Cluster Analysis Assessment in Proposing a Surgical Technique for Benign Prostatic Enlargement

Lazaros TZELVES [a,1,†], Georgios FERETZAKIS [b,c,†], Dimitris KALLES [b],
Ioannis MANOLITSIS [a], Stamatis KATSIMPERIS [a], Themistoklis BELLOS [a],
Marinos BERDEMPES [a], Athanasios ANASTASIOU [d], Dimitrios KOUTSOURIS [d],
Stavroula KOFOPOULOU [d], Vassilios S. VERYKIOS [b], Andreas SKOLARIKOS [a] and
Ioannis VARKARAKIS [a]

[a] *Second Department of Urology, National and Kapodistrian University of Athens,
Sismanogleio General Hospital, Athens, Greece*
[b] *School of Science and Technology, Hellenic Open University, Patras, Greece*
[c] *Sismanogleio General Hospital, Department of Quality Control, Research and
Continuing Education, Marousi, Greece*
[d] *National Technical University of Athens (NTUA)*
†*Equal contribution (1st authors)*

Abstract. Benign prostatic enlargement (BPE) is a common disease in men over 50 years old. The phenotype of patients with BPE is heterogenous, regarding both baseline patient characteristics and disease-related parameters. Treatment can be either medical-conservative or surgical. A great variety of surgical techniques are available for surgical management, with three of the most common being monopolar transurethral resection of the prostate (mTUR-P), bipolar transurethral resection of the prostate (bTUR-P), and bipolar transurethral vaporization of the prostate (bTUVis). The selection of each one of these depends on surgeon reasoning, equipment availability, patient characteristics, and preferences. Since all of these techniques are available in our Urology Department, and surgeons are skilled to perform each one of them, we performed a clustering analysis according to patient pre-operative characteristics, using the k-means algorithm, to compare clustering-related technique assignment with the real-life technique used.

Keywords. Benign prostatic enlargement, artificial intelligence, transurethral prostate resection, clustering analysis

1. Introduction and background

Benign prostatic enlargement is a very common disease, especially in men over 50 years old. The selection of the appropriate surgical method to manage this condition depends on prostate size, patient comorbidities and preferences, surgeon skills, and equipment availability.

---

[1] Corresponding Author, Lazaros Tzelves, MD, PhD(c), Second Department of Urology, National and Kapodistrian University of Athens, Sismanogleio General Hospital, Athens, Greece; E-mail: lazarostzelves@gmail.com.

Machine learning (ML) is a subset of artificial intelligence (AI), which is being increasingly utilized in medicine to make predictions and detect relationships not easily found using conventional statistics. The use of ML in the field of BPE does not have a long history. A recent study [1] assessed the use and applicability of several ML algorithms in predicting surgical outcomes of BPE patients, where it was reported that a Random Forest algorithm outperformed other alternatives in predicting changes in desired outcomes, such as Qmax increase (the speed of urinary stream as measured using a special device) and International Prostate Symptom Score (IPSS) reduction (questionnaire regarding symptoms of BPE) [1]. Clustering is an unsupervised method that works on datasets in which there is no target variable, nor is anything known about the relationship between the observations (i.e., unlabeled data) and aims to uncover subgroups with substantial intra-group similarity and inter-group difference.

Considering the clinical diversity of BPE patients, we designed this study to assess whether a clustering analysis using the k-means algorithm could provide a framework for identifying patient phenotypes suitable for the selection of each surgical technique.

2.   Methods and Materials

Patients suffering from BPE, without any indication for malignancy, who were admitted for operation at our Urology Department between 2017 and 2019, were operated with one of the available methods (mTUR-P, bTUR-P, bTUVis). Each of these three methods is proved to be equivalent in efficacy and safety based on randomized clinical trials [2]. The hospital review board approved the study protocol for using these data. Inclusion criteria were: prostate size >30ml, indication for surgery, no indication or diagnosis for prostate adenocarcinoma, IPSS>7, and Qmax<15 ml/sec. Baseline patient data (age, comorbidities, use of anticoagulants, ASA score), as well as disease-related characteristics (IPSS, Qmax, post-void residual volume, PSA value, haemoglobin, and sodium values), were collected both pre-and postoperatively. In our clustering analysis we used the variables 'Age','PSA','ProstateVolume','Hbbefore','Nabefore', 'PVRbefore', 'Qmaxbefore', 'IPSSbefore'.

To perform the cluster analysis in programming language R [3], we used the '*dplyr*', '*ggpubr*', and '*factoextra*' packages [4-6] and the *k*-means algorithm [7,8]. The *k*-means clustering algorithm takes as input *n* observations and an integer, *k*, which indicates the preferred number of groups into which these *n* observations are partitioned, such that each observation belongs to the cluster (group) with the closest mean. The iterative procedure of the *k*-means algorithm is used to reduce the distance between each observation and the mean of the corresponding cluster.

3.   Results

BPE patient profile varies according to disease parameters, patient comorbidities, and baseline profile. Certain characteristics would lead to proper surgical technique selection to achieve the best surgical outcome in an ideal situation. However, due to the heterogeneous patient profiles, personal preferences, and healthcare system-related parameters (physician skill, equipment availability, economic factors), clinical practice differs, even in the same department.

Our findings suggest that when using *k*-means to categorize patients according to their pre-operative characteristics, no standard pattern is observed regarding technique selection. Specifically, the method used in real-life was commonly different from the predicted. Although these results may initially seem disappointing, we believe that they reflect reality. The output of ML algorithms is based on measurable factors such as age, prostate size, symptom scores, etc. However, in daily clinical practice, physicians and surgeons confront a variety of non-measurable factors such as patient preferences, time limitations, and economic considerations, such as the immediate availability of consumables. All these parameters may contribute to the observed disagreements in our findings and indicate the need for future analysis using additional parameters. In the table below (Table 1) we can see the allocation of the 153 observations (patients) into the three clusters (1,2,3) based on *k*-means relative to the surgical method that the surgeons used. The ideal form of a table like the one below would be each row and each column containing only one non-zero element.

Table 1. The allocation of the 153 observations into the three clusters based on the k-means algorithm

Method	Clusters			Sum
	1	2	3	
bTUR-P	24	15	13	52
mTUR-P	17	16	16	49
bTUVis	13	20	19	52
Sum	54	51	48	153

The following plot (figure 1) has been produced by using the *ggscatter()* function from the *ggpubr* package to visualize the clusters alongside the surgical method that the surgeons used.
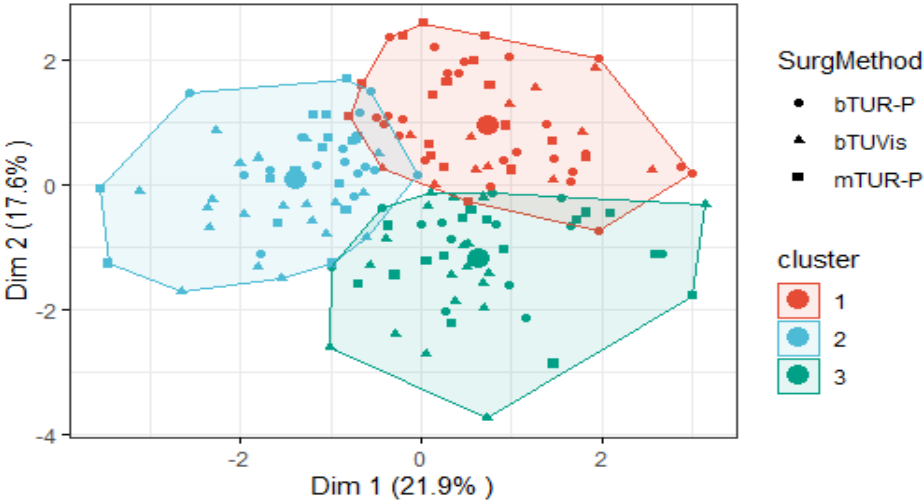

Figure 1. Visualization of the clusters alongside the surgical method used by the surgeons.

The corresponding R code that was used and the plots deduced from this cluster analysis are available in the online appendix [9].

4.   Discussion and Conclusion

A brief discussion of the above findings is due (with respect to Figure 1). One can spot the fairly uniform density of points near the (0,0) origin of axes. Indeed, the rather artificial demand that "3" clusters be detected could have been easily sidelined and one could have done just as well with one cluster only and simply decide to focus on the outliers; those points which are further from the (0,0) origin. Apparently, this is a confirmation of the introductory description that sets the context; surgeons seem at ease to select among the alternative techniques, with no particular preference. Perhaps, then, this suggests that the most interesting aspect of this surgical procedure is the extent to which a department might make a longer-term plan to optimize the non-medical aspects of the actual treatment (for example, by focusing on reducing costs or hospital stays) or, more importantly, to decide to federate its data with other departments, after an anonymization process, to allow for a more comprehensive data analysis. Either way, it should be clear that exploratory data analysis, in the context of applying machine learning techniques, is at the start (and not at the end) of an organizational improvement process.

Acknowledgements

References

[1]   Mourmouris P, Tzelves L, Feretzakis G, Kalles D, et al. The use and applicability of machine learning algorithms in predicting the surgical outcome for patients with benign prostatic enlargement. Which model to use? Arch Ital Urol Androl. 2021 Dec 20;93(4):418-424.

[2]   EAU Guidelines on Non-neurogenic Male LUTS. EAU Guidelines. Edn. presented at the EAU Annual Congress Amsterdam 2022. ISBN 978-94-92671-16-5. 2021

[3]   R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2021. URL https://www.R-project.org/.

[4]   Alboukadel K. ggpubr: 'ggplot2' Based Publication Ready Plots. R package version 0.4.0. 2020.

[5]   Alboukadel K. and Fabian M. (2020). factoextra: Extract and Visualize the Results of Multivariate Data Analyses. R package version 1.0.7.

[6]   Wickham H, Francois R, Henry L, Muller K. dplyr: A Grammar of Data Manipulation. R package version 1.0.7. 2021.

[7]   Hartigan JA, Wong MA. Algorithm AS 136: A K-means clustering algorithm. Applied Statistics. 1979;28:100-108. 10.2307/2346830.

[8]   Forgy EW. Cluster analysis of multivariate data: efficiency vs interpretability of classifications. Biometrics. 1965;21:768-769.

[9]   Feretzakis G, Cluster Analysis assessment in proposing a surgical technique for benign prostatic enlargement.   Available   online: https://www.learningalgorithm.eu/ClusterAnalysisUrolICIMTH2022.html (accessed on 28 April 2022).