di pattern, inclusi alcuni che sono impercettibili al sistema visivo umano. Per questo motivo può essere ben adattata alla segmentazione e alla caratterizzazione delle lesioni per monitorare una malattia o un recupero. [2]

La texture analysis si riferisce a una classe di procedure matematiche e modelli che caratterizzano le variazioni spaziali come mezzo per estrarre informazioni quali quantificazione dei pattern di livelli di grigio, interrelazioni di pixel e proprietà spettrali. [2]

I due scopi primari nella texture analysis sono:

- texture classification = identificazione di una data regione di texture da un set di classi di texture in cui ognuna regione ha caratteristiche uniche;
- texture segmentation = determinazione automatica dei confini tra le varie regioni di texture in un'immagine.

Le features di texture cadono sotto 3 categorie generali: sintattiche, statistiche e spettrali. In questo caso si è interessati ai metodi statistici analizzano la distribuzione spaziale di valori di grigio, calcolando features locali in ogni punto nell'immagine, e derivando un set di statistiche dalle distribuzioni delle features locali. Dipendendo dai numeri di pixels che definiscono la feature locale, i metodi statistici possono essere classificati in statistiche di:

- primo ordine (un pixel);
- secondo ordine (due pixel);
- ordine maggiore (tre o più pixel).

La differenza base sta nel fatto che le statistiche di primo ordine stimano proprietà di valore di pixel individuali, ignorando l'interazione spaziale tra pixels all'interno dell'immagine, invece quelle di secondo ordine e ordine maggiore stimano proprietà di due o più valori di pixel che si verificano in posizioni specifiche. [1] In questo progetto le statistiche di primo ordine sono basate sull'istogramma che rappresenta la distribuzione delle intensità dei livelli di grigio senza tener conto della localizzazione spaziale.

### 3 Dati sperimentali

Sono state analizzate le immagini di fase presenti nel dataset dalla QSM challenge 2016. Nello specifico sono state realizzate misurazioni fatte sia su immagini 2D che su volumi 3D.

I dati, contenuti in un singolo array 4D, sono stati divisi in volumi 3D, corrispondenti alle acquisizioni in 12 orientazioni differenti. A livello di pre-processing, queste immagini sono affini-registrate, 'unwrapped', mascherate (cioè su di esse è applicata una maschera ottenuta dall'FSL Brain Extraction Tool per evitare effetti di confusione risultanti dall'uso di maschere differenti) e il loro campo di sfondo è rimosso.

Come è stato già detto, le immagini sono state acquisite in 12 orientazioni differenti di cui di seguito sono riportate le matrici di rotazione:

$$O_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad O_2 = \begin{pmatrix} 0.9489 & -0.1138 & -0.2942 \\ 0.1298 & 0.9909 & 0.0355 \\ 0.2875 & -0.0719 & 0.9551 \end{pmatrix}$$

$$O_3 = \begin{pmatrix} 0.8717 & -0.2969 & -0.3898 \\ 0.3559 & 0.9305 & 0.0870 \\ 0.3369 & -0.2145 & 0.9168 \end{pmatrix} \quad O_4 = \begin{pmatrix} 0.9423 & 0.2099 & 0.2609 \\ -0.2316 & 0.9712 & 0.0554 \\ -0.2418 & -0.1126 & 0.9638 \end{pmatrix}$$

$$O_5 = \begin{pmatrix} 0.7998 & 0.4379 & 0.4106 \\ -0.4751 & 0.8798 & -0.0127 \\ -0.3668 & -0.1849 & 0.9117 \end{pmatrix} \quad O_6 = \begin{pmatrix} 0.9960 & 0.0862 & -0.0237 \\ -0.0886 & 0.9868 & -0.1353 \\ 0.0117 & 0.1368 & 0.9905 \end{pmatrix}$$

$$O_7 = \begin{pmatrix} 0.9468 & -0.2529 & -0.1988 \\ 0.2366 & 0.9662 & -0.1025 \\ 0.2180 & 0.0500 & 0.9747 \end{pmatrix} \quad O_8 = \begin{pmatrix} 0.8982 & 0.3665 & 0.2426 \\ -0.3541 & 0.9304 & -0.0944 \\ -0.2603 & -0.0012 & 0.9655 \end{pmatrix}$$

$$O_9 = \begin{pmatrix} 0.9996 & -0.0125 & 0.0246 \\ 0.0062 & 0.9709 & 0.2394 \\ -0.0269 & -0.2392 & 0.9706 \end{pmatrix} \quad O_{10} = \begin{pmatrix} 0.9994 & 0.0356 & 0.0028 \\ -0.0343 & 0.9350 & 0.3531 \\ 0.0100 & -0.3529 & 0.9356 \end{pmatrix}$$

$$O_{11} = \begin{pmatrix} 0.9588 & -0.2268 & -0.1709 \\ 0.2709 & 0.9108 & 0.3115 \\ 0.0850 & -0.3450 & 0.9348 \end{pmatrix} \quad O_{12} = \begin{pmatrix} 0.8616 & 0.3923 & 0.3221 \\ -0.4830 & 0.8289 & 0.2824 \\ -0.1562 & -0.3988 & 0.9036 \end{pmatrix}$$

## 4 Immagini 2D

In un primo momento, le misure statistiche sono state eseguite su immagini 2D.

Nello specifico, si sono analizzate due immagini, corrispondenti alla slice con  $z=66$  e  $z=76$  e, per ciascuna di esse, è stata definita una roi  $65 \times 65$  pixel:

- roi 66 = (51:115, 61:125) per l'immagine corrispondente alla slice con  $z=66$ ;
- roi 76 = (51:115, 41:105) per l'immagine corrispondente alla slice con  $z=76$ .

Sono state scelte queste slice perchè sono quelle in cui sono più visibili zone di interesse clinico (riferimenti biologici). Queste sono:

- roi 66 = Red nucleus (rn) e substantia nigra (sn);
- roi 76 = Putamen (pu), caudate (cau), globus pallidus (gp) e white matter (wm).

In realtà per la white matter è stata selezionata una regione della capsula interna.  
Le regioni appena menzionate, sono indicate nella figura 1.

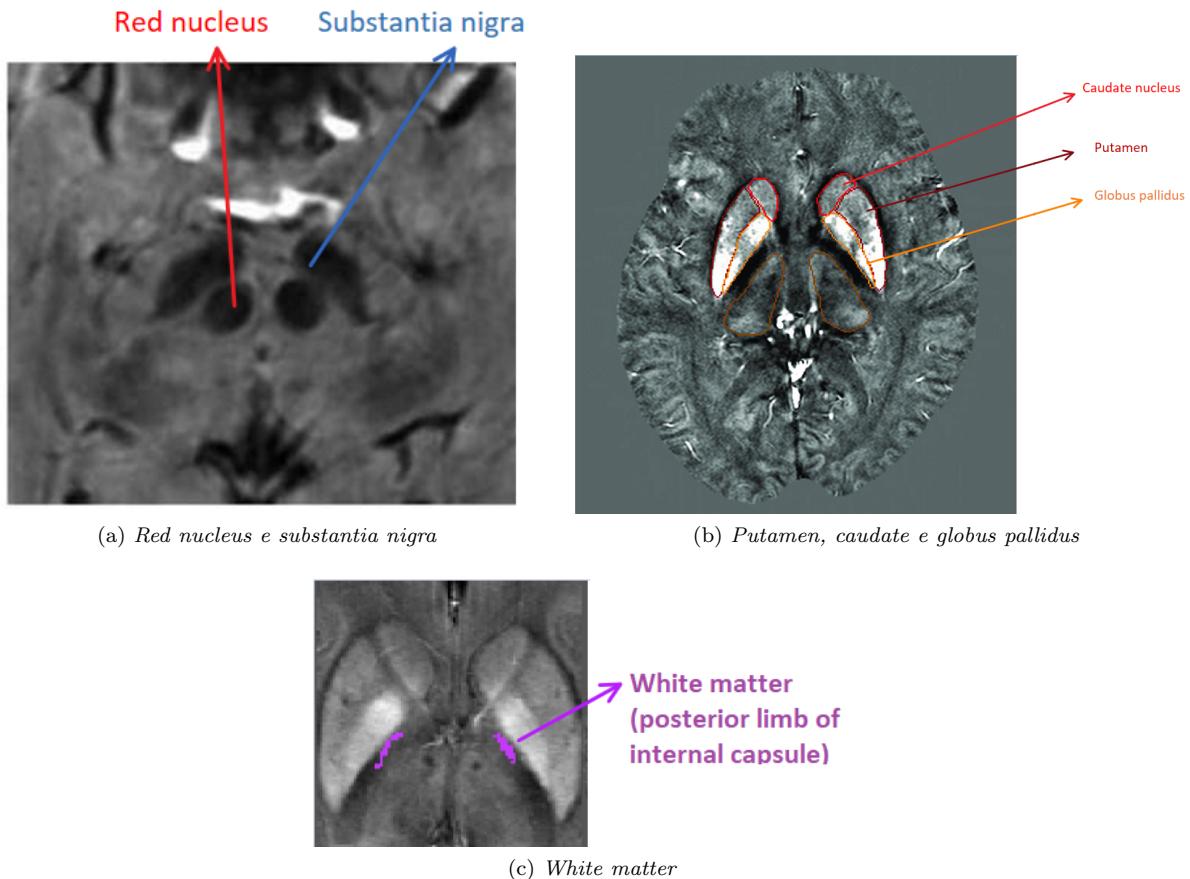


Figura 1: Riferimenti biologici nelle roi 66 e 76 delle slice con  $z=66$  e  $z=76$

## 4.1 Teoria

Prima di analizzare le immagini è stato realizzato lo stretching manuale di ogni istogramma attraverso il tool '*Adjust Contrast*' nell'applicazione '*Image Viewer*' di MATLAB, in quanto lo stretching 'automatico' realizzabile tramite le apposite funzioni *stretchlim* e *imadjust* non avrebbe fornito un'immagine con un contrasto ottimale.

### 4.1.1 ImageJ

Per analizzare le regioni dei riferimenti biologici nelle rispettive roi si è utilizzato il programma ImageJ, che permette di selezionare manualmente l'area di interesse e ottenere da essa delle misure statistiche basate sul corrispondente istogramma.

Sapendo che l'istogramma di un'immagine è formato dall'insieme di  $h$ , vettori colonna aventi ogni entrata indicizzata dai valori di livello di grigio e i cui valori sono i numeri di pixel nella regione di interesse con quel livello di grigio  $i$  e che  $N$  è il numero totale di pixel, si sono ricavate le seguenti misure:

- **Media** = valore medio in scala di grigi, definito da:

$$\mu = \sum_{i=1}^N i \cdot h_i \quad (1)$$

- **Deviazione standard** = deviazione standard dei valori in scala di grigio usati per generare il valore medio, definita da:

$$\sigma = \sqrt{\sum_{i=1}^N (i - \mu)^2 \cdot h_i} \quad (2)$$

Grazie a questi valori, per i riferimenti biologici nella slice con  $z=76$  è stato possibile calcolare il contrasto  $C$ , una misura definita da:

$$C = \frac{I_2 - I_1}{I_1} \quad (3)$$

dove  $I_1$  è la media del 'background' che, in questo caso, è rappresentato dalla white matter o dal putamen, mentre  $I_2$  è la media dell'altro riferimento biologico preso in considerazione.

Per l'errore sulla media è stata considerata la deviazione standard, invece per l'errore sul contrasto si è considerata la seguente formula di propagazione degli errori:

$$\Delta C = \sqrt{\frac{I_2^2}{I_1^4} \cdot \sigma_{I_1}^2 + \frac{1}{I_1^2} \cdot \sigma_{I_2}^2} \quad (4)$$

### 4.1.2 MATLAB

Ogni immagine può essere interpretata come una matrice di pixel, ciascuno avente un determinato valore di intensità in scala di grigi. Questo permette di fare image processing in MATLAB, linguaggio di programmazione ottimizzato per il calcolo vettoriale e matriciale.

A differenza di quanto descritto nella sezione 4.1.1, tramite MATLAB sono state implementate misurazioni eseguite sulle ROI  $65 \times 65$  pixel (roi 66 e roi 76) all'interno delle immagini corrispondenti.

Le misure sono realizzate in MATLAB tramite l'implementazione delle seguenti funzioni:

- **mean2** = calcola la media dei pixel nell'immagine. E' definita come nell'equazione 1.
- **std2** = calcola la deviazione standard dei valori di intensità dei pixel. E' definita come nell'equazione 2.

- **skewness** = calcola la skewness dei valori di intensità dei pixel tramite la formula:

$$s = \frac{1}{\sigma^3} \sum_{i=1}^N [(i - \mu)^3 \cdot h_i]^3 \quad (5)$$

La skewness è una misura dell'asimmetria dei dati rispetto alla media; quindi se è:

- negativa = i dati sono distribuiti più a sinistra che a destra;
- positiva = i dati sono distribuiti più a destra che a sinistra;
- 0 = la distribuzione è simmetrica.

- **kurtosis** = calcola la kurtosis dei valori di intensità dei pixel tramite la formula:

$$k = \frac{1}{\sigma^4} \sum_{i=1}^N [(i - \mu) \cdot h_i]^4 - 3 \quad (6)$$

La kurtosis è una misura che indica quanto i dati siano piccati o piatti rispetto a una distribuzione normale. Quindi, quando è:

- 3 = si ha una distribuzione normale;
- >3 = la distribuzione ha un picco distinto vicino alla media, decade piuttosto velocemente e ha code pesanti;
- <3 = la distribuzione ha un picco piatto vicino alla media.

- **entropy** = calcola l'entropia dei valori di intensità dei pixel, tramite la formula:

$$E = - \sum_{i=1}^N [h_i \cdot \log_2(h_i)] \quad (7)$$

dove  $h_i$  contiene i conteggi relativi all'istogramma normalizzato calcolati tramite la funzione *imhist*. L'entropia è una misura di texture relativa al contenuto informativo (randomicità) di un'immagine e se è:

- alta = tutte le entrate in  $h_i$  hanno grandezza simile;
- bassa = tutte le entrate sono diverse.

In aggiunta a queste misure statistiche di prim'ordine, si sono realizzate misure statistiche di ordini maggiori. Nello specifico, per ciascuna immagine è stata calcolata la matrice di co-occorrenza (GLCM). Questa è una matrice di dipendenza spaziale dei livelli di grigio che caratterizza la texture di un'immagine calcolando quanto spesso paia di pixel con valori specifici e con una specifica relazione spaziale si trovano in un'immagine. Sia  $P$  la matrice GLCM di un'immagine  $I(x,y)$  con  $x = 1, \dots, M$  e  $y = 1, \dots, N$  e sia  $N_g$  numero di livelli di grigio quantizzati in  $I$ .  $P(i,j)$  rappresenta il numero di volte in cui i pixel con livello di grigio  $i$  sono confinati con i pixel con livello di grigio  $j$  e la GLCM calcolata ha dimensioni  $N_g \times N_g$ . L'entrata  $(i,j)$  della GLCM normalizzata è uguale a:

$$p(i,j) = \frac{P(i,j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j)} \quad (8)$$

Sono definite anche le seguenti quantità:

$$\mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{N_g} p(i,j) \quad \mu_j = \sum_{i=1}^{N_g} j \sum_{j=1}^{N_g} p(i,j) \quad (9)$$

$$\sigma_i = \sum_{i=1}^{N_g} (i - \mu_i)^2 \sum_{j=1}^{N_g} p(i, j) \quad \sigma_j = \sum_{i=1}^{N_g} (j - \mu_j)^2 \sum_{j=1}^{N_g} p(i, j) \quad (10)$$

dove  $\mu_i$  e  $\mu_j$  corrispondono alla media sulla riga  $i$  o colonna  $j$  e  $\sigma_i$  e  $\sigma_j$  corrispondono alla varianza sulla riga  $i$  e colonna  $j$ .

Dalla GLCM sono state estratte le seguenti misure:

- **contrast** = misura la quantità di cambiamenti locali in un'immagine, cioè produce la misura di un contrasto di intensità tra un pixel e il suo vicinato ed è definita da:

$$Contrast = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i - j)^2 p(i, j) \quad (11)$$

Il contrasto è:

- 0 per un'immagine costante;
- grande quando c'è un grande quantità di variazione e, in questo caso, si parla di feature grossolana;
- piccolo quando c'è una piccola quantità di variazione e, in questo caso, si parla di texture acuta.

- **correlation** = misura quanto correlato è un pixel al suo vicinato ed è definita da:

$$Correlation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{(i - \mu_i)(j - \mu_j)p(i, j)}{\sigma_i \sigma_j} \quad (12)$$

La correlazione è la misura delle dipendenze lineari dei toni di grigio nell'immagine (misura della linearità dell'immagine) ed è:

- -1 quando si ha una correlazione perfetta negativa;
- 1 quando si ha una correlazione perfetta positiva;
- alta se l'immagine contiene una quantità considerevole di strutture lineari.

- **energy** = chiamata anche uniformità o momento angolare secondo ed è definita da:

$$Energy = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [p(i, j)]^2 \quad (13)$$

L'energia è:

- 1 quando l'immagine è costante;
- più grande quanto più è omogenea l'immagine.

- **homogeneity** = misura la somiglianza dei pixel ed è definita da:

$$Homogeneity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{p(i, j)}{1 + |i - j|} \quad (14)$$

L'omogeneità è:

- 1 se la GLCM è diagonale;
- grande se le texture locali hanno solo cambiamenti minimi;
- piccolo quando  $P_d(i, j)$  tende ad essere clusterizzato attorno alla diagonale principale.

## 4.2 Risultati

Nelle figure 2 e 3 sono mostrate, per ogni orientazione e per le slice con  $z=66$  e con  $z=76$  rispettivamente, le immagini risultanti dopo lo stretching dell'istogramma. Tutte le misurazioni relative alle immagini bidimensionali sono state realizzate su di esse.

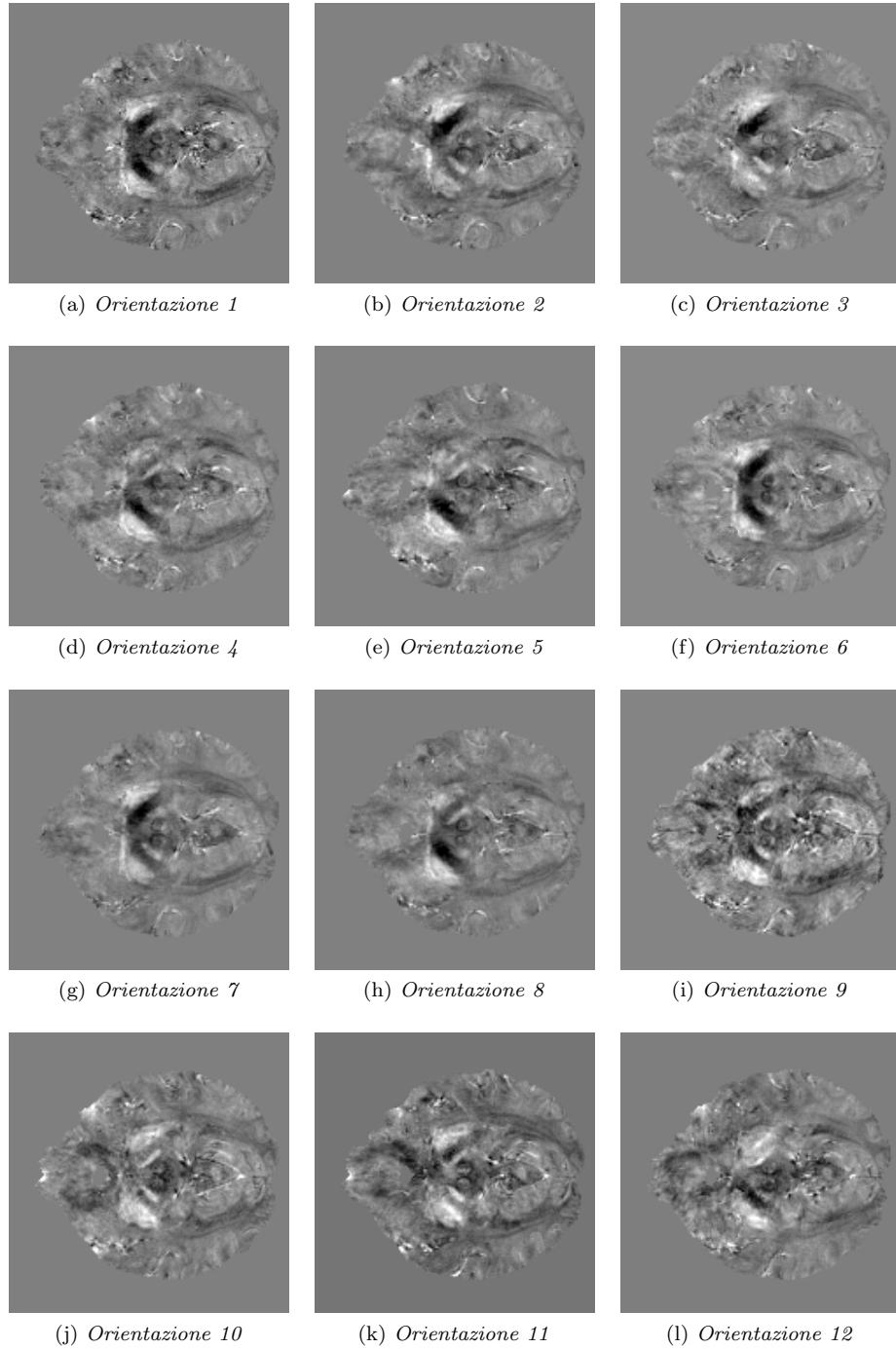


Figura 2: Immagini di fase con  $z=66$ , sottoposte a histogram stretching

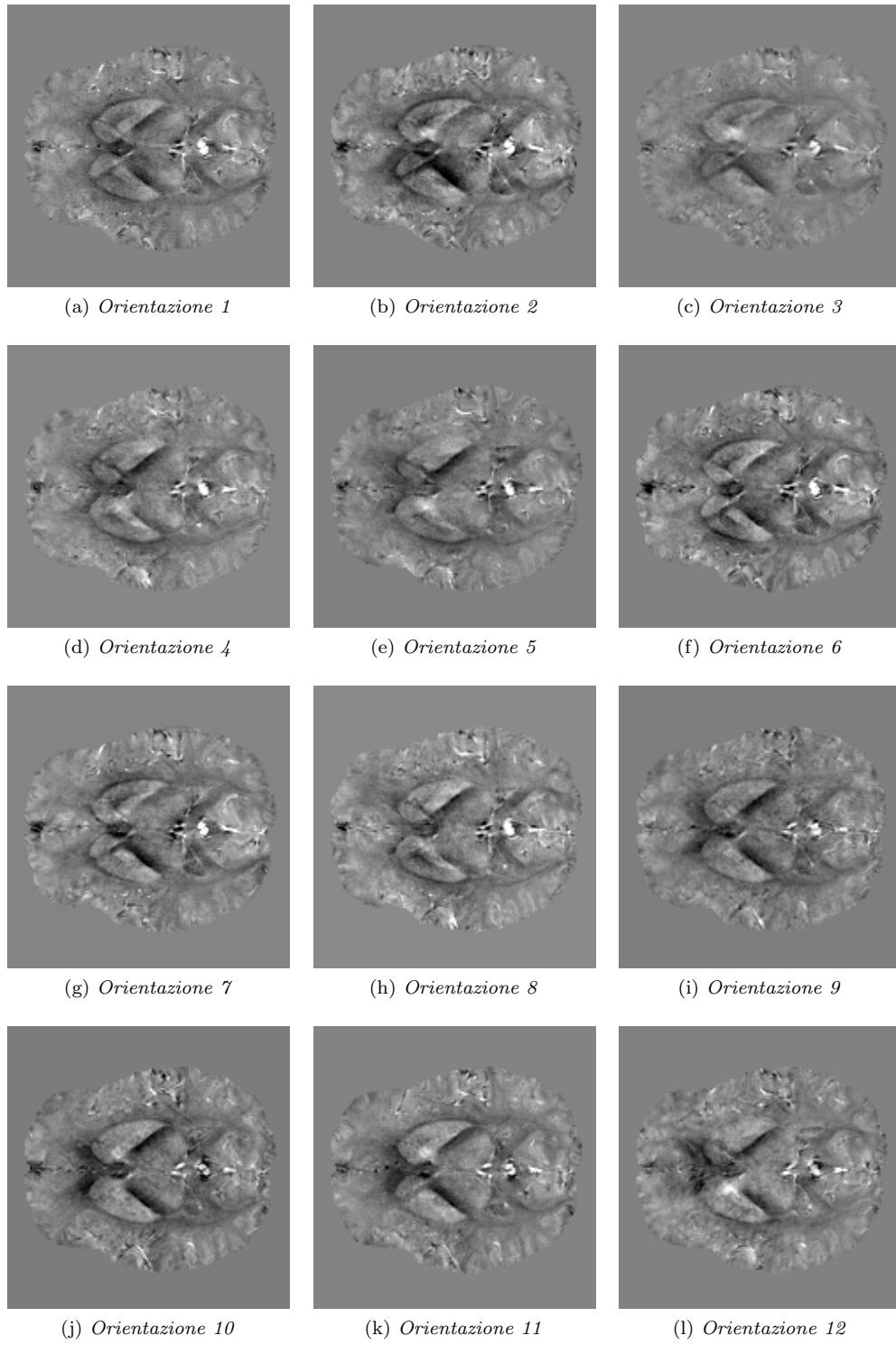


Figura 3: Immagini di fase con  $z=76$ , sottoposte a histogram stretching

#### 4.2.1 ImageJ

Nelle figure 4, 5 e 6 sono riportati i risultati ottenuti con ImageJ per le misure statistiche di primo ordine relative ai riferimenti biologici ottenuti selezionando le corrispondenti ROI sulle immagini 2 e 3, come già spiegato.

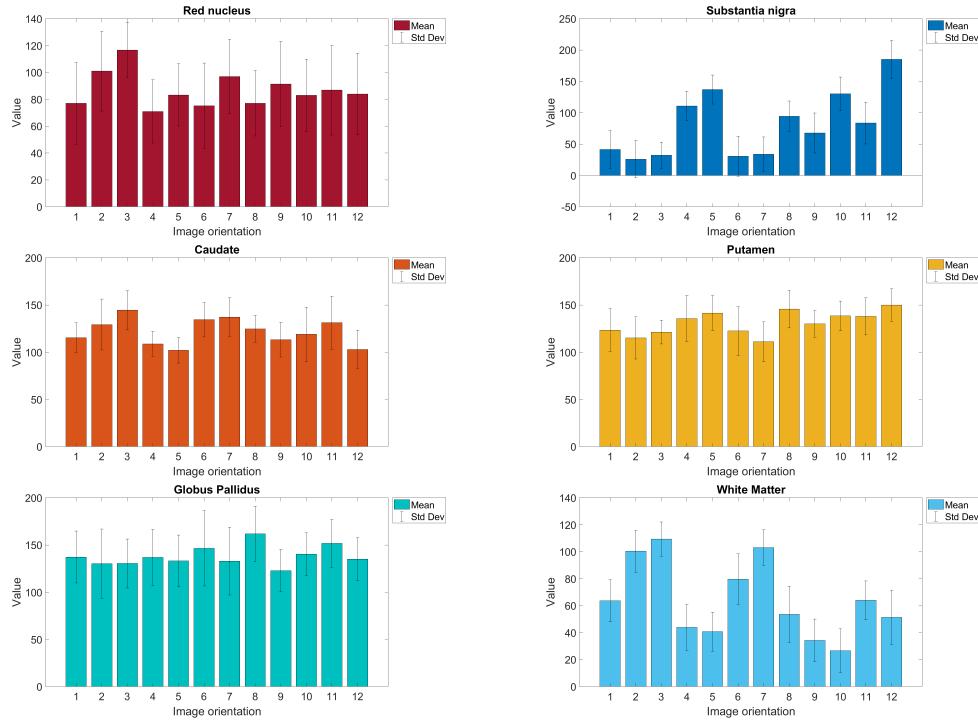


Figura 4: Media e deviazione standard dei riferimenti biologici

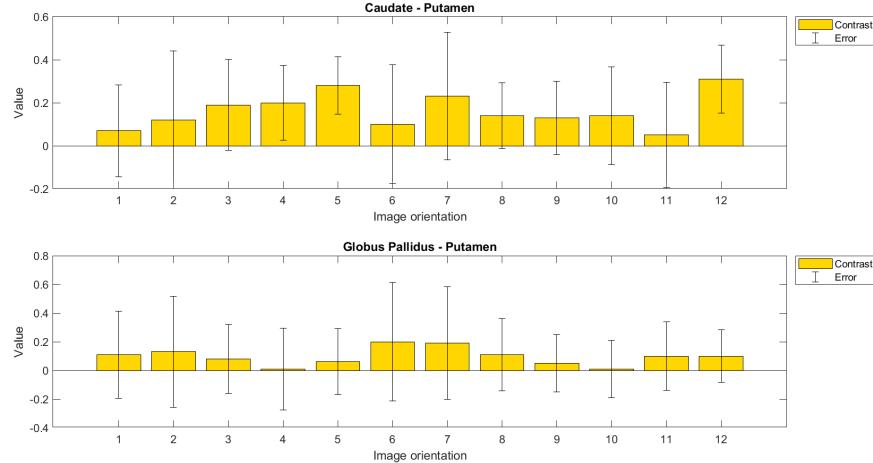


Figura 5: Contrasto dei riferimenti biologici rispetto al Putamen

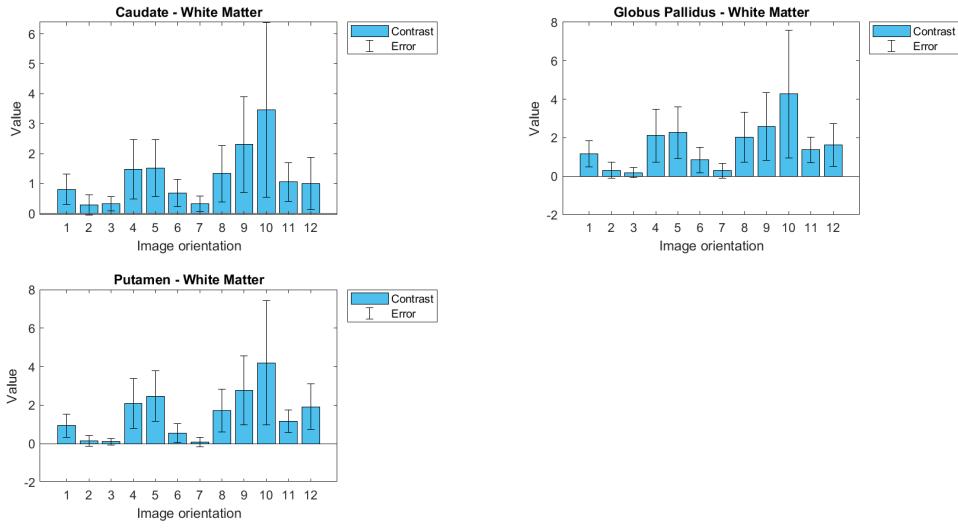


Figura 6: Contrasto dei riferimenti biologici rispetto alla White Matter

#### 4.2.2 MATLAB

In appendice A sono riportati gli histogrammi su cui è stato già fatto lo stretching relativi alle roi 66 e 76 (uno per ogni orientazione diversa).

In figura 7 e 8 sono mostrate le misure statistiche di prim'ordine ottenute dalle roi di interesse.

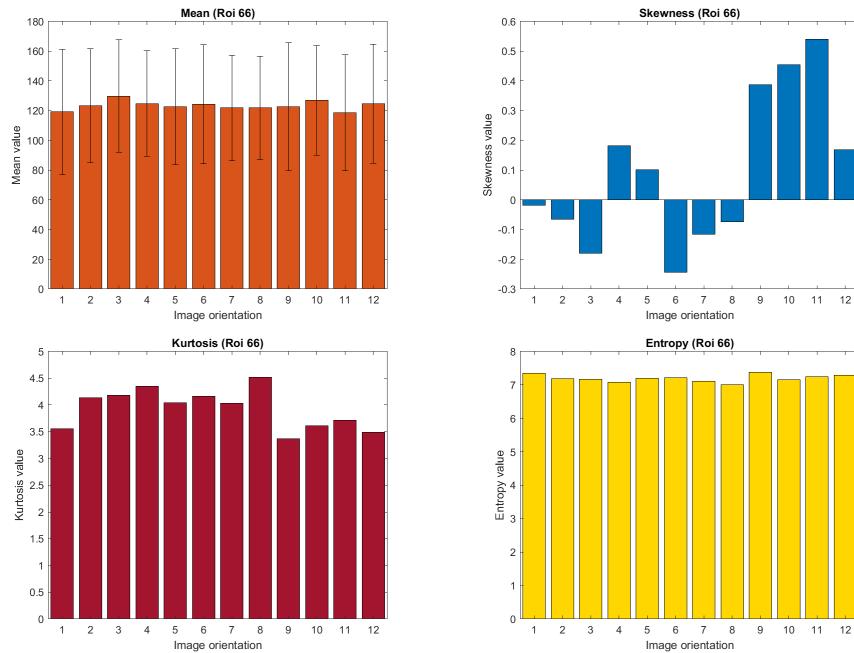


Figura 7: Misure statistiche di prim'ordine per la roi 66 (Immagine on z=66)

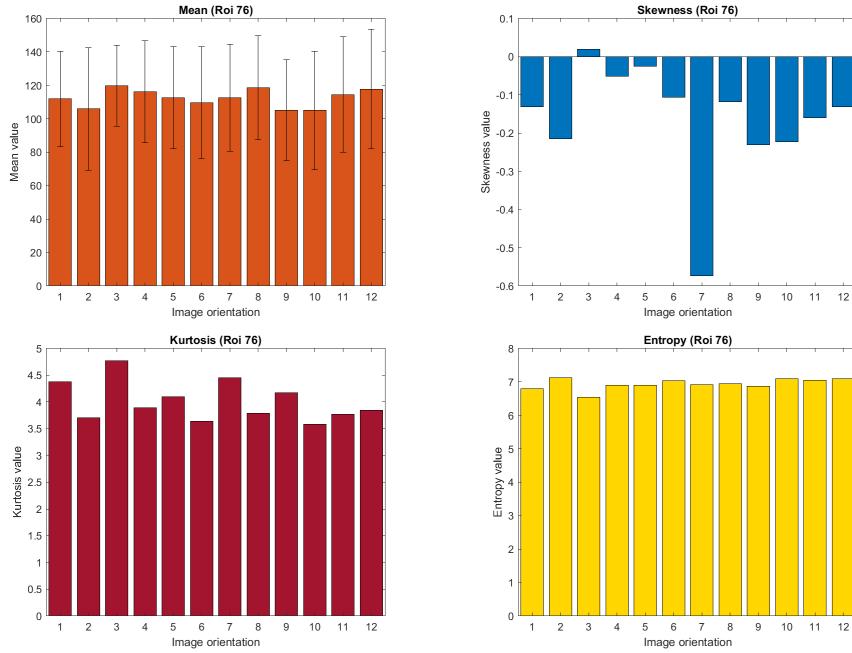


Figura 8: Misure statistiche di prim'ordine per la roi 76 (Immagine on z=76)

In appendice B sono riportate le matrici di co-occorrenza relative alle roi 66 e 76 presenti nelle immagini 2D con  $z=66$  e  $z=76$ .

Nelle figure 9 e 10 sono riportate le misure ottenute dalle rispettive matrici di co-occorrenza.

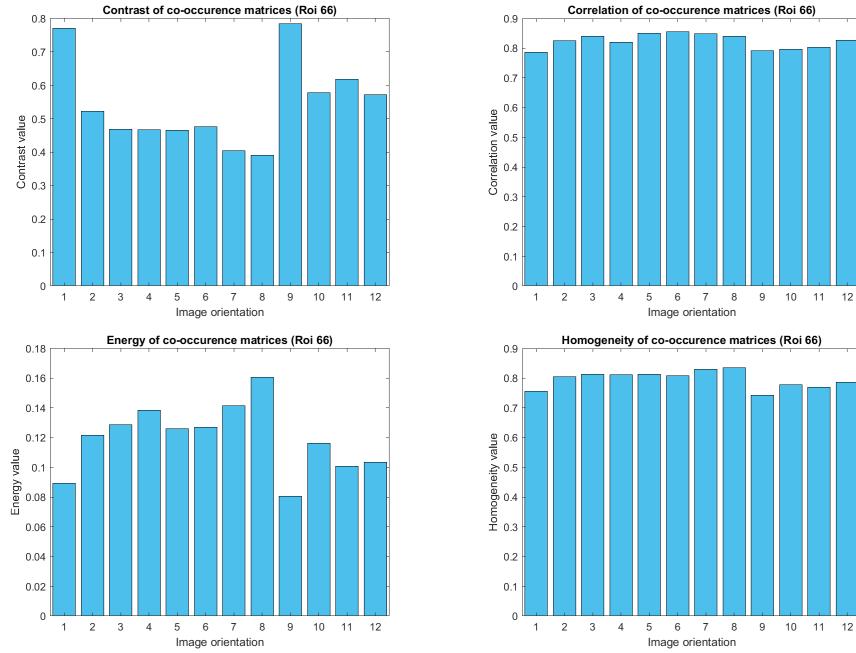


Figura 9: Misure statistiche della GLCM della roi 66 (Immagine con z=66)

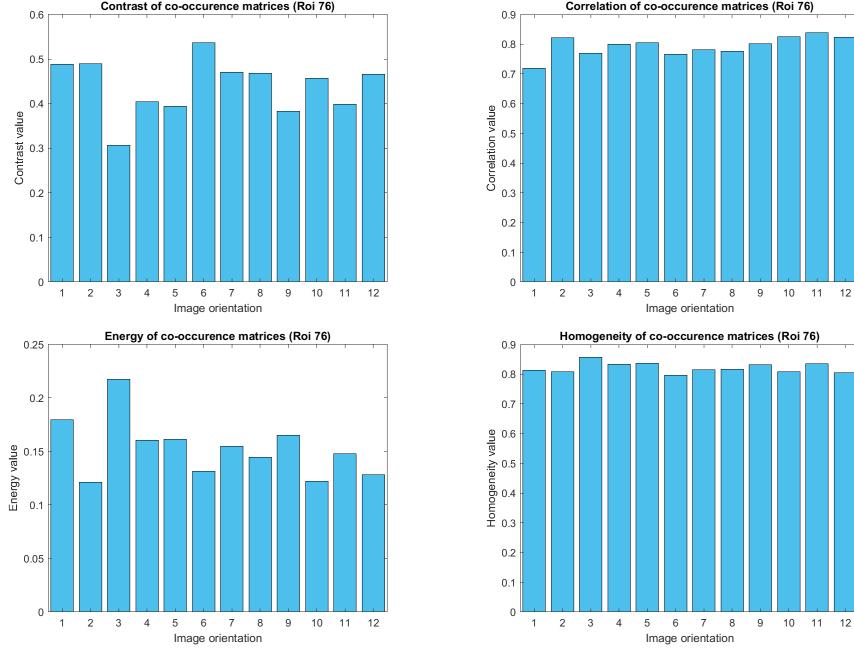


Figura 10: Misure statistiche della GLCM della roi 76 (Immagine con z=76)

## 5 Immagini 3D

Prima di ricavare le misure statistiche dalle immagini 3D, queste sono state normalizzate. Per quanto riguarda la visualizzazione, questa è possibile tramite diversi tool di MATLAB come:

- **3D image viewer and slicer** di Jun Tan;
- **Medical Image Reader and Viewer** di Josh Schaefferkoetter.

Le diverse visualizzazioni con questi due tool sono mostrate in appendice C.

### 5.1 Teoria

Per ricavare le misure statistiche, è stato utilizzato il pacchetto **RADIOMICS** di Martin Vallières. Questo permette di fare misure statistiche di texture analysis 3D. Gli input alla funzione *prepareVolume* che consente di estrarre le variabili *ROIonly*, *ROIonly2* e *levels* sono:

- *volume* = array 3D contenente l'immagine da analizzare;
- *mask* = array 3D delle stesse dimensioni di 'volume'. La maschera contiene 1 nella regione di interesse e 0 al di fuori di essa;
- *scantype* = stringa che specifica il tipo di immagine analizzata;
- *pixelW* = valore numerico che specifica la risoluzione in-piano (mm) del 'volume';
- *sliceS* = valore numerico che specifica la spaziatura della slice del 'volume';

- $R$  = valore numerico che specifica il rapporto dei pesi per i coefficienti di un filtro passa-banda sul peso del resto dei coefficienti (HHH e LLL). Per  $R=1$  non viene applicato il filtro wavelet passa-basso.
- $scale$  = valore numerico che specifica la scala alla quale 'volume' è ricampionato isotropicamente (mm). Se viene fornita la stringa 'pixelW', il volume è ricampionato isotropicamente alla risoluzione in-piano iniziale;
- $textType$  = stringa che specifica per quale tipo di texture 'volume' deve essere preparato. Può essere:
  - 'Global' = volume preparato per il calcolo delle texture features globali;
  - 'Matrix' = volume preparato per calcolo delle texture features basate sulle matrici (GLCM, GLRLM, GLZM, NGTDM).
- $quantAlgo$  = stringa che specifica l'algoritmo di quantizzazione da usare su 'volume' (solo quando 'textType' è impostato su 'Matrix'). Può essere:
  - 'Equal' = quantizzazione di uguale probabilità;
  - 'Lloyd' = quantizzazione Lloyd-Max;
  - 'Uniform' = quantizzazione uniforme.
- $Ng$  = intero che specifica il numero di livelli di grigio nel processo di quantizzazione (solo quando 'textType' è impostato su 'Matrix').

Di seguito sono mostrati i valori impostati nel codice utilizzato, per misure basate su matrici:

```
[ROIonly, levels]=prepareVolume(n_tot_3d, BW_v, 'MRscan', 1.0625, 1.0714, 1,
                                'pixelW', 'Matrix', 'Uniform', 256);
```

Di seguito sono mostrati i valori impostati nel codice utilizzato, per misure globali:

```
[ROIonly2] = prepareVolume(n_tot_3d, BW_v, 'MRscan', .0625, 1.0714, 1, 'pixelW',
                            'Global');
```

Il pacchetto **RADIOMICS** di Martin Vallières calcola le seguenti matrici e da esse è possibile estrarre diverse misure statistiche:

- **Gray-Level Co-occurrence Matrix (GLCM)** = questa matrice prende in considerazione le disposizioni delle paia di voxel, come è stato già detto nel caso 2D. In questo caso anziché avere un'immagine  $I(x,y)$ , si ha un volume quantizzato  $V(x,y,z)$ .  $P(i,j)$  rappresenta il numero di volte in cui i voxel con livello di grigio  $i$  sono confinanti con i voxel con livello di grigio  $j$  in  $V$  e  $N_g$  rappresenta il numero pre-definito del set di livelli di grigio quantizzati in  $V$ . Per un volume  $V$  è calcolata solo una GLCM di dimensione  $N_g \times N_g$  sommando simultaneamente le frequenze delle co-occorrenze di tutti i voxel con i loro 26 vicini connessi nello spazio 3D. Un'entrata  $(i,j)$  della GLCM normalizzata è quella definita nell'equazione 8 e con essa sono definite  $\mu_i$  e  $\mu_j$ ,  $\sigma_i$  e  $\sigma_j$  come nelle equazioni 9 e 10.

Dalla matrice GLCM sono state estratte le seguenti misure:

- a. **Energy** = misura l' uniformità delle paia di voxel (chiamata anche momento angolare secondo), definita come nell'equazione 13.  
Un valore maggiore di energia implica che ci sono più istanze di coppie di valori di intensità nell'immagine che confinano tra di loro a frequenze più alte.
- b. **Contrast** = misura le variazioni locali nella GLCM, definito come nell'equazione 11.  
Un valore maggiore di contrasto è correlato con una maggiore disparità nei valori di intensità tra voxel vicini.
- c. **Correlation** = misura la dipendenza lineare dei livelli di grigio nella GLCM, definita come nell'equazione 12.  
La correlazione è un valore tra 0 (non correlato) e 1 (perfettamente correlato) che mostra la dipendenza lineare dei livelli di grigio nei rispettivi voxel nella GLCM.

- d. **Homogeneity** = misura l'omogeneità delle paia di voxel, definita come nell'equazione ??.  
L'omogeneità è una misura della somiglianza in valori di intensità per voxel vicini. E' una misura di omogeneità locale che aumenta con minor contrasto nella finestra.
- e. **Variance** = misura della diffusione della distribuzione sulla media ed è definita come:

$$Variance = \frac{1}{N_g \times N_g} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [(i - \mu_i)^2 p(i, j) + (j - \mu_j)^2 p(i, j)] \quad (15)$$

- f. **Sum Average** = misura la relazione tra occorrenze di coppie con valori di intensità più bassi e occorrenze di coppie con valori di intensità più alti ed è definita come:

$$Sum\ Average = \frac{1}{N_g \times N_g} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [i \cdot p(i, j) + j \cdot p(i, j)] \quad (16)$$

- g. **Entropy<sub>log2</sub>** = misura la randomicità delle paia di voxel, definita come:

$$Entropy_{log2} = - \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) \cdot log_2(p(i, j)) \quad (17)$$

- h. **Dissimilarity** = misura la variazione di intensità locale ovvero la differenza assoluta media tra le coppie confinanti ed è definita come:

$$Dissimilarity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} |i - j| \cdot p(i, j) \quad (18)$$

Un valore più grande è correlato a una maggiore disparità nei valori di intensità tra voxel vicini.

- i. **Auto Correlation** = misura la grandezza della finezza e della grossolanità della texture ed è definita come:

$$Auto\ Correlation = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i, j) i j \quad (19)$$

- **Gray-Level Run-Length Matrix (GLRLM)** = questa matrice dà la misura di *run* (=lunghezza, in numero di pixel, di pixel consecutivi con lo stesso valore di grigio) omogenei per ogni livello di grigio. Sia P la matrice GLRLM di un volume quantizzato V(x,y,z) con voxel di dimensione isotropica. P(i,j) rappresenta il numero di runs di livelli di grigio i e lunghezza j in V, N<sub>g</sub> rappresenta il numero di livelli di grigio quantizzati e L rappresenta la lunghezza del run più lungo in V. Per un volume V è calcolata solo una GLRLM di dimensioni N<sub>g</sub> × L<sub>r</sub> sommando simultaneamente tutte possibili lunghezze di run più lunghe nelle 13 direzioni dello spazio 3D. L'entrata (i,j) della GLRLM normalizzata è definita come:

$$p(i, j) = \frac{P(i, j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{L_r} P(i, j)} \quad (20)$$

Sono definite anche le seguenti quantità:

$$\mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{L_r} p(i, j) \quad \mu_j = \sum_{j=1}^{L_r} j \sum_{i=1}^{N_g} p(i, j) \quad (21)$$

Dalla GLRLM sono state estratte le seguenti misure:

- a. **SRE**, Short-Run Emphasis = rappresenta la distribuzione di piccoli run omogenei in un'immagine, definita come:

$$SRE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \frac{p(i,j)}{j^2} \quad (22)$$

Un valore SRE più alto indica run di lunghezza minore e textures più fini.

- b. **LRE**, Long-Run Emphasis = rappresenta la distribuzione di lunghi runs omogenei in un'immagine, definita come:

$$LRE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} j^2 p(i,j) \quad (23)$$

Un valore LRE più alto indica runs di lunghezza maggiore e textures più grossolane.

- c. **GLN**, Gray-Level Non-uniformity = misura la somiglianza delle intensità dei livelli di grigio nell'immagine ed è definita come:

$$GLN = \sum_{i=1}^{N_g} \left( \sum_{j=1}^{L_r} p(i,j) \right)^2 \quad (24)$$

Un valore GLN più basso è correlato con una somiglianza più alta nei valori di intensità.

- d. **RLN**, Run-Length Non-uniformity = misura la somiglianza delle lunghezze dei runs attraverso l'immagine ed è definita come:

$$RLN = \sum_{j=1}^{L_r} \left( \sum_{i=1}^{N_g} p(i,j) \right)^2 \quad (25)$$

Un valore RLN minore indica più omogeneità lungo le lunghezze dei run.

- e. **RP**, Run Percentage = misura la grossolanità della texture prendendo il rapporto del numero di runs e il numero di voxels nella ROI, definita come:

$$RP = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{L_r} p(i,j)}{\sum_{j=1}^{L_r} j \sum_{i=1}^{N_g} p(i,j)} \quad (26)$$

- f. **LGRE**, Low Gray-level Run Emphasis = indica la distribuzione di runs di bassi livelli di grigio, definita come:

$$LGRE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \frac{p(i,j)}{i^2} \quad (27)$$

Un valore LGRE più alto indica una concentrazione maggiore di valori bassi di livello di grigio.

- g. **HGRE**, High Gray-level Run Emphasis = misura la distribuzione di runs di alti livelli di grigio, definita come:

$$HGRE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} i^2 p(i,j) \quad (28)$$

Un valore HGRE più alto indica una concentrazione maggiore di valori alti di livello di grigio.

- h. **SRHGE**, Short-Run High Gray-level Emphasis = misura la distribuzione di runs omogenei corti con bassi livelli di grigio, definita come:

$$SRHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \frac{i^2 p(i,j)}{j^2} \quad (29)$$

- i. **LRLGE**, Long-Run Low Gray-level Emphasis = misura la distribuzione di runs omogenei lunghi con bassi livelli di grigio, definita come:

$$LRLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} \frac{j^2 p(i,j)}{i^2} \quad (30)$$

- j. **GLV**, Gray-Level Variance = misura la varianza in livelli di grigio per i runs, definita come:

$$GLV = \frac{1}{N_g \times L_r} \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} (i \cdot p(i,j) - \mu_i)^2 \quad (31)$$

- k. **RLV**, Run-Length Variance = misura la varianza in lunghezza dei runs per i runs, definita come:

$$RLV = \frac{1}{N_g \times L_r} \sum_{i=1}^{N_g} \sum_{j=1}^{L_r} (j \cdot p(i,j) - \mu_j)^2 \quad (32)$$

- **Gray-Level Size Zone Matrix (GLSZM)** = matrice che fornisce informazione sulla dimensione delle zone omogenee per ogni livello di grigio. Una zona di livello di grigio è definita come il numero di voxel connessi che condividono la stessa intensità di livello di grigio.

Sia  $P$  la matrice GLSZM di un volume quantizzato  $V(x,y,z)$  con voxel di dimensione isotropica.  $P(i,j)$  rappresenta il numero di zone 3D di livello di grigio  $i$  e di dimensione  $j$  in  $V$ ,  $N_g$  rappresenta il numero di livelli di grigio quantizzato in  $V$  e  $L_z$  rappresenta la dimensione della zona più grande in  $V$ . Per un volume  $V$  è calcolata una GLSZM di dimensione  $N_g \times L_z$ . V sommando tutte le possibili zone più grandi, con le zone costruite da vicini 26-connessi dello stesso livello di grigio nello spazio 3D (un voxel può essere parte di una sola zona). L'entrata  $(i,j)$  della GLSZM è definita come:

$$p(i,j) = \frac{P(i,j)}{\sum_{i=1}^{N_g} \sum_{j=1}^{L_z} P(i,j)} \quad (33)$$

Inoltre, sono state definite anche le seguenti quantità:

$$\mu_i = \sum_{i=1}^{N_g} i \sum_{j=1}^{L_z} p(i,j) \quad \mu_j = \sum_{j=1}^{L_z} j \sum_{i=1}^{N_g} p(i,j) \quad (34)$$

Dalla matrice GLSZM sono state estratte le seguenti misure:

- a. **SZE**, Short-Zone Emphasis = indica la distribuzione di zone omogenee piccole in un'immagine, definita come:

$$SZE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} \frac{p(i,j)}{j^2} \quad (35)$$

Un valore maggiore di SZE indica zone di dimensione minore e texture più fini.

- b. **LZE**, Long-Zone Emphasis = indica la distribuzione di zone omogenee lunghe in un'immagine, definita come:

$$LZE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} j^2 p(i,j) \quad (36)$$

Un valore maggiore indica zone di dimensione maggiore e texture più grossolane.

c. **GLN**, Gray-Level Non-uniformity = definita come:

$$GLN = \sum_{i=1}^{N_g} \left( \sum_{j=1}^{L_z} p(i, j) \right)^2 \quad (37)$$

d. **ZSN**, Zone-Size Non-uniformity = definita come:

$$ZSN = \sum_{j=1}^{L_z} \left( \sum_{i=1}^{N_g} p(i, j) \right)^2 \quad (38)$$

Un valore ZSN minore indica più omogeneità nei volumi delle size-zone.

e. **ZP**, Zone Percentage = misura l'omogeneità della distribuzione di volumi zone size in un immagine tra i livelli di grigio osservati ed è definita come:

$$ZP = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{L_z} p(i, j)}{\sum_{j=1}^{L_z} j \sum_{i=1}^{N_g} p(i, j)} \quad (39)$$

f. **LGZE**, Low Gray-level Zone Emphasis = misura la distribuzione di zone a bassi livelli di grigio, definita come:

$$LGZE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} \frac{p(i, j)}{i^2} \quad (40)$$

Un valore LGZE maggiore indica una percentuale maggiore di valori minori di livelli di grigio e size zone.

g. **HGZE**, High Gray-level Zone Emphasis = misura la distribuzione di zone ad alti livelli di grigio, definita come:

$$HGZE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} i^2 p(i, j) \quad (41)$$

Un valore HGZE maggiore indica una percentuale maggiore di valori maggiori di livelli di grigio e size zone.

h. **SZLGE**, Short-Zone Low Gray-level Emphasis = misura la distribuzione di piccole zone omogenee con bassi livelli di grigio, definita come:

$$SZLGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} \frac{p(i, j)}{i^2 j^2} \quad (42)$$

i. **SZHGE**, Short-Zone High Gray-level Emphasis = misura la distribuzione di piccole zone omogenee con alti livelli di grigio, definita come:

$$SZHGE = \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} \frac{i^2 p(i, j)}{j^2} \quad (43)$$

j. **GLV**, Gray-Level Variance = misura la varianza in intensità di livelli di grigio per le zone ed è definita come:

$$GLV = \frac{1}{N_g \times L_z} \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} (i \cdot p(i, j) - \mu_i)^2 \quad (44)$$

k. **ZSV**, Zone-Size Variance = misura la varianza in volumi zone-size per le zone ed è definita come:

$$ZLV = \frac{1}{N_g \times L_z} \sum_{i=1}^{N_g} \sum_{j=1}^{L_z} (j \cdot p(i, j) - \mu_j)^2 \quad (45)$$

- **Neighborhood Gray-Tone Difference Matrix (NGTDM)** = matrice che corrisponde alla differenza di livelli di grigio tra un voxel e i suoi 26 vicini in 3 dimensioni. Sia  $P(i)$  la matrice NGTDM di un volume quantizzato  $V(x,y,z)$  con voxel di dimensione isotropica.  $P(i)$  rappresenta la somma delle differenze di livelli di grigio tra tutti i voxel con livello di grigio e il livello di grigio medio dei loro 26-connessi vicini nello spazio 3D.  $N_g$  rappresenta il numero predefinito del set di livelli di grigio quantizzati in  $V$  e  $(N_g)_{eff}$  è il numero effettivo di livelli di grigio in  $V$  con  $(N_g)_{eff} < N_g$  (sia il vettore dei valori di livello di grigio in  $V$  denotato come  $\mathbf{g} = g(1), g(2), \dots, g(N_g)$ : alcuni livelli di grigio ad esclusione di  $g(1)$  e  $g(N_g)$  potrebbero non apparire in  $V$  a causa di schemi di quantizzazione diversi). Una NGTDM di dimensione  $N_g \times 1$  è calcolata per volume  $V$ . La  $i^{th}$  entrata del NGTDM è quindi definita come:

$$P(i) = \begin{cases} \sum_{\text{all voxels } \in N_i} |i - \bar{A}_i| & \text{if } N_i > 0 \\ 0 & \text{if } N_i = 0 \end{cases} \quad (46)$$

dove  $N_i$  è il set di tutti i voxel con livello di grigio  $i$  in  $V$  (includendo la regione periferica),  $N_i$  è il numero di voxel con livello di grigio  $i$  in  $V$  e  $\bar{A}_i$  è il livello di grigio medio dei vicini 26-connessi attorno a un voxel centrale con livello di grigio  $i$  e collocati alla posizione  $(j, k, l)$  in  $V$  così che:

$$\bar{A}_i = \bar{A}(j, k, l) = \frac{\sum_{m=-1}^{m=1} \sum_{n=-1}^{n=1} \sum_{o=-1}^{o=1} w_{m,n,o} \cdot V(j+m, k+n, l+o)}{\sum_{m=-1}^{m=1} \sum_{n=-1}^{n=1} \sum_{o=-1}^{o=1} w_{m,n,o}} \quad (47)$$

dove

$$w_{m,n,o} = \begin{cases} 1 & \text{if } |j-m| + |k-n| + |l-o| = 1 \\ \frac{1}{\sqrt{2}} & \text{if } |j-m| + |k-n| + |l-o| = 2 \\ \frac{1}{\sqrt{3}} & \text{if } |j-m| + |k-n| + |l-o| = 3 > 0 \\ 0 & \text{if } V(j-m, k-n, l-o) \text{ is undefined.} \end{cases} \quad (48)$$

E' anche definita la seguente quantità:

$$n_i = \frac{N_i}{N} \quad (49)$$

dove  $N$  è il numero totale di voxel in  $V$ .

Dalla matrice NGTDM sono state estratte le seguenti misure:

- a. **Coarseness** = misura la differenza media tra il centro del voxel e il suo vicinato, è un indicazione del tasso di variazione spaziale ed è definita come:

$$Coarseness = \left[ \epsilon + \sum_{i=1}^{N_g} n_i P(i) \right]^{-1} \quad (50)$$

dove  $\epsilon$  è un piccolo numero usato per prevenire che 'coarseness' diventi infinita. Un valore più alto indica un tasso di variazione spaziale minore e una texture più uniforme localmente.

- b. **Contrast** = misura il cambiamento di intensità spaziale ma è anche dipendente dal range dinamico complessivo di livello di grigio ed è definito come:

$$Contrast = \left[ \frac{1}{(N_g)_{eff}[(N_g)_{eff} - 1]} \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} n_i n_j \cdot (i - j)^2 \right] \cdot \left[ \frac{1}{N} \sum_{i=1}^{N_g} P(i) \right] \quad (51)$$

Il contrasto è alto quando sia il range dinamico che il tasso di variazione spaziale sono alti.

c. **Busyness** = misura il cambiamento da un pixel al suo vicinato.

$$Busyness = \frac{\sum_{i=1}^{N_g} n_i P(i)}{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i \cdot n_i - j \cdot n_j)}, \quad n_i \neq 0, n_j \neq 0 \quad (52)$$

Un alto valore di *Busyness* indica un'immagine 'busy' con rapidi cambiamenti di intensità tra i pixel e il loro vicinato.

d. **Complexity** = definita come:

$$complexity = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \frac{|i - j| |n_i P(i) + n_j P(j)|}{N(n_i + n_j)}, \quad n_i \neq 0, n_j \neq 0 \quad (53)$$

Un'immagine è considerata complessa quando ci sono alcune componenti primitive in essa, ovvero l'immagine non è uniforme e ci sono molti cambiamenti rapidi nell'intensità del livello di grigio.

e. **Strength** = è una misura dei primitivi in un'immagine ed è definita come:

$$strength = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (n_i + n_j)(i - j)^2}{[\epsilon + \sum_{i=1}^{N_g} P(i)]}, \quad n_i \neq 0, n_j \neq 0 \quad (54)$$

dove  $\epsilon$  è un piccolo numero che evita che strength diventi infinito.

Il suo valore è alto quando i primitivi sono facilmente definiti e visibili, cioè un'immagine con un cambiamento di intensità lento ma più grandi differenze grossolane nelle intensità del livello di grigio.

Il pacchetto **RADIOMICS** permette anche di calcolare misure statistiche globali. A questo proposito, sia  $P$  un istogramma di prim'ordine di un volume  $V(x,y,z)$  con dimensione di voxel isotropica.  $P(i)$  rappresenta il numero di voxel con livello di grigio  $i$  e  $N_g$  rappresenta il numero di set di bin di livelli di grigio per  $P$ . La  $i^{th}$  entrata dell'istogramma normalizzato è quindi definita come:

$$p(i) = \frac{P(i)}{\sum_{i=1}^{N_g} P(i)} \quad (55)$$

Da qui è possibile estrarre le seguenti features globali:

a. **Variance**, definita come:

$$\sigma^2 = \sum_{i=1}^{N_g} (i - \mu)^2 p(i) \quad (56)$$

b. **Skewness**, definita come:

$$s = \sigma^{-3} \sum_{i=1}^{N_g} (i - \mu)^2 p(i) \quad (57)$$

c. **Kurtosis**, definita come:

$$k = \sigma^{-4} \sum_{i=1}^{N_g} [(i - \mu)^4 p(i)] - 3 \quad (58)$$

## 5.2 Risultati

Le misure statistiche 3D sono state realizzate considerando:

- per un 'Volume' di dimensioni  $160 \times 160 \times 160$  (volume della singola orientazione):
  - 'Mask' relativa alla roi66 (volume di dimensioni  $65 \times 65 \times 160$ );
  - 'Mask' relativa alla roi76 (volume di dimensioni  $65 \times 65 \times 160$ );
- per un 'Volume' di dimensioni  $65 \times 65 \times 160$  relativo alla roi66:
  - 'Mask' relativa al Red Nucleus;
  - 'Mask' relativa alla Substantia Nigra;
- per un 'Volume' di dimensioni  $65 \times 65 \times 160$  relativo alla roi76:
  - 'Mask' relativa al Caudate;
  - 'Mask' relativa al Globus Pallidus;
  - 'Mask' relativa al Putamen;
  - 'Mask' relativa alla White Matter;

Tutte le aree corrispondenti ai riferimenti biologici sono state selezionate manualmente in MATLAB tramite le funzioni *imellipse* (solo per il Red Nucleus) e *imfreehand* (per tutti gli altri). Nella figura 11 sono mostrate le aree selezionate.

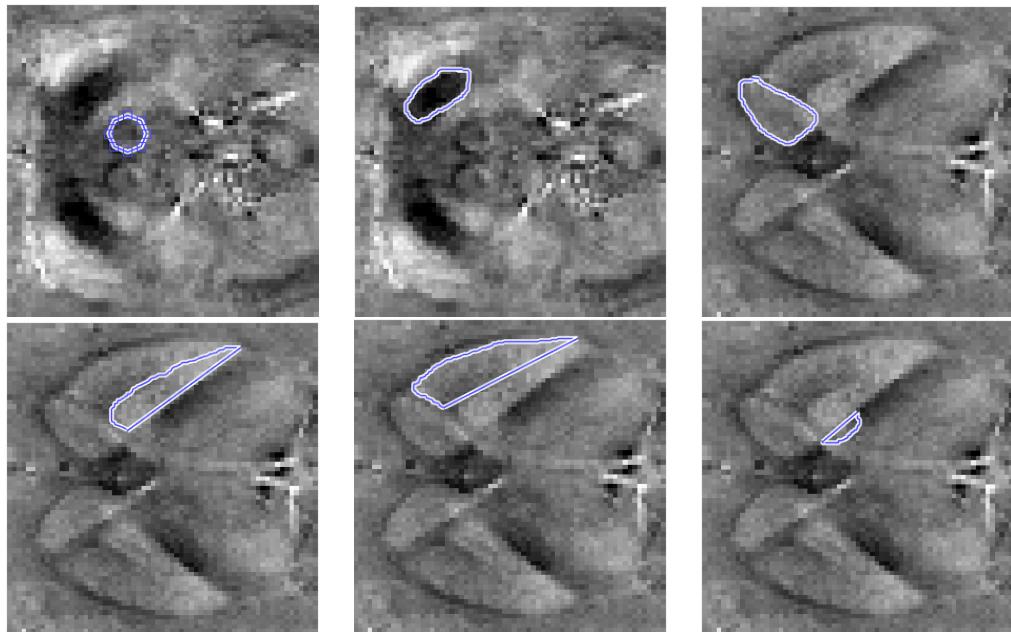


Figura 11: Alto sx: Red nucleus, Alto centrale: Substantia nigra, Alto dx: Caudate, Basso sx: Globus pallidus, Basso centrale: Putamen, Basso dx: White Matter

I risultati ottenuti per la roi66 sulle misurazioni basate su matrici sono mostrati nelle figure 12, 13, 14, 15 mentre quelle globali sono mostrate nella figura 16.

Per i risultati relativi a tutte le altre roi, si veda l'appendice D.

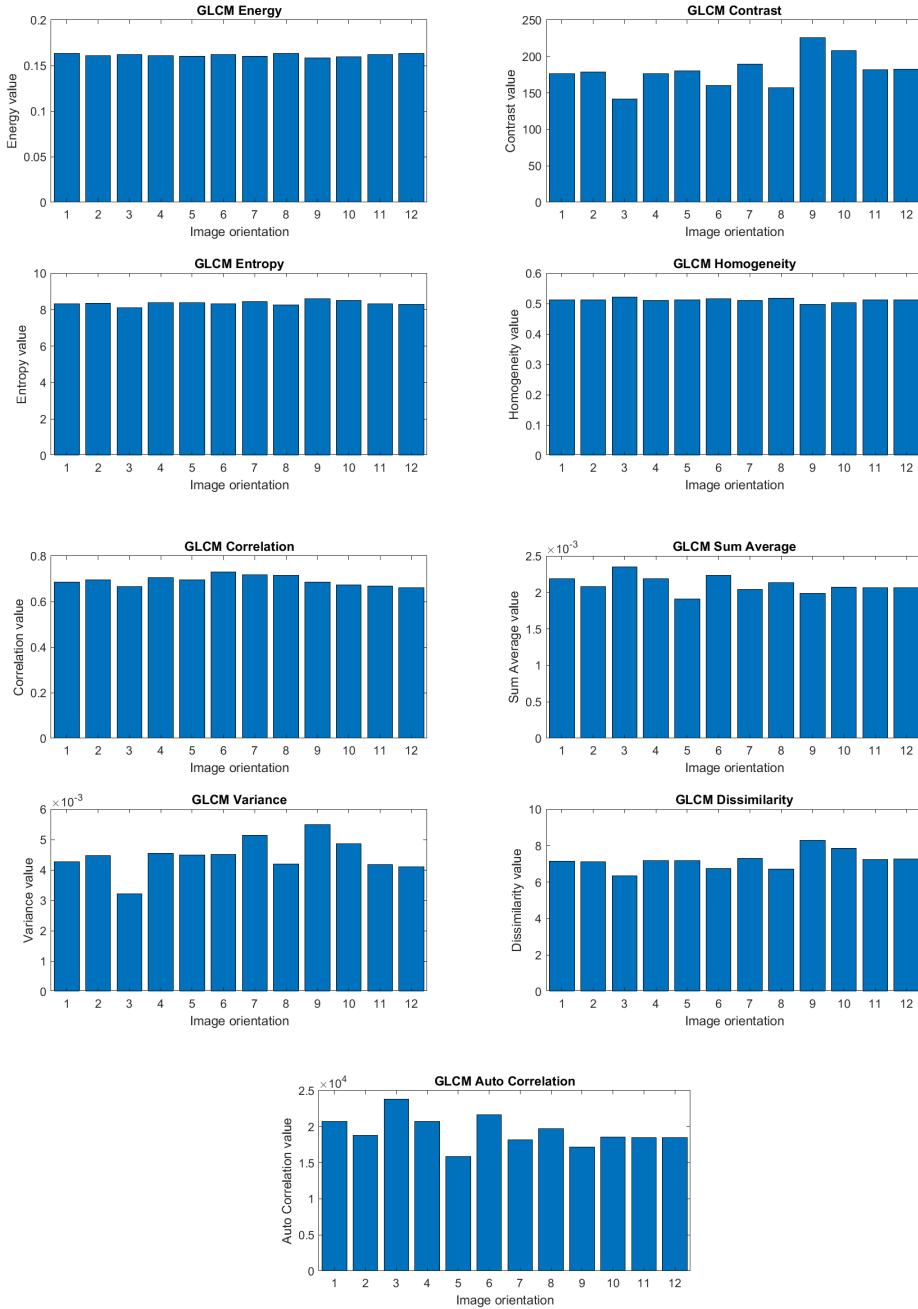


Figura 12: Misure statistiche della GLCM della roi 66 (Volume 3D)

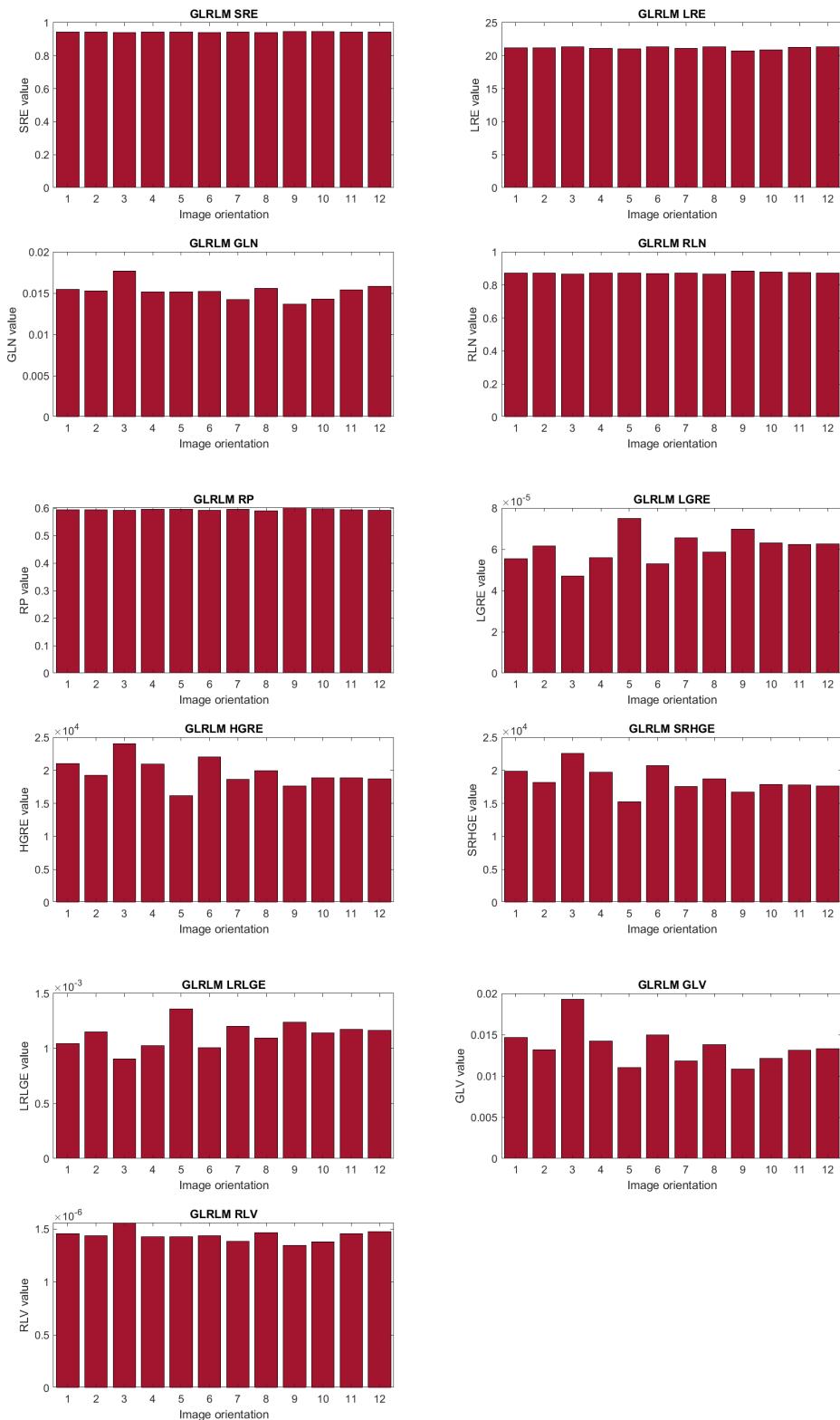


Figura 13: Misure statistiche della GLMRL della roi 66 (Volume 3D)

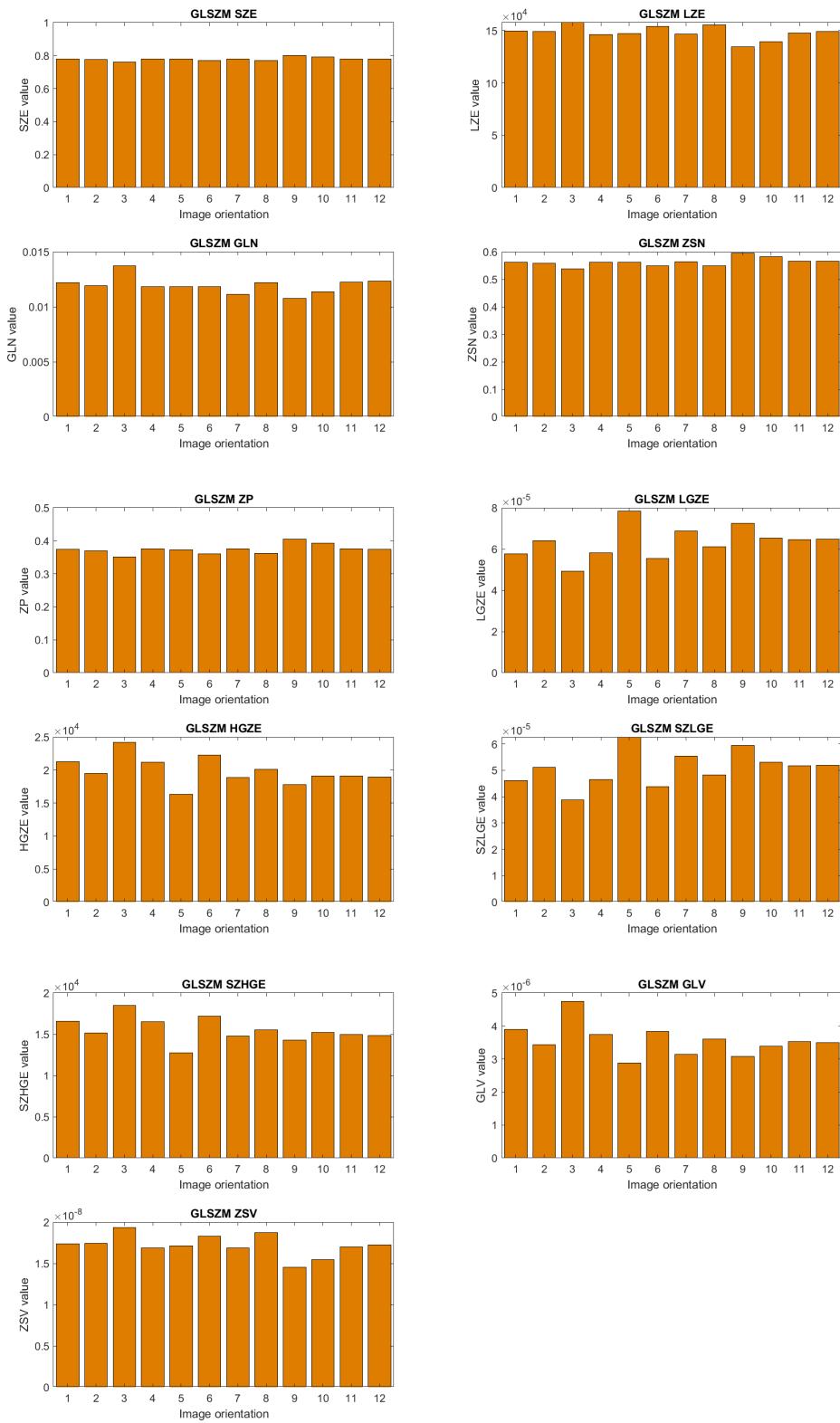


Figura 14: Misure statistiche della GLSzm della roi 66 (Volume 3D)

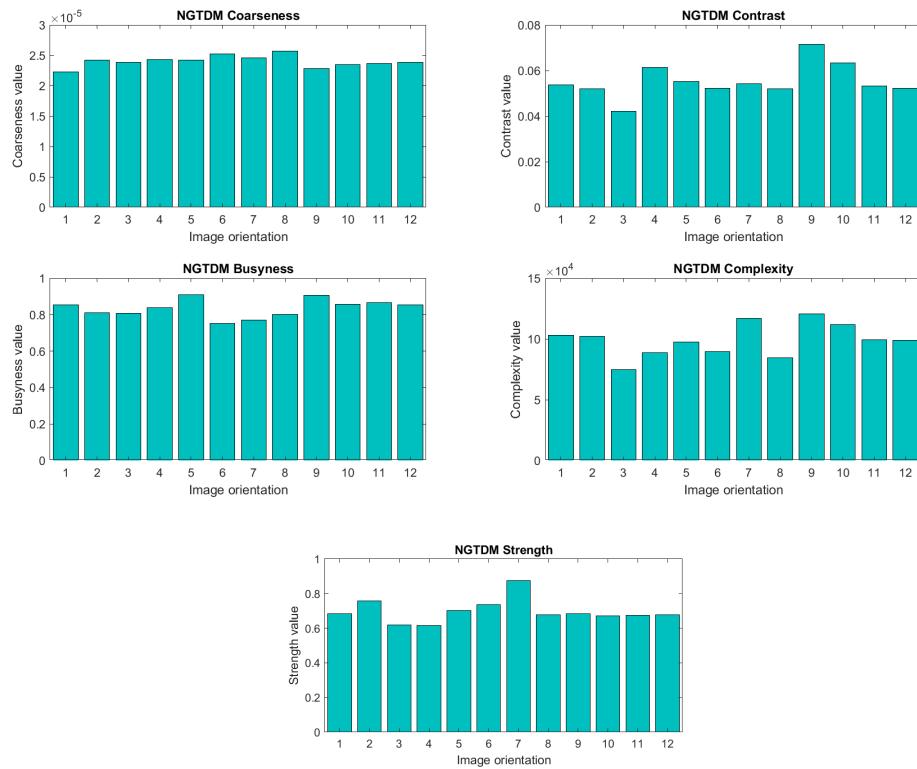


Figura 15: Misure statistiche della NGTDM della roi 66 (Volume 3D)

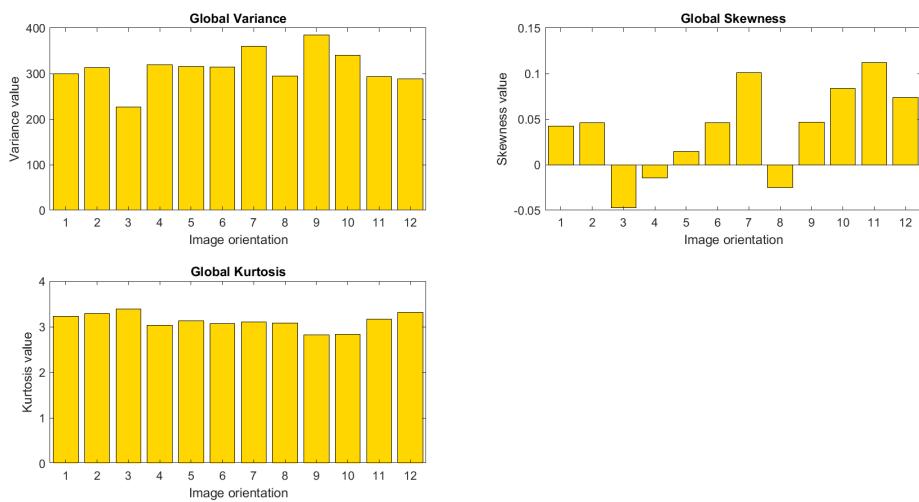


Figura 16: Misure statistiche globali della roi 66 (Volume 3D)

### 5.3 Confronto tra le ROI

Per capire in che modo varino e come si distribuiscano le misurazioni 3d (basate su matrici e globali) rispetto alle singole ROI di interesse biologico, ovvero *Red Nucleus*, *Substantia Nigra*, *Caudate*, *Globus Pallidus*, *Putamen* e *White Matter*, sono state realizzate, tenendo fissa l'orientazione per ogni grafico:

- le figure 17, 18, 19 e 20 in cui è possibile notare l'andamento dei valori relativi alle misure basate su matrici nelle varie ROI;
- le figure 21, 22, 23 e 24 in cui è possibile notare l'andamento dei valori relativi alle misure globali nelle varie ROI.

Nel primo caso sull'asse delle y è stata usata una scala logaritmica poichè le misurazioni differiscono tra loro di diversi ordini di grandezza. Questo permette la rappresentazione di tutte le misurazioni ma non risalta la loro differenza a livello locale.

Inoltre, le features numerate sull'asse delle x corrispondono alle features basate su matrici e, in ordine crescente sono:

1	<i>glcm<sub>ene</sub></i>	10	<i>glrlm<sub>sre</sub></i>	21	<i>glszm<sub>sze</sub></i>	32	<i>ngtdm<sub>coa</sub></i>
2	<i>glcm<sub>con</sub></i>	11	<i>glrlm<sub>tre</sub></i>	22	<i>glszm<sub>lze</sub></i>	33	<i>ngtdm<sub>con</sub></i>
3	<i>glcm<sub>ent</sub></i>	12	<i>glrlm<sub>gln</sub></i>	23	<i>glszm<sub>gln</sub></i>	34	<i>ngtdm<sub>bus</sub></i>
4	<i>glcm<sub>hom</sub></i>	13	<i>glrlm<sub>rln</sub></i>	24	<i>glszm<sub>zsn</sub></i>	35	<i>ngtdm<sub>com</sub></i>
5	<i>glcm<sub>cor</sub></i>	14	<i>glrlm<sub>rp</sub></i>	25	<i>glszm<sub>zp</sub></i>	36	<i>ngtdm<sub>str</sub></i>
6	<i>glcm<sub>sa</sub></i>	15	<i>glrlm<sub>lgre</sub></i>	26	<i>glszm<sub>lgze</sub></i>		
7	<i>glcm<sub>var</sub></i>	16	<i>glrlm<sub>hgre</sub></i>	27	<i>glszm<sub>hgze</sub></i>		
8	<i>glcm<sub>dis</sub></i>	17	<i>glrlm<sub>srhge</sub></i>	28	<i>glszm<sub>szlge</sub></i>		
9	<i>glcm<sub>ac</sub></i>	18	<i>glrlm<sub>lrhge</sub></i>	29	<i>glszm<sub>szhge</sub></i>		
		19	<i>glrlm<sub>glv</sub></i>	30	<i>glszm<sub>glv</sub></i>		
		20	<i>glrlm<sub>rlv</sub></i>	31	<i>glszm<sub>zsv</sub></i>		

Nel secondo caso non è stata usata la scala logaritmica sull'asse delle y sia poichè è possibile visualizzare tutte le misurazioni in scala lineare sia poichè queste misurazioni presentano anche dei valori negativi che rendono impossibile la rappresentazione in scala logaritmica.

Le features numerate in ordine crescente sull'asse delle x corrispondono alle features globali e sono:

1	$g_v$
2	$g_s$
3	$g_k$

Per facilitare la gestione di queste misurazioni, per ogni ROI sono state create 2 matrici del tipo:

- $t_{cau}$  = avente dimensioni  $12 \times 36$  dove il numero delle righe corrisponde alle orientazioni, mentre quello delle colonne corrisponde alle misurazioni 'basate su matrici';
- $tg_{cau}$  = avente dimensioni  $12 \times 3$  dove il numero delle righe corrisponde alle orientazioni, mentre quello delle colonne corrisponde alle misurazioni 'globali'.

Complessivamente sono state create 12 matrici, salvate come file MATLAB nella cartella **tot** utili sia per creare i plot spiegati in questa sezione che per la PCA, spiegata in sezione 6.1.

Di conseguenza, in questa sezione ogni figura rappresenta l'andamento di 6 vettori, uno per ogni ROI di interesse biologico, aventi lunghezza 36 nel caso di misurazioni 'basate su matrici' e 3 nel caso di misurazioni 'globali'.

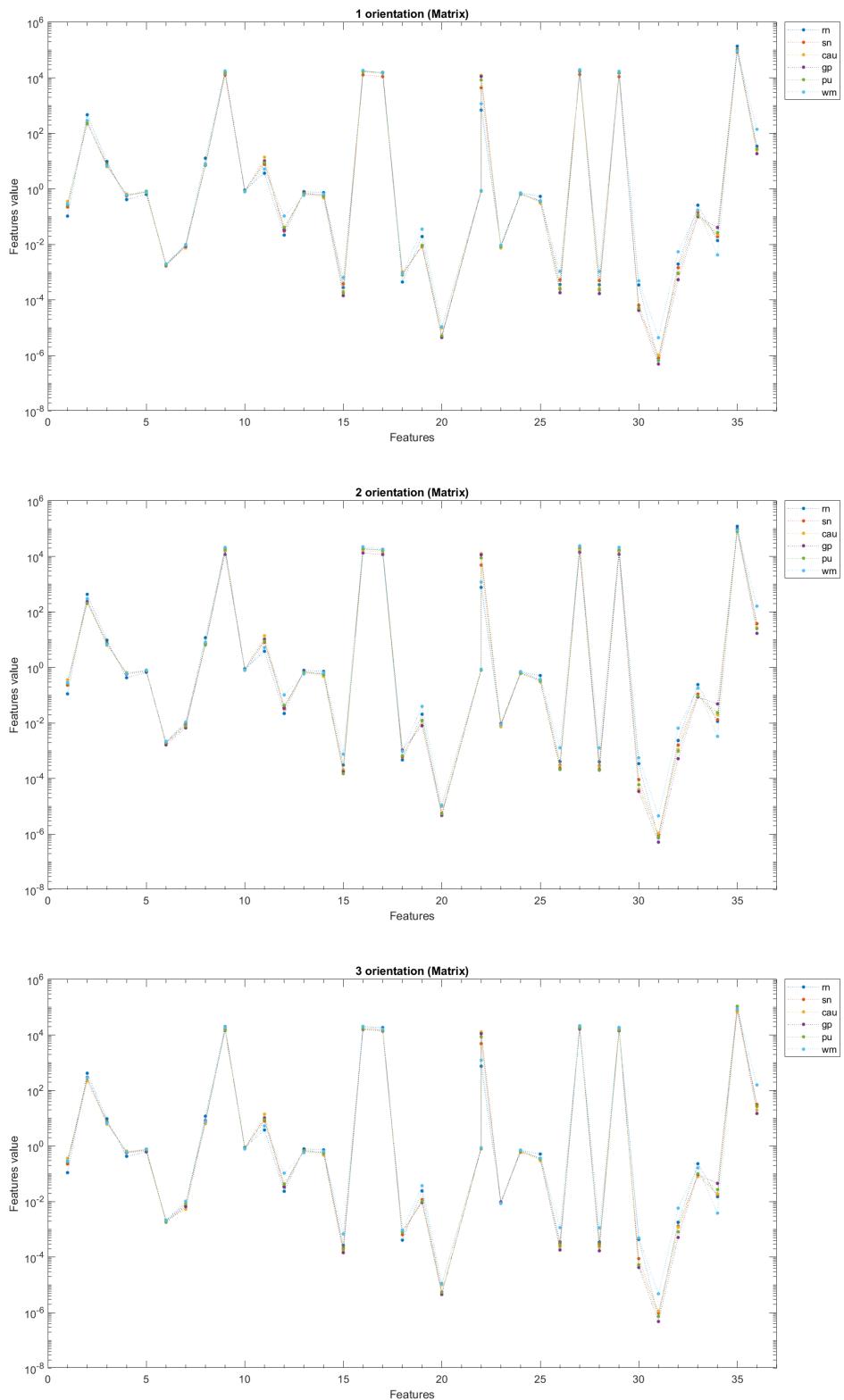


Figura 17: Misure basate su matrici per le ROI di interesse biologico: Orientazioni 1,2,3

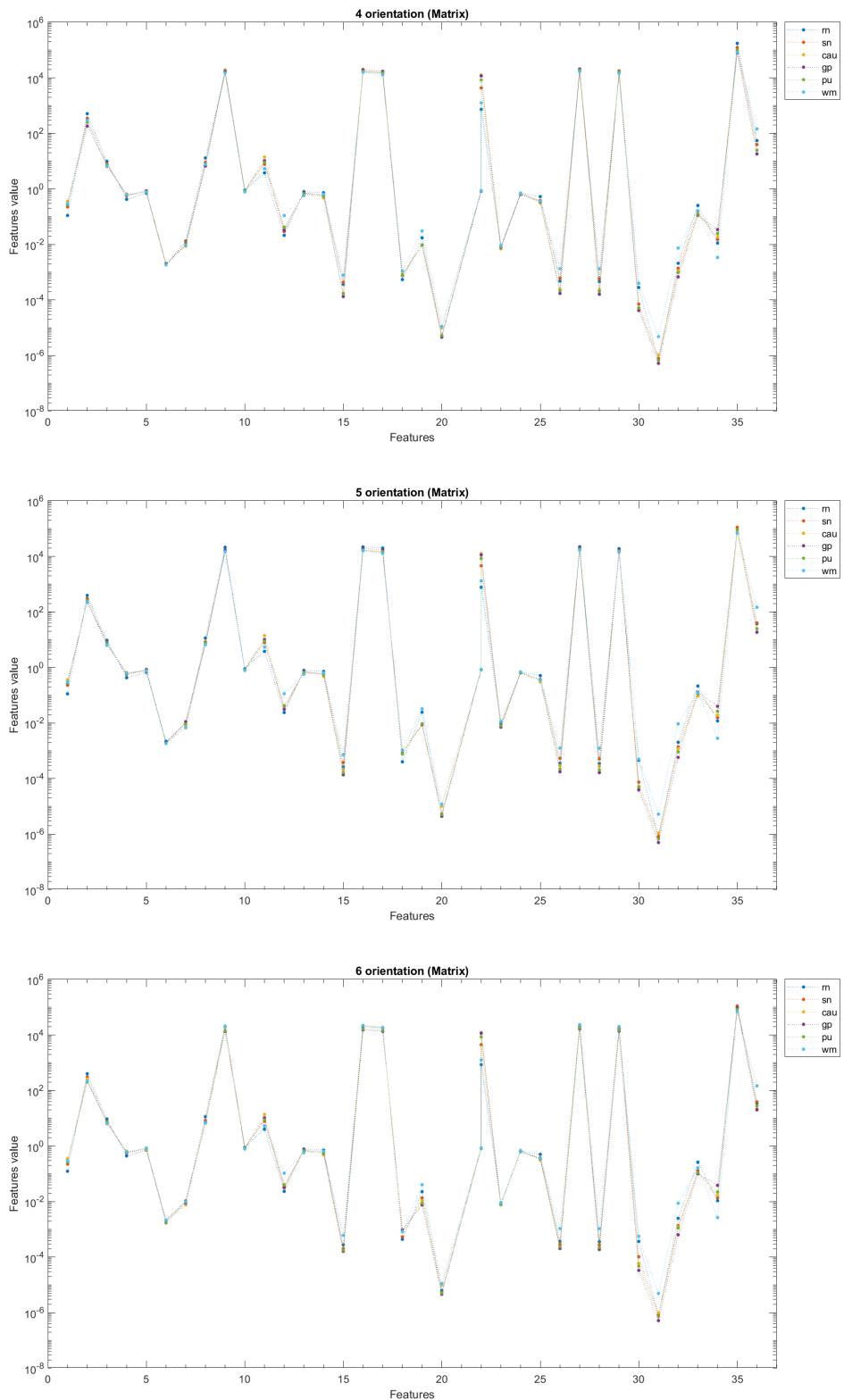


Figura 18: Misure basate su matrici per le ROI di interesse biologico: Orientazioni 4,5,6

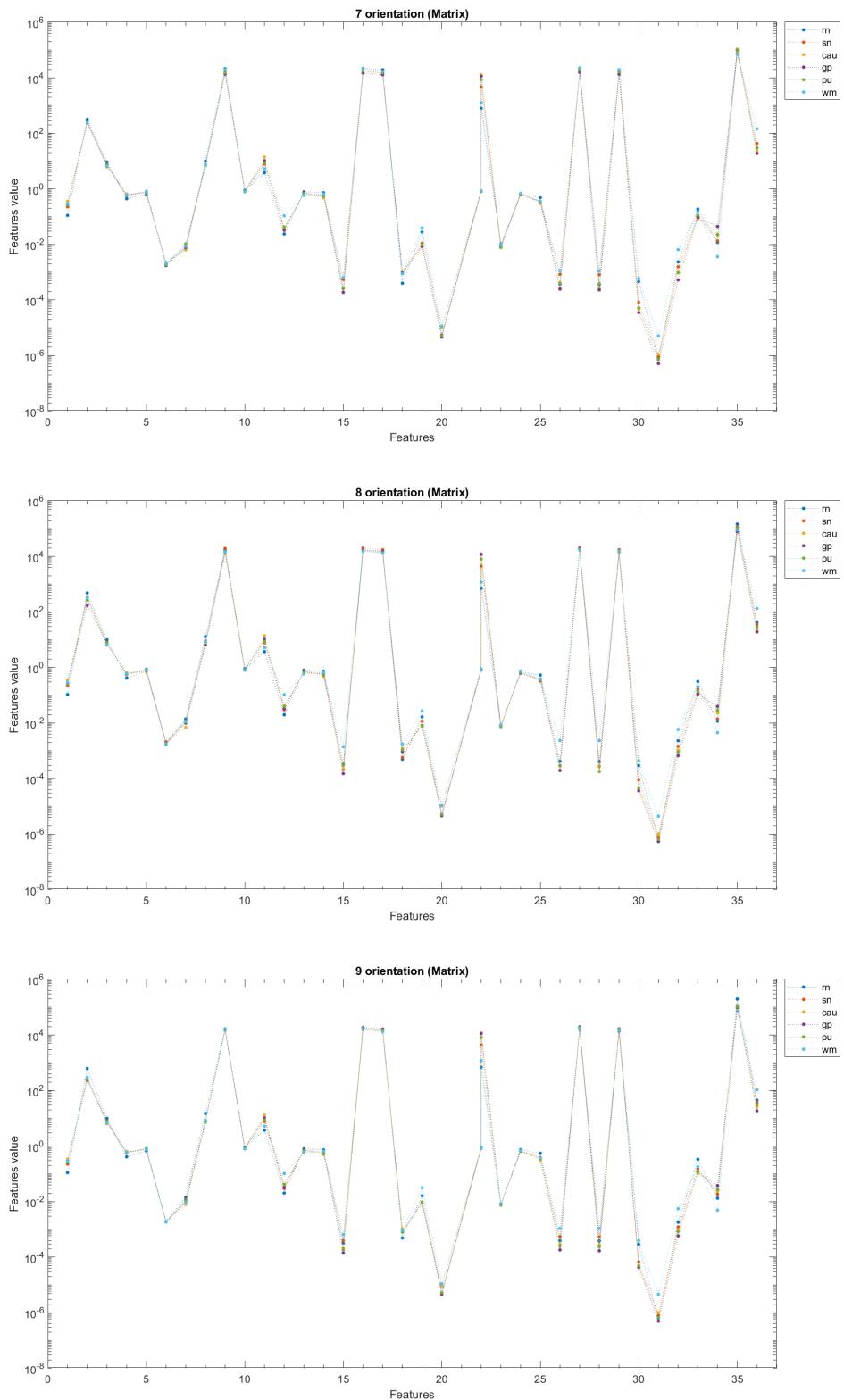


Figura 19: Misure basate su matrici per le ROI di interesse biologico: Orientazioni 7,8,9

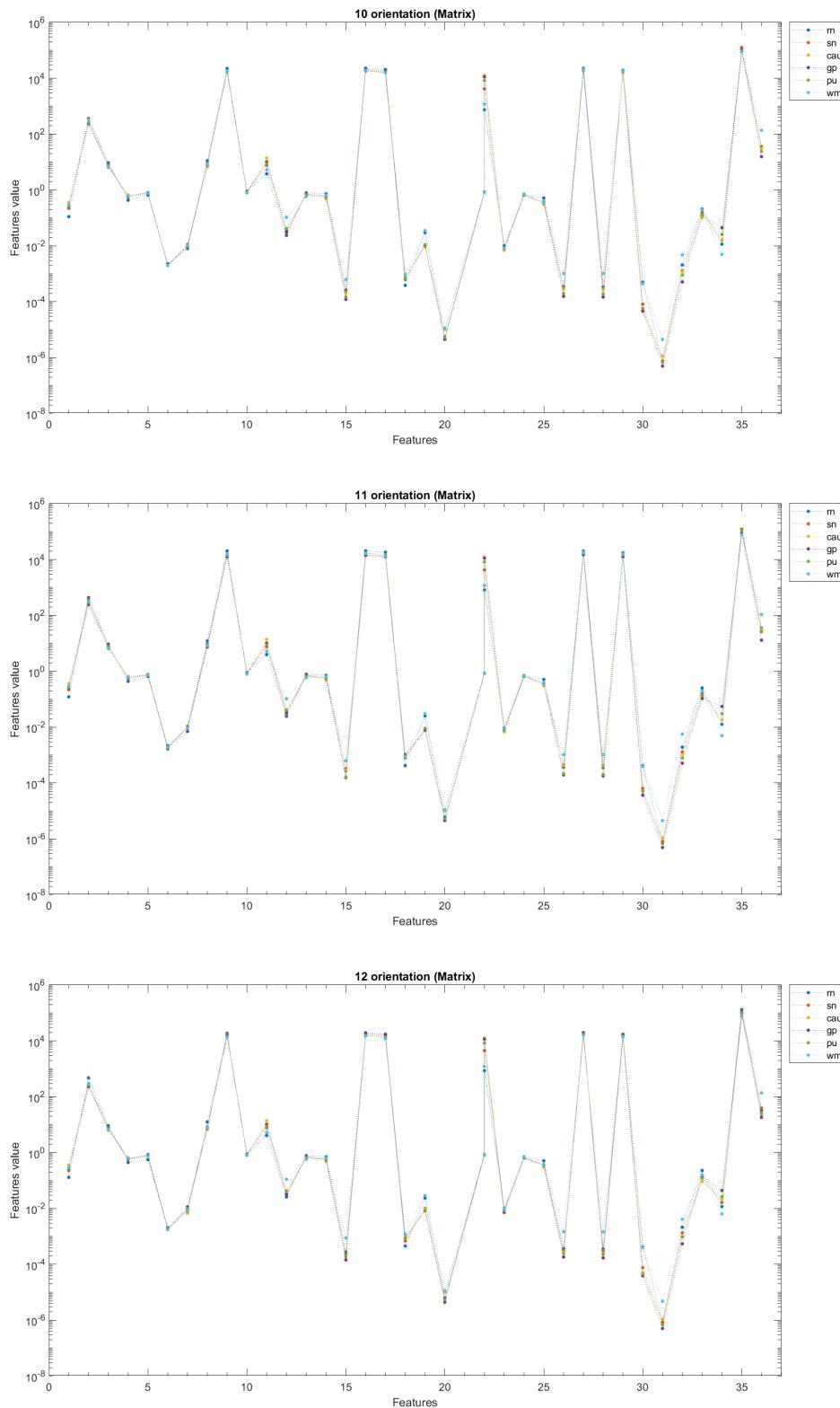


Figura 20: Misure basate su matrici per le ROI di interesse biologico: Orientazioni 10,11,12

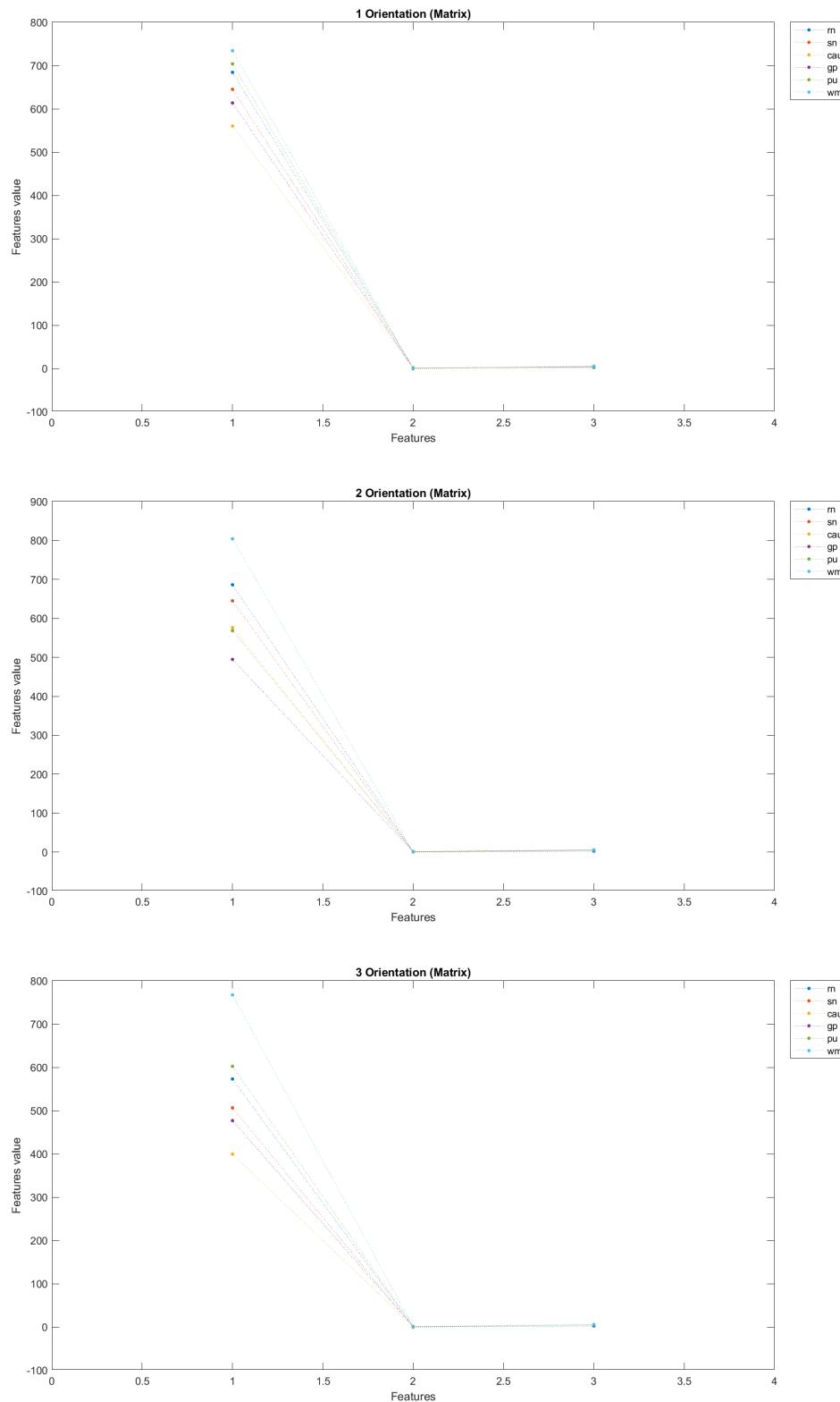


Figura 21: Misure globali per le ROI di interesse biologico: Orientazioni 1,2,3

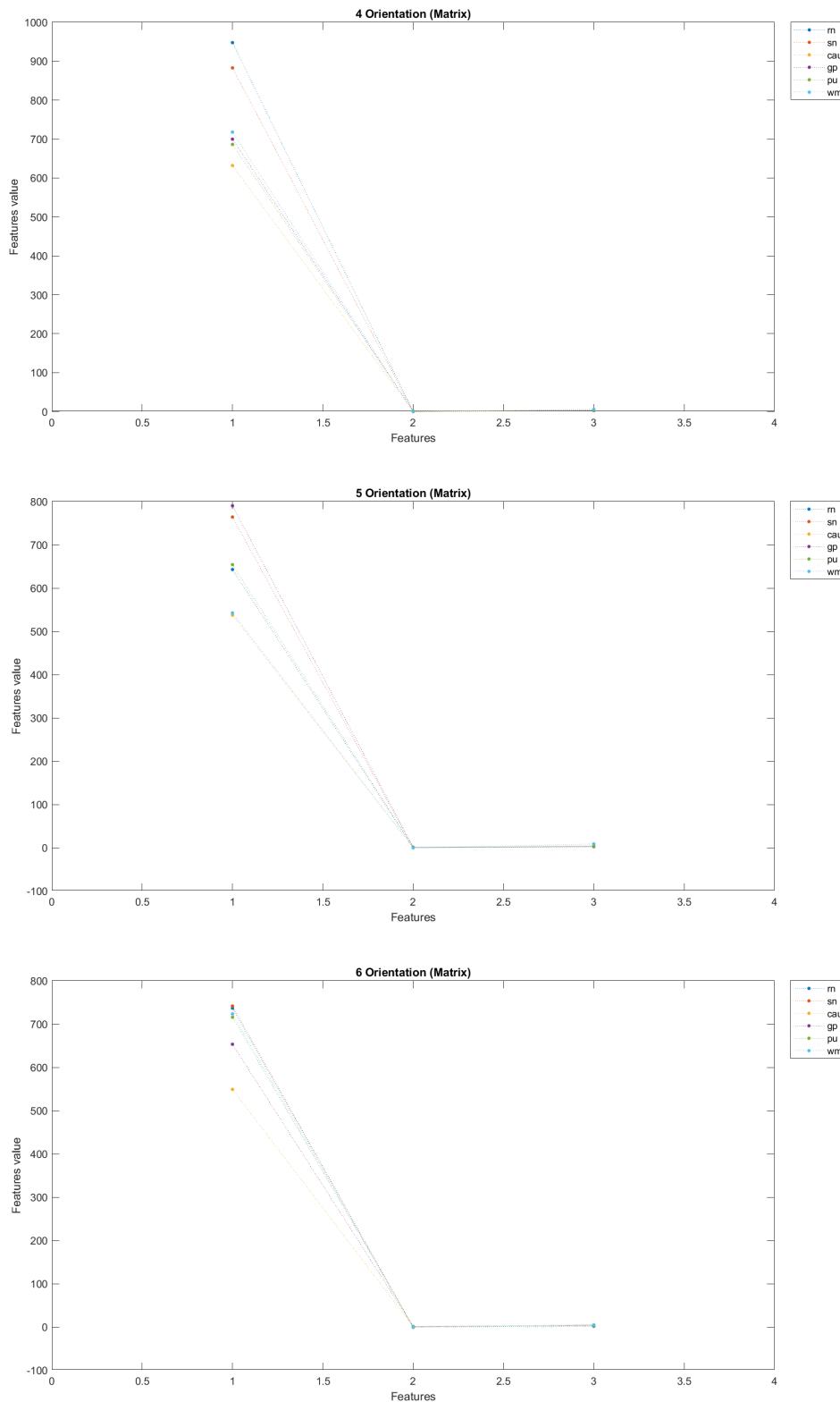


Figura 22: Misure globali per le ROI di interesse biologico: Orientazioni 4,5,6

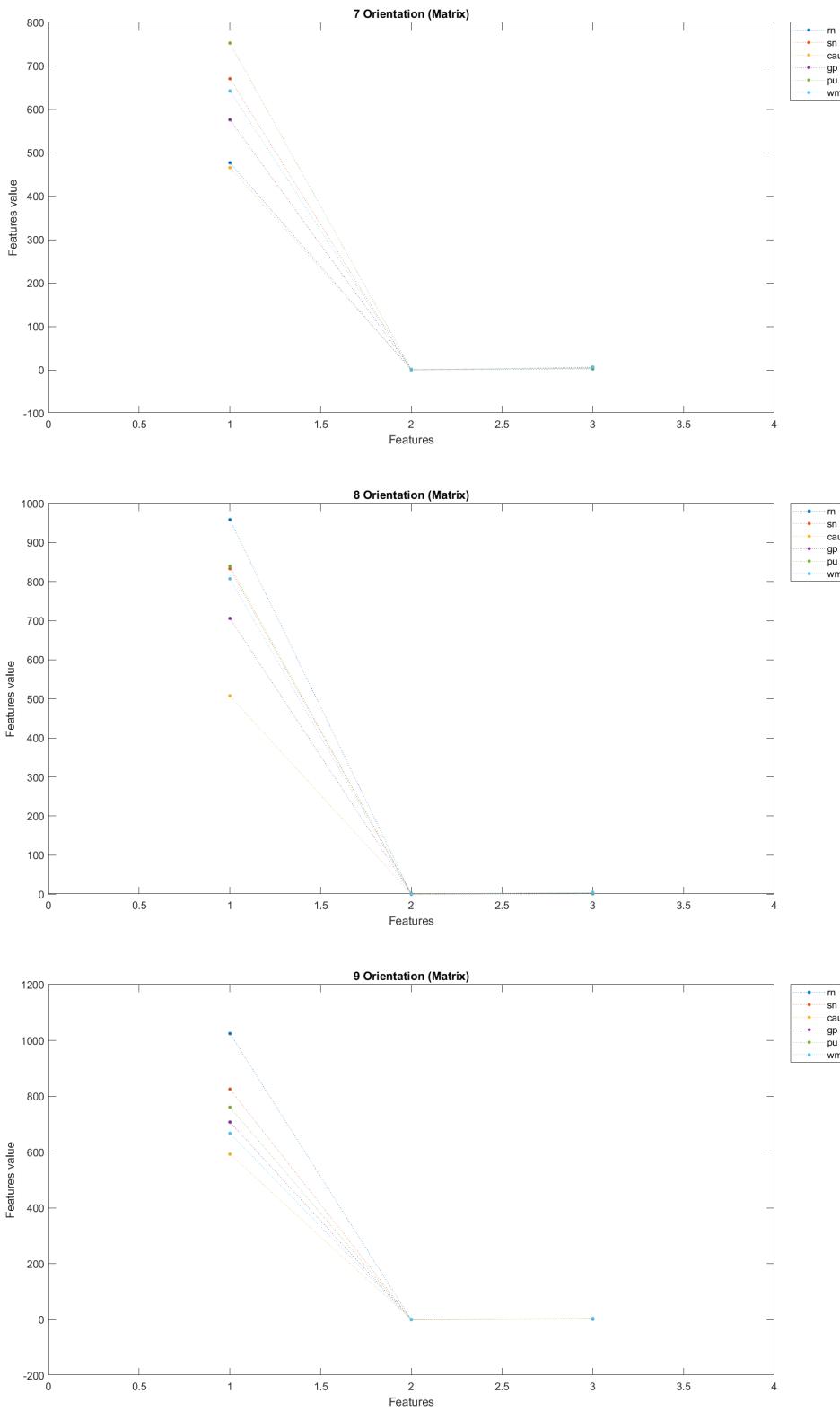


Figura 23: Misure globali per le ROI di interesse biologico: Orientazioni 7,8,9

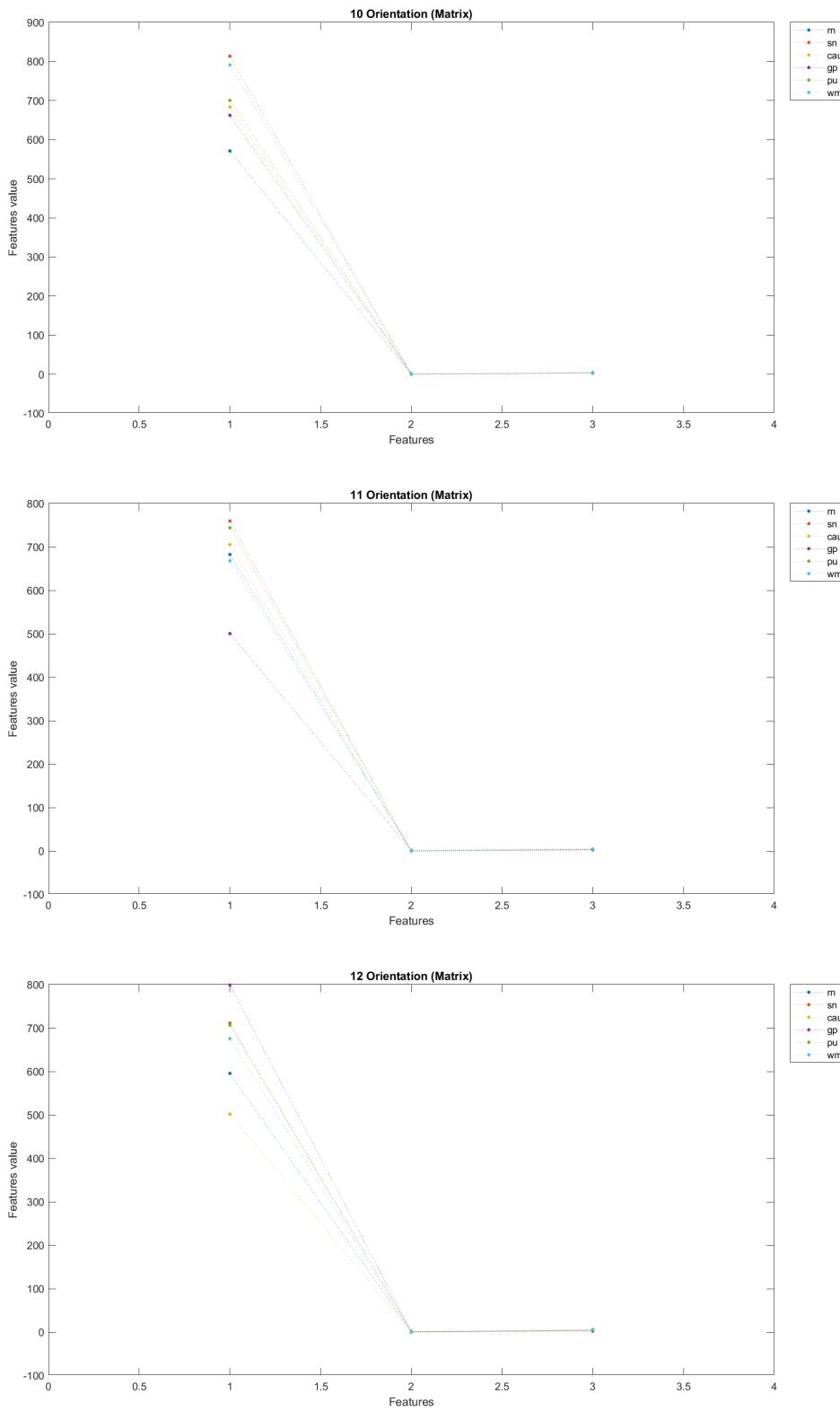


Figura 24: Misure globali per le ROI di interesse biologico: Orientazioni 10,11,12

## 6 Valutazione dei risultati

Le features ottenute nelle misurazioni descritte e mostrate nelle sezioni 4 e 5 possono essere implementate all'interno di reti neurali. A questo fine, per fare una prima riduzione, si sono individuate le features che variassero molto da un'orientazione all'altra, tenendo conto del fatto che in caso contrario, non si riuscirebbero ad esaminare le differenze tra le immagini e quindi a distinguerle e classificarle.

Si è fatto riferimento alla test t di Student che è un test per la verifica di ipotesi su un valore medio. Nello specifico questo verifica se il valore medio di una distribuzione si discosta significativamente da un certo valore di riferimento.

Tramite la funzione *ttest* implementata in MATLAB sono stati calcolati i valori h e p per ogni feature assumendo come valore atteso il valore della feature ottenuto nella prima orientazione, essendo quella 'standard'. In particolare:

- h è una variabile che assume valore:
  - 1 quando il *ttest* rigetta l'ipotesi nulla che i dati vengano da una distribuzione con una media predefinita con livello di significatività del 5%;
  - 0 quando il *ttest* non rigetta l'ipotesi nulla.
- p è una variabile che esprime in maniera quantitativa di quanto la distribuzione si discosti o meno dalla media ed assume valore:
  - <0.05 quando  $h = 0$ ;
  - >0.05 quando  $h = 1$ .

I risultati ottenuti sono rappresentati nelle tabelle 1 e 2. Per non appesantire la lettura delle tabelle sono riportati i risultati ottenuti dei valori p di cui quelli in grassetto indicano le features che variano nelle orientazioni quindi non sono variabili omogenee.

	Roi66	Roi76
Mean	<0.001	<b>0.711</b>
Skewness	<b>0.157</b>	<b>0.491</b>
Kurtosis	0.004	0.005
Entropy	<0.001	0.012
GLCM		
Contrast	<0.001	0.017
Correlation	<0.001	<0.001
Energy	<0.001	0.006
Homogeneity	<0.001	<b>0.159</b>

Tabella 1: Tabella dei valori p per le misure ottenute dalle immagini 2D

	Roi66	Roi66 rn	Roi66 sn	Roi76	Roi76 cau	Roi76 gp	Roi76 pu	Roi76 wm
<b>GLCM</b>								
Energy	0.001	0.002	0.001	<0.001	<b>0.946</b>	<b>0.084</b>	<b>0.064</b>	0.031
Contrast	<b>0.568</b>	<b>0.325</b>	<b>0.068</b>	0.015	<b>0.566</b>	0.005	<b>0.419</b>	<b>0.723</b>
Correlation	<b>0.358</b>	0.029	<b>0.374</b>	0.034	<b>0.279</b>	<b>0.354</b>	<b>0.704</b>	<b>0.422</b>
Homogeneity	<b>0.931</b>	0.001	<b>0.630</b>	0.004	<b>0.105</b>	<b>0.945</b>	0.018	<b>0.101</b>
Variance	<b>0.385</b>	<b>0.105</b>	<b>0.346</b>	0.040	<b>0.986</b>	0.013	<b>0.695</b>	<b>0.881</b>
Sum Average	0.042	0.005	<0.001	<b>0.081</b>	<b>0.254</b>	<b>0.187</b>	<b>0.243</b>	<b>0.507</b>
Entropy	<b>0.275</b>	<b>0.593</b>	0.015	<b>0.317</b>	<b>0.967</b>	<b>0.895</b>	<b>0.427</b>	<b>0.980</b>
Dissimilarity	<b>0.697</b>	<b>0.137</b>	<b>0.088</b>	0.016	<b>0.220</b>	0.019	<b>0.993</b>	<b>0.863</b>
Auto Correlation	0.048	0.004	<0.001	<b>0.075</b>	<b>0.239</b>	<b>0.174</b>	<b>0.311</b>	<b>0.587</b>
<b>GLRLM</b>								
SRE	<b>0.420</b>	0.001	<b>0.144</b>	0.008	<b>0.220</b>	<b>0.136</b>	0.035	0.001
LRE	<b>0.300</b>	0.002	0.002	<0.001	<b>0.433</b>	0.014	0.012	0.001
GLN	<b>0.437</b>	<b>0.057</b>	<b>0.372</b>	<b>0.489</b>	0.050	<b>0.354</b>	<b>0.206</b>	<b>0.431</b>
RLN	<b>0.430</b>	<0.001	<b>0.251</b>	0.009	<b>0.201</b>	<b>0.207</b>	0.037	<0.001
RP	<b>0.183</b>	0.001	0.007	<0.001	<b>0.362</b>	0.023	0.017	<0.001
LGRE	0.027	<b>0.417</b>	<b>0.052</b>	<b>0.074</b>	<b>0.644</b>	<b>0.436</b>	<b>0.280</b>	<b>0.109</b>
HGRE	0.041	0.002	<0.001	<b>0.071</b>	<b>0.439</b>	<b>0.206</b>	<b>0.506</b>	<b>0.699</b>
SRHGE	0.039	0.003	<0.001	<b>0.078</b>	<b>0.542</b>	<b>0.258</b>	<b>0.562</b>	<b>0.737</b>
LRLGE	0.037	<b>0.984</b>	0.001	<b>0.310</b>	<b>0.444</b>	<b>0.735</b>	<b>0.069</b>	<b>0.088</b>
GLV	<b>0.121</b>	0.030	0.001	<b>0.078</b>	<b>0.072</b>	<b>0.453</b>	<b>0.061</b>	<b>0.264</b>
RLV	<b>0.276</b>	0.007	<b>0.097</b>	<b>0.106</b>	<b>0.154</b>	<b>0.426</b>	0.017	0.007
<b>GLSZM</b>								
SZE	<b>0.974</b>	0.015	<b>0.945</b>	0.004	0.001	<b>0.641</b>	0.002	<b>0.548</b>
LZE	<b>0.469</b>	0.001	<b>0.151</b>	0.002	0.030	<b>0.238</b>	0.012	0.001
GLN	<b>0.273</b>	<b>0.714</b>	0.009	<b>0.427</b>	<b>0.624</b>	<b>0.179</b>	<b>0.407</b>	<b>0.354</b>
ZSN	<b>0.962</b>	0.016	<b>0.952</b>	0.004	0.001	<b>0.629</b>	0.002	<b>0.625</b>
ZP	<b>0.982</b>	0.002	<b>0.952</b>	0.008	0.007	<b>0.781</b>	0.041	0.004
LGZE	0.026	<b>0.090</b>	<b>0.105</b>	<b>0.072</b>	<b>0.470</b>	<b>0.585</b>	<b>0.263</b>	<b>0.079</b>
HGZE	0.043	0.002	<0.001	<b>0.071</b>	<b>0.450</b>	<b>0.353</b>	<b>0.538</b>	0.914
SZLGE	0.032	<b>0.078</b>	<b>0.153</b>	0.042	<b>0.490</b>	<b>0.984</b>	<b>0.289</b>	<b>0.084</b>
SZHGE	0.029	0.002	<0.001	<b>0.140</b>	<b>0.637</b>	<b>0.438</b>	<b>0.356</b>	<b>0.839</b>
GLV	0.034	<b>0.114</b>	0.001	<b>0.220</b>	<b>0.707</b>	<b>0.815</b>	0.027	<b>0.373</b>
ZSV	<b>0.638</b>	0.001	<b>0.541</b>	0.005	0.012	<b>0.506</b>	0.027	0.003
<b>NGTDM</b>								
Coarseness	<0.001	0.047	<b>0.072</b>	<0.001	<0.001	<b>0.299</b>	<b>0.120</b>	<b>0.075</b>
Contrast	<b>0.478</b>	<b>0.397</b>	<b>0.276</b>	<b>0.954</b>	<b>0.563</b>	<b>0.341</b>	<b>0.062</b>	<b>0.977</b>
Busyness	<b>0.192</b>	0.001	<0.001	<b>0.093</b>	0.003	<b>0.427</b>	<b>0.241</b>	<b>0.808</b>
Complexity	<b>0.333</b>	<b>0.350</b>	<0.001	<0.001	0.030	<b>0.442</b>	0.002	0.003
Strength	<b>0.457</b>	<b>0.184</b>	<0.001	0.016	<b>0.925</b>	<b>0.151</b>	<b>0.138</b>	<b>0.796</b>
<b>GLOBALE</b>								
Variance	<b>0.280</b>	<b>0.555</b>	0.013	<b>0.269</b>	<b>0.966</b>	<b>0.963</b>	<b>0.428</b>	<b>0.338</b>
Skewness	<b>0.875</b>	0.001	0.004	<b>0.655</b>	<b>0.317</b>	<b>0.420</b>	<b>0.111</b>	<b>0.184</b>
Kurtosis	<b>0.069</b>	<b>0.824</b>	<b>0.431</b>	<b>0.079</b>	<b>0.859</b>	<b>0.581</b>	<b>0.182</b>	<b>0.805</b>

Tabella 2: Tabella dei valori  $p$  per le misure ottenute dalle immagini 3D

## 6.1 PCA

La PCA (Analisi delle Componenti Principali) è una tecnica usata per la semplificazione dei dati nell'ambito della statistica multivariata. Il suo scopo è quello di ridurre il numero di variabili che descrivono un insieme di dati ad un numero minore, limitando il più possibile la perdita di informazioni. [6]

Se si considera un insieme di  $n$  osservazioni su  $p$  variabili, la variabilità totale del sistema può essere quantificata da un numero più piccolo delle  $p$  componenti principali ovvero  $k$  componenti (con  $k < p$ ) che le sostituiscono. In questo modo il set originario dei dati viene trasformato in un set di  $n$  misurazioni su  $k$  componenti principali.

Nello specifico la PCA usa una trasformazione ortogonale per convertire il set di osservazioni di variabili possibilmente correlate (di cui ogni entità prende vari valori numerici) in un set di valori di variabili non correlate linearmente (ovvero le componenti principali).

Questa trasformazione è definita in modo che la prima componente principale abbia la più grande varianza possibile e che ogni componente successiva abbia la più alta varianza possibile sotto la condizione che sia ortogonale a quelle precedenti.

In conclusione, questa trasformazione lineare delle variabili proietta le variabili originarie in un nuovo sistema cartesiano in cui vengono ordinate in ordine decrescente di varianza e i vettori risultanti (ognuno combinazione lineare delle variabili e contenente  $n$  osservazioni) sono un set di basi ortogonali non correlate.

La riduzione della complessità del sistema avviene limitandosi ad analizzare le 4 componenti principali, per varianza, delle nuove variabili.

La PCA è stata realizzata in MATLAB tramite la funzione *pca*, implementata in linee di codice simili alla seguente:

```
[coeff_cau , score_cau , latent_cau , ~ , explained_cau]=pca( t_cau );
```

dove:

- *coeff* = fornisce i coefficienti della componente principale per la matrice dei dati di dimensioni  $n \times p$ . Le righe di X corrispondono alle osservazioni, quindi nel nostro caso alle orientazioni, e le colonne corrispondono alle variabili, ovvero alle misurazioni (basate su matrici o globali). La matrice dei coefficienti è di dimensioni  $p \times p$ . Ogni colonna contiene i coefficienti per una componente principale e le colonne sono in ordine decrescente della varianza della componente. Di default, la PCA centra i dati e usa l'algoritmo SVD (di decomposizione di singolo valore);
- *score* = fornisce i punteggi delle componenti principali, ovvero le rappresentazioni di X nello spazio delle componenti principali. Le righe di score corrispondono alle osservazioni e le colonne alle componenti;
- *latent* = fornisce le varianze delle componenti principali, ovvero gli autovalori della matrice di covarianza di X;
- *explained* = fornisce la percentuale della varianza totale esplicitata da ogni componente principale.

Nelle figure sono rappresentati sia i coefficienti delle componenti principali ortonormali per ogni variabile che i loro scores. Tutte le 36 o 3 variabili sono rappresentate tramite un vettore la cui direzione e lunghezza indicano quanto quella variabile contribuisca alle due componenti principali.

La prima componente principale è sull'asse orizzontale mentre la seconda componente principale è sull'asse verticale.

Questo plot 2-D include anche un punto per ognuna delle 12 orientazioni le cui coordinate indicano la valutazione di ogni osservazione per le due componenti principali nel plot. Per esempio, i punti vicini al bordo sinistro del plot hanno le valutazioni più basse per la prima componente principale. I punti sono scalati rispetto al massimo valore di valutazione e alla lunghezza del massimo coefficiente, quindi solo le loro posizioni relative possono essere determinate dal plot.

Nei grafici mostrati in figura 25 e 26 sono mostrati i risultati della PCA per le 6 ROI di interesse biologico rispettivamente per le 36 misurazioni basate su matrici e per le 3 misurazioni globali.

Le percentuali di variabilità delle componenti principali, ovvero le direzioni lungo cui si registra la massima

dispersione dei dati, sono state verificate a posteriori con il parametro *explained* e tramite esse è stato regolato il numero di componenti principali da raffigurare. Nello specifico, si è usato un plot 2D (a 2 componenti principali) in tutti i casi in cui PC1 o PC1+PC2 superassero il 95% di variabilità. In caso contrario è stata aggiunta una terza componente (PC3) e si è usato un plot 3D (a 3 componenti principali). I risultati ottenuti sono riportati nelle tabelle 3 e 4.

	RN	SN	CAU	GP	PU	WM
PC1	99.32	91.46	86.52	95.06	80.29	70.80
PC2	0.67	8.48	13.18	4.79	19.59	29.04

Tabella 3: Percentuale di variabilità delle PCs nelle 36 misure basate su matrici

	RN	SN	CAU	GP	PU	WM
PC1	99.99	99.99	99.99	99.99	99.99	99.98

Tabella 4: Percentuale di variabilità delle PCs nelle 3 misure globali

Da un’analisi visiva dei grafici è possibile osservare che tra le features ’basate su matrici’ quelle che danno maggiore contributo (anche se in maniera diversa) alle due componenti principali nelle varie ROI sono:

- Componente 1:  $ngtdm_{con}$ ;
- Componente 2:  $glszm_{hgze}$ ,  $glrlm_{hgze}$ ,  $glcm_{ac}$ ,  $glszm_{szhge}$ ,  $glszm_{srhge}$ .

Da un’analisi visiva dei grafici è possibile osservare che tra le features ’globali’ quelle che danno maggiore contributo (anche se in maniera diversa) alle due componenti principali nelle varie ROI sono:

- Componente 1:  $g_v$ ;
- Componente 2:  $g_s$  (per CAU),  $g_k$  (per tutte le altre ROI).

In realtà è possibile notare come le features appena menzionate per le misurazioni basate su matrici siano quelle aventi i più grandi valori in termini di ordini di grandezza. Quindi, in appendice E è mostrata l’analisi PCA in cui sono stati raggruppati i dati per 3 intervalli di ordini di grandezza differenti.

Inoltre per svolgere un’analisi completa, in appendice F è mostrata l’analisi PCA per le features ’migliori’ in comune alle 6 ROI di interesse biologico.

Nei grafici mostrati in figura 26 sono mostrati i risultati della PCA per le 6 ROI di interesse biologico per le 3 misurazioni globali.

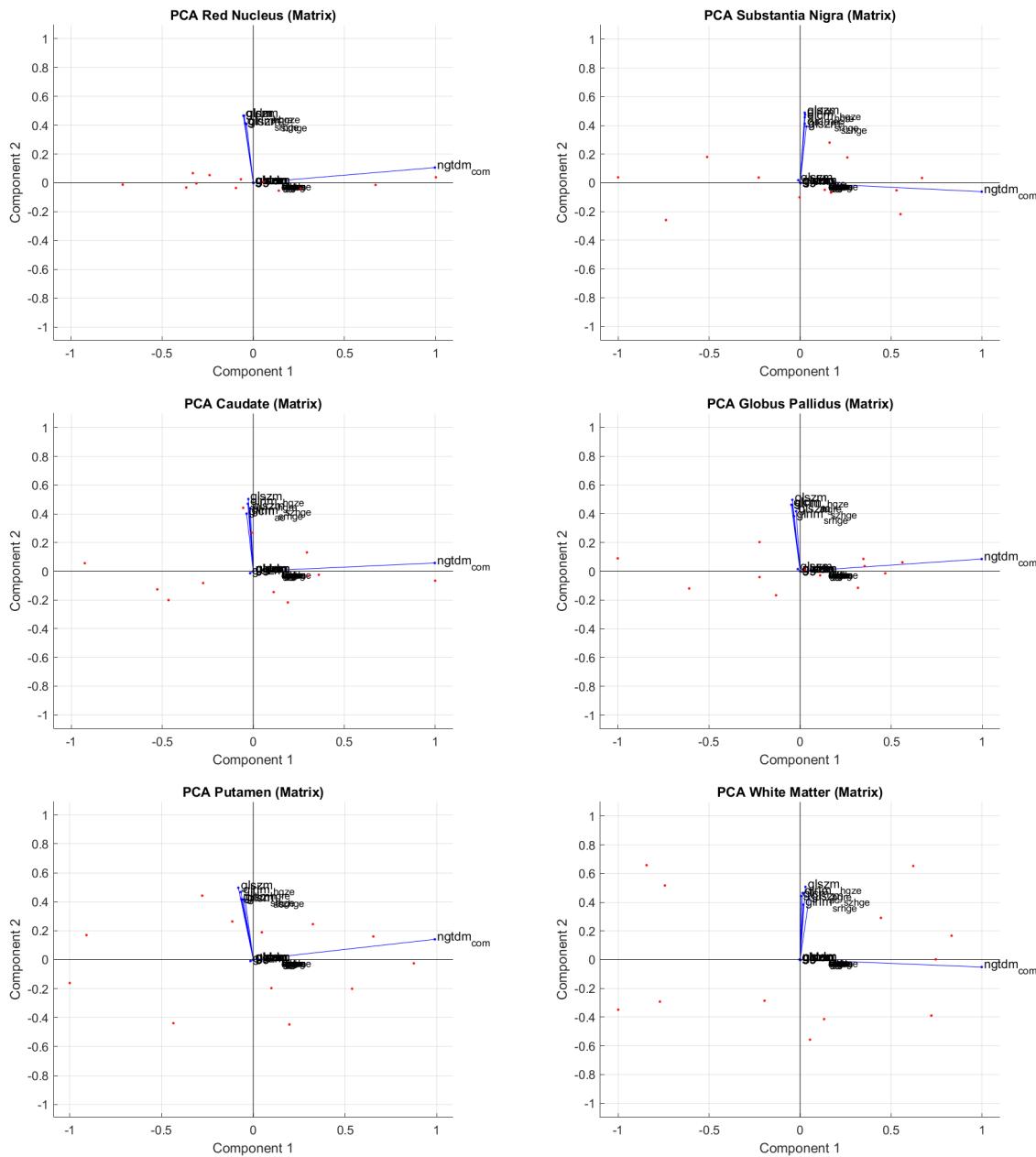


Figura 25: PCA per le misure basate su matrici per le ROI di interesse biologico

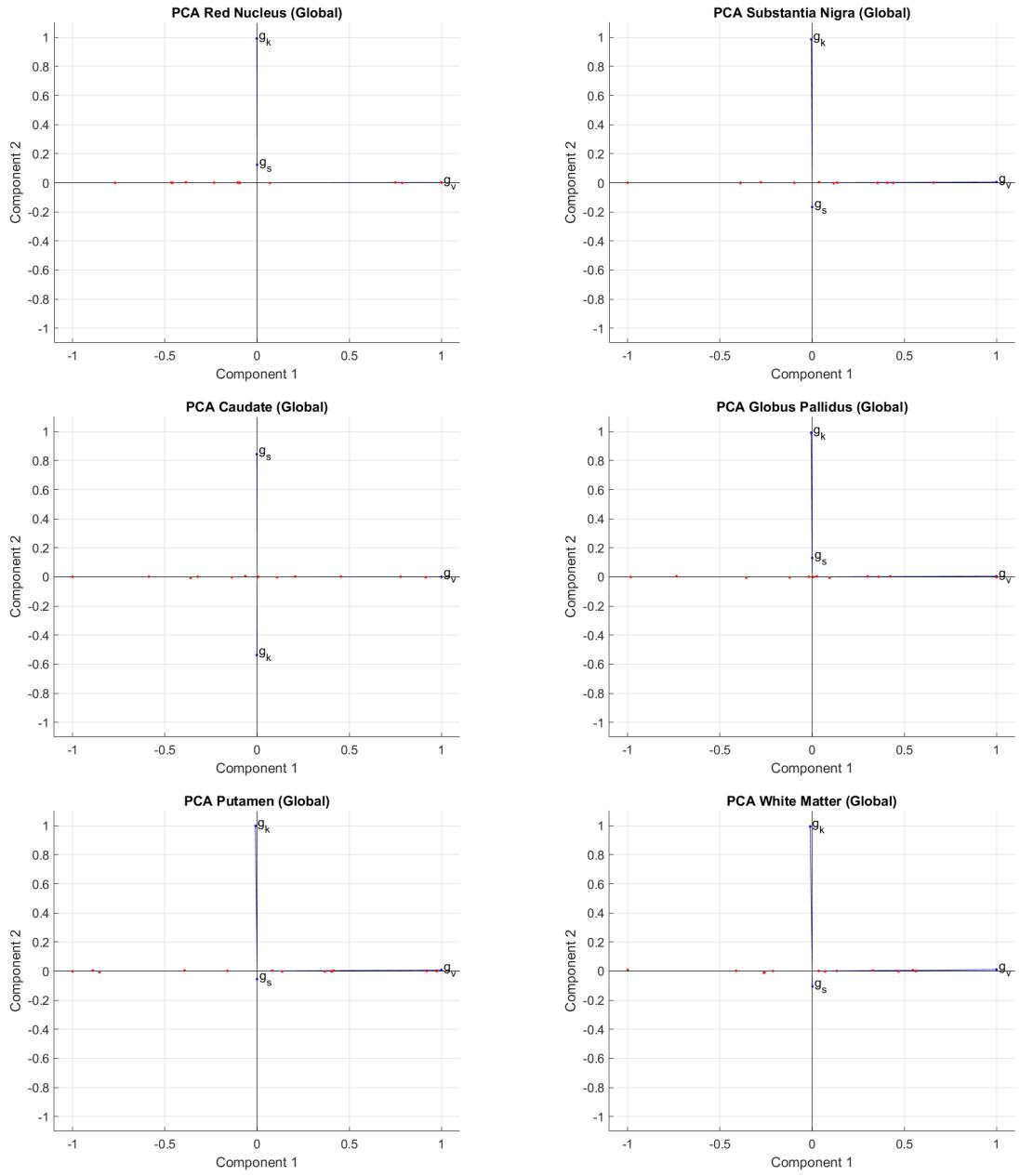


Figura 26: PCA per le misure globali per le ROI di interesse biologico

## 7 Conclusioni e lavori futuri

Questo progetto ha previsto l'implementazione e la discussione di vari approcci statistici per la descrizione e l'analisi delle texture in immagini di risonanza magnetica di fase. La texture analysis rappresenta una tecnica di valutazione quantitativa che può essere usata potenzialmente in aggiunta a imaging diagnostico. Infine, estraendo features tramite la texture analysis si potrebbero creare reti neurali per la classificazione di immagini/tessuti, tenendo conto dei risultati ottenuti.

Allo stesso modo, tenendo conto delle features "costanti", si potrebbero individuare con maggior precisione ROI e VOI all'interno di immagini di risonanza magnetica di fase, permettendo di velocizzare il processo di segmentazione o di individuare anomalie.

## A Appendice

Di seguito sono riportati gli istogrammi su cui è stato già fatto lo stretching, per le 12 orientazioni e per le roi 66 e 76, rispettivamente relative alla slice con  $z=66$  e  $z=76$ .

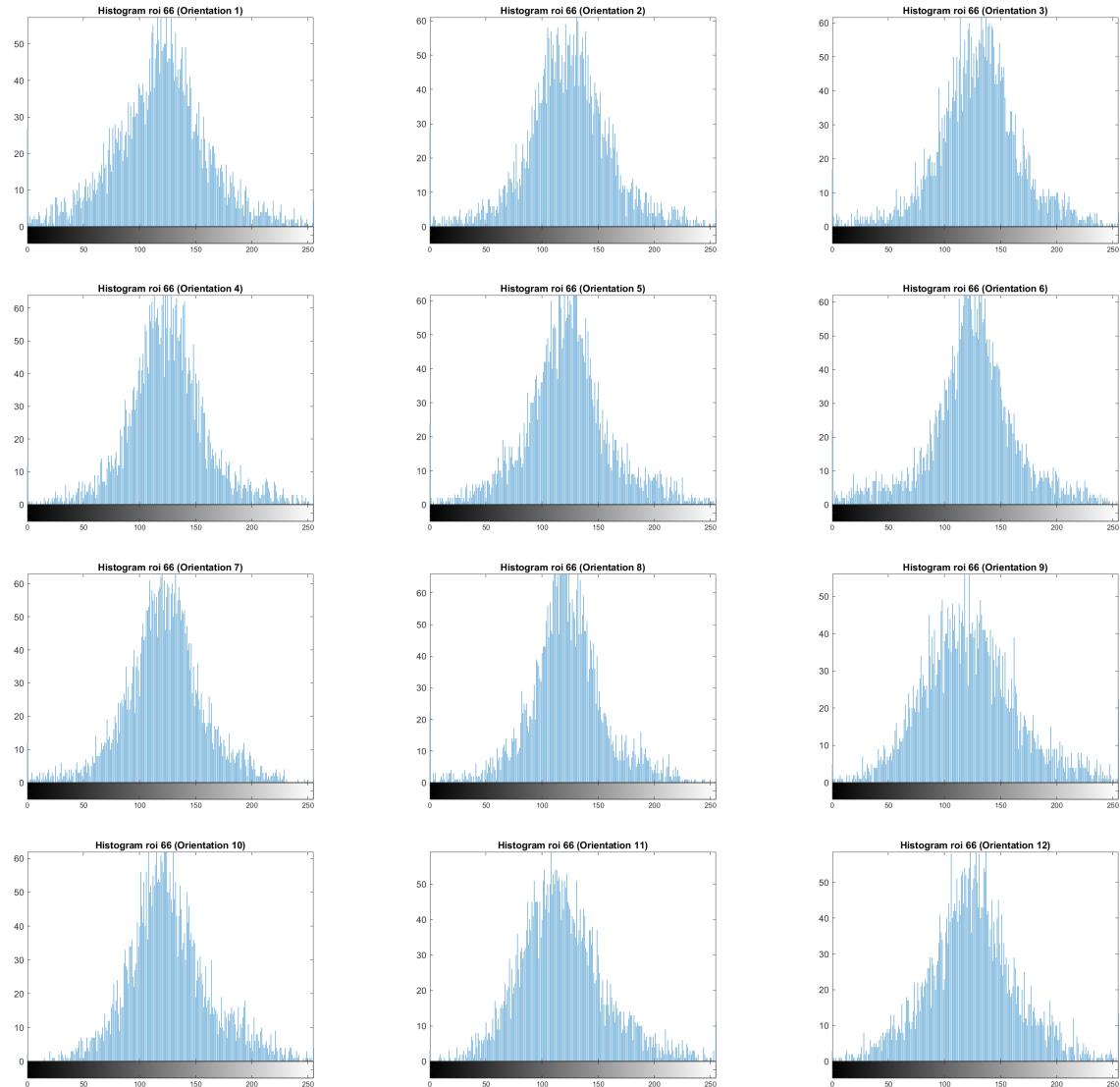


Figura 27: Istogrammi della roi 66 (Immagine con  $z=66$ )

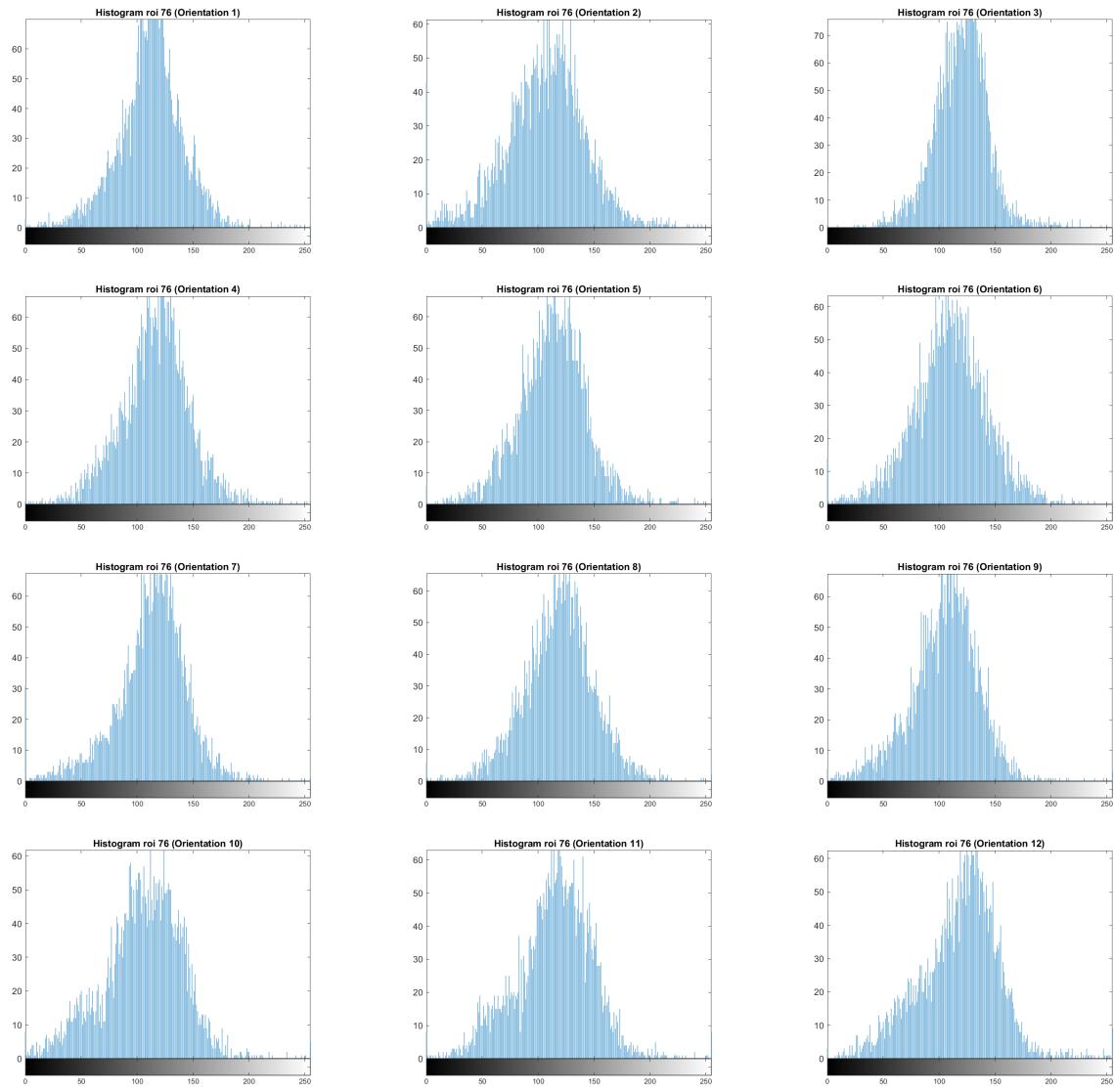


Figura 28: Istogrammi della roi 76 (Immagine con z=76)

## B Appendice

Di seguito sono mostrate le matrici di co-occorrenza.

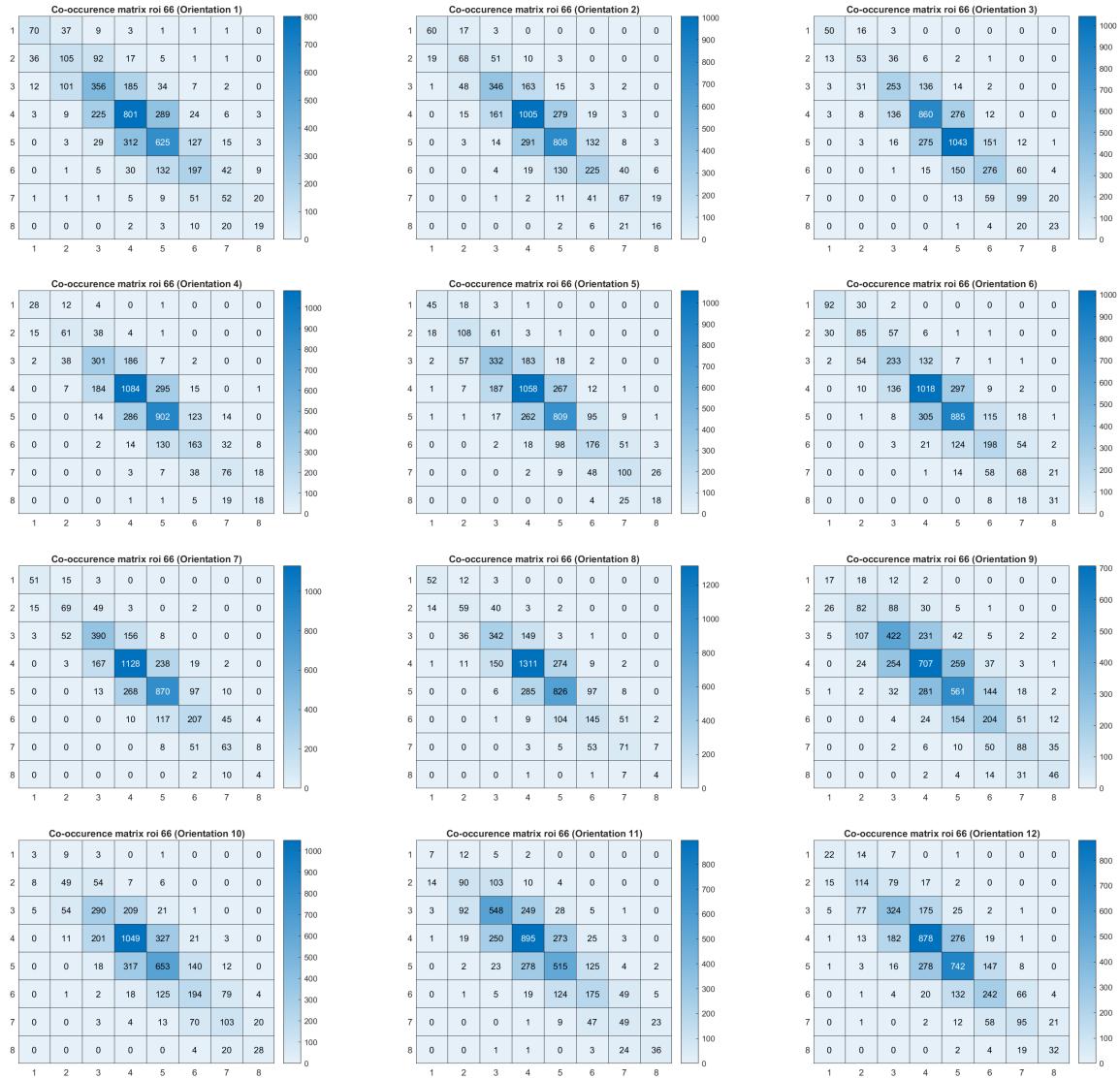


Figura 29: Matrici di co-occorrenza della roi 66 (Immagine con z=66)

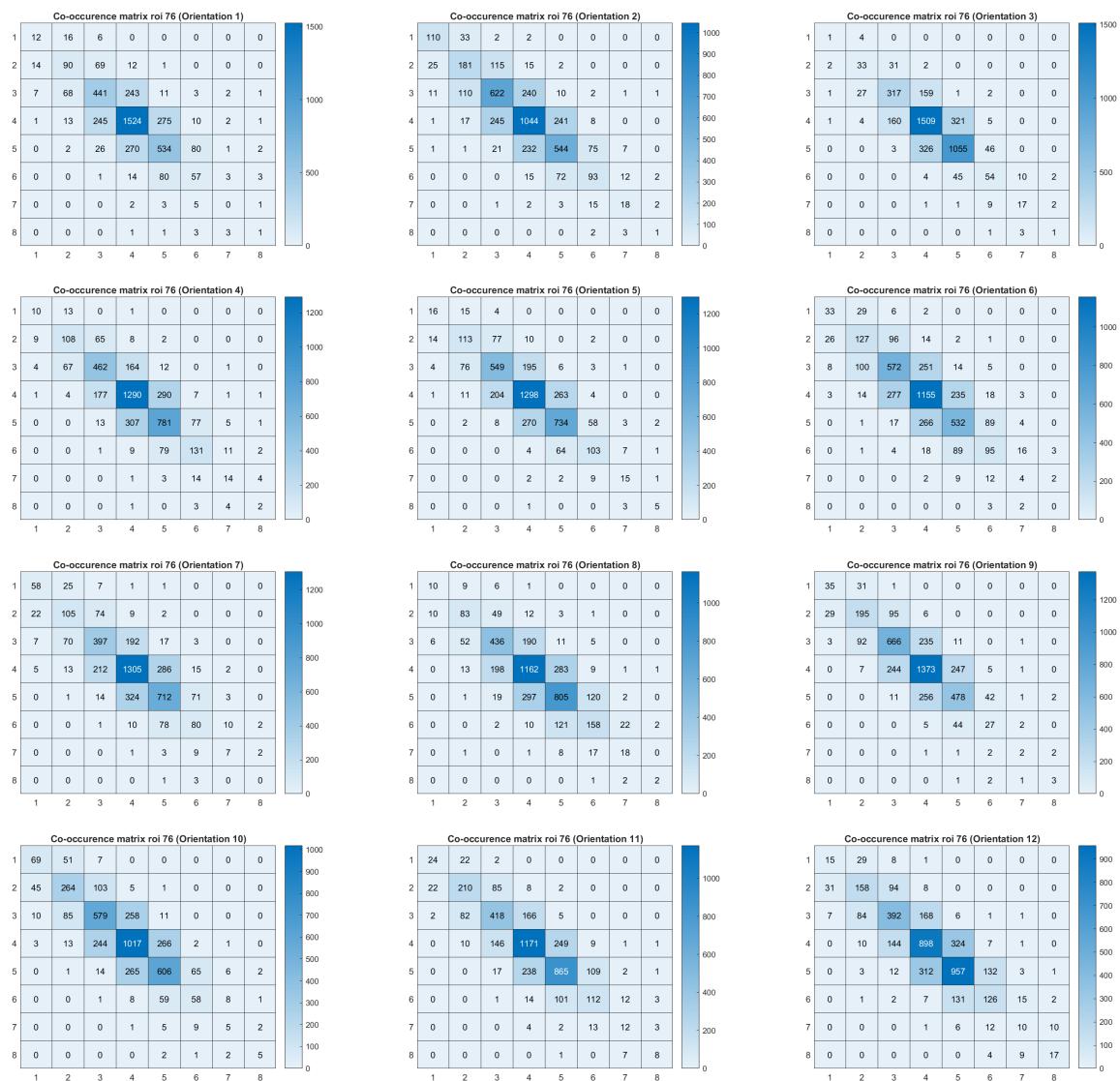


Figura 30: Matrici di co-occorrenza della roi 76 (Immagine con z=76)

## C Appendice

Di seguito sono mostrati gli screen relativi ai tools che consentono la visualizzazione di immagini mediche 3D in MATLAB e che sono stati utilizzati in questo progetto.

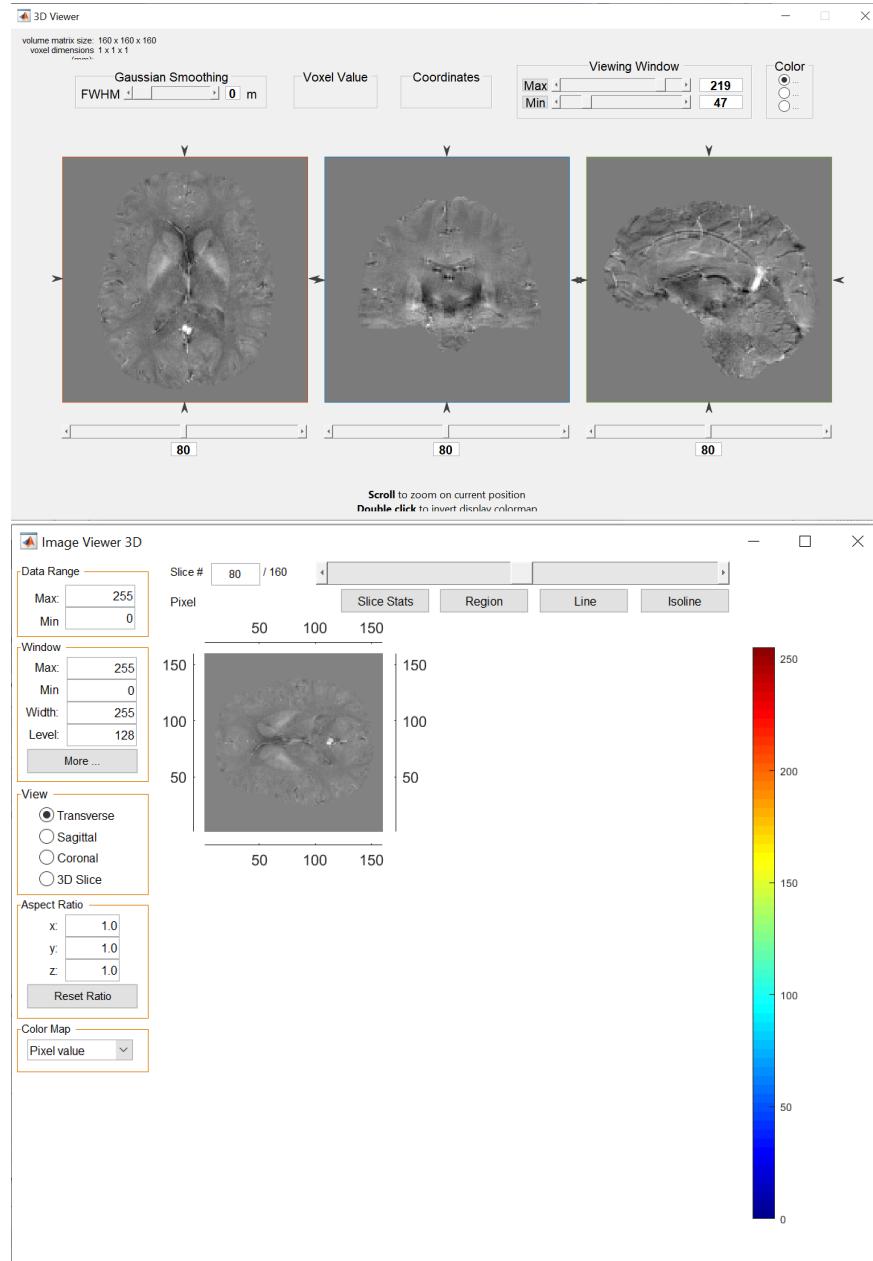


Figura 31: In alto lo screen relativo al tool *Medical Image Reader and Viewer*, in basso lo screen relativo al tool *3D image viewer and slicer*

## D Appendice

Di seguito sono mostrati i risultati ottenuti usando il pacchetto RADIOMICS di MATLAB per la roi76 e i riferimenti biologici sia della roi66 che della roi76.

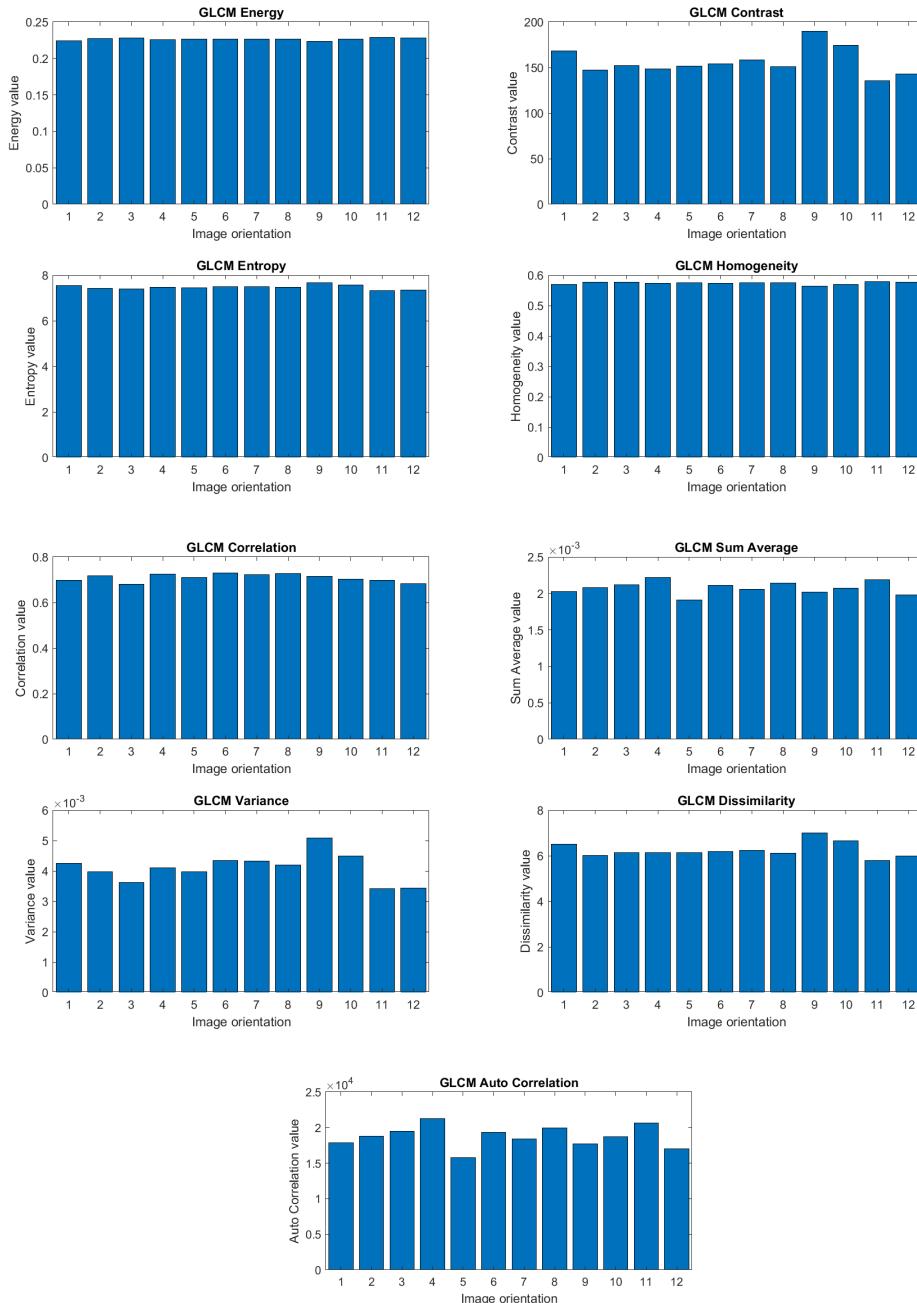


Figura 32: Misure statistiche della GLCM della roi 76 (Volume 3D)

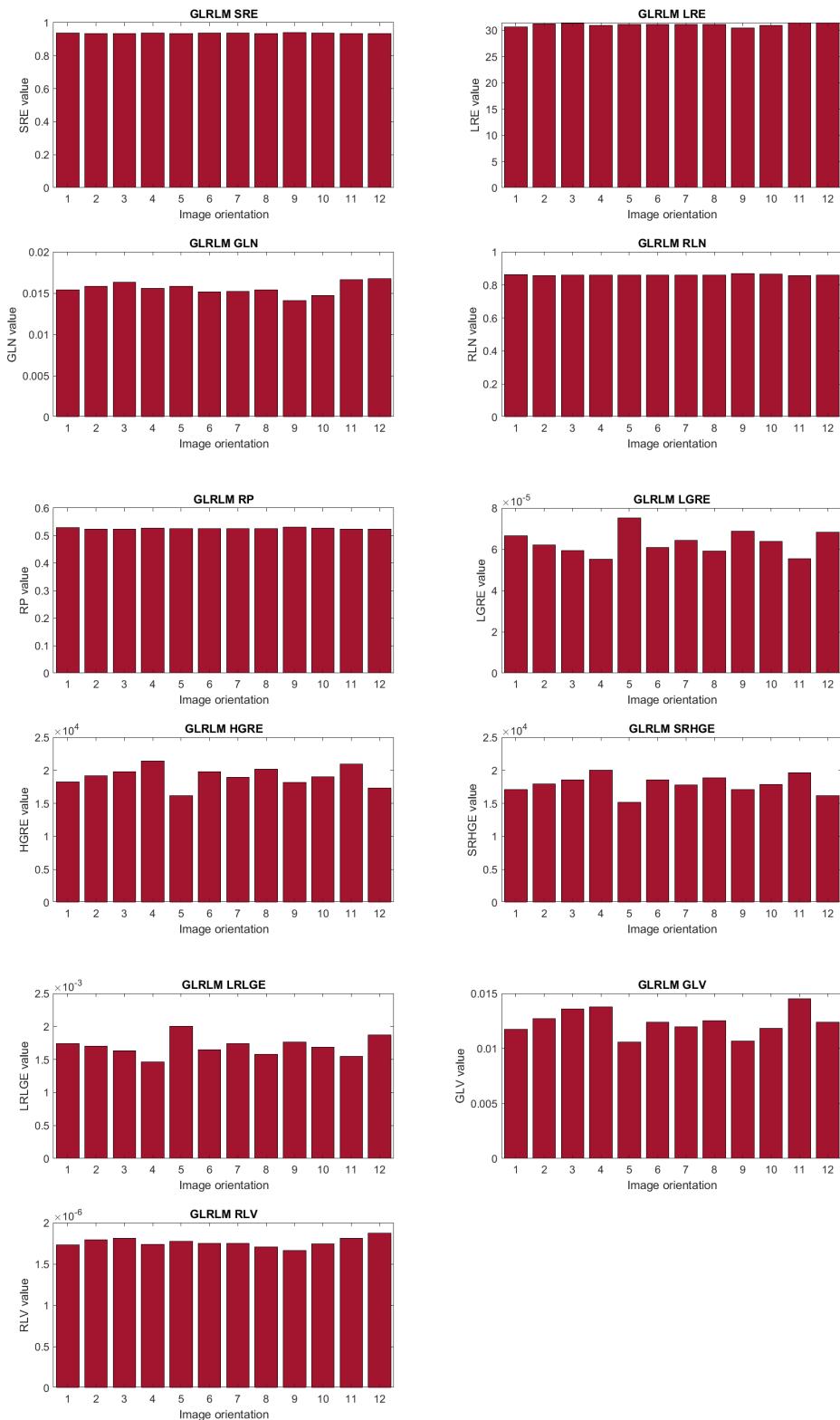


Figura 33: Misure statistiche della GLMRL della roi 76 (Volume 3D)

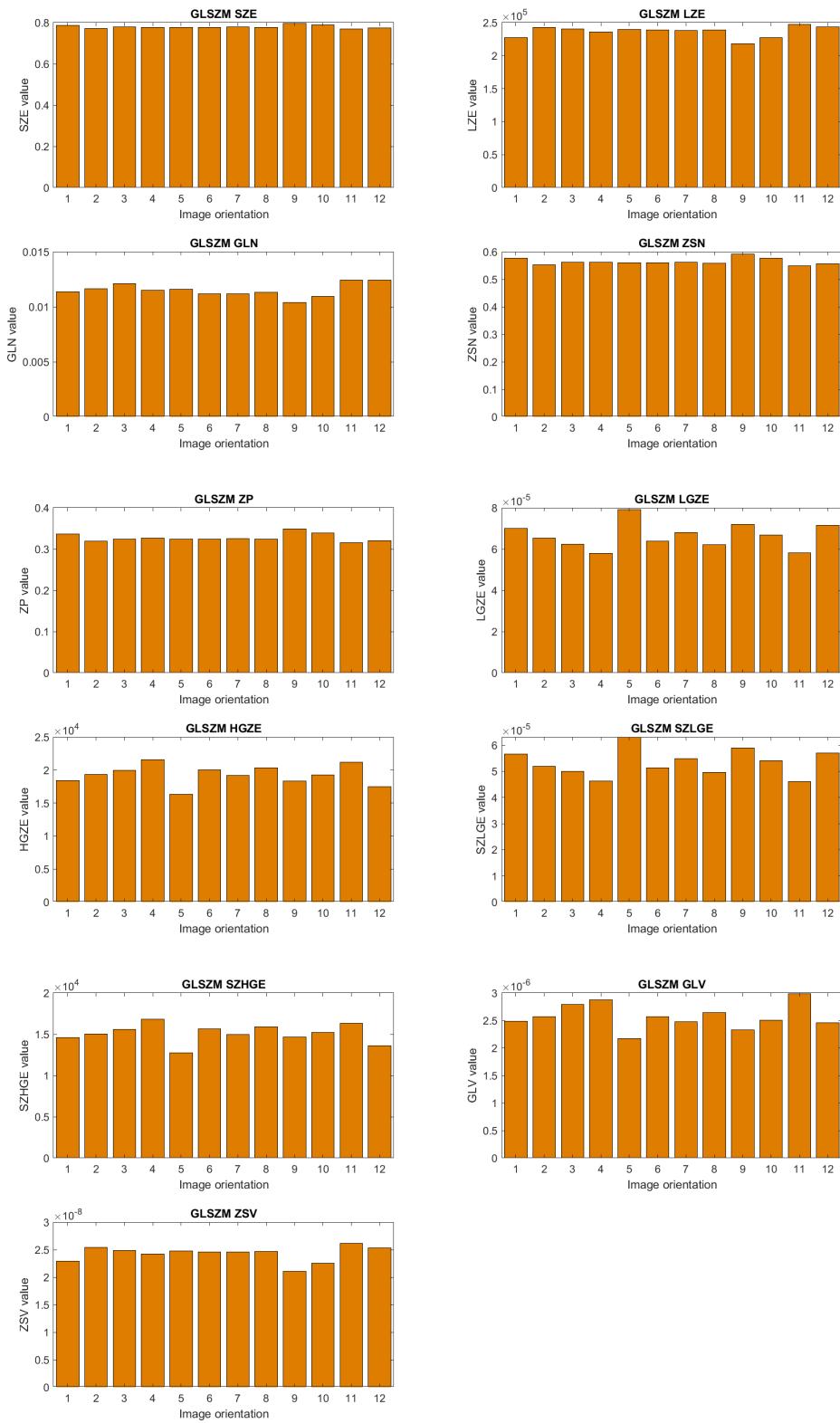


Figura 34: Misure statistiche della GLSzm della roi 76 (Volume 3D)

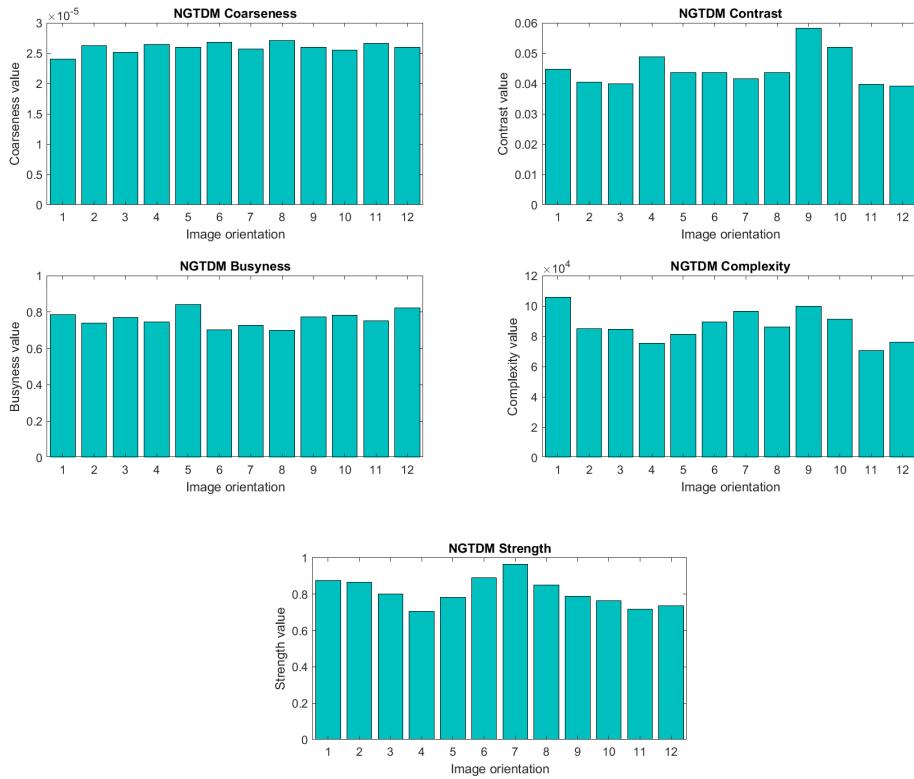


Figura 35: Misure statistiche della NGTDM della roi 76 (Volume 3D)

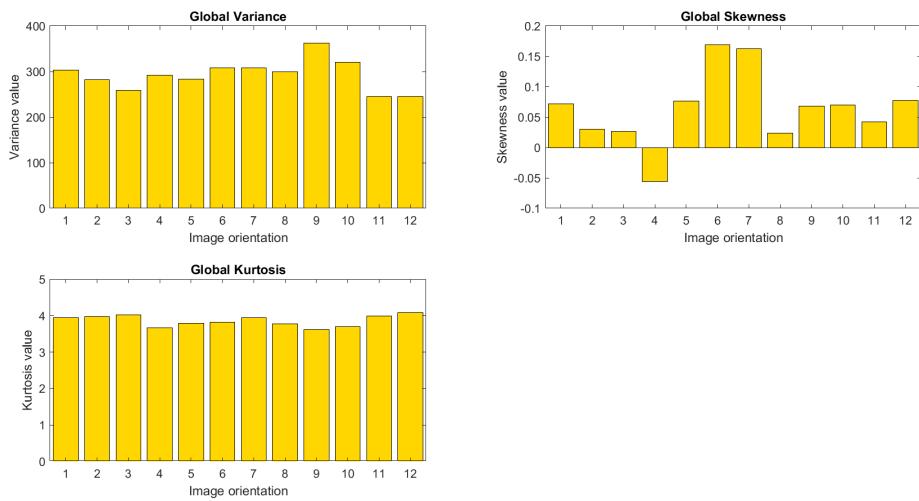


Figura 36: Misure statistiche globali nella roi 76 (Volume 3D)

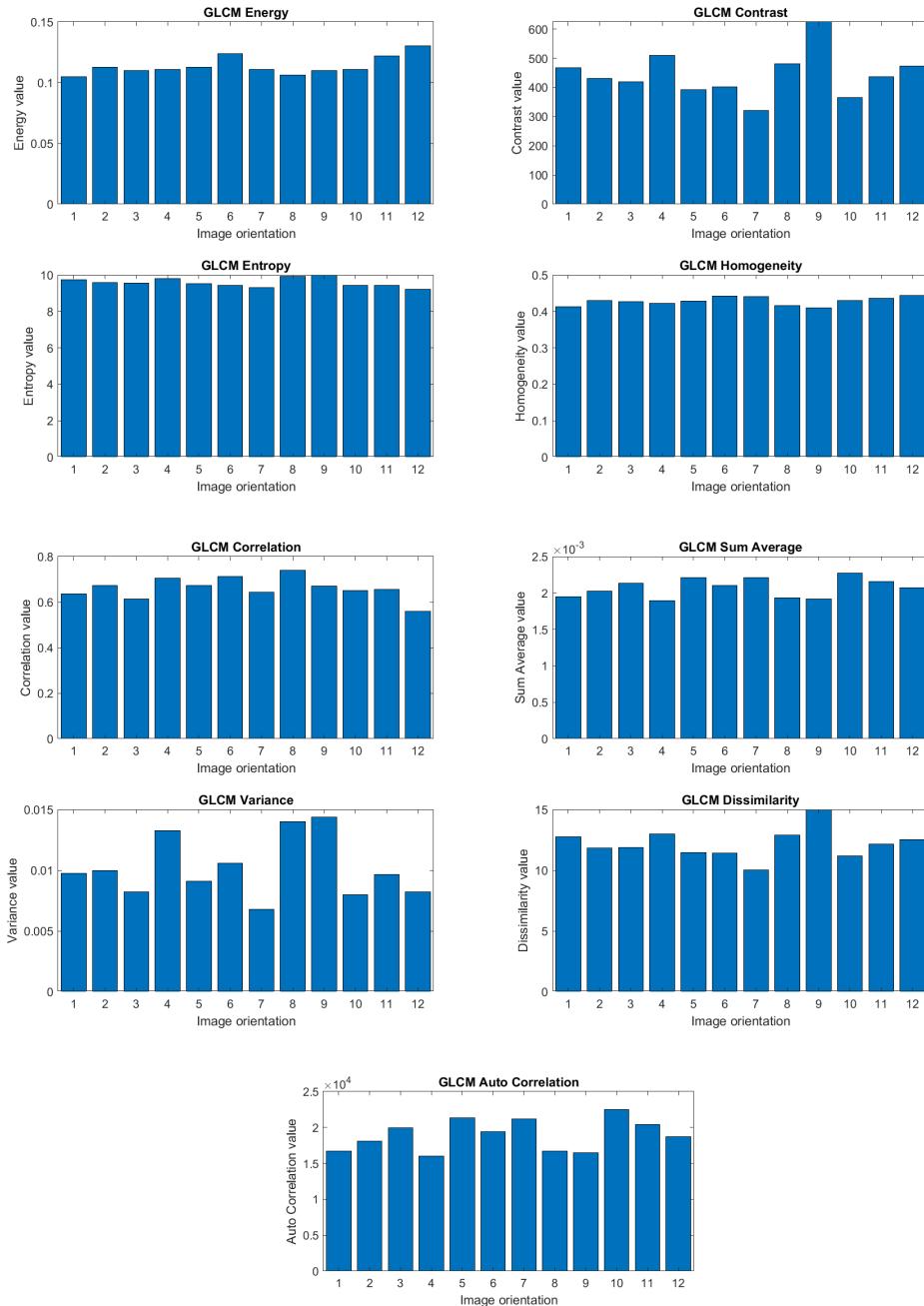


Figura 37: Misure statistiche della GLCM del Red Nucleus nella roi 66 (Volume 3D)

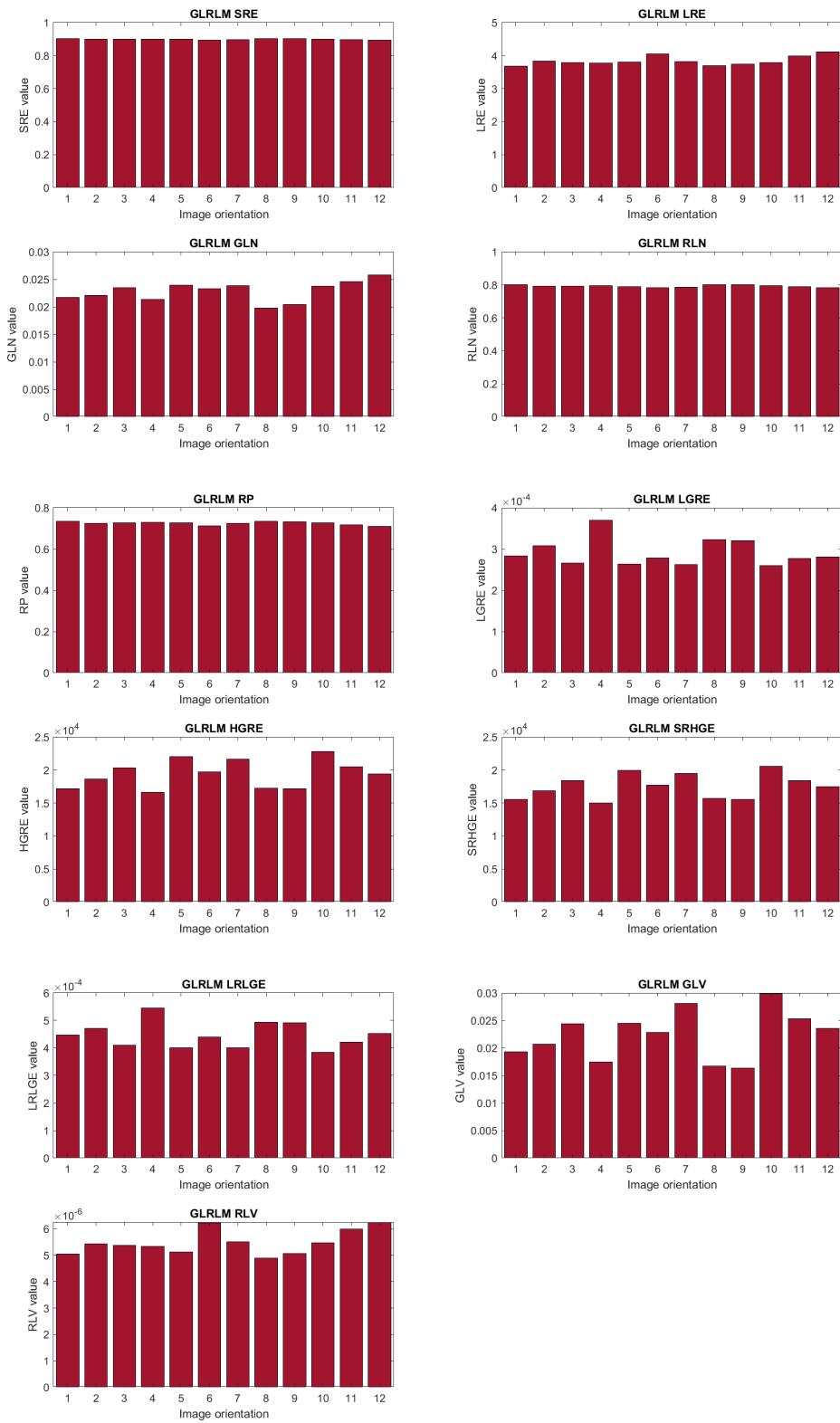


Figura 38: Misure statistiche della GLMRL del Red Nucleus nella roi 66 (Volume 3D)

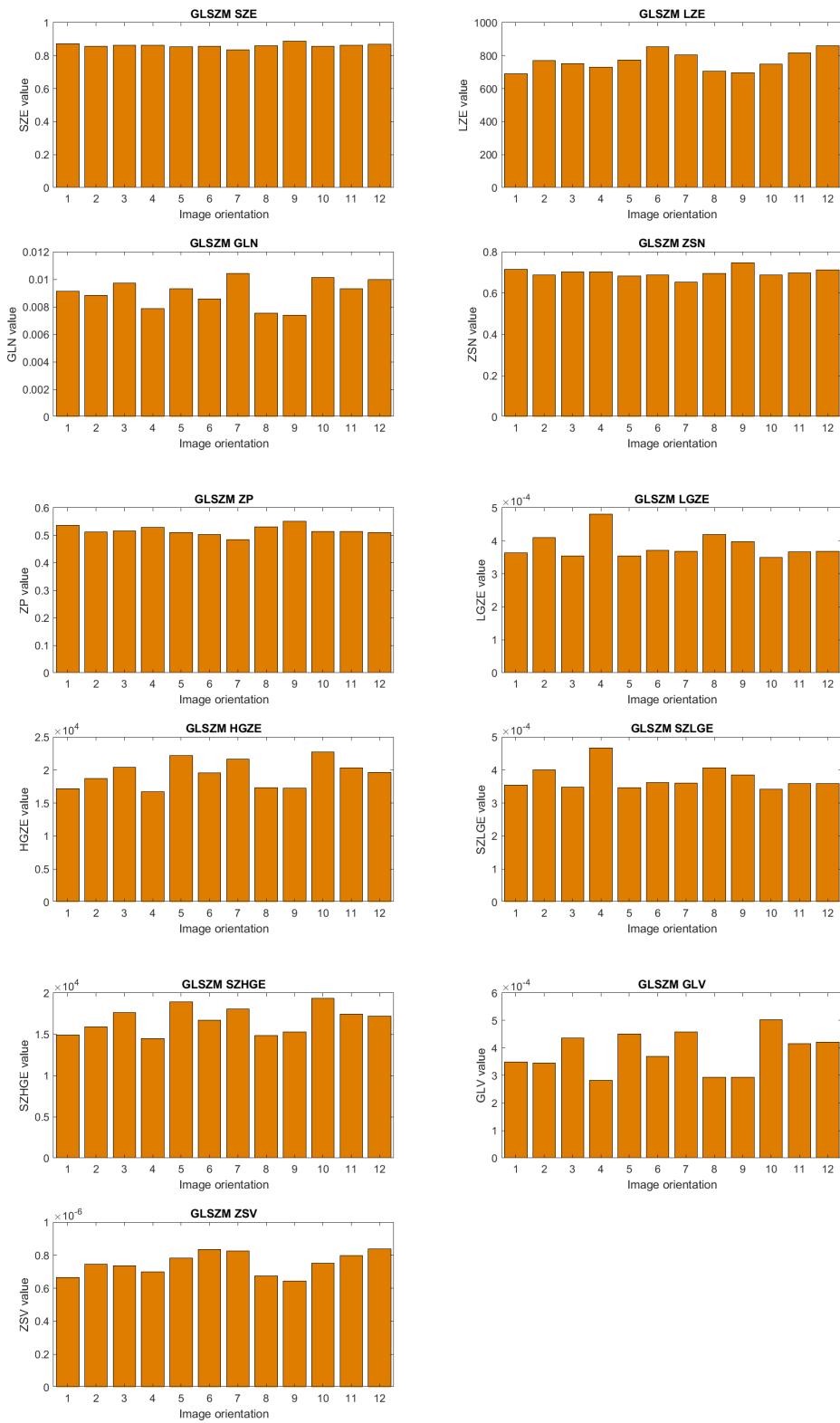


Figura 39: Misure statistiche della GLSzm del Red Nucleus nella roi 66 (Volume 3D)

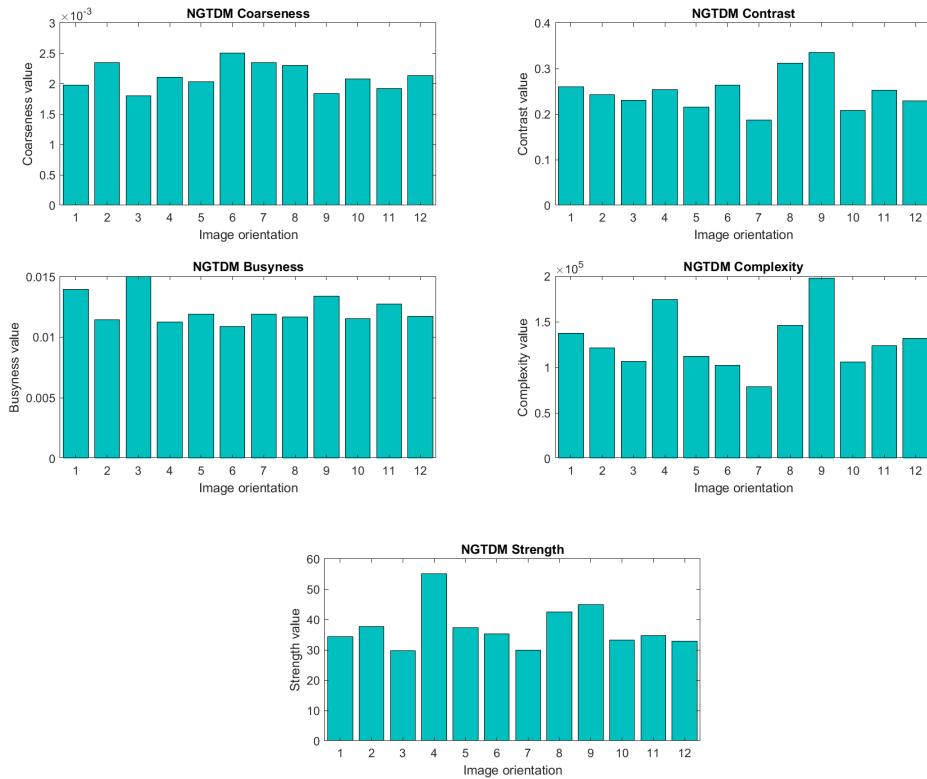


Figura 40: Misure statistiche della NGTDM del Red Nucleus nella roi 66 (Volume 3D)

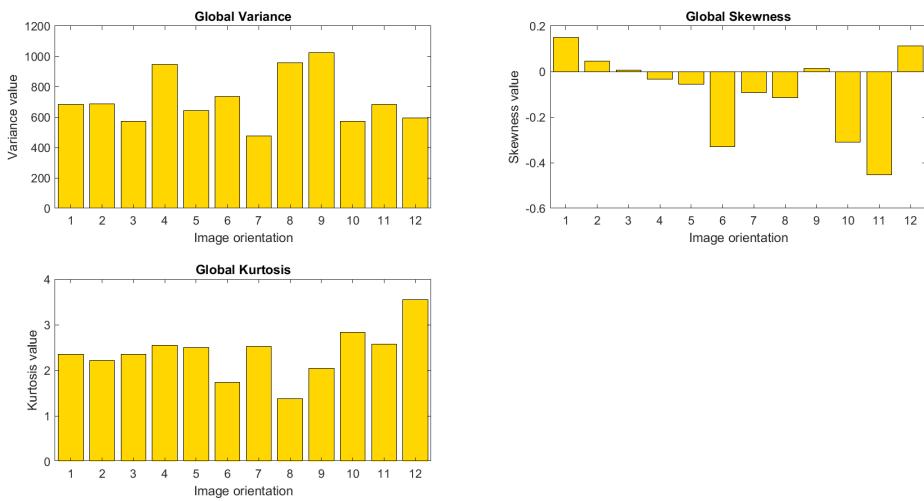


Figura 41: Misure statistiche globali del Red Nucleus nella roi 66 (Volume 3D)

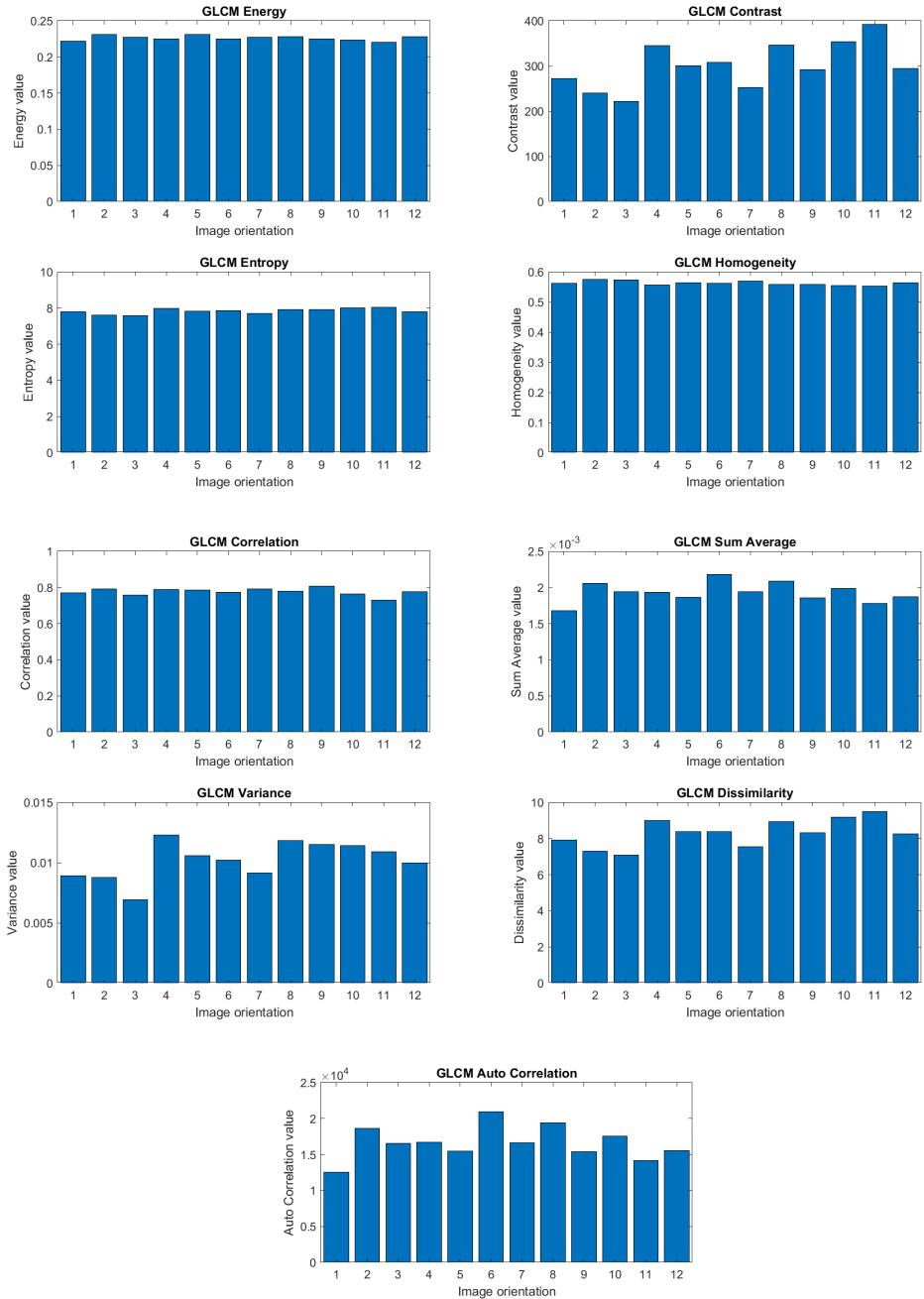


Figura 42: Misure statistiche della GLCM della Substantia Nigra nella roi 66 (Volume 3D)

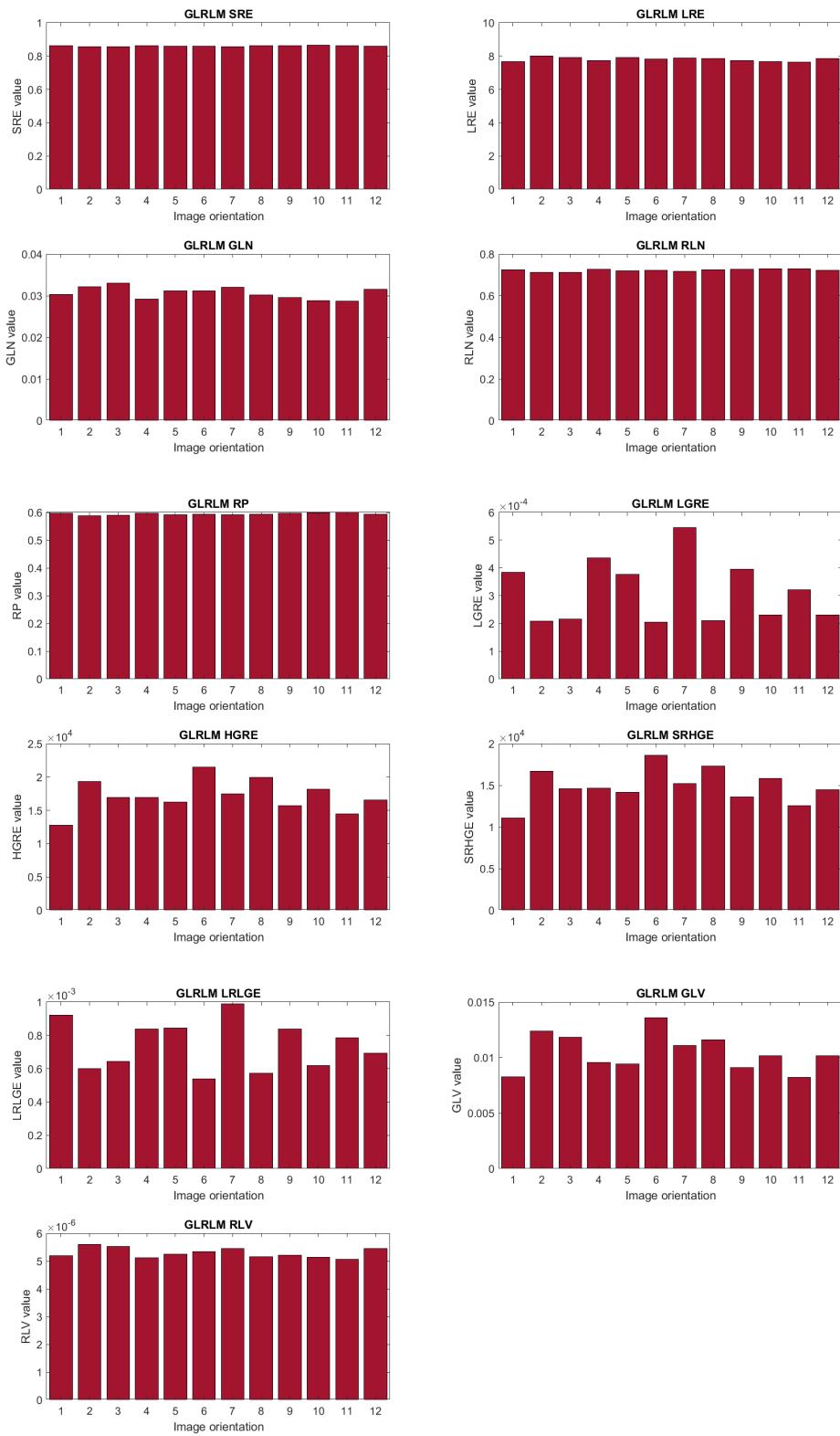


Figura 43: Misure statistiche della GLMRL della Substantia Nigra nella roi 66 (Volume 3D)

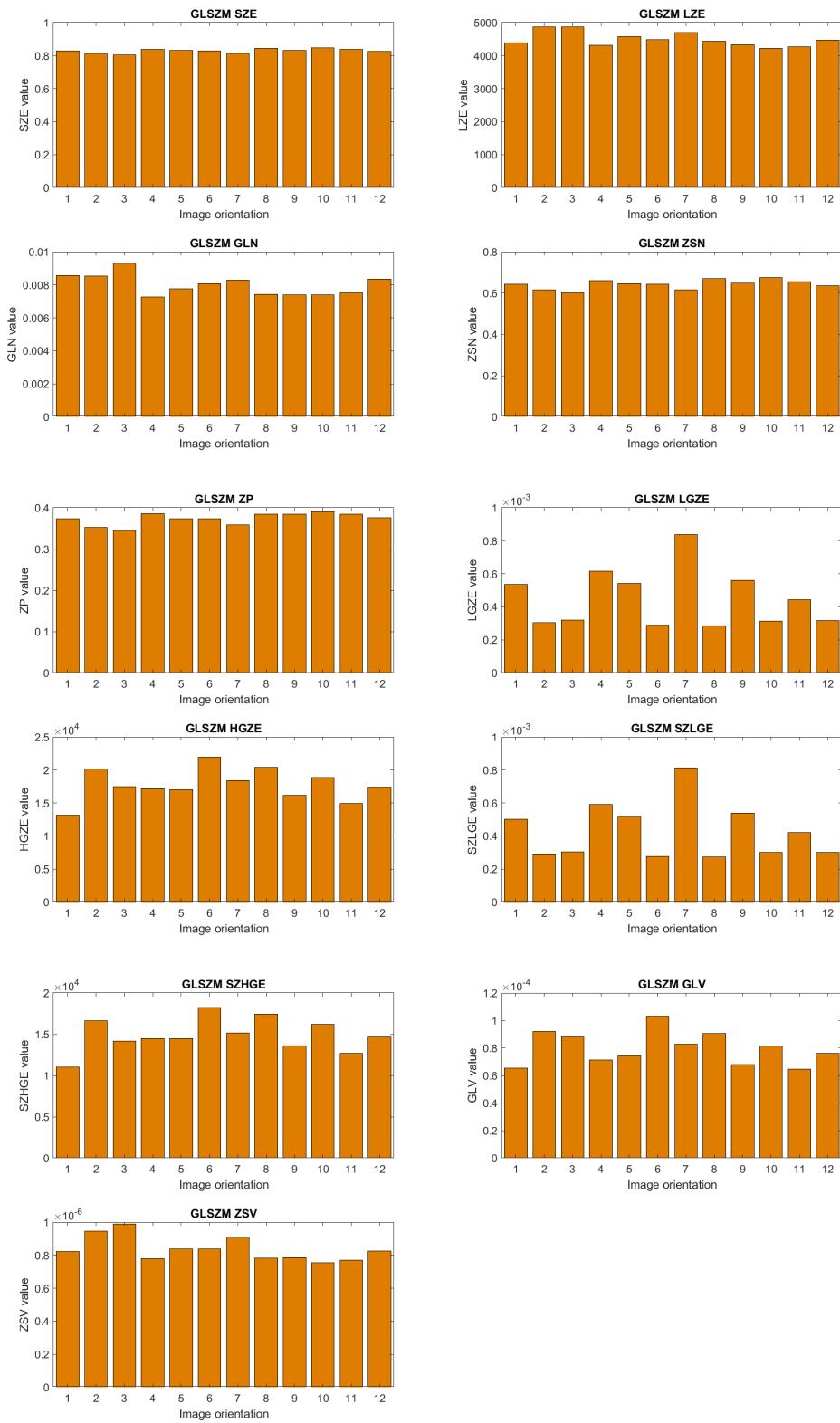


Figura 44: Misure statistiche della GLSzm della Substantia Nigra nella roi 66 (Volume 3D)

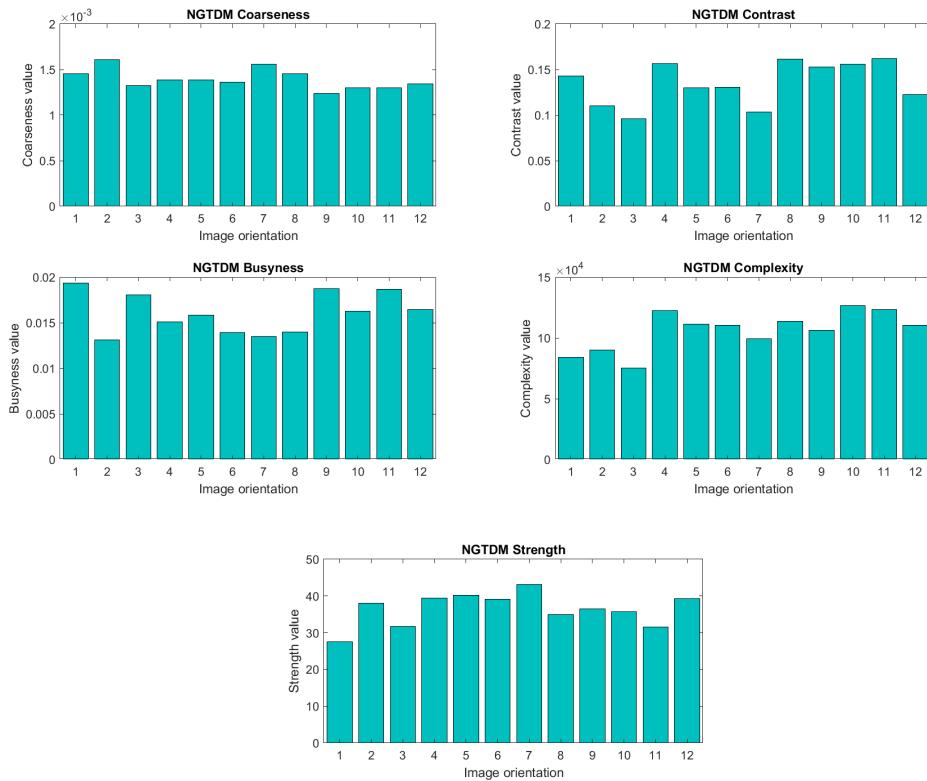


Figura 45: Misure statistiche della NGTDM della Substantia Nigra nella roi 66 (Volume 3D)

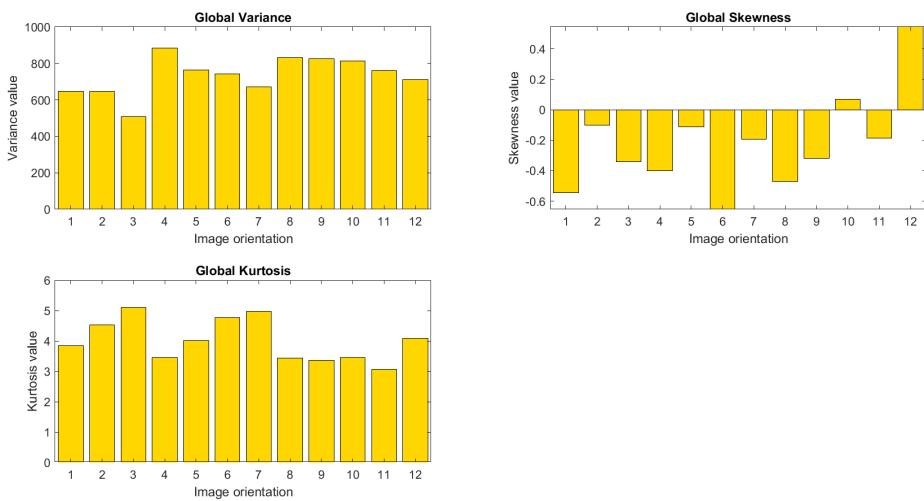


Figura 46: Misure statistiche globali della Substantia Nigra nella roi 66 (Volume 3D)

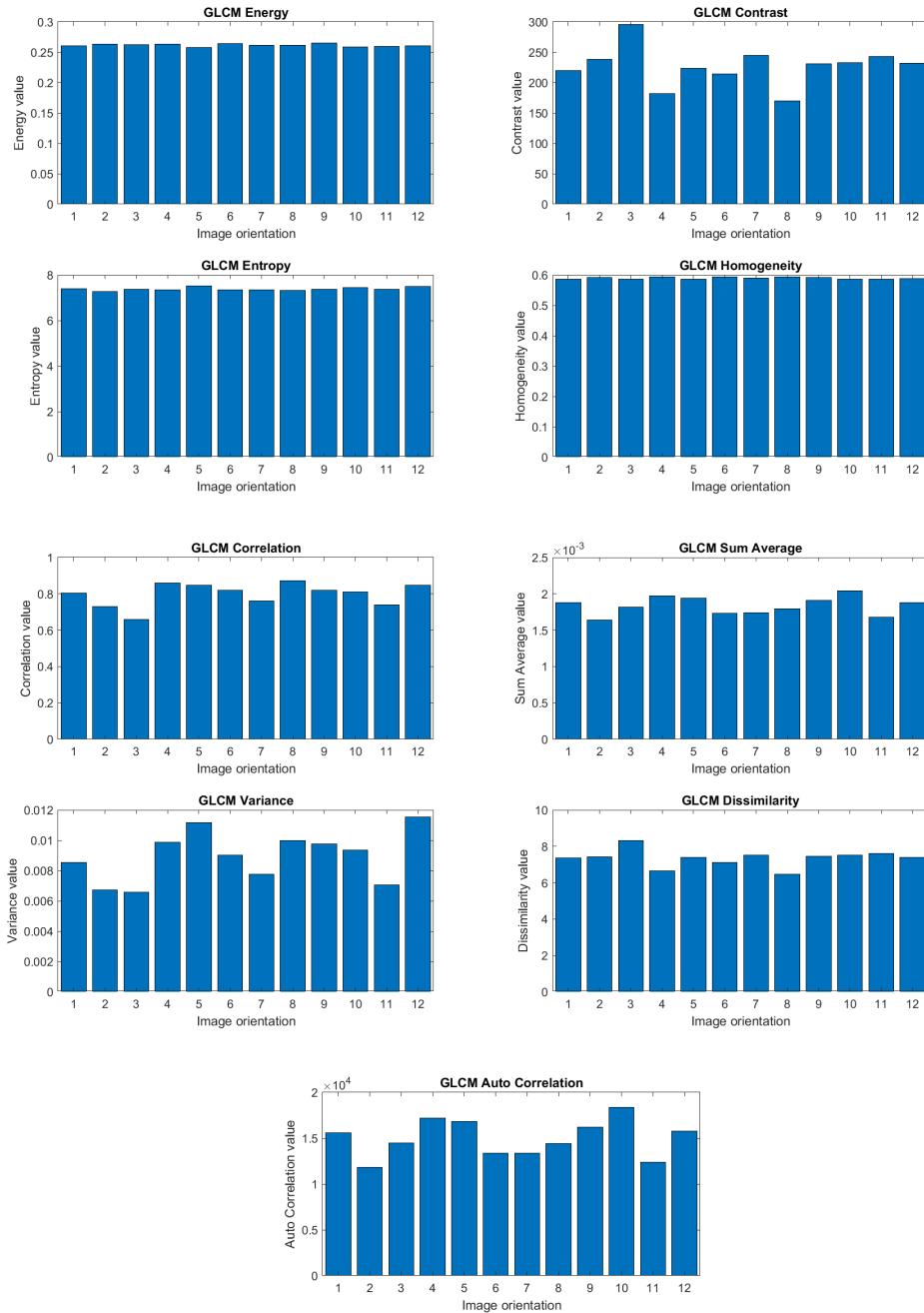


Figura 47: Misure statistiche della GLCM del Putamen nella roi 76 (Volume 3D)

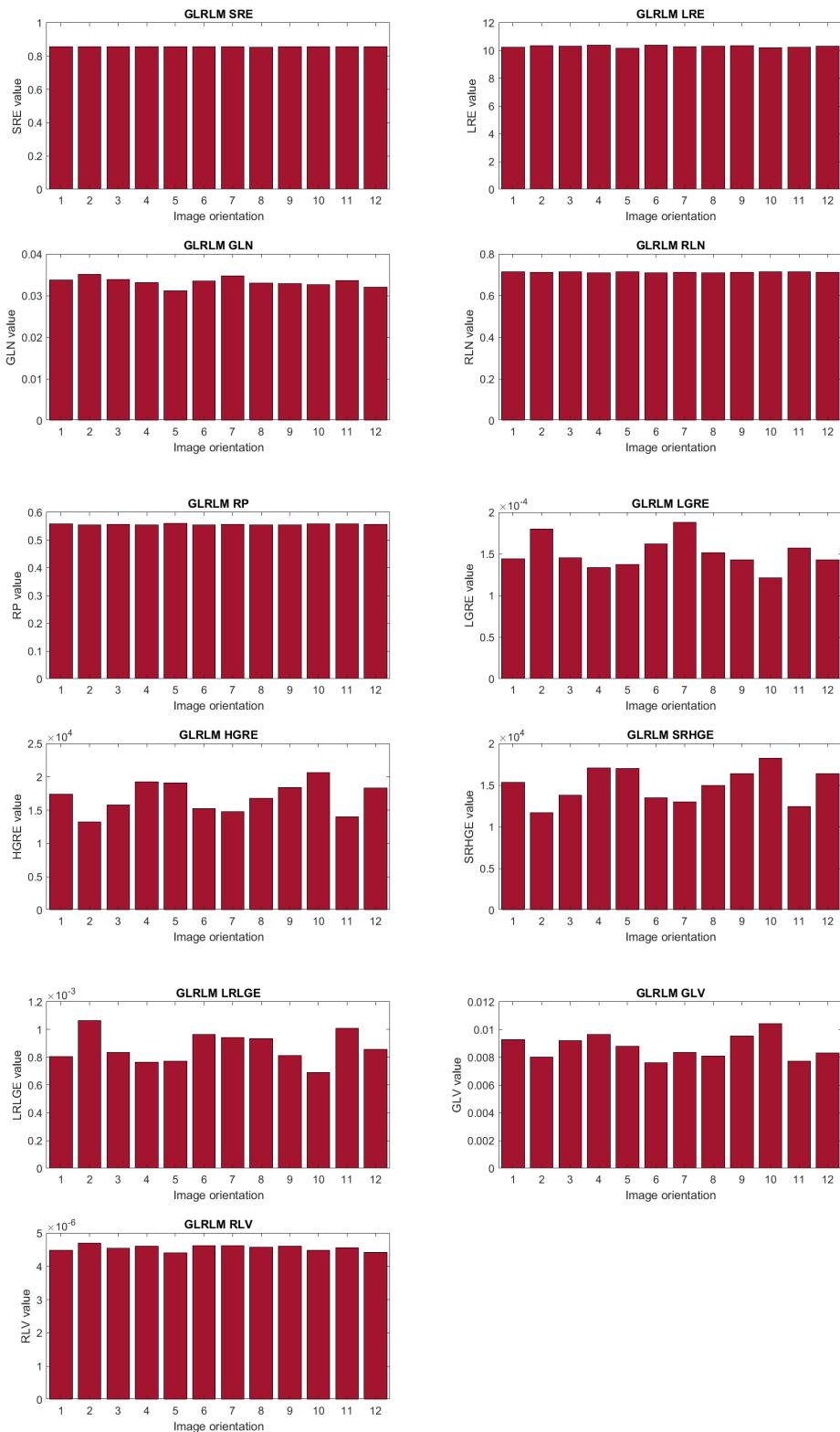


Figura 48: Misure statistiche della GLMRL del Putamen nella roi 76 (Volume 3D)

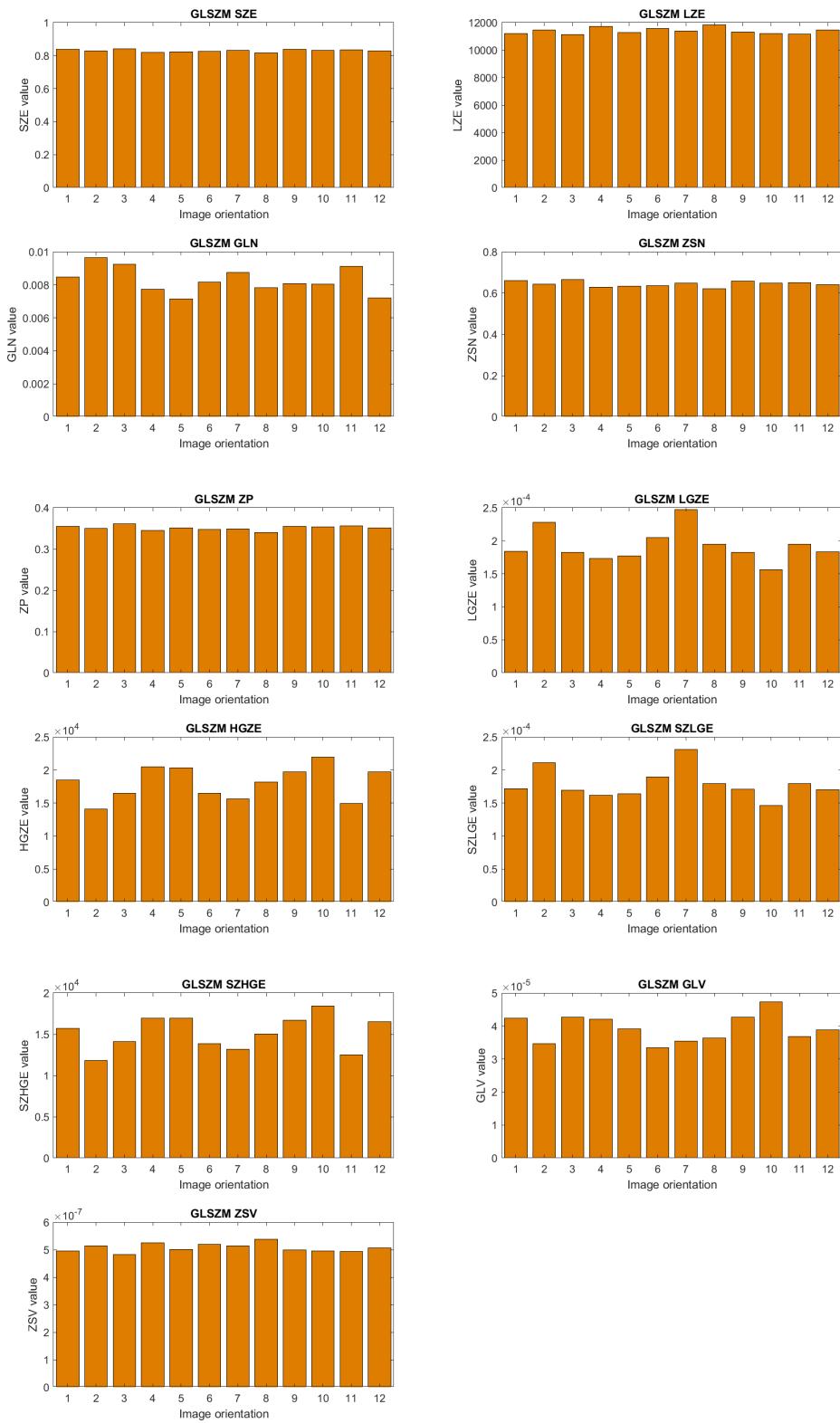


Figura 49: Misure statistiche della GLSzm del Putamen nella roi 76 (Volume 3D)

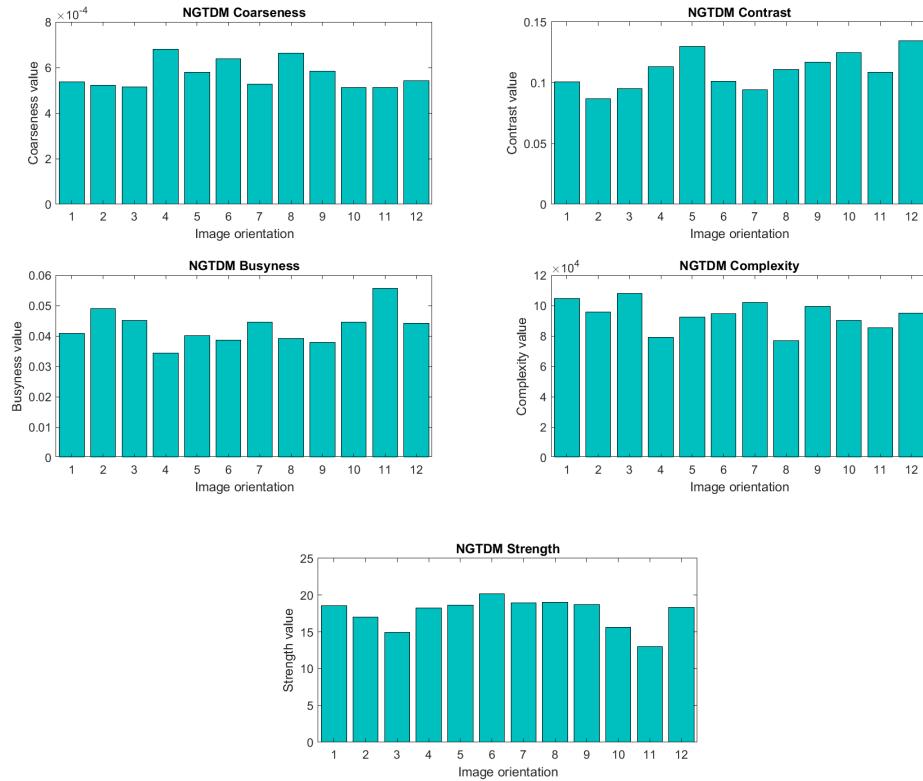


Figura 50: Misure statistiche della NGTDM del Putamen nella roi 76 (Volume 3D)

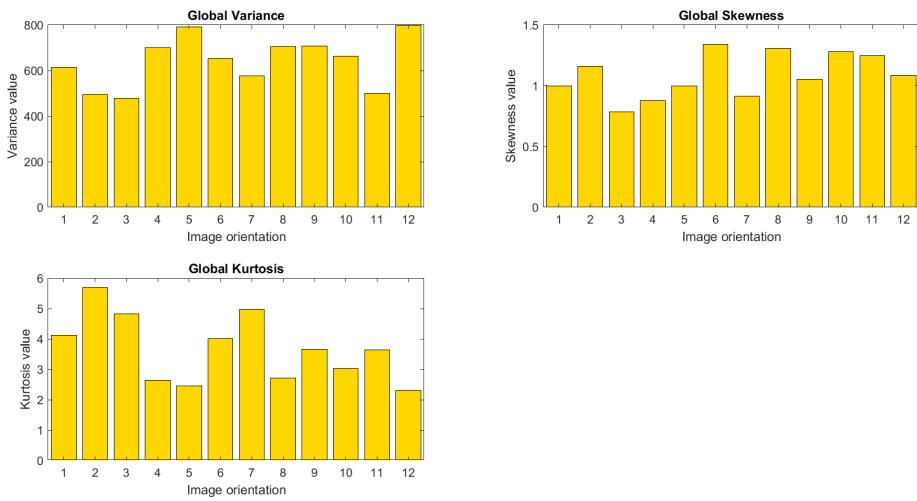


Figura 51: Misure statistiche globali del Putamen nella roi 76 (Volume 3D)

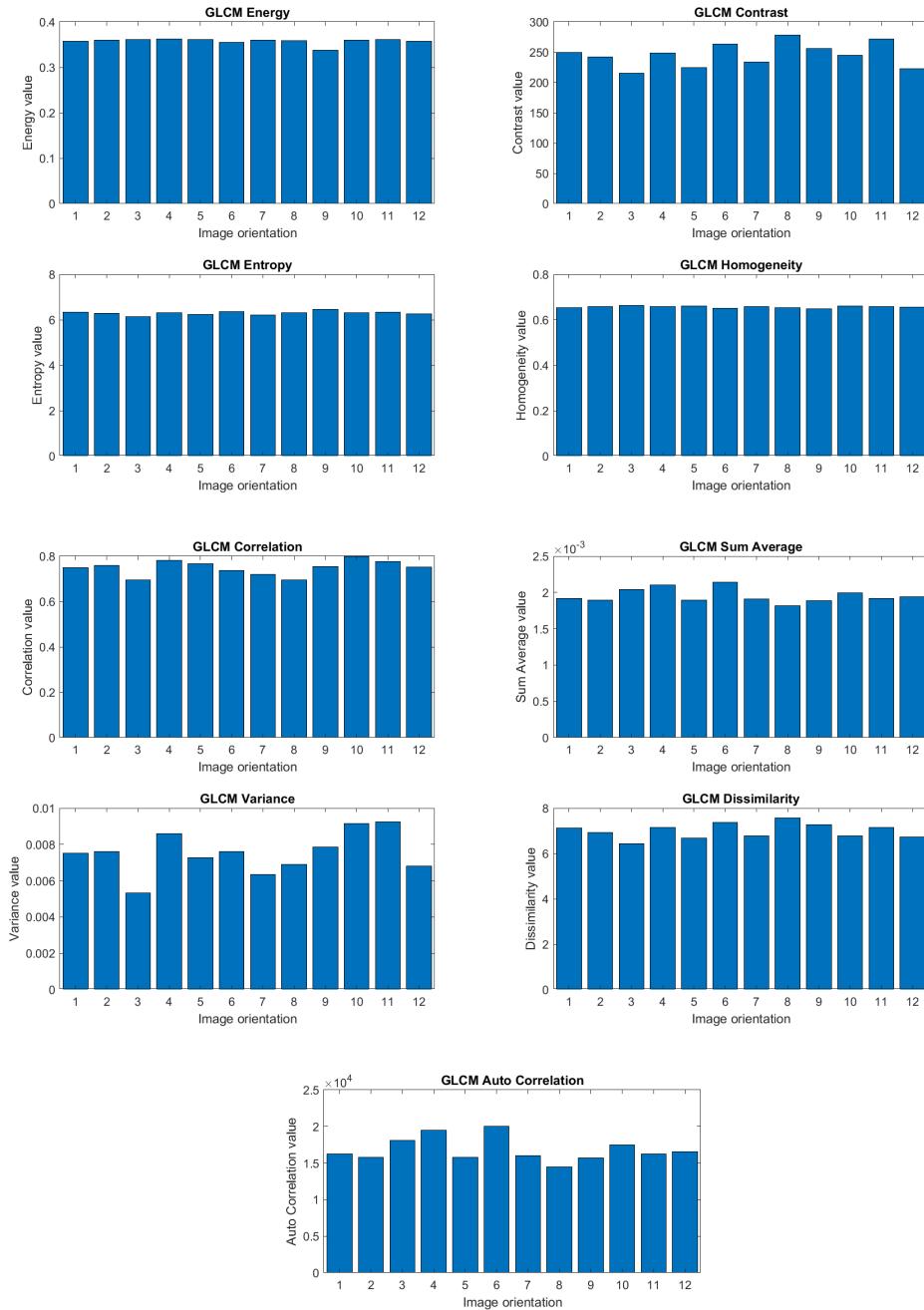


Figura 52: Misure statistiche della GLCM del Caudate nella roi 76 (Volume 3D)

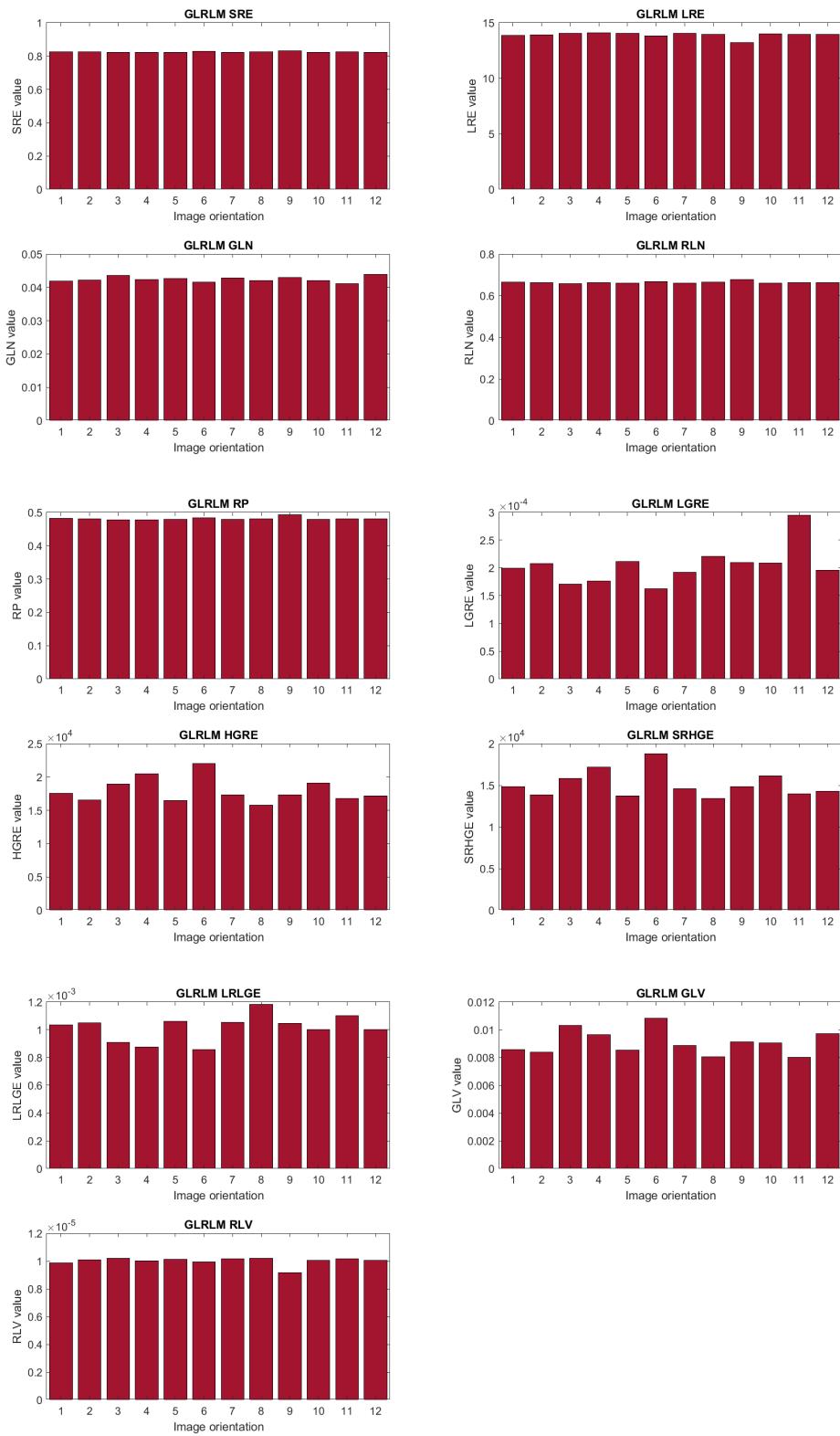


Figura 53: Misure statistiche della GLMRL del Caudate nella roi 76 (Volume 3D)

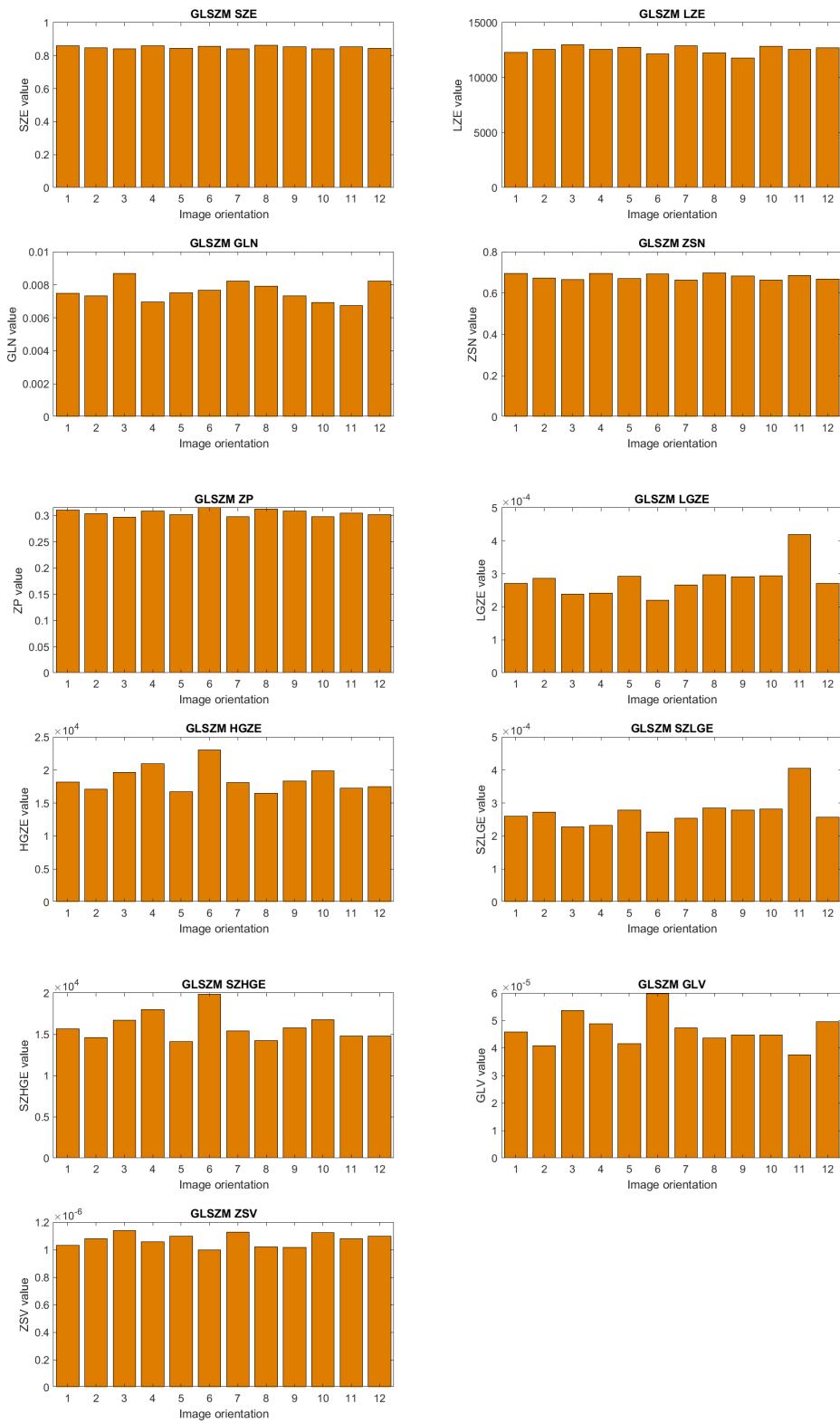


Figura 54: Misure statistiche della GLSzm del Caudate nella roi 76 (Volume 3D)

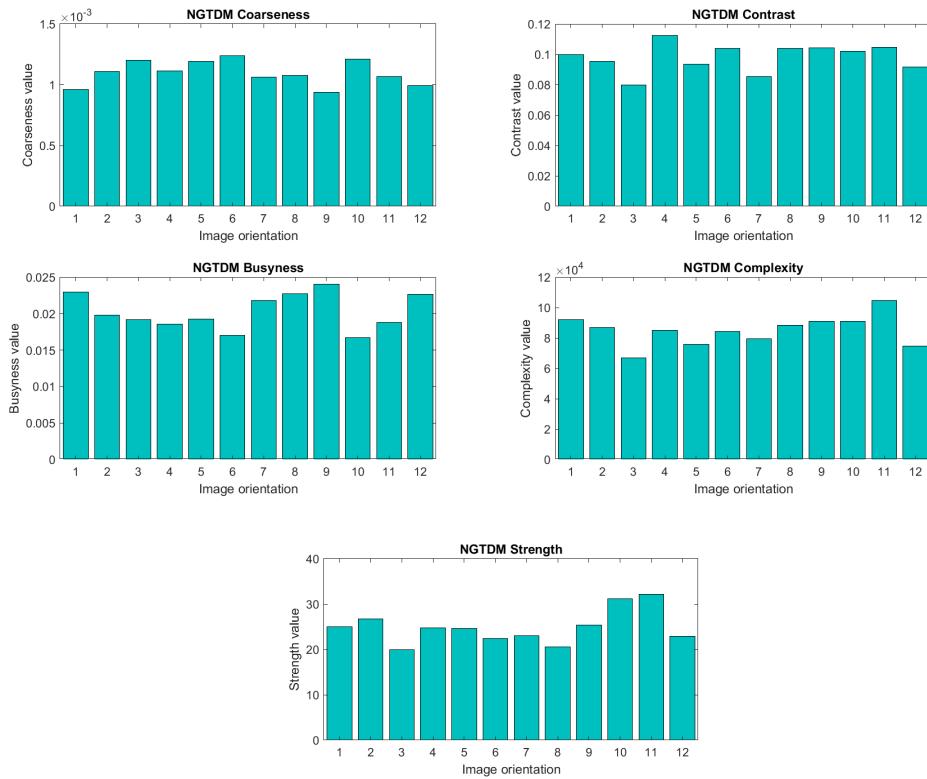


Figura 55: Misure statistiche della NGTDM del Caudate nella roi 76 (Volume 3D)

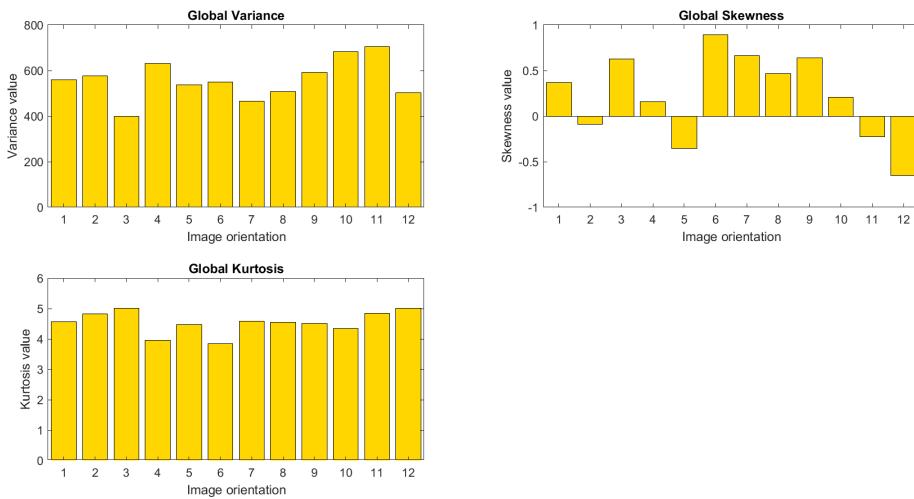


Figura 56: Misure statistiche globali del Caudate nella roi 76 (Volume 3D)

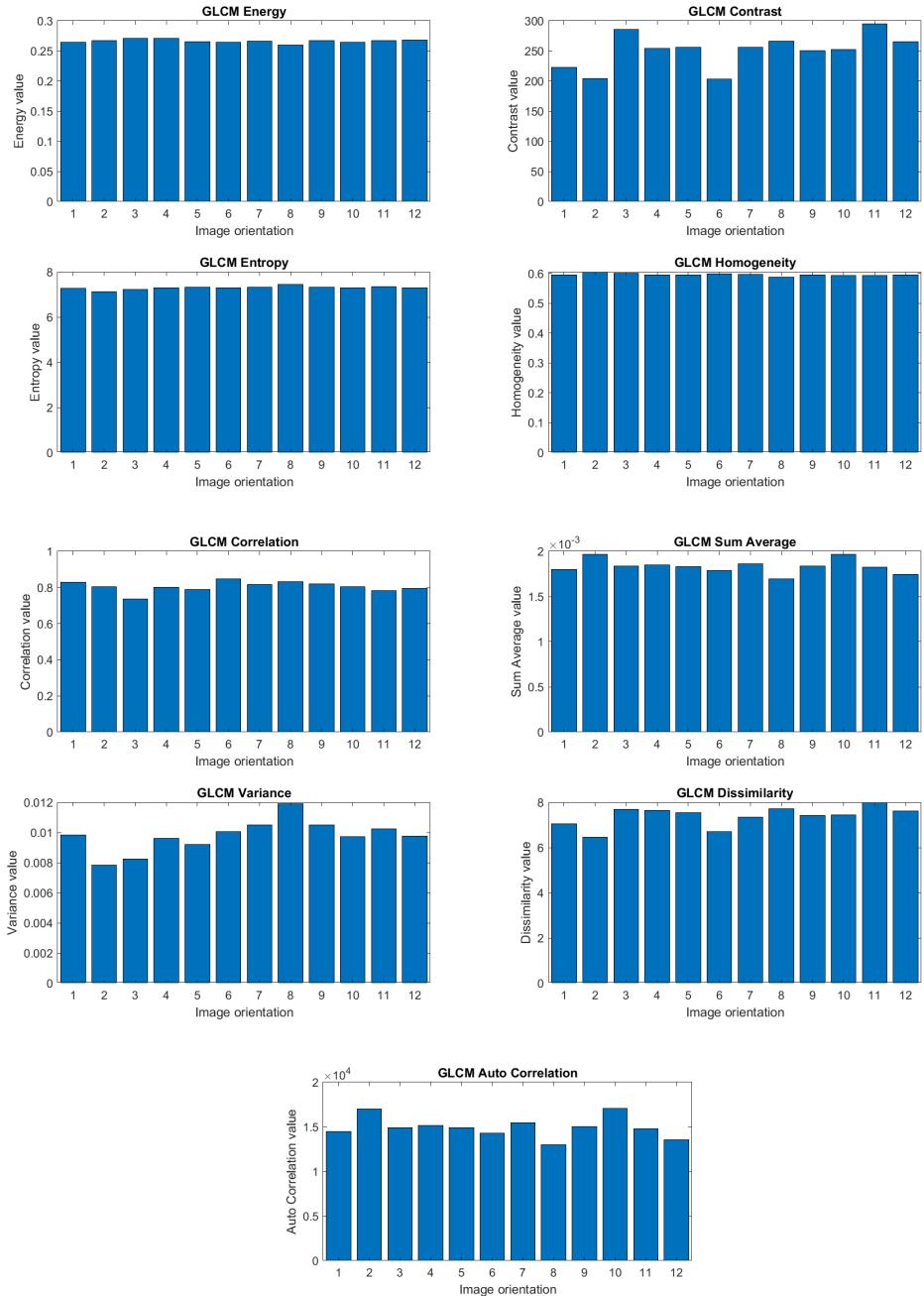


Figura 57: Misure statistiche della GLCM del Globus Pallidus nella roi 76 (Volume 3D)

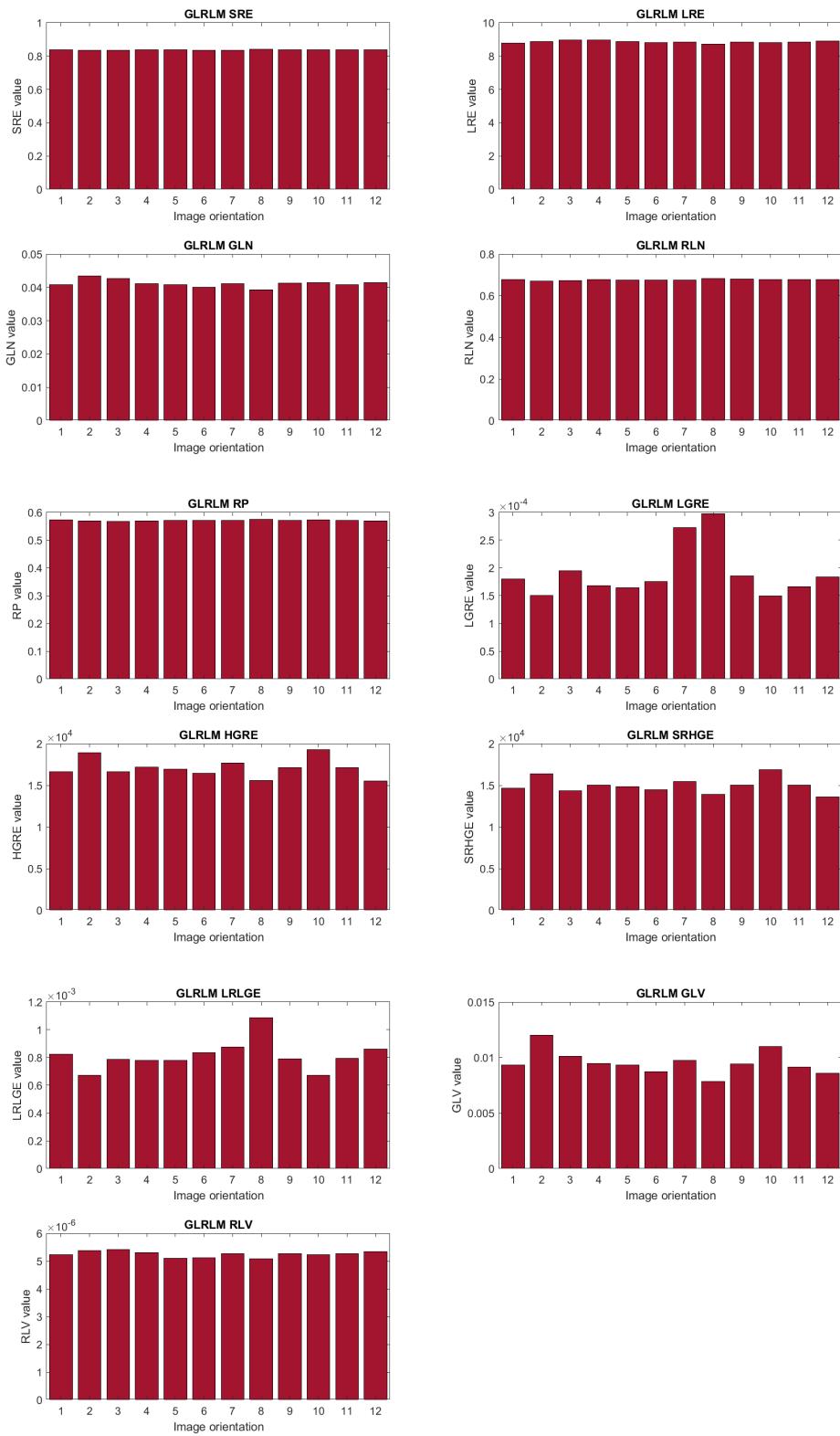


Figura 58: Misure statistiche della GLMRL del Globus Pallidus nella roi 76 (Volume 3D)

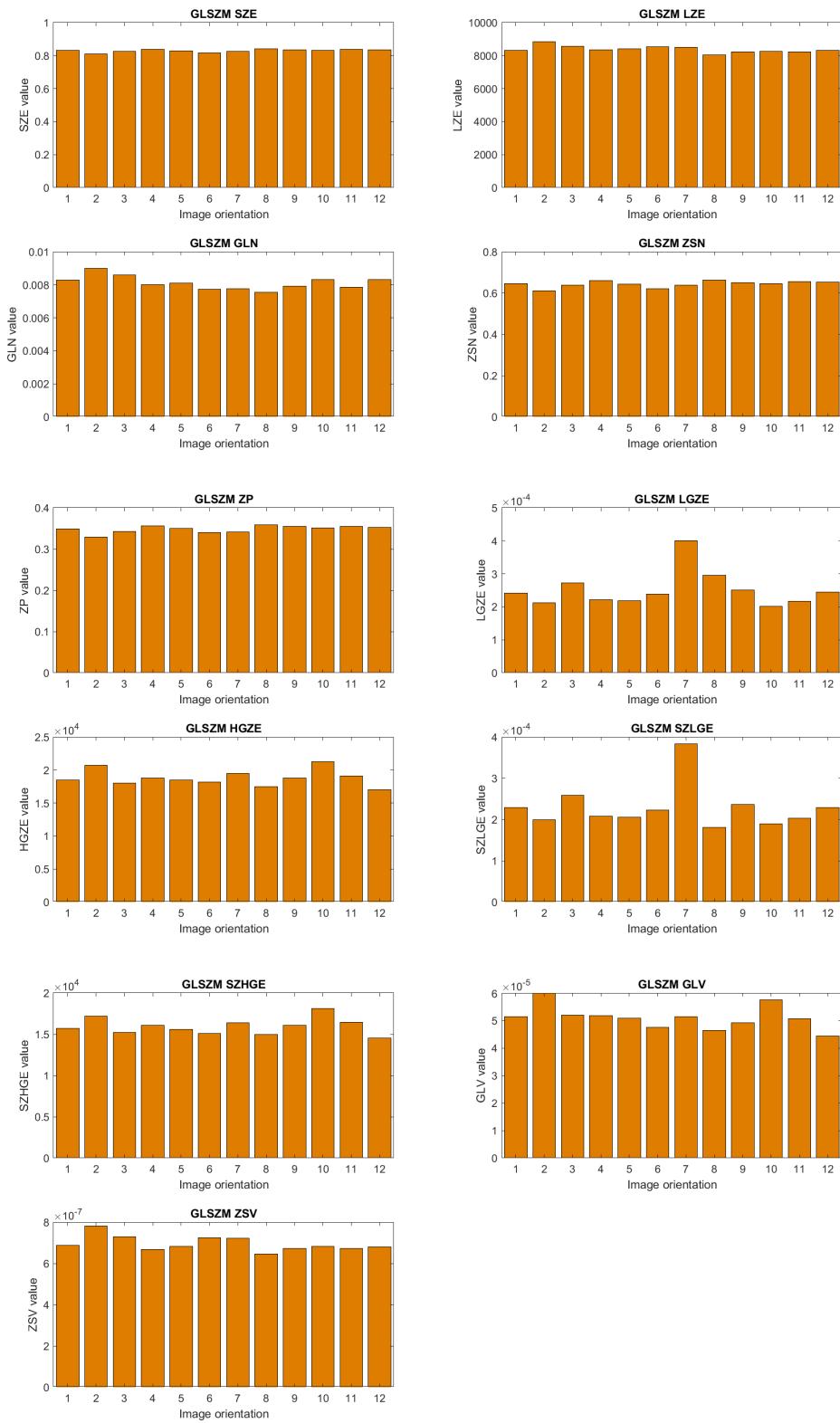


Figura 59: Misure statistiche della GLSzm del Globus Pallidus nella roi 76 (Volume 3D)

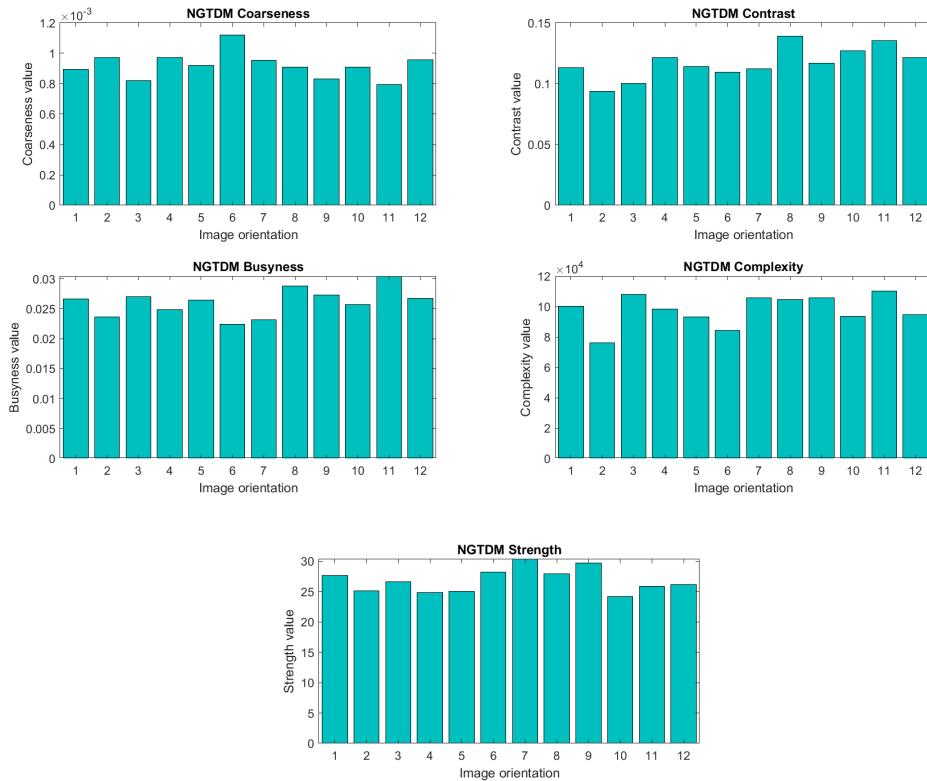


Figura 60: Misure statistiche della NGTDM del Globus Pallidus nella roi 76 (Volume 3D)

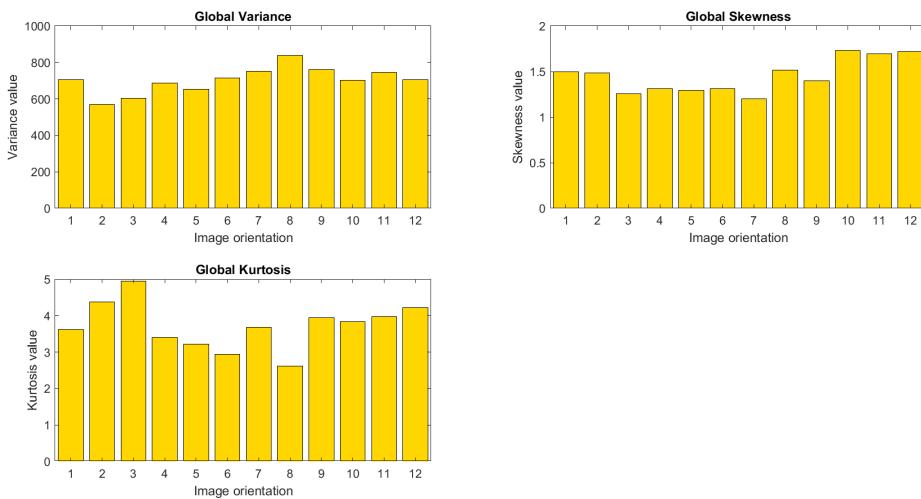


Figura 61: Misure statistiche globali nella roi 76 (Volume 3D)

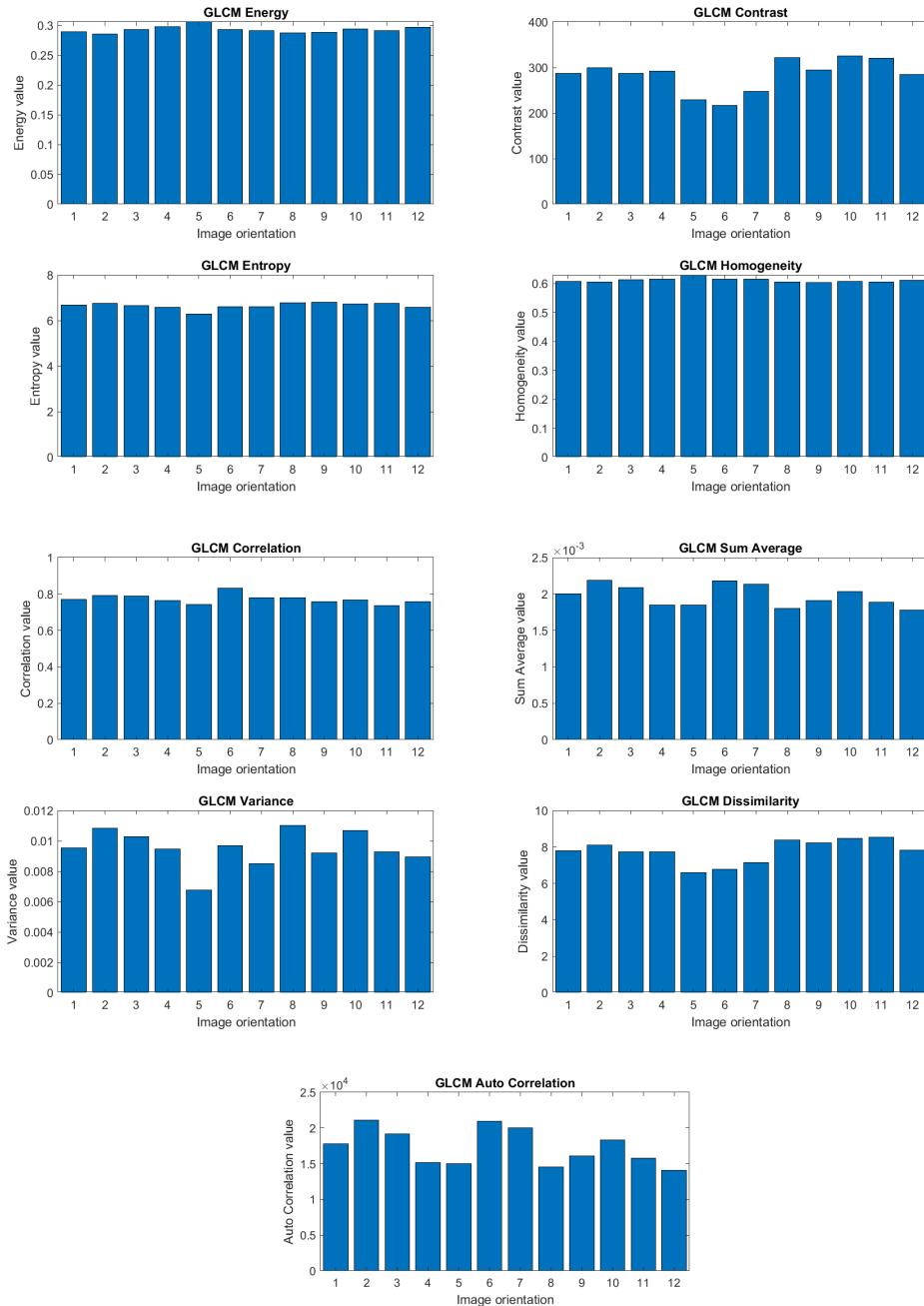


Figura 62: Misure statistiche della GLCM della White Matter nella roi 76 (Volume 3D)

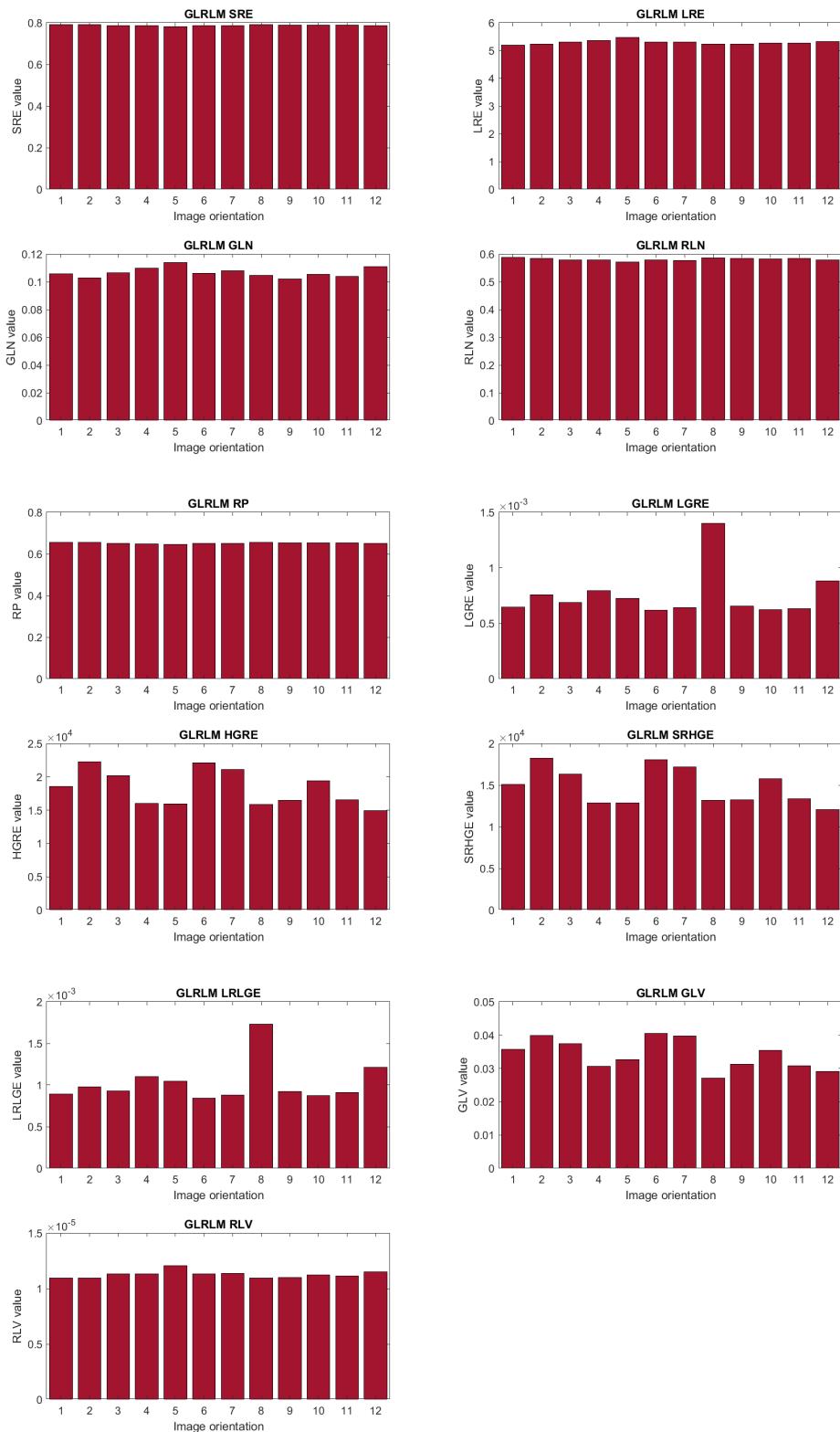


Figura 63: Misure statistiche della GLMRL della White Matter nella roi 76 (Volume 3D)

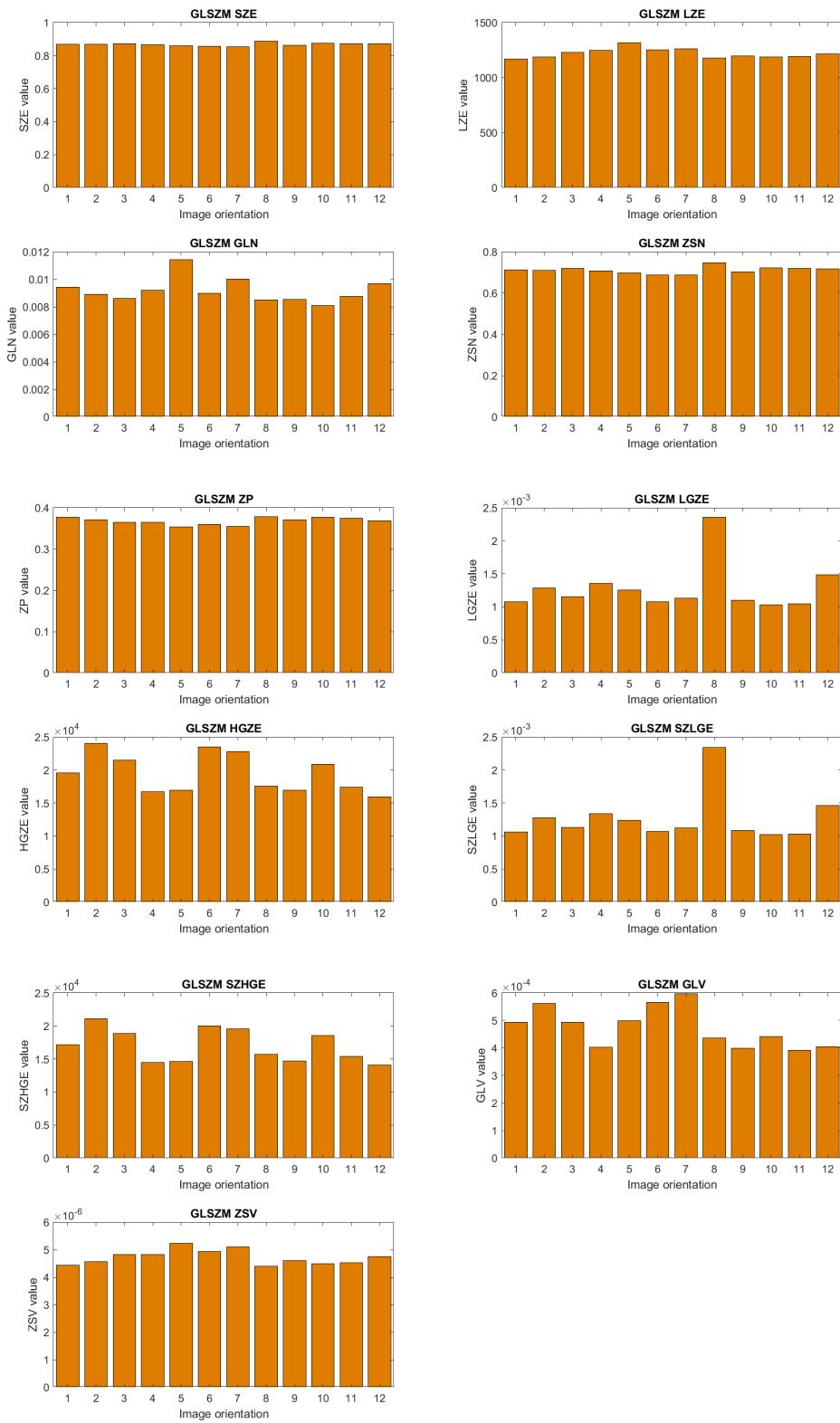


Figura 64: Misure statistiche della GLSzm della White Matter nella roi 76 (Volume 3D)

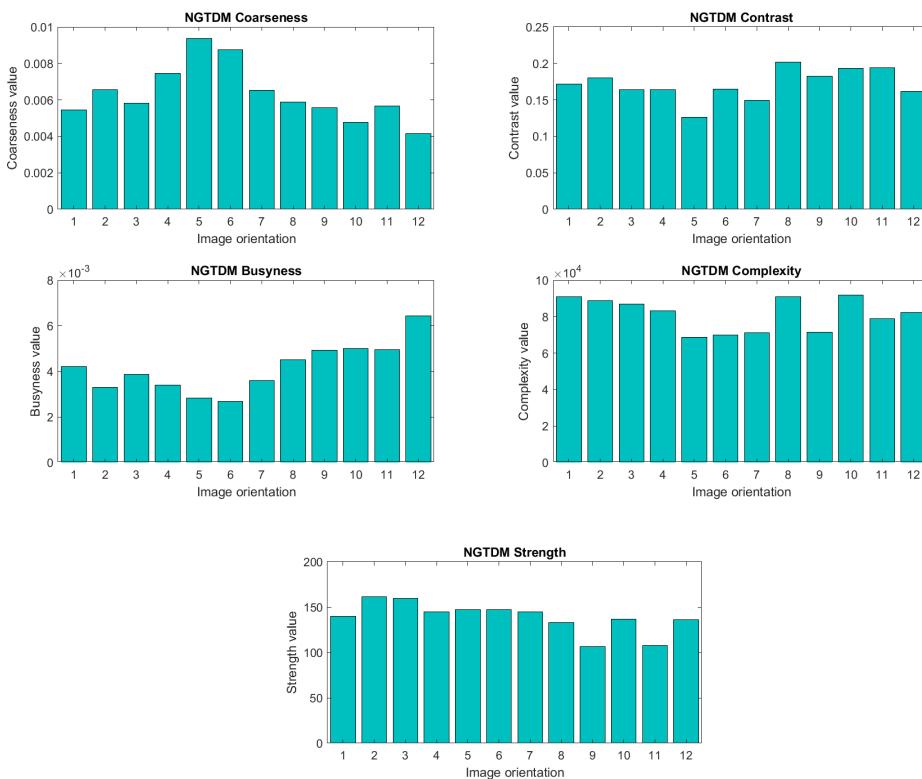


Figura 65: Misure statistiche della NGTDM della White Matter nella roi 76 (Volume 3D)

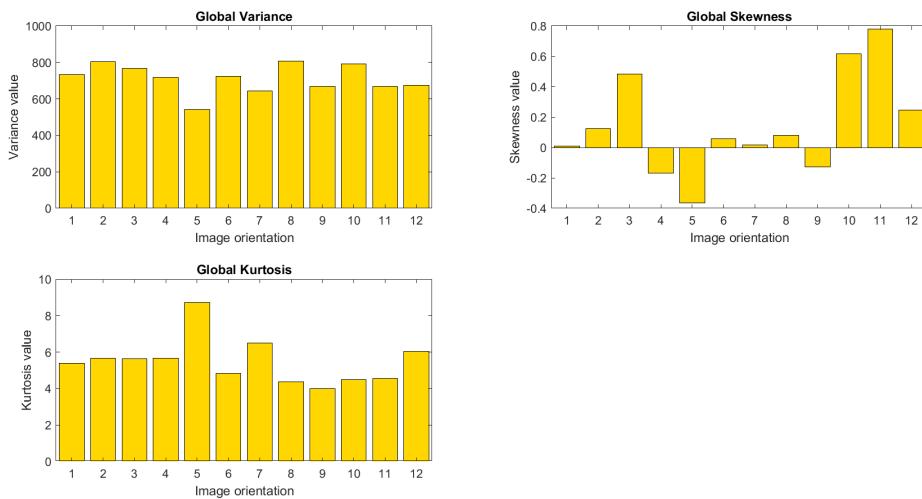


Figura 66: Misure statistiche globali della White Matter nella roi 76 (Volume 3D)

## E Appendice: PCA per diversi ordini di grandezza

Come accennato nella sezione 6.1 nelle figure 67, 68, 69, 70, 71, 72 sono mostrati i risultati della PCA ottenuti usando 3 diversi intervalli di ordini di grandezza:  $10^{-6} - 10^{-3}$ ,  $10^{-3} - 1$  e  $1 - 10^4$  per le misurazioni 'basate su matrici'.

Anche in questo caso, il numero di componenti principali è stato controllato a posteriori, tenendo conto del parametro *explained* i cui risultati sono riportati nella tabella 5 in cui al nome delle ROI è associato il pedice:

- $rl$  per indicare l'intervallo  $10^{-6} - 10^{-3}$ ;
- $r$  per indicare l'intervallo  $10^{-3} - 1$ ;
- $rh$  per indicare l'intervallo  $1 - 10^4$ .

	$RN_{rl}$	$SN_{rl}$	$CAU_{rl}$	$GP_{rl}$	$PU_{rl}$	$WM_{rl}$
PC1	100	100	100	100	100	100
	$RN_r$	$SN_r$	$CAU_r$	$GP_r$	$PU_r$	$WM_r$
PC1	61.98	75.77	70.08	58.66	93.45	51.11
PC2	31.85	20.59	23.03	38.29	4.06	38.92
PC3	4.68		5.79			7.51
	$RN_{rh}$	$SN_{rh}$	$CAU_{rh}$	$GP_{rh}$	$PU_{rh}$	$WM_{rh}$
PC1	99.47	93.13	88.52	96.10	83.06	75.14
PC2		6.85	11.35		16.89	24.79

Tabella 5: Percentuale di variabilità delle PCs sulle misure basate su matrici raggruppate per diversi intervalli di ordini di grandezza

Come è possibile notare nelle figure 67, 68, 69, 70, 71, 72 le features che contribuiscono di più alle componenti principali sono:

- Intervallo  $10^{-6} - 10^{-3}$ :
  - Componente 1:  $glcm_{ac}$ ;
  - Componente 2:  $ngtdm_{coa}$  (per RN, CAU, PU e WM);  $glszmlgze$  e  $glszm_{sgze}$  (per SN);  $glszmlgze$  e  $glrlm_{lrlge}$  (per GP).
- Intervallo  $10^{-3} - 1$ : riportate in tabella 6.
- Intervallo  $1 - 10^4$ :
  - Componente 1:  $ngtdm_{com}$ ;
  - Componente 2:  $glszm_{hgze}, glrlm_{hgre}, glrlm_{srhge}$  e  $glszm_{szhge}$ .

	<i>RN</i>	<i>SN</i>	<i>CAU</i>	<i>GP</i>	<i>PU</i>	<i>WM</i>
<b>PC1</b>	$ngtdm_{con}$	$ngtdm_{con}$	$glcm_{cor}$	$glcm_{cor}$	$glcm_{cor}$	$glszm_{zsn}$
		$glszm_{zsn}$				
<b>PC2</b>	$glszm_{zp}$	$glcm_{cor}$	$glszm_{zsn}$	$glcm_{hom}$	$ngtdm_{con}$	$glcm_{cor}$
	$glszm_{zsn}$			$glszm_{zp}$		
	$glrlm_{rp}$			$glszm_{zsn}$		
<b>PC3</b>	$glcm_{hom}$		$glrlm_{rln}$	$ngtdm_{con}$		$glszm_{zsn}$
	$glcm_{cor}$		$glrlm_{rp}$			$ngtdm_{con}$
	$glcm_{ene}$					

Tabella 6: Features che hanno maggiore contributo nelle componenti principali, comprese nell’intervallo  $10^{-3} - 1$

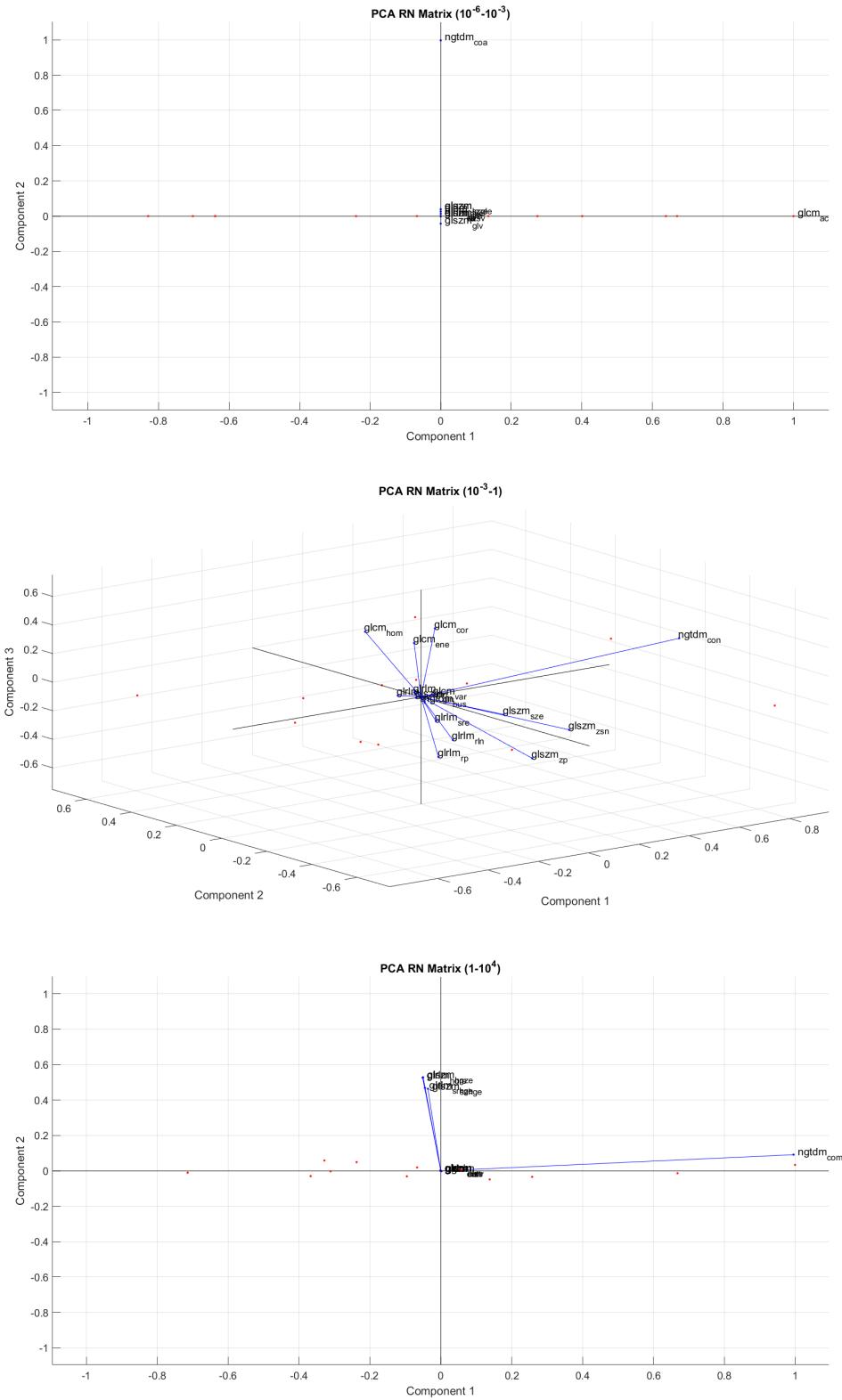


Figura 67: PCA della ROI rn per i 3 intervalli di ordini di grandezza

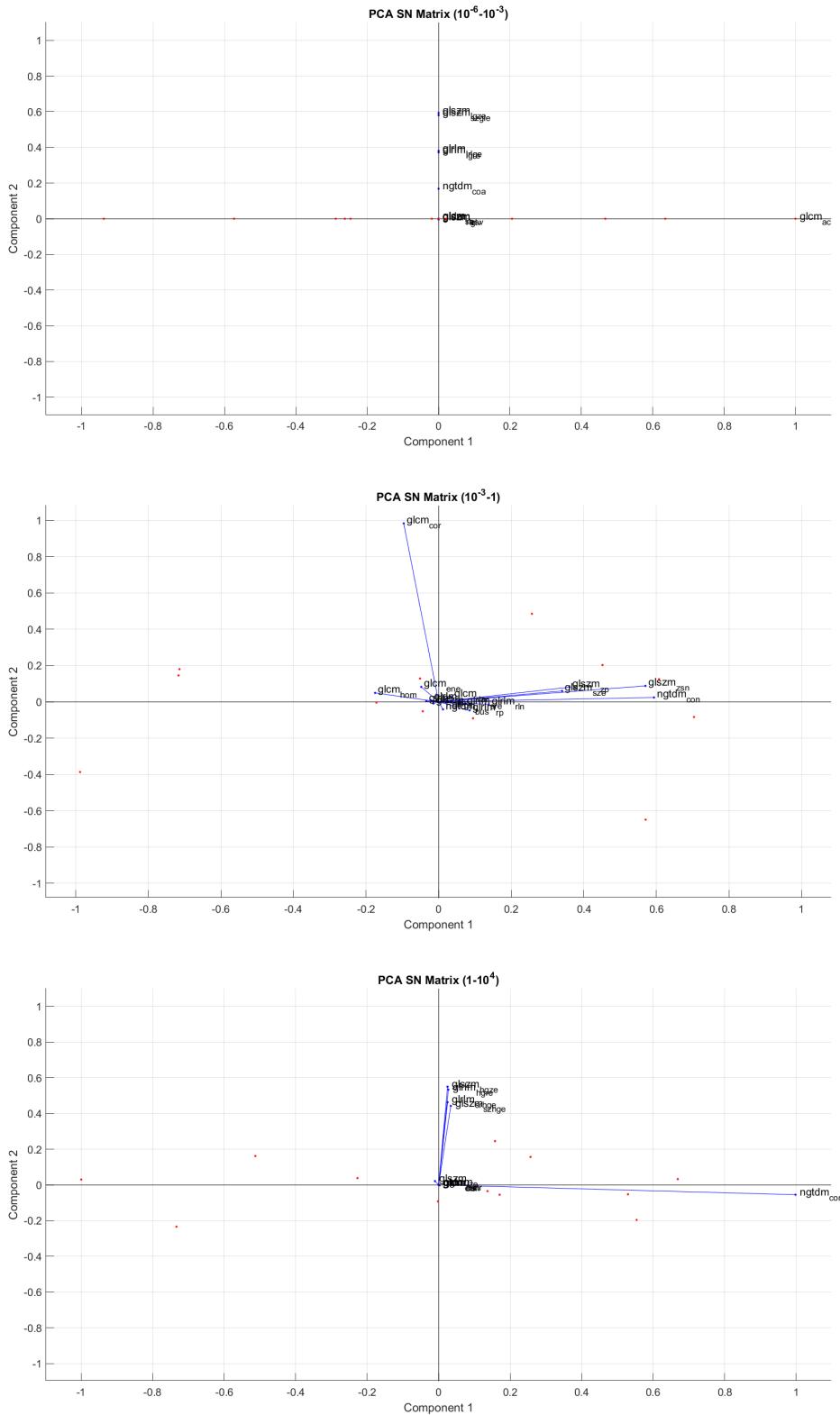


Figura 68: PCA della ROI sn per i 3 intervalli di ordini di grandezza

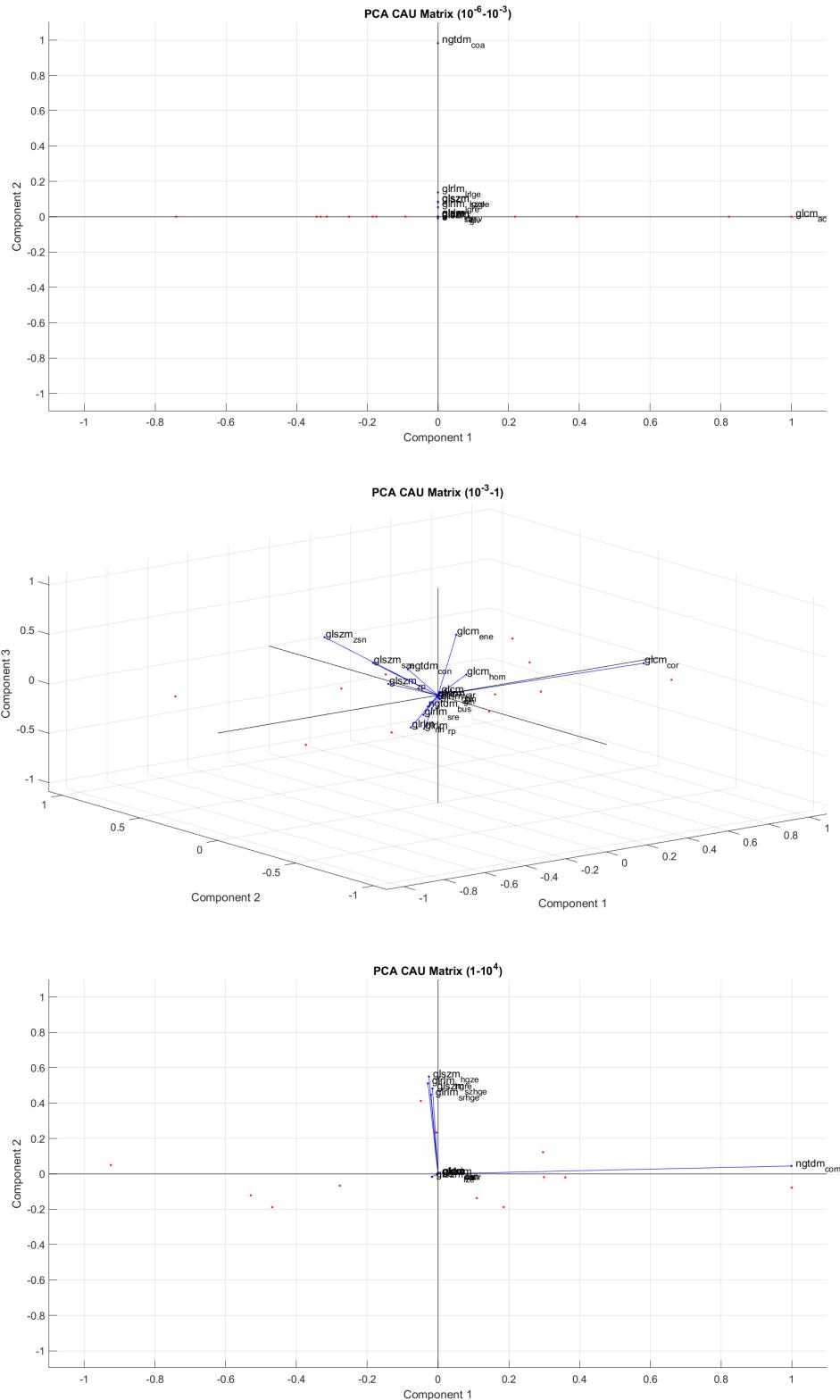


Figura 69: PCA della ROI cau per i 3 intervalli di ordini di grandezza

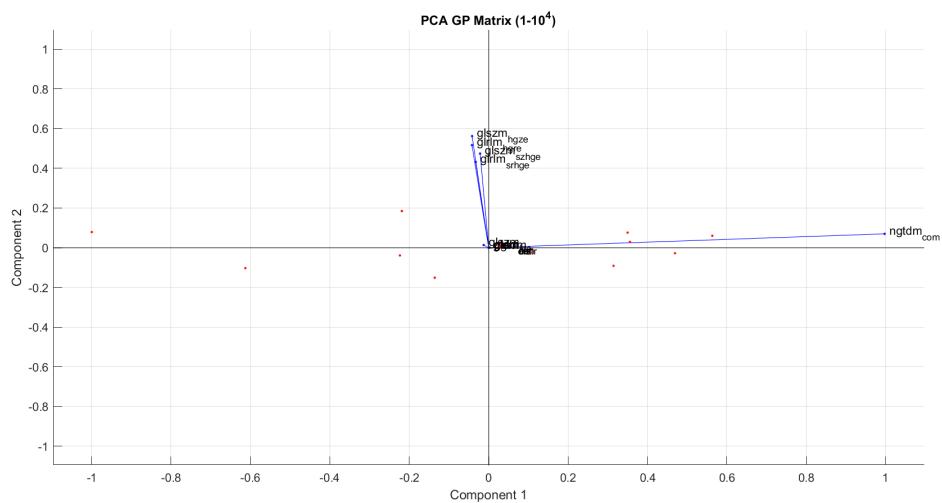
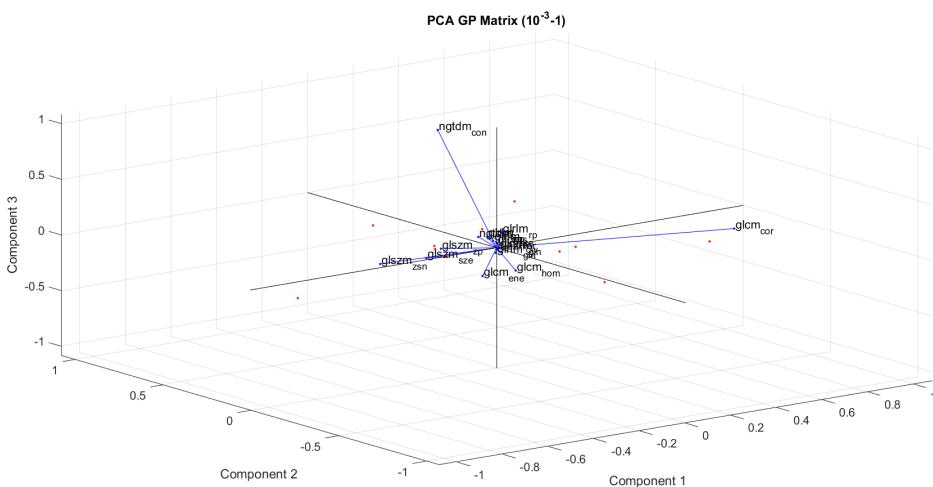
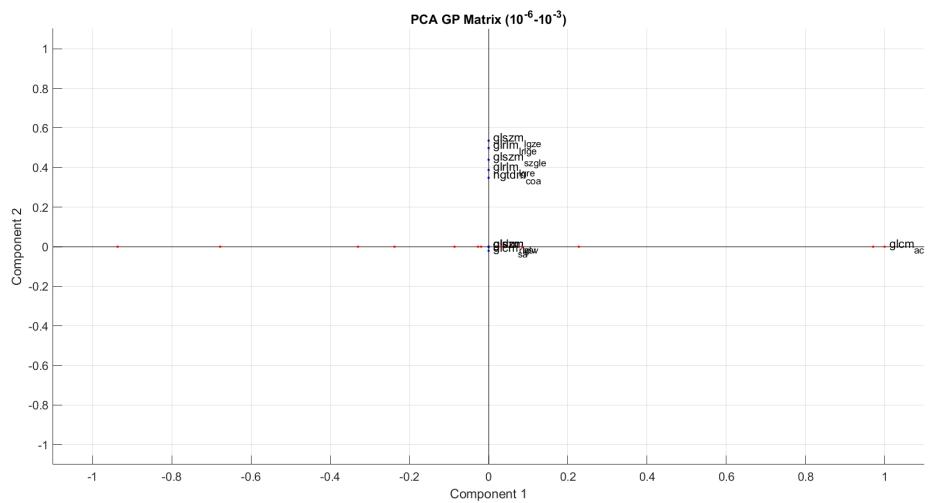


Figura 70: PCA della ROI gp per i 3 intervalli di ordini di grandezza

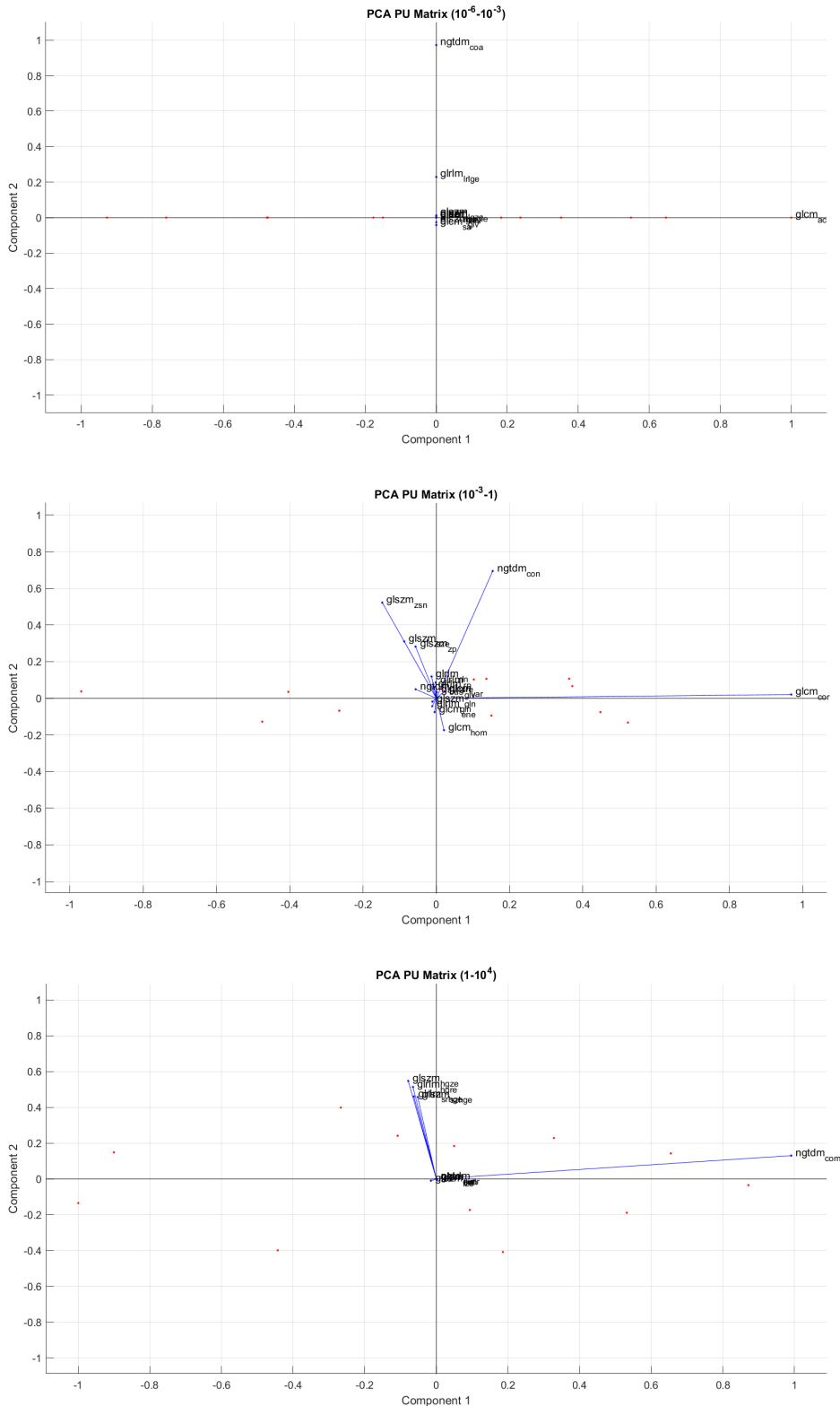


Figura 71: PCA della ROI pu per i 3 intervalli di ordini di grandezza

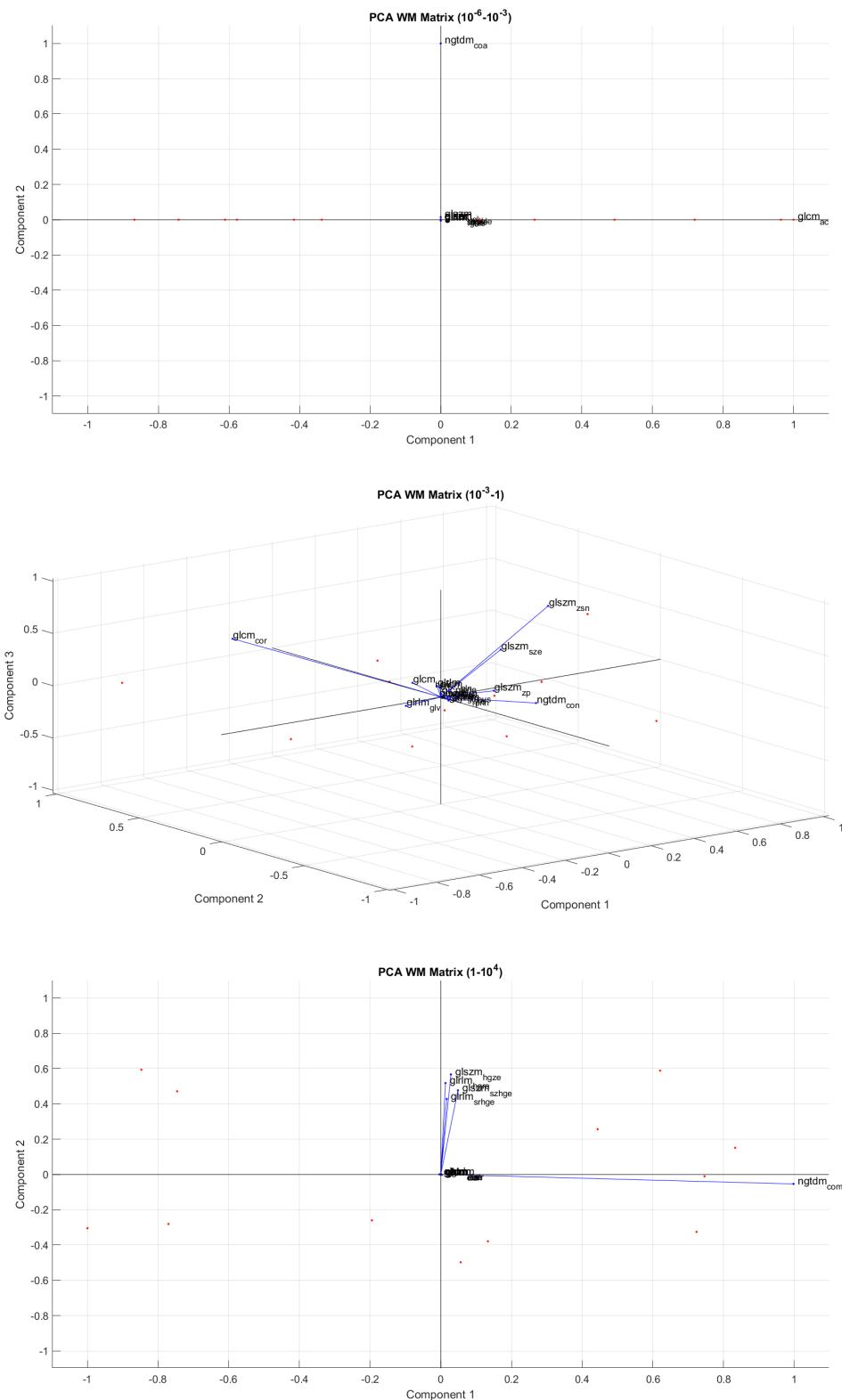


Figura 72: PCA della ROI wm per i 3 intervalli di ordini di grandezza

## F Appendice: PCA features migliori in comune

In questa sezione si sono prese in considerazioni le features 'migliori', ovvero quelle aventi i valori in grassetto nella 2 per tutte le ROI dei riferimenti biologici. In questo modo è possibile selezionare le features che variano in tutte le regioni.

Le 'migliori' features risultanti sono  $glrlmlgze$ ,  $glszmlgze$ ,  $glszm_szlge$ ,  $ngtdm_{con}$  e  $ngtdm_{str}$  e i risultati ottenuti tramite la PCA sono mostrati nelle figure 73 e 74.

Come spiegato nella sezione 6.1, il numero di componenti principali è stato controllato a posteriori, tenendo conto del parametro *explained* i cui risultati sono riportati nella tabella 7.

	<i>RN</i>	<i>SN</i>	<i>CAU</i>	<i>GP</i>	<i>PU</i>	<i>WM</i>
PC1	99.99	99.99	99.99	99.99	99.99	99.99

Tabella 7: Percentuale di variabilità delle PCs sulle migliori features tra le misure basate su matrici

Come è possibile notare nelle figure 73, 74 le features che contribuiscono di più alle componenti principali sono:

- Componente 1:  $ngtdm_{str}$ ;
- Componente 2:  $ngtdm_{con}$ .

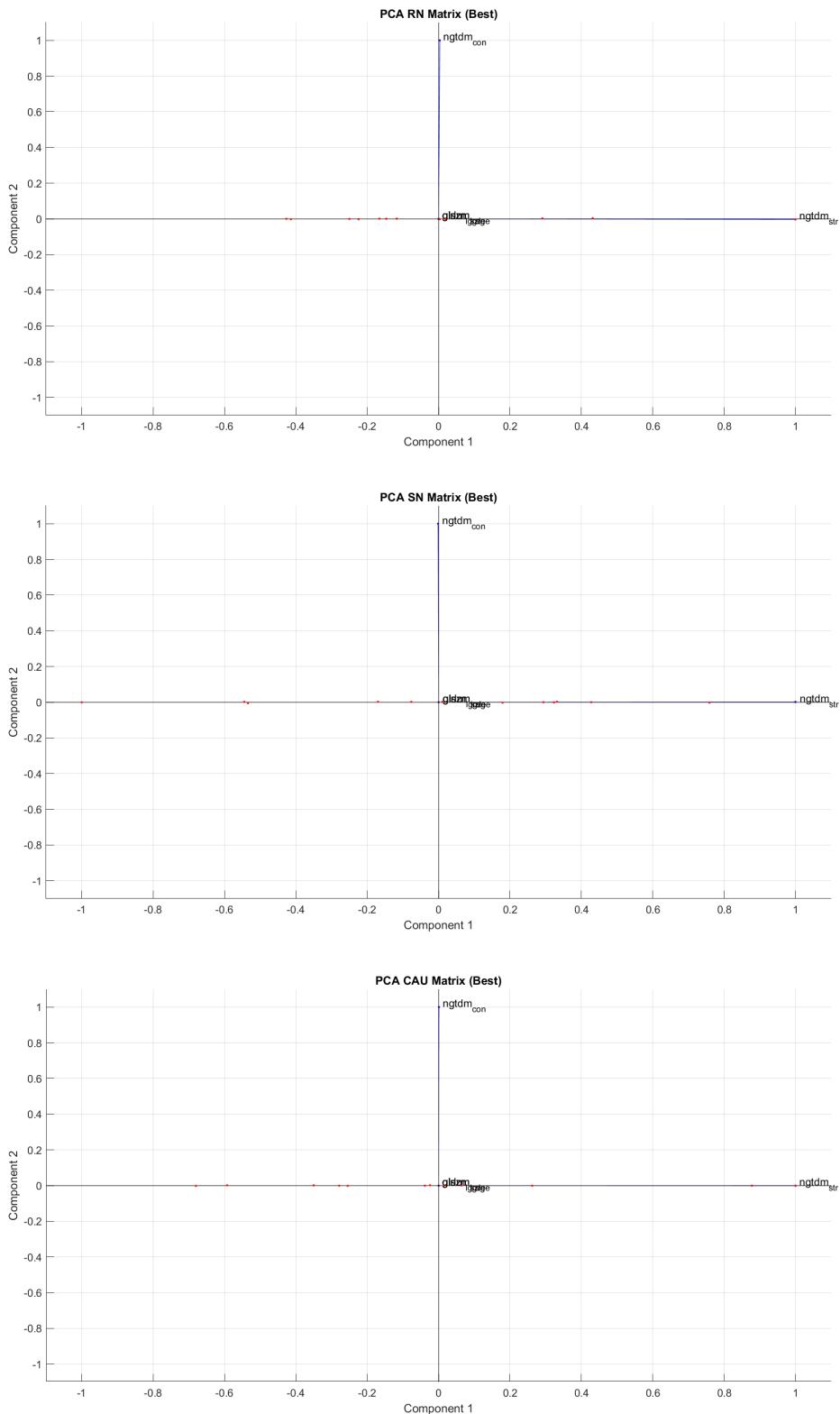


Figura 73: PCA delle features migliori: ROI RN, SN e CAU

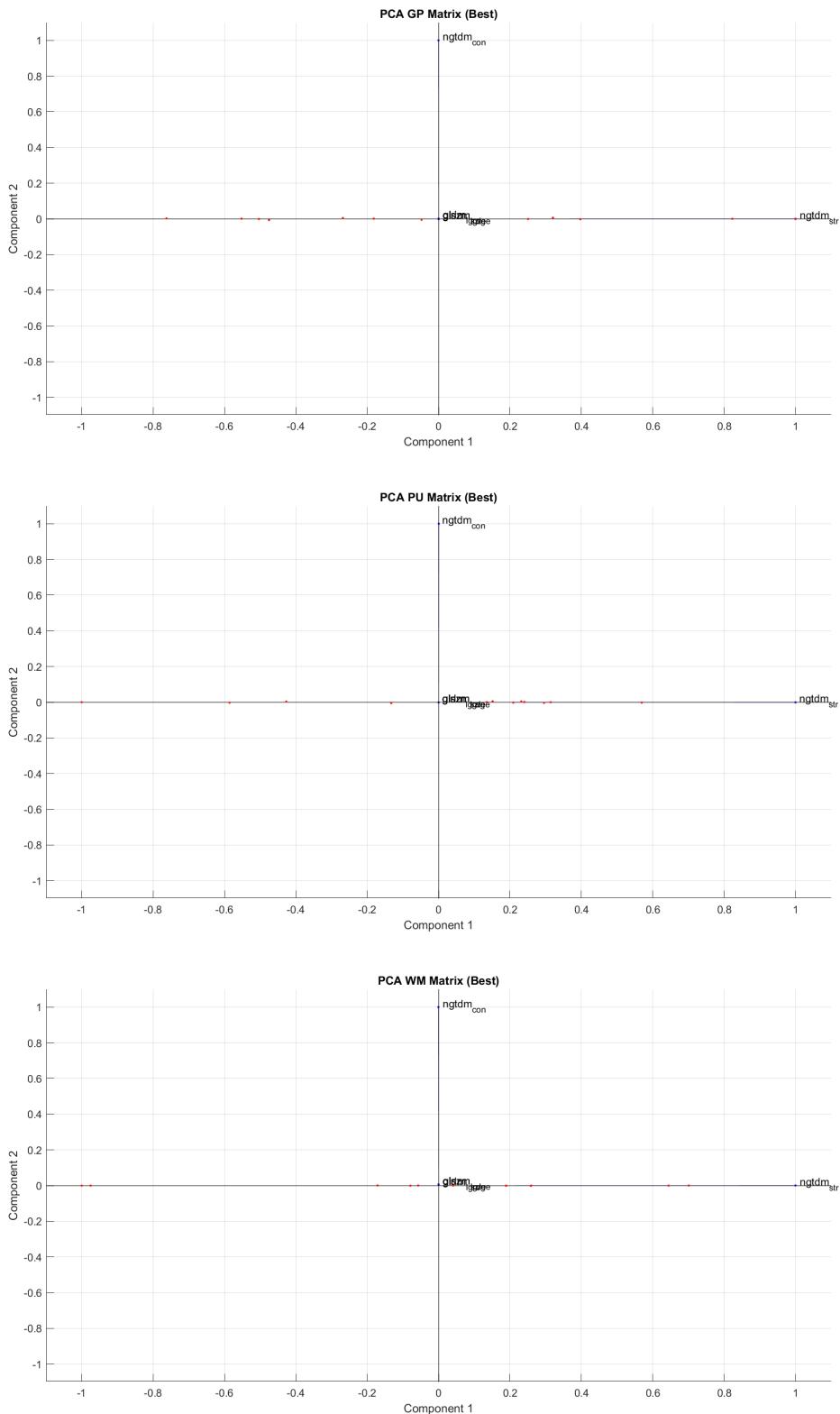


Figura 74: PCA delle features migliori: ROI GP, PU e WM

## Riferimenti bibliografici

- [1] "Statistical Texture Analysis"; G. N. Srinivasan, G. Shobha; 2008
- [2] "Texture Analysis: A Review of Neurologic MR Imaging Applications"; A. Kassner, R.E. Thornhill; 2009
- [3] "Characterization of PET/CT images using texture analysis: the past, the present... any future?"; Hatt, M. et al.; 2016
- [4] "A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities."; M. Vallieres, C.R. Freeman, S.R. Skamene, I.E. Naqa; 2015
- [5] "Textural Features for Image Classification"; R.M. Haralick, K. Shanmugam, I. Dinstein; 1973
- [6] "Data Mining and Statistics for Decision Making"; S. Tufféry; 2011