



Knowledge discovery from existing radiotherapy patient databases based on unsupervised learning methodology

D. Tewatia¹ and R. Tolakanahalli²

¹ University of Wisconsin - Madison, Madison WI

² Miami Cancer Institute, Baptist Health South Florida Miami, FL



INTRODUCTION

Data in radiotherapy patient databases have increased many fold in past decade. Vast knowledge in these databases can augment recent radiotherapy treatment planning efforts of knowledge based planning

AIM

To explore an unsupervised learning methodology for harnessing knowledge from existing radiotherapy patient databases.

METHOD

- Spatial and texture based parameters were computed for 9 randomly selected patient's planning target volumes (PTVs) and surrounding critical Organs At Risk (OARs).
- Similarity metrics were computed using a hierarchical clustering technique and patients were group them in their respective clusters.
- Pearson correlation based distance matrix was computed. Under this approach two pairs are close to each other if their feature vectors are highly correlated.
- A heatmap and dendrogram were generated to visualize this correlation.

**Dendrogram : Tree like representation to demonstrate the inter-relationship of similar feature vectors*

**Heat map : Color coded map to identify patterns.*

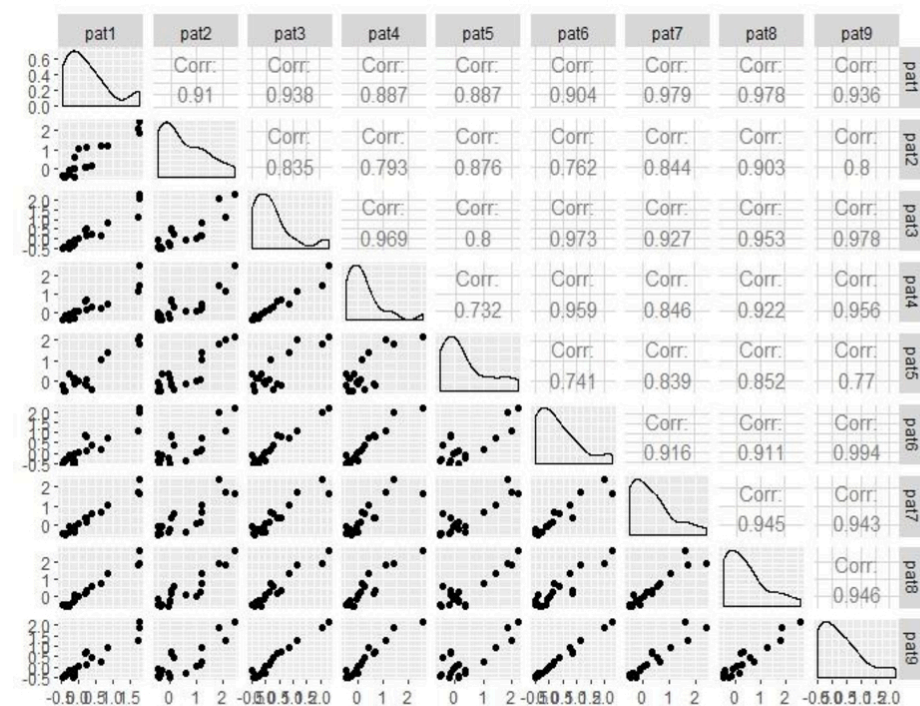


Figure 1: Bladder Correlation Coefficients for nine patients

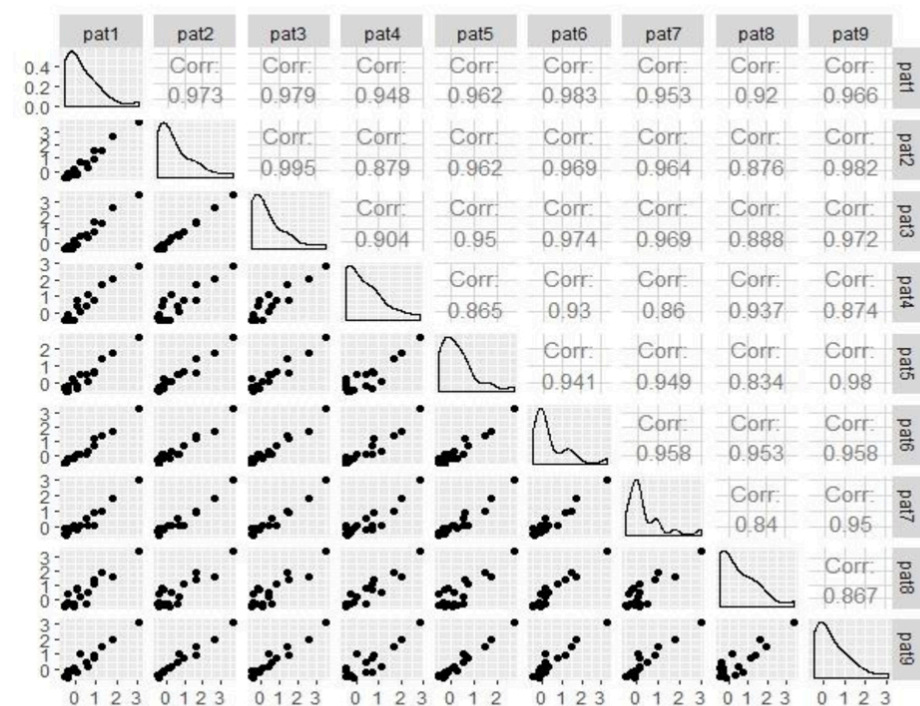


Figure 3: Rectum Correlation Coefficients for nine patients

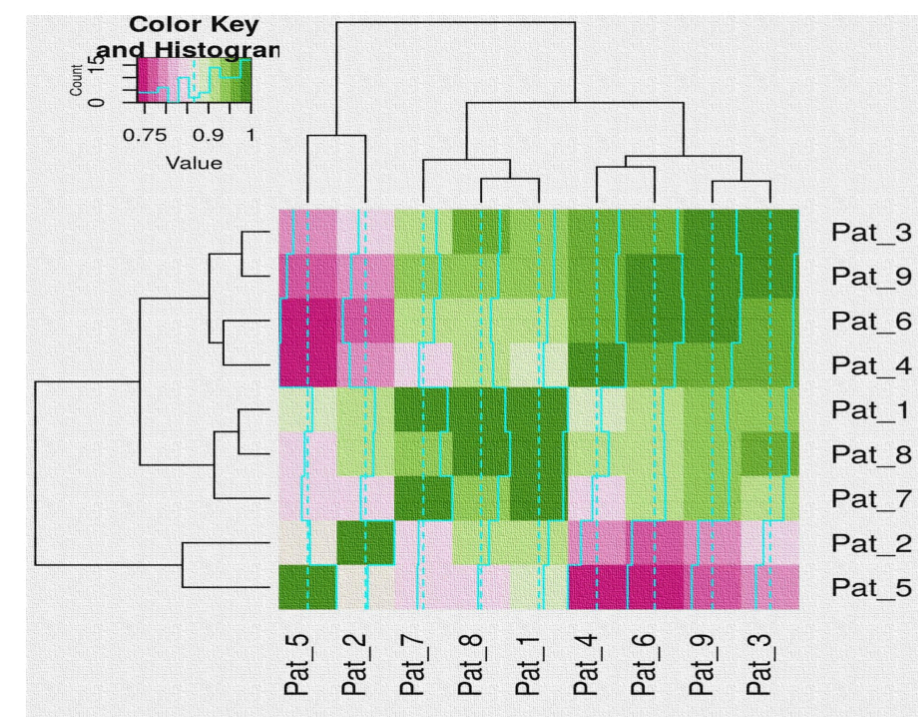


Figure 2: Bladder Dendrogram and heat map for nine patients

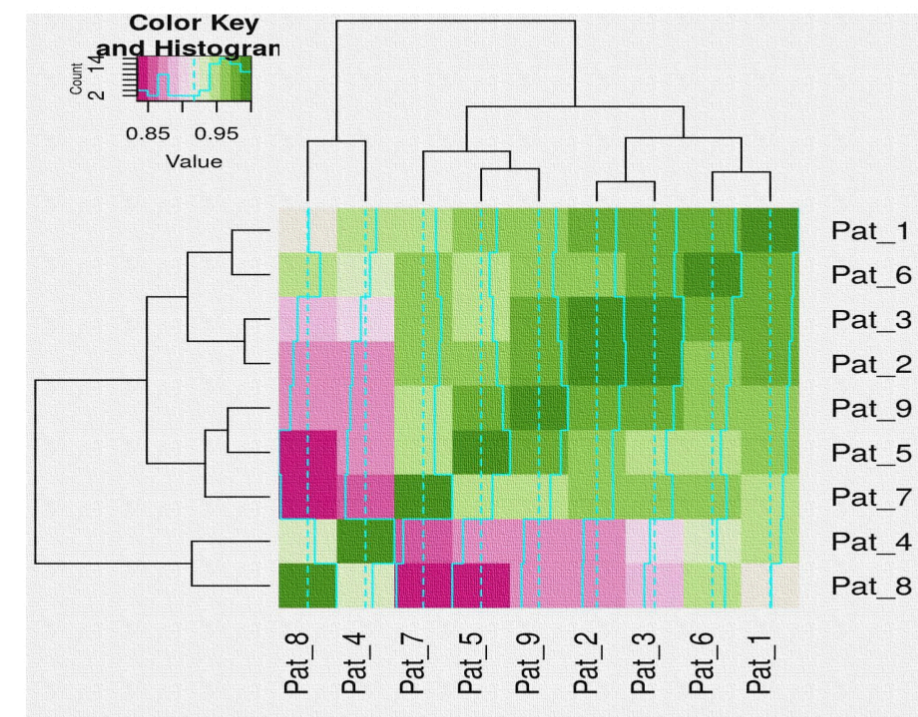


Figure 4: Rectum Dendrogram and heat map for nine patients

RESULTS

- Correlation index varied from 0.73 to 0.99 for bladder as shown in Figure 1.
- Each patient's bladder contour correlation with other eight patients is visualized easily in dendrogram and heatmap as shown in figure 2.
- Figure 3 and 4 shows similar metrics for Rectum OAR.

DISCUSSION

- Main advantage of hierarchical clustering is that there is no need to specify number of clusters upfront. The number of clusters are driven based on data available in each database.
- The data visualization techniques shown here enhance the deduction of each patient data correlation with the other remaining patients in a given database.

CONCLUSIONS

This study demonstrates that hierarchical clustering technique is a very promising technique for automated knowledge extraction from existing radiotherapy databases. This approach is perfect for radiotherapy patient databases where every patient is unique and knowledge keeps increasing with each new case added to existing databases

CONTACT INFORMATION

Dinesh Tewatia
tewatia@wisc.edu

Ranjini Tolakanahalli
ranjinit@gmail.com