

Cluster-Based Toxicity Estimation of Osteoradionecrosis via Unsupervised Machine Learning: Moving Beyond Single Dose-Parameter Normal Tissue Complication Probability by Using Whole Dose-Volume Histograms for Cohort Risk Stratification

Running Title: Cluster-based toxicity estimation of ORN

Authors:

Syedmohammadhossein Hosseinian ^{a,*}, Mehdi Hemmati ^{b,c}, Cem Dede ^c, Travis C. Salzillo ^c, Lisanne V. van Dijk ^d, Abdallah S. R. Mohamed ^c, Stephen Y. Lai ^e, Andrew J. Schaefer ^b, Clifton D. Fuller ^{b,c}

^a Department of Mechanical and Materials Engineering, University of Cincinnati, Cincinnati, Ohio, USA

^b Department of Computational Applied Mathematics & Operations Research, Rice University, Houston, Texas, USA

^c Department of Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

^d Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, Groningen, Netherlands

^e Department of Head and Neck Surgery, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

*Corresponding author:

Syedmohammadhossein Hosseinian, Ph.D.
Department of Mechanical and Materials Engineering
University of Cincinnati
P.O. Box 210072
Cincinnati, Ohio, USA 45221-0072
E-mail: s.hosseinian@uc.edu

Conflict of interest statement: The authors declare no conflicts of interest.

Funding statement: Dr. Schaefer and Fuller received/receive related funding and salary support from National Science Foundation (NSF)/National Institutes of Health (NIH) National Cancer Institute (NCI) via the NSF/NIH Smart and Connected Health Program (R01 CA257814). Dr. Fuller received/receives related support from the National Institute of Dental and Craniofacial Research (NIDCR) Establishing Outcome Measures for Clinical Studies of Oral and Craniofacial Diseases and Conditions award (R01DE025248). Dr. Fuller receives grant and infrastructure support from MD Anderson Cancer Center via: the Charles and Daneen Stiefel Center for Head and Neck Cancer Oropharyngeal Cancer Research Program; the Program in Image-guided Cancer Therapy; and the NIH/NCI Cancer Center Support Grant (CCSG) Radiation Oncology and Cancer Imaging Program (P30CA016672). Dr. Fuller has received unrelated direct industry grant/in-kind support, honoraria, and travel funding from Elekta AB, and has served as a consulting capacity for Varian/Siemens Healthineers, Philips Medical Systems, and Oncospace, Inc.

Abstract

Purpose:

Given the limitations of extant models for normal tissue complication probability estimation for osteoradionecrosis (ORN) of the mandible, the purpose of this study was to enrich statistical inference by exploiting structural properties of data and provide a clinically reliable model for ORN risk evaluation through an unsupervised-learning analysis.

Materials and Methods:

The analysis was conducted on retrospective data of 1,259 head and neck cancer (HNC) patients treated at the University of Texas MD Anderson Cancer Center between 2005 and 2015. The (structural) clusters of mandibular dose-volume histograms (DVHs) were identified through the K-means clustering method. A soft-margin support vector machine (SVM) was used to determine the cluster borders and partition the dose-volume space. The risk of ORN for each dose-volume region was calculated based on the clinical risk factors and incidence rates.

Results:

The K-means clustering method identified six clusters among the DVHs. Based on the first five clusters, the dose-volume space was partitioned almost perfectly by the soft-margin SVM into distinct regions with different risk indices. The sixth cluster overlapped the others entirely; the region of this cluster was determined by its envelopes. These regions and the associated risk indices provide a range of constraints for dose optimization under different risk levels.

Conclusion:

This study presents an unsupervised-learning analysis of a large-scale data set to evaluate the risk of mandibular ORN among HNC patients. The results provide a visual risk-assessment tool (based on the whole DVH) and a spectrum of dose constraints for radiation planning.

Keywords: Radiation injuries, Osteoradionecrosis, Normal tissue complication probability, Clustering, Unsupervised machine learning

1. Introduction

Osteoradionecrosis (ORN) of the mandible is a debilitating side effect of radiation therapy for head and neck cancer (HNC) patients [1, 2, 3, 4, 5]. ORN is commonly defined as non-healing bone for a period of at least three months due to exposure to radiation [6, 7, 8]. Despite its relatively low prevalence [9, 10, 11], ORN may severely affect the quality of life of surviving HNC patients; it is often accompanied by pain, dysaesthesia, dysgeusia, and difficulties in mastication and speech, and in case of progression, it may lead to infection, bone fracture, and intraoral or extra-oral fistulae [1, 6, 12, 13, 14]. With the improved life expectancy of HNC patients, especially among young adults with human papillomavirus associated tumors [11], the importance of ORN prevention strategies in modern practice is spotlighted more than ever before. This requires developing reliable risk evaluation models for ORN to guide dose optimization/adaptation in radiation therapy planning.

Normal tissue complication probability (NTCP) is a widely accepted prediction tool for radiation toxicity [15, 16, 17, 18, 19]. NTCP models aim to provide the likelihood of treatment-induced complications based on the radiation dose delivered to organs at risk (OARs) as well as other clinical risk factors. The current paradigm of NTCP modeling for HNCs relies on supervised-learning (classification) methods, most popularly the classical *logistic regression* method [20]. During the past decade, a vast number of classification-based NTCP models have been proposed for various HNC treatment-induced complications [11, 21, 22, 23, 24, 25, 26, 27, 28, 29]. The classification methods, however, have a major limitation for NTCP modeling—due to *multicollinearity* of dose features—which undermines clinical reliability of these models. Distribution of radiation dose over an OAR is commonly represented by a dose-volume histogram (DVH). As an OAR's DVH depicts a *distribution*, the dose-related features for these models (i.e., DVH values) are highly correlated. As a result, the entire DVH curve is summarized into a single number (e.g., mean dose), which completely disregards the *distribution* of radiation dose and the fact that clinically different DVH curves—with different toxicity outcomes—may share a feature value. Single-parameter evaluation of dose distributions has been reported as a caveat of early NTCP models (e.g., LKB) for clinical use [16], yet it has persisted throughout the subsequent data-driven models, due to the multicollinearity issue.

In addition, the best performance of classification methods concerns *separable* data [30]. In general, these methods aim to learn the best separator of two training classes (i.e., safe radiation plans vs. those leading to complications) from observed clinical data. They perform well even if the data is softly separable (i.e., with a relatively few violations), but in case of inseparable (i.e., highly mixed) classes, they generally do not perform better than a *naive Bayes* classifier. Naive Bayes is a simple classifier whose predictions mainly rely on the observed incidence rates [30]. The effect of data inseparability on classification accuracy is shown in Appendix A. Based on available data of mandibular ORN development among HNC patients [31], there exists evidence that the two classes of ORN and non-ORN patients are highly inseparable by radiation dose to the mandible; see Appendix B. The inseparability observation is aligned with an evidence-based pathogenesis theory of ORN, suggesting that radiation arteritis leads to the development of a hypocellular, hypovascular, and hypoxic environment, and consequently ORN [6, 7]. This implies that a decisive factor in ORN development is the radiation dose received by mandibular arteries, including arterioles and capillaries. Capturing the dose delivered to such a fine network is impractical with the current technologies. On the other hand, DVHs summarize the spatially

(3D) distributed radiation dose into one-dimensional data. For statistical inference, such a “summary” is a sufficient statistic only if radiosensitivity of the OAR is spatially homogeneous. However, homogeneous distribution of arteries across the mandible is an unrealistic assumption. This suggests that, while mandibular DVHs are the best available clinical indicator of ORN development, they generally do not contain sufficient information to separate the two classes (ORN vs. non-ORN) rigorously as intended in the advanced classification methods.

The aforementioned limitations of supervised-learning methods motivate alternative analyses to build more informative and clinically reliable models for evaluation of the risk of ORN. This paper presents an unsupervised-learning approach for ORN risk assessment and the results of the analysis on retrospective data of a large cohort of HNC patients.

2. Materials and Methods

2.1. Data

The analysis was conducted on a data set compiled and made publicly available by van Dijk et al. [31]. The data is composed of anonymized (retrospective) clinical records and the mandibular DVHs of 1,259 HNC patients treated by radiation therapy alone or in combination with surgery and/or chemotherapy at the University of Texas MD Anderson Cancer Center between 2005 and 2015, following an institutional review board approval (RCR030800). The clinical records include age, gender, cancer subsite, T-stage, N-stage, pre-radiation (within 6 weeks before treatment) dental extraction status, smoking status and pack-years, chemotherapy treatment, definitive vs. postoperative radiation therapy, and the mandible bone volume. The reported DVH values include D2%, D5% to D95% in 5% increments, D97%, D98%, and D99% as well as V5Gy to V70Gy in 5 Gy increments; the minimum, maximum, and mean dose (Gy) delivered to the mandible are also included. Out of the 1,259 patients, 1,086 (86.3%) have been reported ORN-free during the minimum of 12 months of post-therapy follow-up; 173 (13.7%) patients are reported to have developed ORN (grade I to IV) during the follow-up period. A complete description of the data, including more information about the cohort and treatments as well as the details of data extraction and processing, has been presented in [11]. The data is publicly available at <https://doi.org/10.6084/m9.figshare.13568207>.

2.2. Analysis

This paper presents a secondary analysis of the data set described above; van Dijk et al. [11] have used this data to train a multivariate logistic regression NTCP model for mandibular ORN (any grade). Their univariate analysis, followed by AIC-based forward stepwise selection, has identified D30% and pre-radiation dental extraction (PDE) status as the model features; PDE is a nominal (binary) feature indicating no/edentulous vs. dental extractions. While their statistical analysis rigorously identifies radiation intensity and dental extraction as significant predictive factors of ORN development, due to multicollinearity of the dose features, the DVH curve is reduced to D30% in their model. Besides, our analysis shows that the performance of the proposed logistic regression model is almost equivalent to that of naive Bayes (AUC of 0.77 for logistic regression vs. 0.75 for naive Bayes); see Appendix C. This confirms that, due to inseparability of classes, the logistic regression predictions are mainly driven by the ORN incidence rates with respect to the model features. Based on this observation, we present a

cluster-based analysis that provides more clinical insight about the risk of ORN development as a function of radiation intensity and dental extraction status.

Our method takes advantage of the structural properties of mandibular DVHs to divide them into disjoint subgroups (of similar radiation plans), each spanning a distinct region of the dose-volume space. It then renders a risk index for each region based on the observed incidence rate of ORN (depending on PDE status) in the associated DVH subgroup. The analysis was conducted in three steps, described below. Steps 1 and 2 were performed using the scikit-learn library of Python [32].

Step 1: The K-means clustering method [30] was employed to identify the inherent clusters of the mandibular DVHs. To this end, each DVH curve was represented as an array of dose values (i.e., $Dx\%$ for $x\%$ ranging from 2% to 99%); the inertia (within-cluster sum-of-squares) criterion was used to identify an optimal number of clusters.

Step 2: To partition the dose-volume space with respect to these clusters, the shared borders of adjacent clusters were identified by a soft-margin support vector machine (SVM) [30]. The analysis was conducted pointwise (for each $Dx\%$ individually) to obtain smooth borders among the clusters with minimal violations.

Step 3: For each region of the dose-volume space, spanned by an identified cluster, the ORN incidence rate (within the cluster) per PDE status was reported as the associated risk index. As a result, each dose-volume region was associated with two risk indices, for no/edentulous and dental extractions, separately.

3. Results

The K-means clustering method identified 5-6 clusters among the mandibular DVHs. In general, as the number of clusters (K) increases, the clustering inertia decreases; an optimal number of clusters is identified by the value of K after which the inertia curve behaves linearly. Figure 1 illustrates the inertia plot of our analysis. Based on this figure, $K = 5$ is optimal; however, the sixth cluster identified by the clustering method (through $K = 6$) is structurally different from the first five clusters and provides additional insight. Therefore, we considered both cases.

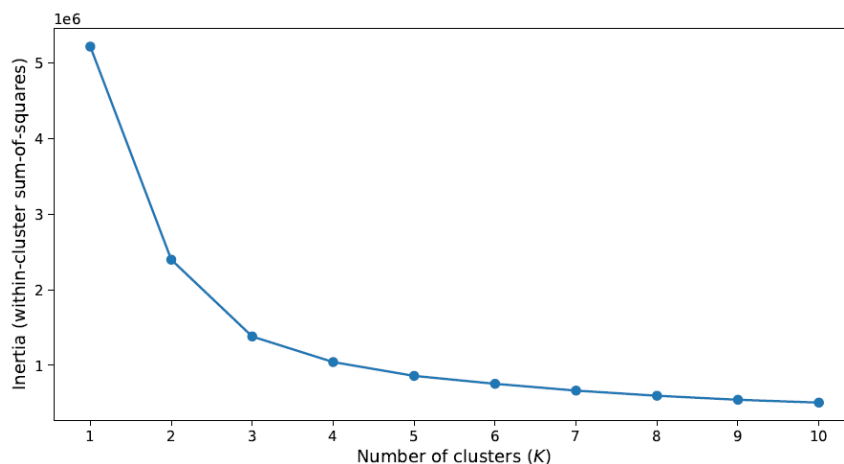


Figure 1: Inertia plot of the K-means clustering method

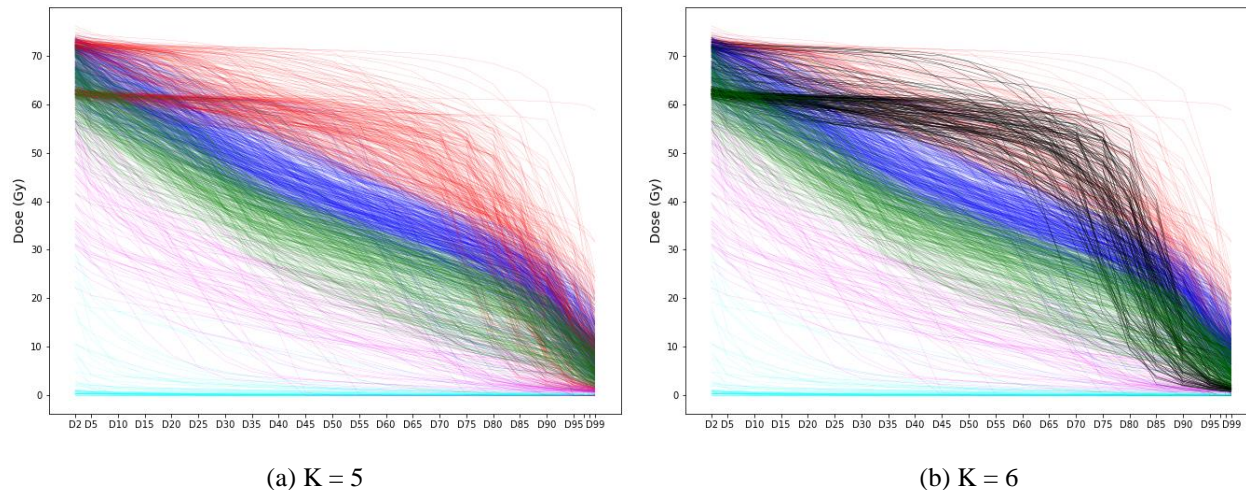


Figure 2: Visualization of the identified clusters

Figure 2 illustrates the identified clusters by the K-means clustering method, color-coded for better visualization. We have also represented DVHs directly by $Dx\%$ values (in Gy) rather than the conventional dose-volume axes, to increase readability. An important distinction between the results for $K = 5$ and $K = 6$ concerns the sixth cluster, depicted in black in Figure 2b. This cluster is mainly composed of a group of DVHs with high dose (above 60 Gy) almost constantly imposed over most of the mandible volume. While this group is visually distinguishable from the other clusters, with $K = 5$, it has been identified as a part of the top cluster, illustrated in red, implying it is most similar to that cluster structurally.

As the next step of the analysis, a soft-margin SVM was employed to identify the cluster borders and partition the dose-volume space; SVM renders the best separator of two classes. As the analysis was performed pointwise, the best separating point (with minimum violations) was identified for every pair of adjacent clusters for each $Dx\%$. With $K = 5$, as the black cluster was considered a part of the red cluster and highly overlapping with the others in the most left and right parts of the plot, SVM failed to identify its border in the $[D2\%, D5\%]$ and $[D90\%, D99\%]$ intervals. However, with $K = 6$ and exclusion of the sixth (black) cluster, SVM was able to identify the cluster borders and partition the dose-volume space almost perfectly. For the sixth cluster, we determined the corresponding dose-volume region by its envelopes (i.e., pointwise minimum and maximum dose). Figure 3 illustrates the cluster borders by dashed black lines. As the choice of $K = 6$ provided a more accurate partition of the dose-volume space, we present the final step of the analysis with six clusters.

For each part of the dose-volume space, characterized by the borderlines of the identified clusters, two risk indices were calculated based on the observed incidence rate of ORN in the corresponding cluster per PDE status. Table 1 reports the observed frequency of ORN development among the patients for each cluster per PDE status. In this table, $PDE = 0$ indicates no/edentulous dental extractions, and $PDE = 1$ concerns performed dental extractions. We have labeled the first five clusters according to their order from the lower left to the upper right corners of the DVH plot in Figure 3b; cluster No. 6 is the overlapping cluster, illustrated in black in Figures 3b and 3c. The columns “ORN incidence” indicate the number of patients with a

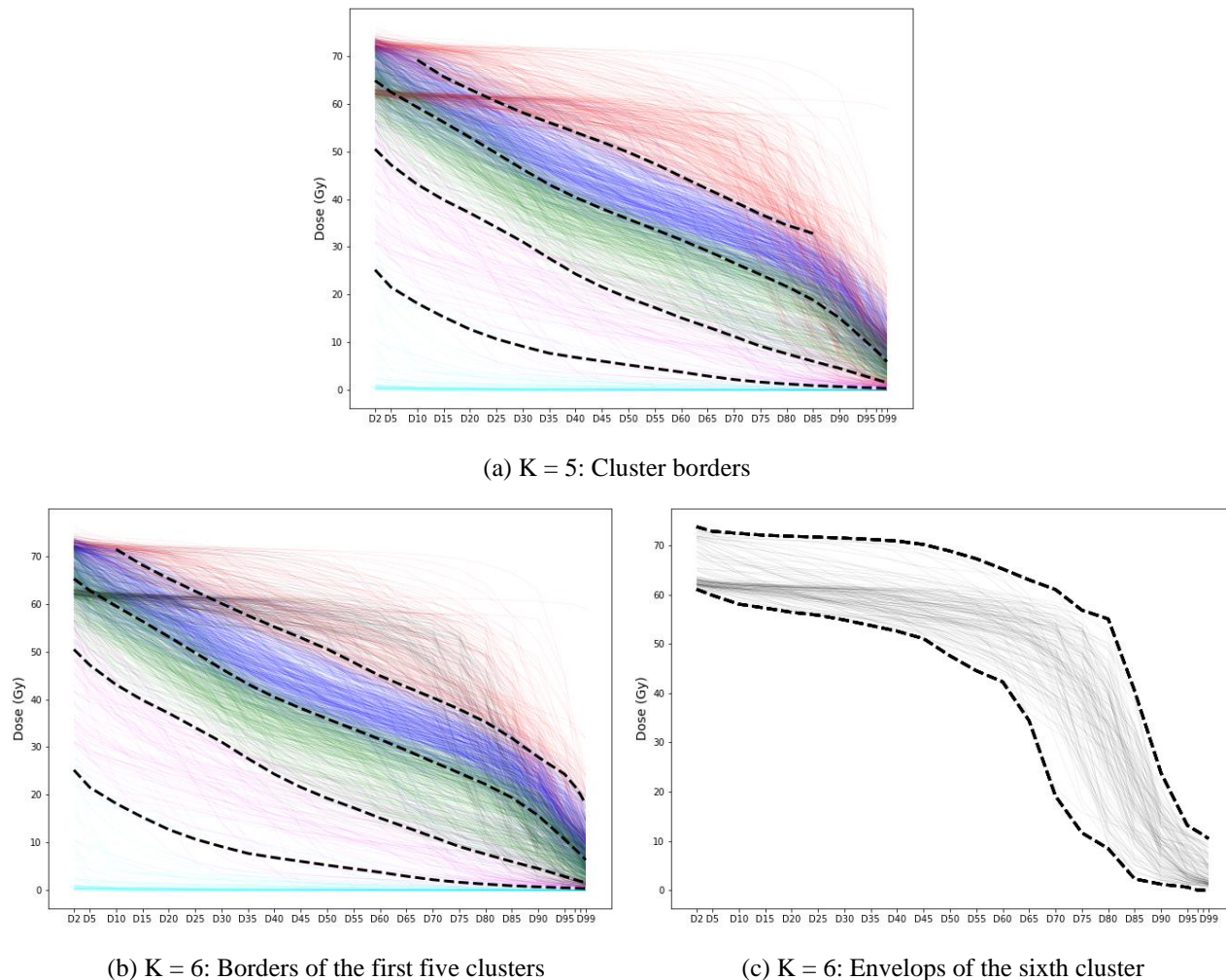


Figure 3: Cluster borders identified by a soft-margin SVM

Table 1: ORN incidence rates among the identified clusters per PDE status

Cluster No.	PDE = 0		PDE = 1	
	ORN incidence	Risk index (r)	ORN incidence	Risk index (r)
1	0 out of 58	0.0 %	0 out of 9	0.0 %
2	2 out of 68	2.9 %	0 out of 8	0.0 %
3	8 out of 273	2.9 %	10 out of 86	11.6 %
4	39 out of 318	12.3 %	34 out of 144	23.6 %
5	31 out of 118	26.3 %	14 out of 47	29.8 %
6	21 out of 82	25.6 %	14 out of 48	29.2 %

PDE = Pre-radiation dental extraction (0 means no/edentulous and 1 means dental extractions.)

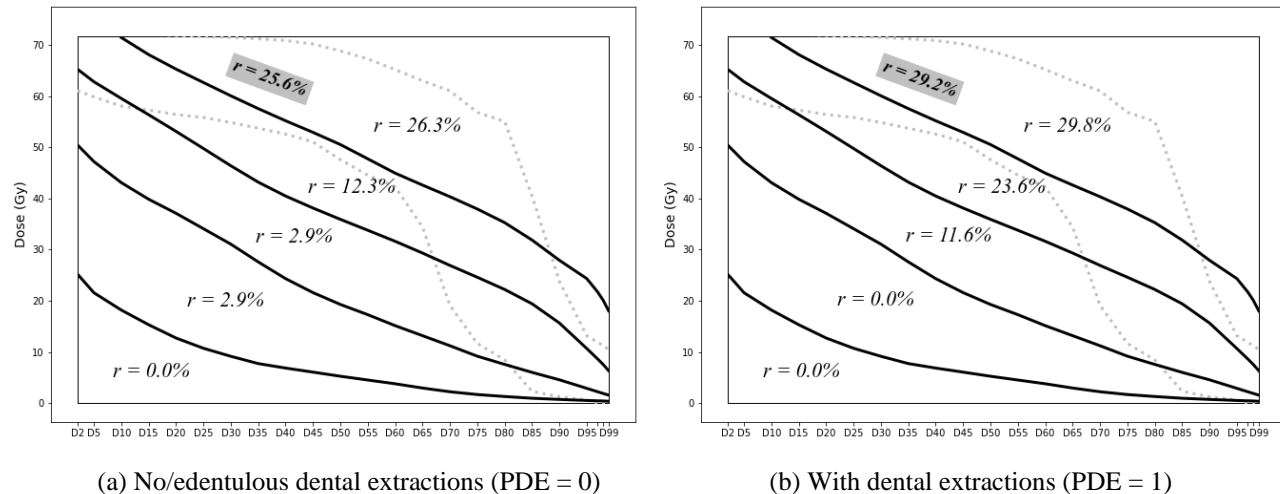


Figure 4: Risk indices of dose-volume regions

positive ORN record out of the total number of patients in that cluster with the same PDE status, and “Risk index (r)” is the corresponding ratio. Figure 4 shows the risk indices on the dose-volume regions identified in the previous step; each region spans the area between two consecutive cluster borders, except for the region of cluster No. 6, which was determined by its envelopes. The borders of the sixth cluster are depicted by dotted lines in Figure 4.

The results clearly show significant correlations between the risk of ORN and radiation intensity as well as PDE status, which is aligned with the univariate analysis results and confirms the relevance of structural clusters to the risk of ORN development. As mentioned earlier, cluster No. 6 (black) is structurally most similar to cluster No. 5 (red); it was identified as a part of the red cluster with $K = 5$. The results also show that these two clusters have very similar risk indices in both PDE categories (25.6% vs. 26.3% for PDE = 0 and 29.2% vs. 29.8% for PDE = 1). This further reveals the merit of the analysis results. Our method provides a visual tool for ORN risk evaluation of radiation plans based on the region that its mandibular DVH falls in and the patient’s PDE status. As the risk is characterized by the whole DVH (not a single-value summary), the results provide a spectrum of dose constraints (i.e., limit on various $Dx\%$ values for dose optimization) to ensure that the risk of ORN development stays below a certain level.

4. Discussion

The main advantage of the presented analysis is evaluation of radiation plans based on the whole DVH. As mentioned earlier, due to the multicollinearity issue, DVHs are reduced to a single-value summary in classification-based NTCP models, which disregards the fact that clinically different DVH curves—with different toxicity outcomes—may share a feature value. This can be observed by the toxicity outcome for the patients whose D30% is in the range of 55 Gy to 60 Gy, in our data set. According to van Dijk et al. [11] model, the risk of ORN for these patients (per PDE status) is relatively the same, as their mandibular DVHs have close D30% values. However, the actual treatment outcomes for these patients are different and arguably determined by the shape of DVHs. Figure 5 shows these patients’ DVHs; the incidence of ORN among the blue DVHs is 13.6% (3 out of 22) and among the black DVHs is 23.5% (12 out of 51). Other than this

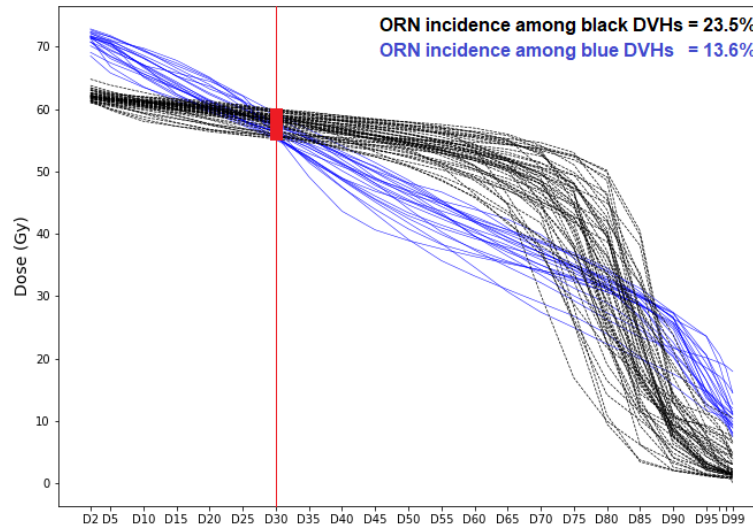


Figure 5: Different ORN incidences among DVHs with the same D30% value

important difference, the van Dijk et al. [11] and our models predict similar risks for these patients, i.e., 13.0%-19.0% by van Dijk et al. [11] compared with 12.3%-25.6% (of the blue and black clusters) by our model for PDE = 0 and 22.4%-31.2% by van Dijk et al. [11] compared with 23.6%-29.2% (of the blue and black clusters) by our model for PDE = 1.

Our motivation for this analysis method was the observation of inseparability of classes (i.e., ORN vs. non-ORN) by measurable features; see Appendix B. Incidence rate is the unique minimum-variance unbiased estimator (MVUE) for the mean of a Bernoulli random variable representing the hidden (unmeasurable) features. In the absence of information about the main separating features (i.e., highly inseparable classes by measurable features), such an estimator offers the most that can be learned from the data. This is supported by the close performances of logistic regression and naive Bayes classifiers on our data; see Appendix C. Those estimates will be naturally more accurate if the data is divided into smaller categories (clusters) characterized by statistically significant (measurable) features, to consider their correlations with the outcome. Thus, the presented analysis not only considers the whole DVH (not a single-value summary), but also enjoys a simple yet robust statistical-inference foundation.

The proposed approach has its own limitations to be considered. First, the choice of the number of clusters (K) in the K-means method is relatively subjective. While the inertia criterion provides straightforward guidance to identify the number of clusters from the mathematical perspective, our analysis showed that such a recommendation is not necessarily optimal from the application standpoint; repeating the analysis with a few different values of K and comparing and contrasting the results will lead to well-informed conclusions. Second, confidence on the risk indices rendered by our model directly relies on the number of observations for each cluster, independent of the others. For example, the inference of $r = 0.0\%$ for the first and second regions of the dose-volume space in Figure 4b is made based on 9 and 8 observations, respectively. While this estimate is acceptable for these low-dose regions, in general, such small samples are highly sensitive to noise. Finally, ORN has low prevalence, which makes the classes of ORN and non-ORN patients unbalanced. This poses a challenge to statistical inference, in general, both supervised- and unsupervised-learning methods.

Visualization tools play a key role in clinical interpretability and reliability of mathematical models. In our analysis, visualization of the identified clusters by the K-means clustering method uncovered the distinction between the sixth and first five clusters and led to an improved partitioning of the dose-volume space. Visualization of the intermediate results may also provide additional insight for radiation planning. For example, the sixth (black) cluster may be named the cluster of *post-operative radiation plans*; more than 77% of the patients in this cluster have received post-operative radiation therapy while this percentage for the other clusters remains below 20%. The special shape of this cluster is easily detectable to oncologists, which immediately reveals the relatively high prevalence of ORN among this group of patients. Availability of a visualization dashboard to illustrate every step of the process (Figures 1 to 3)—rather than just the final result (Figure 4)—can greatly improve clinical reliability of the proposed analysis for risk evaluation of ORN (with different data sets) and other complications.

5. Conclusion

This paper presents the results of an unsupervised-learning analysis for mandibular ORN risk evaluation on retrospective data of a large cohort of HNC patients treated by radiation therapy at the University of Texas MD Anderson Cancer Center between 2005 and 2015. The presented analysis provides a visual tool for ORN risk assessment and a spectrum of dose constraints on the mandible for radiation planning with respect to this complication. The unsupervised-learning approach is extendable to other radiation therapy side effects to improve the current state of NTCP modeling.

References

- [1] A. S. Jacobson, D. Buchbinder, K. Hu, and M. L. Urken. Paradigm shifts in the management of osteoradionecrosis of the mandible. *Oral Oncology*, 46(11):795–801, 2010.
- [2] I. J. Lee, W. S. Koom, C. G. Lee, Y. B. Kim, S. W. Yoo, K. C. Keum, G. E. Kim, E. C. Choi, and I. H. Cha. Risk factors and dose-effect relationship for mandibular osteoradionecrosis in oral and oropharyngeal cancer patients. *International Journal of Radiation Oncology, Biology, Physics*, 75(4):1084–1091, 2009.
- [3] C. G. Murray, J. Herson, T. E. Daly, and S. Zimmerman. Radiation necrosis of the mandible: A 10 year study. Part I. Factors influencing the onset of necrosis. *International Journal of Radiation Oncology, Biology, Physics*, 6(5):543–548, 1980.
- [4] M. S. Teng and N. D. Futran. Osteoradionecrosis of the mandible. *Current Opinion in Otolaryngology & Head and Neck Surgery*, 13(4):217–221, 2005.
- [5] C. J. Tsai, T. M. Hofstede, E. M. Sturgis, A. S. Garden, M. E. Lindberg, Q. Wei, S. L. Tucker, and L. Dong. Osteoradionecrosis and radiation dose to the mandible in patients with oropharyngeal cancer. *International Journal of Radiation Oncology, Biology, Physics*, 85(2):415–420, 2013.

- [6] A. Chronopoulos, T. Zarra, M. Ehrenfeld, and S. Otto. Osteoradionecrosis of the jaws: Definition, epidemiology, staging and clinical and radiological findings. A concise review. *International Dental Journal*, 68(1):22–30, 2018.
- [7] R. E. Marx. A new concept in the treatment of osteoradionecrosis. *Journal of Oral and Maxillofacial Surgery*, 41(6):351–357, 1983.
- [8] P. Pitak-Arnop, R. Sader, K. Dhanuthai, P. Masaratana, C. Bertolus, A. Chaine, J-C. Bertrand, and A. Hemprich. Management of osteoradionecrosis of the jaws: An analysis of evidence. *European Journal of Surgical Oncology*, 34(10):1123–1134, 2008.
- [9] B. M. Beadle, K-P. Liao, M. S. Chambers, L. S. Elting, T. A. Buchholz, K. Kian Ang, A. S. Garden, and B. A. Guadagnolo. Evaluating the impact of patient, tumor, and treatment characteristics on the development of jaw complications in patients treated for oral cancers: A SEER–Medicare analysis. *Head & Neck*, 35(11):1599–1605, 2013.
- [10] T. Reuther, T. Schuster, U. Mende, and A. Kübler. Osteoradionecrosis of the jaws as a side effect of radiotherapy of head and neck tumour patients—A report of a thirty year retrospective review. *International Journal of Oral and Maxillofacial Surgery*, 32(3):289–295, 2003.
- [11] L. V. van Dijk, A. A. Abusaif, J. Rigert, M. A. Naser, K. A. Hutcheson, S. Y. Lai, C. D. Fuller, A. S. Mohamed, et al. Normal Tissue Complication Probability (NTCP) prediction model for osteoradionecrosis of the mandible in patients with head and neck cancer after radiation therapy: Large-scale observational cohort. *International Journal of Radiation Oncology, Biology, Physics*, 111(2):549–558, 2021.
- [12] B. A. Jereczek-Fossa and R. Orecchia. Radiotherapy-induced mandibular bone complications. *Cancer Treatment Reviews*, 28(1):65–74, 2002.
- [13] S. Otto, C. Pautke, T. Van den Wyngaert, D. Niepel, and M. Schiødt. Medication-related osteonecrosis of the jaw: Prevention, diagnosis and management in patients with cancer and bone metastases. *Cancer Treatment Reviews*, 69:177–187, 2018.
- [14] D. Yang, F. Zhou, X. Fu, J. Hou, L. Lin, Q. Huang, and C. H. Yeh. Symptom distress and interference among cancer patients with osteoradionecrosis of jaw: A cross-sectional study. *International Journal of Nursing Sciences*, 6(3):278–282, 2019.
- [15] X. A. Li, M. Alber, J. O. Deasy, A. Jackson, K. K. Jee, L. B. Marks, M. K. Martel, C. Mayo, V. Moiseenko, A. E. Nahum, et al. The use and QA of biologically related models for treatment planning: Short report of the TG-166 of the therapy physics committee of the AAPM. *Medical Physics*, 39(3):1386–1409, 2012.
- [16] L. B. Marks, E. D. Yorke, A. Jackson, R. K. Ten Haken, L. S. Constine, A. Eisbruch, S. M. Bentzen, J. Nam, and J. O. Deasy. Use of normal tissue complication probability models in the clinic. *International Journal of Radiation Oncology, Biology, Physics*, 76(3):S10–S19, 2010.
- [17] G. Palma, S. Monti, M. Conson, R. Pacelli, and L. Cella. Normal tissue complication probability (NTCP) models for modern radiation therapy. In *Seminars in Oncology*, volume 46, pages 210–218. Elsevier, 2019.

- [18] G. J. Kutcher and C. Burman. Calculation of complication probability factors for non-uniform normal tissue irradiation: The effective volume method gerald. *International Journal of Radiation Oncology, Biology, Physics*, 16(6):1623–1630, 1989.
- [19] A. Niemierko. A generalized concept of equivalent uniform dose (EUD). *Medical Physics*. 26(6):1100, 1999. (Abstract)
- [20] S. Stieb, A. Lee, L. V. Van Dijk, S. Frank, C. D. Fuller, and P. Blanchard. NTCP modeling of late effects for head and neck cancer: A systematic review. *International Journal of Particle Therapy*, 8(1):95–107, 2021.
- [21] M. Bakhshandeh, B. Hashemi, S. R. M. Mahdavi, A. Nikoofar, and A. Vasheghani, M. and Kazemnejad. Normal tissue complication probability modeling of radiation-induced hypothyroidism after head-and-neck radiation therapy. *International Journal of Radiation Oncology, Biology, Physics*, 85(2):514–521, 2013.
- [22] I. Beetz, C. Schilstra, F. R. Burlage, P. W. Koken, P. Doornaert, H. P. Bijl, O. Chouvalova, C. R. Leemans, G. H. de Bock, M. E. Christianen, et al. Development of NTCP models for head and neck cancer patients treated with three-dimensional conformal radiotherapy for xerostomia and sticky saliva: the role of dosimetric and clinical factors. *Radiotherapy and Oncology*, 105(1):86–93, 2012.
- [23] I. Beetz, C. Schilstra, A. van der Schaaf, E. R. van den Heuvel, P. Doornaert, P. van Luijk, A. Vissink, B. F. van der Laan, C. R. Leemans, H. P. Bijl, et al. NTCP models for patientrated xerostomia and sticky saliva after treatment with intensity modulated radiotherapy for head and neck cancer: the role of dosimetric and clinical factors. *Radiotherapy and Oncology*, 105(1):101–106, 2012.
- [24] F. Buettner, A. B. Miah, S. L. Gulliford, E. Hall, K. J. Harrington, S. Webb, M. Partridge, and C. M. Nutting. Novel approaches to improve the therapeutic index of head and neck radiotherapy: an analysis of data from the PARSPORT randomised phase III trial. *Radiotherapy and Oncology*, 103(1):82–87, 2012.
- [25] J. A. Dean, K. H. Wong, H. Gay, L. C. Welsh, A-B. Jones, U. Schick, J. H. Oh, A. Apte, K. L. Newbold, S. A. Bhide, et al. Incorporating spatial dose metrics in machine learning-based normal tissue complication probability (NTCP) models of severe acute dysphagia resulting from head and neck radiotherapy. *Clinical and Translational Radiation Oncology*, 8:27–39, 2018.
- [26] J. A. Dean, K. H. Wong, L. C. Welsh, A-B. Jones, U. Schick, K. L. Newbold, S. A. Bhide, K. J. Harrington, C. M. Nutting, and S. L. Gulliford. Normal tissue complication probability (NTCP) modelling using spatial dose metrics and machine learning methods for severe acute oral mucositis resulting from head and neck radiotherapy. *Radiotherapy and Oncology*, 120(1):21–27, 2016.
- [27] M. Morimoto, H. P. Bijl, A. van der Schaaf, C-J. Xu, R. J. Steenbakkers, O. Chouvalova, Y. Yoshioka, T. Teshima, and J. A. Langendijk. Development of normal tissue complication probability model for trismus in head and neck cancer patients treated with radiotherapy: The role of dosimetric and clinical factors. *Anticancer Research*, 39(12):6787–6798, 2019.
- [28] L. van den Bosch, A. van der Schaaf, H. P. van der Laan, F. J. Hoebbers, O. B. Wijers, J. G. van den Hoek, K. G. Moons, J. B. Reitsma, R. J. Steenbakkers, E. Schuit, et al.

- Comprehensive toxicity risk profiling in radiation therapy for head and neck cancer: A new concept for individually optimised treatment. *Radiotherapy and Oncology*, 157:147–154, 2021.
- [29] K. Wopken, H. P. Bijl, A. van der Schaaf, H. P. van der Laan, O. Chouvalova, R. J. Steenbakkers, P. Doornaert, B. J. Slotman, S. F. Oosting, M. E. Christianen, et al. Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer. *Radiotherapy and Oncology*, 113(1):95–101, 2014.
- [30] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*, volume 112. Springer, 2013.
- [31] L. V. van Dijk, C. D. Fuller, and A. S. Mohamed. Dose-volume histogram (DVH) parameters of the mandible for normal tissue complication probability modelling. 2021. Available at: <https://doi.org/10.6084/m9.figshare.13568207>.
- [32] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

Appendix A: Effect of Data Inseparability on Classification Accuracy

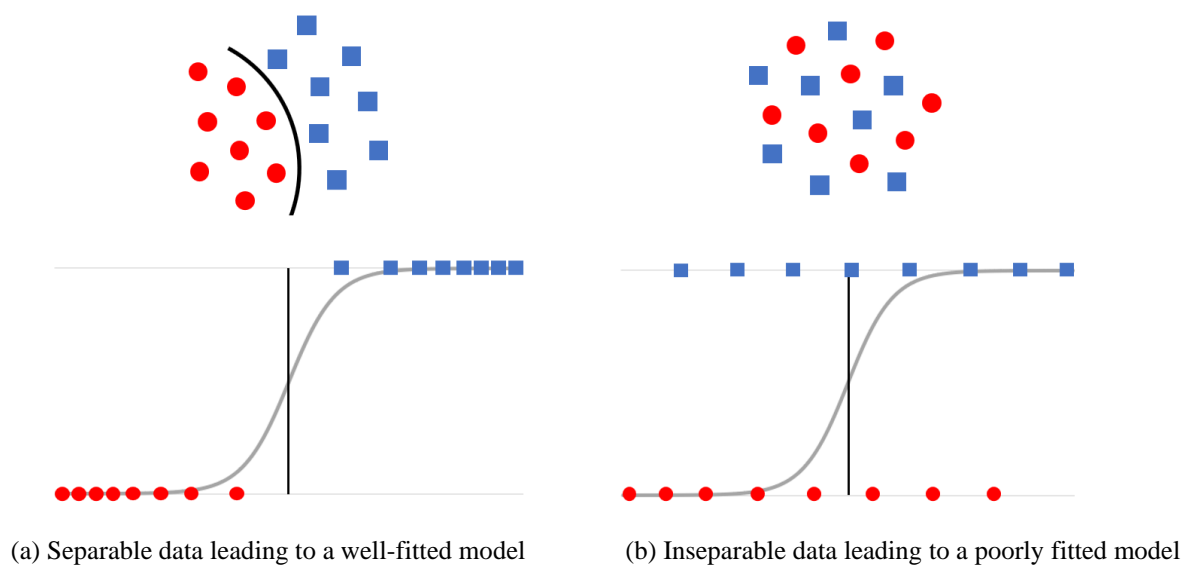


Figure A.1: Data inseparability and its effect on logistic regression performance

Appendix B: ORN Data Inseparability by DVH Features

The following figures show that the two classes of ORN-negative and ORN-positive patients in our data set are inseparable by D30% and pre-radiation dental extraction (PDE) status. Given only two variables, this could be shown on a single plot. However, for better visualization and providing more insight on the structure of the data, we illustrate the data inseparability in smaller groups; that is, for each pair of DVH cluster and PDE status separately. Evidently, if the whole data set is separable, it remains separable when divided in smaller groups. Also, for each patient, we have plotted D30% against the standard deviation (SD) of the delivered dose, for visualization purposes only, so all data points are not located on a horizontal line.

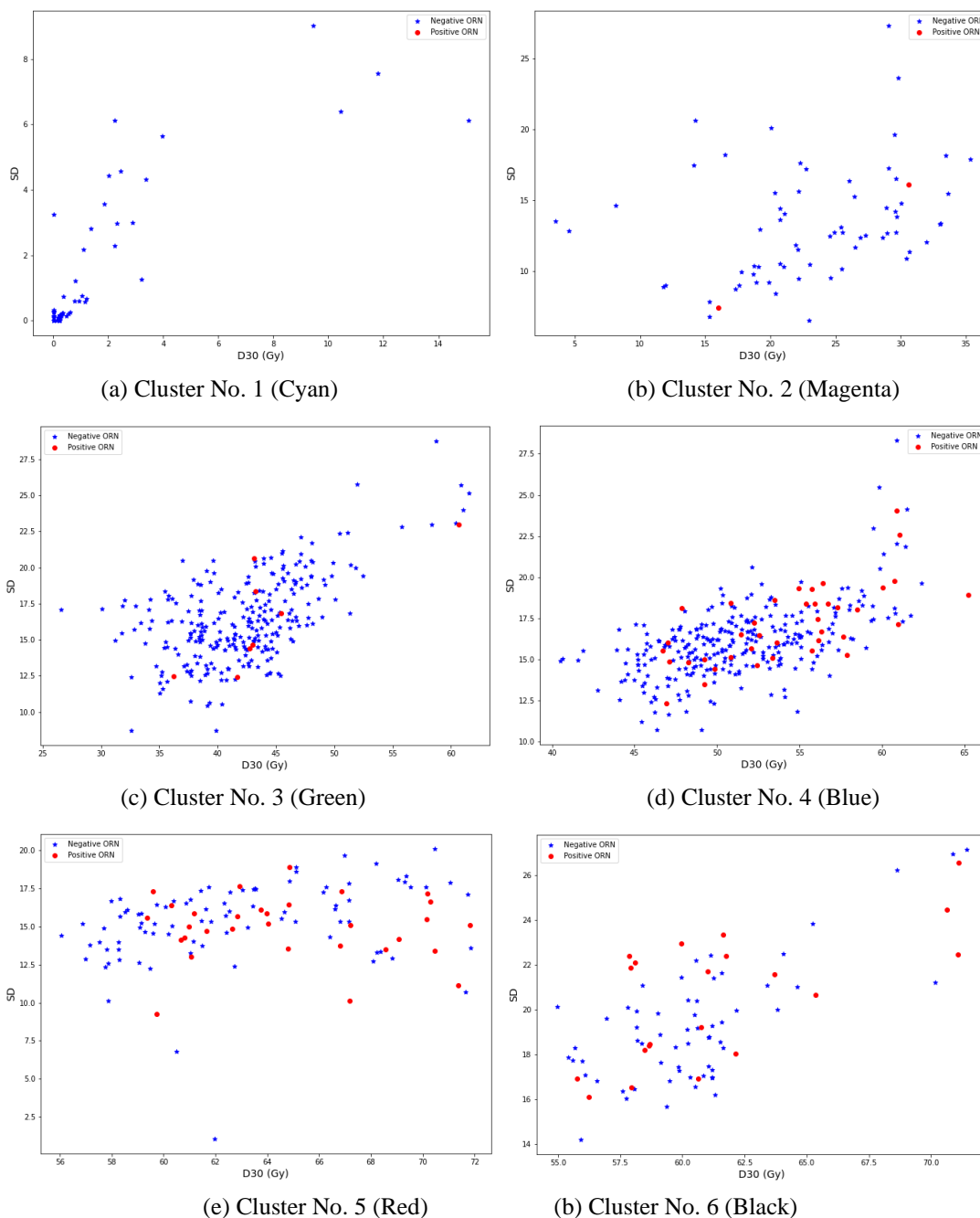


Figure B.1: Data inseparability within the clusters for PDE = 0

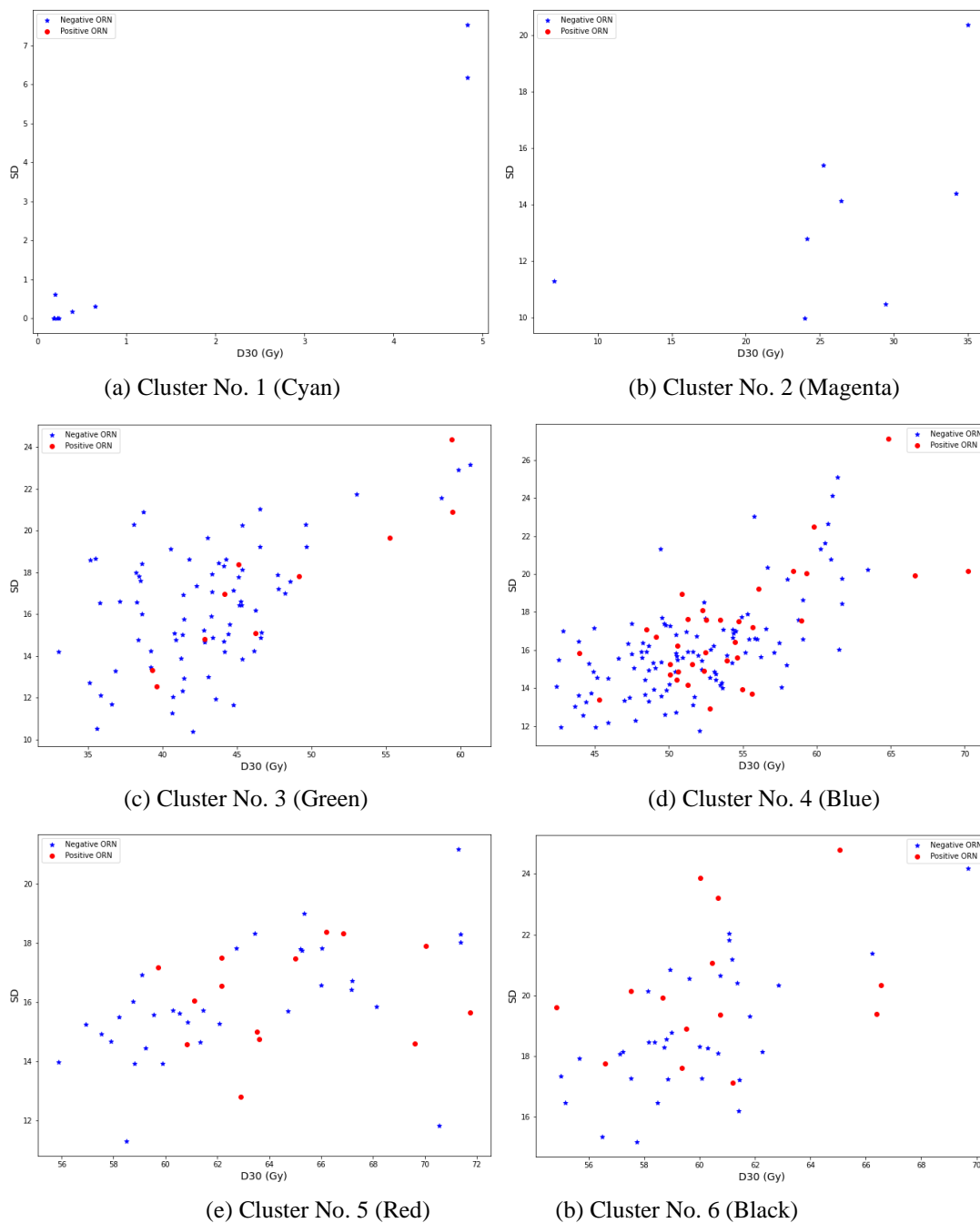
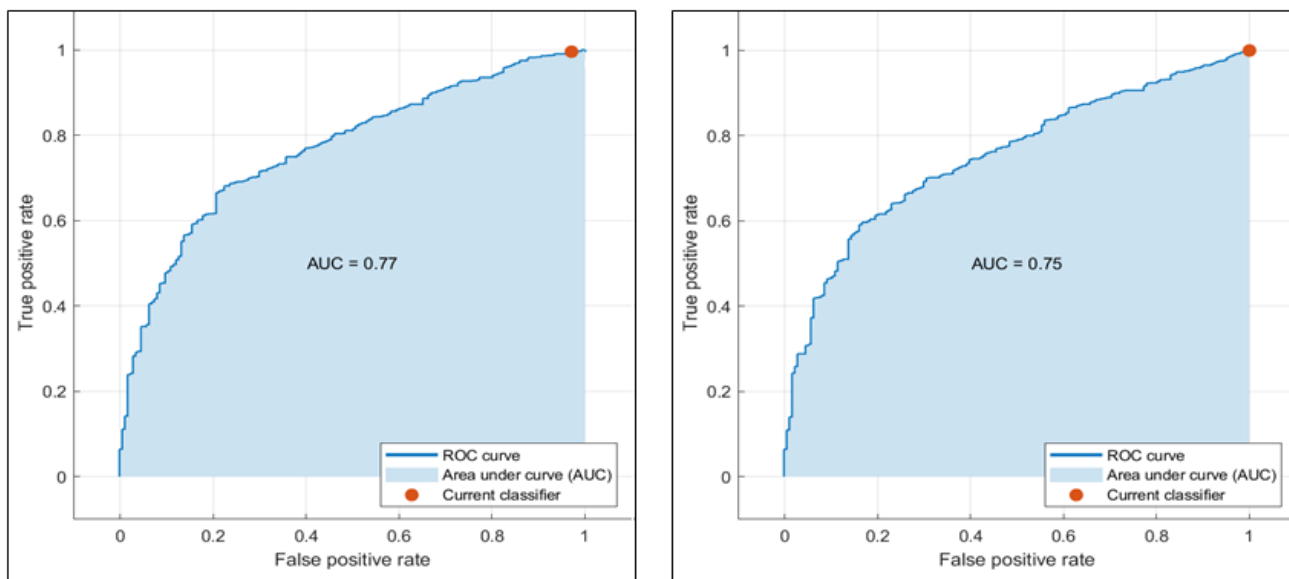


Figure B.2: Data inseparability within the clusters for PDE = 1

Figures B.1 and B.2 show that, in all D30%-PDE categories, the class of ORN-negative patients is dominant, and the two classes are highly mixed (inseparable). As we have used SD (y axis) for better visualization, the data inseparability by D30% (for each PDE status) is implied by nonexistence of vertical separators of “Negative ORN” and “Positive ORN” points. However, the fact that these classes cannot be separated by other lines/curves implies that including SD as a feature is not expected to improve the performance of a logistic regression model. These figures also reveal that introducing “DVH cluster” as an independent feature to a logistic regression model is not likely to significantly improve its performance, due to inseparability of data in all D30%-cluster-PDE categories. Finally, the existence of high (linear) correlations among $D_x\%$ features implies that replacing D30% with any other DVH feature (including the mean dose) will not improve the inseparability.

Appendix C: Logistic Regression vs. Naive Bayes



(a) Logistic Regression

(b) Naive Bayes

Figure C.1: Comparison of the performances of logistic regression and naive Bayes on the data (The features include D30% and PDE status for both models.)