Volume 19, No 1 | International Journal of Radiation Research, January 2021

Clustering of nasopharyngeal carcinoma intensity modulated radiation therapy plans based on k-means algorithm and geometrical features

Z. Zhou[1*], J. Li[1,2], J. Tu[1,2], R. Xin[1,2], W. Zhang[3], D. Wu[1]

[1]*State Key Laboratory of Mechanics and Control of Mechanical Structures, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[2]*Department of Nuclear Science and Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China*
[3]*Cancer Hospital of Tianjin Medical University, Tianjin 300060, China*

► ***Original article***

Corresponding authors:
Zhengdong Zhou, Ph.D.,
E-mail:
zzd_msc@nuaa.edu.cn

Revised: June 2019
Accepted: August 2019
Int. J. Radiat. Res., January 2021; 19(1): 13-21
DOI: 10.29252/ijrr.19.1.13

ABSTRACT

Background: The design of intensity modulated radiation therapy (IMRT) plans is difficult and time-consuming. The retrieval of similar IMRT plans from the IMRT plan dataset can effectively improve the quality and efficiency of IMRT plans and automate the design of IMRT planning. However, the large IMRT plans datasets will bring inefficient retrieval result. ***Materials and Methods:*** An intensity-modulated radiation therapy (IMRT) plan clustering method based on k-means algorithm and geometrical features is proposed to improve the retrieval efficiency from the IMRT plan dataset. The proposed method could benefit future automatic IMRT planning based on prior knowledge. In this study, a collection dataset including 100 cases of nasopharyngeal carcinoma IMRT plans was employed in the clustering experiment. The geometrical features of each cluster center were used to qualitatively predict the dosimetric characteristics of organs at risk (OARs) and compared with practical results. ***Results:*** Experimental results demonstrate that the tested dataset can be well clustered using the proposed method. The predicted dosimetric characteristics of OARs for each cluster agree well with their practical results, and the difficulty of IMRT planning for each cluster can be derived. ***Conclusion:*** The proposed IMRT plan clustering method can bring great benefit to the new cases of IMRT planning.

Keywords: *Intensity-modulated radiation therapy (IMRT) planning, clustering of IMRT plans, IMRT plan*

INTRODUCTION

The design of intensity modulated radiation therapy (IMRT) plans is usually time-consuming trial-and-error process and requires the expertise of experienced medical physicists and oncologists. When designing a IMRT for a new patient, medical physicists try to retrieve similar cases that they have encountered in the past along with their radiation therapy plans and try to map them to the new case. The retrieval of similar intensity-modulated radiation therapy (IMRT) plans from the IMRT plan dataset is crucial to guide the design of IMRT planning for new cases [1-2]. This process can effectively improve the quality and efficiency of IMRT plans and automate the design of IMRT planning. Many researches have demonstrated that the quality and efficiency of IMRT planning can be improved by IMRT plan retrieval. Petrovic *et al.* [1] developed two adaptation approaches to adapt the beam number and beam angles suggested in the retrieved case. Schlaefer *et al.* [2] presented a case-based approach to generate candidate beams from a database of previous treatment plans. Zhou *et al.* [3] proposed an

[Downloaded from ijrr.com on 2026-11-10]
[DOR: 20.1001.1.23223243.2021.19.1.2.3]
[DOI: 10.29252/ijrr.19.1.13]

effective calculation method for an overlap volume histogram descriptor for IMRT plan retrieval. Wu *et al.* [4] utilized the overlap-volume histogram-driven planning method to improve robotic stereotactic body radiation therapy plan quality and planning efficacy. Zhou *et al.* [5] proposed a method for quality control of IMRT planning based on prior knowledge and novel measures derived from both OVHs and DVHs. Yang *et al.* [6] proposed an overlap-volume-histogram based method for rectal dose prediction and automated treatment planning in the external beam prostate radiotherapy. Good *et al.* [7] proposed a knowledge-based approach to improving and homogenizing intensity modulated radiation therapy planning quality among treatment centers. Petit *et al.* [8] used shape-based treatment plan optimization for IMRT of pancreatic adenocarcinoma to increase organ sparing. Kazhdan *et al.* [9] proposed a shape relationship descriptor to retrieval similar cases for radiation therapy planning. Despite their benefits, large datasets lead to low retrieval efficiency. Therefore, the effective retrieval of similar cases in a large-scale IMRT plan dataset is challenging and demanding.

In order to improve the retrieval efficiency of IMRT plan, the k-means [10] algorithm and geometrical features are proposed for the clustering of IMRT plans in the present study, where the OVH geometrical feature descriptors are used as the similarity measures, K-means clustering algorithm is used to cluster the cases in the IMRT plan dataset.

In the proposed method, the cluster of a new case is first identified. Then, the similarities between the new case and the cases in this cluster are calculated, and similar cases are retrieved. This process can certainly improve the retrieval efficiency by avoiding useless calculation between the new case and the cases in other clusters. The purpose of this study is to evaluate the performance of the proposed method for the clustering of the IMRT plans. To the best of our knowledge, this study is the first report that focused on the clustering of IMRT plans for improving retrieval efficiency.

MATERIALS AND METHODS

K-means algorithm

K-means is a popular partitional clustering algorithm [10]. For a given dataset, k-means outputs a disjoint set of clusters $\boldsymbol{S} = \{\boldsymbol{S}_1, \boldsymbol{S}_2, \ldots, \boldsymbol{S}_K\}$ and a centroid c_k for each cluster $\boldsymbol{S}_k$. Each centroid c_k is set to have the smallest sum of distances to all elements in $\boldsymbol{S}_k$, making each centroid c_k a good general representation of $\boldsymbol{S}_k$.

Similar to any other algorithms in machine learning, k-means has its inherent weaknesses. For instance, the k-means algorithm is susceptible to the number of clusters, isolated points, and initial clustering centers. One practical solution is to select the initial cluster center based on the density of data points. The underlying idea is to select the data points with large density as the initial center of the cluster and the points with small density as the isolated points.

Geometrical features

For the design of IMRT planning of nasopharyngeal carcinoma (NPC), many organs at risk (OARs), such as brainstem, spinal cord, left and right parotid glands, and left and right crystals, should be considered, as shown in figure 1.

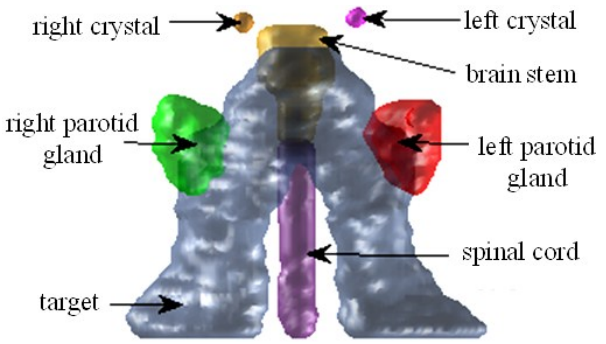


Figure 1. Nasopharyngeal carcinoma target and OARs.

The overlapped volume histogram (OVH) is a commonly used descriptor in the investigation of radiotherapy plan retrieval. The OVH descriptor can describe the geometric relationship between the target and the OARs [3, 4]. However, given the small size and simple shape of the left and right crystals, the signed

[Downloaded from ijrr.com on 2026-11-10]

[DOR: 20.1001.1.23223243.2021.19.1.2.3]

[DOI: 10.29252/ijrr.19.1.13]

distances from their centers to the boundary of the plan target volume (PTV) substitute the OVH descriptors as their geometric descriptors. Therefore, for a PTV T and OAR O, the OVH descriptor of O with respect to T at distance d can be defined as equation (1) [9].

$$\mathrm{OVH}_{O,T}(d) = \frac{|\{p \in O | d(p,T) \le d\}|}{|O|} \quad (1)$$

where $|O|$ is the volume of the OAR; $d(p,T)$ is the signed distance from a point p in the OAR to the boundary of the PTV; $\{p \in O | d(p,T) \ell d\}$ represents a subset of OAR whose distance to the tumor is less than d.

For NPC IMRT plans, six feature attributes are defined as shown in table 1.

Table 1. Feature attributes of nasopharyngeal carcinoma IMRT plan dataset.

Index of feature attributes	Name	Definition
1	OVH_1	OVH descriptor for the brain stem OAR
2	OVH_2	OVH descriptor for the spinal cord OAR
3	OVH_3	OVH descriptor for the left parotid gland OAR
4	OVH_4	OVH descriptor for the right parotid gland OAR
5	d_1	Signed distance from the center of the left crystal to the boundary of the PTV
6	d_2	Signed distance from the center of the right crystal to the boundary of the PTV

Similarity measure

Similarity regarding OVH descriptor

The OVH descriptor is represented as a curve, as shown in Figure 2, where d is the signed distance from a point p in the OAR to the boundary of the PTV, OVH represents the overlapped volume histogram as defined in equation (1). For cases i and j, the similarity between the q-th (q = 1,2,.....,4) OVH descriptors $OVH_{i,q}$ and $OVH_{j,q}$ can be represented with the area $s_{i,j,q}^{OVH}$ between the two curves[3], as shown in the shadow region in figure 2. The smaller the area between the two OVH curves, the more similar the geometric features of the two radiotherapy plans.

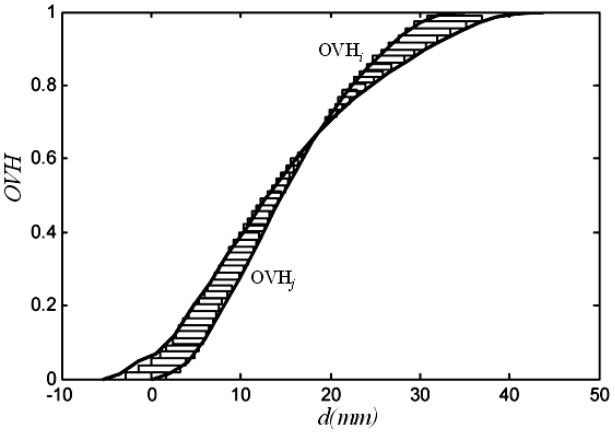


Figure 2. Area between the OVHi and OVHj which are the OVH curves of case i and j.

Similarity regarding the signed distance

For cases i and j, let the signed distances from the center of the left and right crystals to the boundary of the PTV be $d_{i,q}$ and $d_{j,q}$ (q = 1,2). Then, the similarity is expressed by equation (2).

$$s_{i,j,q}^{\mathrm{Len}} = | d_{i,q} - d_{j,q} | \quad (q = 1, 2) \quad (2)$$

Among the signed distances, di,q represents the q-th feature attributes for case i, and dj,q represents the q-th feature attributes of case j.

Similarity for IMRT plan retrieval

In this study, six feature attributes were used to formulate the similarity for IMRT plan retrieval. In general, their values are relatively different. For the collected dataset used in this paper, we calculated the similarities of six feature attributes between two arbitrary cases and obtained their mean similarities, as shown in table 2. Results show that the mean similarities vary significantly in terms of their attribute type.

Table 2. Mean similarities of six feature attributes of 100 IMRT plans

Feature attribute	OVH_1	OVH_2	OVH_3	OVH_4	d_1	d_2
Mean values of similarities	3.02	3.92	5.47	4.95	8.79	8.51

Data preprocessing should be considered before applying any data exploration algorithm,

[Downloaded from ijrr.com on 2026-11-10]
[DOR: 20.1001.1.23223243.2021.19.1.2.3]
[DOI: 10.29252/ijrr.19.1.13]

especially distance-based methods [11]. Normalization is a typical data transformation method that may increase the accuracy and efficiency of mining procedures that involve distance measurements. Such transformation ensures that all attributes have an equal weight.

To normalize the above six attributes, we let n be the number of cases. For cases i and j, the normalized similarity $s^{O}_{i,j,q}$ between the q-th OVH descriptors $OVH_{i,q}$ and $OVH_{j,q}$ or the normalized similarity $s^{L}_{i,j,q}$ between $d_{i,q}$ and $d_{j,q}$ can be represented as equation (3) and (4).

$$s^{O}_{i,j,q} = \frac{s^{OVH}_{i,j,q} - \min(s^{OVH}_{l,p,q})}{\max(s^{OVH}_{l,p,q}) - \min(s^{OVH}_{l,p,q})}, \quad (l = 1,2,...,n;\ p = 1,2,...,n;\ q = 1,2,...,4) \quad (3)$$

$$s^{L}_{i,j,q} = \frac{s^{Len}_{i,j,q} - \min(s^{Len}_{l,p,q})}{\max(s^{Len}_{l,p,q}) - \min(s^{Len}_{l,p,q})}, \quad (l = 1,2,...,n;\ p = 1,2,...,n;\ q = 1,2) \quad (4)$$

The summation similarity $s_{i,j}$ for cases i and j can be expressed as equation (5).

$$s_{i,j} = \sum_{q=1}^{4} s^{O}_{i,j,q} + \sum_{q=1}^{2} s^{L}_{i,j,q} \quad (5)$$

To facilitate the clustering of IMRT plans, the similarity $s_{i,j}$ is used to replace the Euclidean distance in representing the similarity measure of clustering.

To evaluate the clustering performance, we use the parameter SE expressed in equation (6).

$$SE = \sum_{k=1}^{K} \sum_{i \in \boldsymbol{S}_k} s_{i,k}, \quad (6)$$

where K is the number of clusters, $\boldsymbol{S}_k$ is the k-th cluster, and $s_{i,k}$ is the distance from case i to the center c_k of $\boldsymbol{S}_k$, as defined in equation (6).

Clustering steps

In this study, 100 cases of NPC IMRT plans were experimented. The clustering steps are described as follows:

I. Preprocessing. Six feature attributes of all IMRT plans of NPC were calculated as the test dataset X:
X={xi|xi=(OVHi,1,..., OVHi,4, di,1, di,2), i=1,2,...,100}.

II. Clustering. The k-mean algorithm was used for clustering the test dataset X, and si,j was employed as the similarity measure. The clustering steps are listed as follows:

1) Detect the isolated points in the dataset X and remove them;
2) Let the initial cluster number k = 1 and the maximum number of clusters K = √n = 10;
3) Select the data points with densities ranked from 1 to k as the k initial cluster centers, complete the clustering of k clusters, and calculate ES;
4) Let k = k+1, repeat step (3) until k> K;
5) Analyze the change in SE at different numbers of clusters and determine the optimal number of clusters and their center points.

RESULTS

Clustering results

The test dataset $\boldsymbol{X}$ was clustered in accordance with the clustering procedure in Section 2.3. First, the data point radius r was calculated, which is equal to the mean value of the median value for each row of data in the distance matrix $\boldsymbol{X_D}=[s_{i,k}]$. The data densities of each data point within the radius r were calculated, as shown in figure 3(a), where the X axis is the serial number of the 100 cases, the Y axis is the number of cases within the radius r at each case. From this figure, four data points with zero data densities, which were isolated points, were found. Thus, they were removed from the following clustering procedure. Then, we let k = 1,2,...,10, selected the initial centers based on point densities, completed the clustering procedure, and calculated the SE for each k value. The change in SE at different numbers of clusters is shown in figure 3(b). The 3D reconstruction of 8 selected cases in each cluster at k = 3 is shown in figure 4. They are ordered from small to large according to the distance from each data point to its cluster center. As shown in figure 4, the cases can be well clustered at k = 3. Figure 5 shows the 3D reconstruction for the test dataset of isolated cases, which are quite different in shape from each other and also quite different in shape from those cases in figure 4.

[Downloaded from ijrr.com on 2026-11-10]
[DOR: 20.1001.1.23223243.2021.19.1.2.3]
[DOI: 10.29252/ijrr.19.1.13]

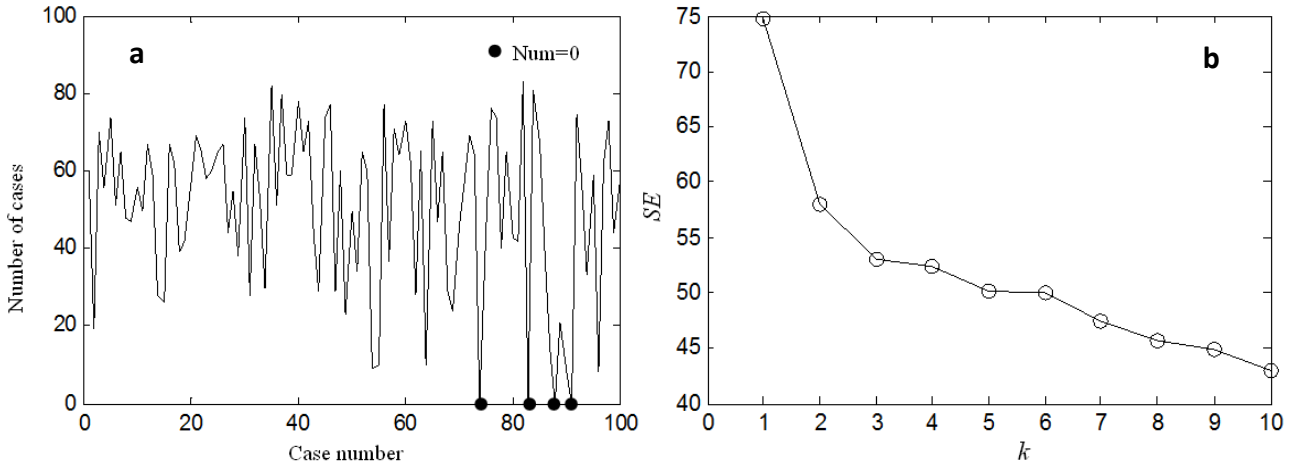


Figure 3. Data density of each case and SE at different numbers of clusters. **(a)** Data density of each case within radius r. The Black spots refer to the isolated cases whose data density is 0. **(b)** SE value at different numbers of clusters from 1 to 10

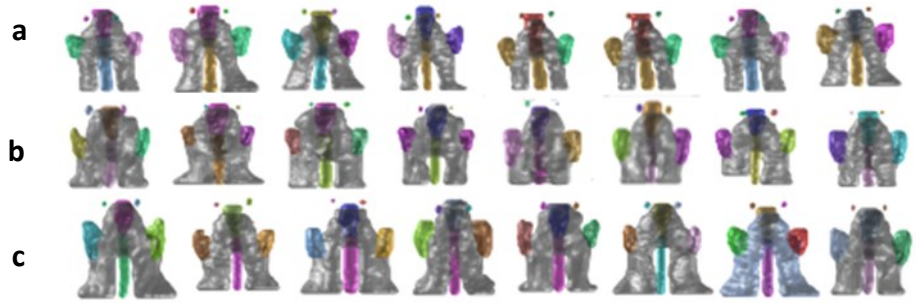


Figure 4. Clustering results for the test dataset **X** at k = 3, 96 cases in total. Cases in each cluster are ordered from small to large according to the distance from each data point to its cluster center. **(a)** 3D reconstruction of 8 cases selected from 34 cases in the first cluster. **(b)** 3D reconstruction of 8 cases selected from 33 cases in the second cluster. **(c)** 3D reconstruction of 8 cases selected from 29 cases in the third cluster.

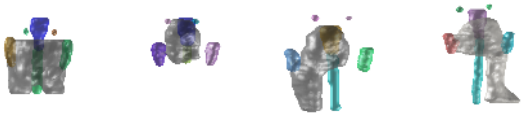


Figure 5. 3D reconstruction of 4 isolated cases which are quite different from each other and the cases in figure 4.

Dosimetry clustering analysis

To evaluate the clustering performance, we used the geometrical features of each cluster center to predict the dosimetric characteristics of the OARs for each cluster. The difficulty of IMRT planning for each cluster was derived and compared with the practical results.

In general, the closer the distance from the OAR to the PTV is, the greater the dose received by the OAR. Therefore, the dosimetric characteristics of the OARs for each cluster can be predicted by analyzing the geometrical features of the cluster centers, which could provide valuable information for the IMRT planning of each case in its corresponding cluster.

Figure 6 shows the OVH_1–OVH_4 of the three clustering centers, in which the black curve, the dotted line, and the long-dashed line are the clustering centers of the first, second, and third clusters, respectively. Table 3 shows the fifth and sixth attributes of the cluster centers, and values represent the average signed distance from the OAR to the PTV in each cluster.

In the test dataset *X*, the first feature attribute

[Downloaded from ijrr.com on 2026-11-10]
[DOR: 20.1001.1.23223243.2021.19.1.2.3]
[DOI: 10.29252/ijrr.19.1.13]

refers to the OVH descriptor of the brainstem with respect to the PTV. The maximum dose received by the brainstem, being a serial organ, should be restricted in IMRT planning. The maximum dose received by the brainstem is significantly affected by the minimum distance $d_{\min}^{b}$ from the brainstem to the PTV. Thus, $d_{\min}^{b}$ is used for the dose rank prediction of the brainstem. Figure 6(a) shows that the $d_{\min}^{b}$ of the second cluster center is larger than that of the first cluster center and that the $d_{\min}^{b}$ of the first cluster center is larger than that of third cluster center. Thus, we can predict that the brainstems of the cases in the second, first, and third clusters should receive the smallest, medium, and largest doses, respectively.

The second feature attribute refers to the OVH descriptor of the spinal cord with respect to the PTV. The maximum dose received by the spinal cord, being a serial organ, should be restricted in IMRT planning. The maximum dose received by the spinal cord is significantly affected by the minimum distance $d_{\min}^{s}$ from the spinal cord to the PTV. Thus, $d_{\min}^{s}$ is used for the dose rank prediction of the spinal cord. As shown in figure 6(b), the $d_{\min}^{s}$ of the first cluster center is larger than that of the third cluster center, and the $d_{\min}^{s}$ of the third cluster center is larger than that of the second cluster center. Thus, we can predict that the spinal cords for the cases in the first, third, and second clusters should receive the smallest, medium, and largest doses, respectively.

The third and fourth feature attributes refer to the OVH descriptors for the left and right parotid glands with respect to the PTV, respectively. The V_{30} or mean doses of the left and right parotid glands, being parallel organs, should be restricted in IMRT planning, where V_{30} indicates the volume percentage receiving greater than or equal to 30 Gy to the total volume. Thus, the OVH curve is used for the dose rank prediction of the left and right parotid glands.

In figure 6(c), the OVH_3 curve of the first cluster center is at the left most part, the OVH_3 curve of the third cluster center is at the middle, and the OVH_3 curve of the second cluster center is at the rightmost part. This condition implies that the first cluster center is the farthest to the PTV, the third cluster center is the second farthest to the PTV, and the second cluster center is the closest to the PTV. Thus, we can predict that among the left parotid glands for the three clusters, the left parotid glands for the cases in the first, third, and second clusters should receive the smallest, medium, largest doses, respectively.

Similarly, we can also predict from figure 6 (d) that among the right parotid glands for the three clusters, the right parotid glands for the cases in the first, third, and second clusters should receive the smallest, medium, and largest doses, respectively.

In table 3, the fifth attribute refers to the nearest distance from the center of the left crystal to the PTV. Among the three distances, the distance of the first cluster is the smallest, that of the third cluster is the medium, and that of the second cluster is the largest. Thus, we can predict that among the left crystals of the three clusters, the left crystals for the cases in the first, third, and second clusters should receive the smallest, medium, and largest doses, respectively.

In table 3, the sixth attribute refers to the nearest distance from the center of the right crystal to the PTV. Among the three distances, the distance of the first cluster is the smallest, that of the third cluster is the medium, and that of the second cluster is the largest. Thus, we can predict that among the right crystals of the three clusters, the right crystals for the cases in the first, third, and second clusters should receive the smallest, medium, and largest doses, respectively.

According to the above analysis, the doses received by the OARs in the three clusters can be predicted as small, medium, and large, respectively, as shown in table 4.

To evaluate the accuracy of the above prediction, we calculated the maximum doses received by the brain stems, spinal cords, and left and right crystals, as well as the average doses of the parotid glands for the cases in the three clusters. These results are shown in table 5.

[Downloaded from ijrr.com on 2026-11-10]

[DOR: 20.1001.1.23223243.2021.19.1.2.3]

[DOI: 10.29252/ijrr.19.1.13]

The table shows that the ranks of the dose received by the spinal cords, the crystals, and the parotid glands in the three clusters are the same as the predicted results. However, the ranks of the dose received by the brain stems in the three clusters are different from the predicted values.

In summary, the proposed clustering method can qualitatively predict the dose rank. On the basis of this prediction, the difficulty in IMRT planning for the cases in the three clusters can be summarized as follows. The IMRT planning of the cases in the first cluster is the simplest, that in the third cluster is moderate, and that in the second cluster is the most difficult.

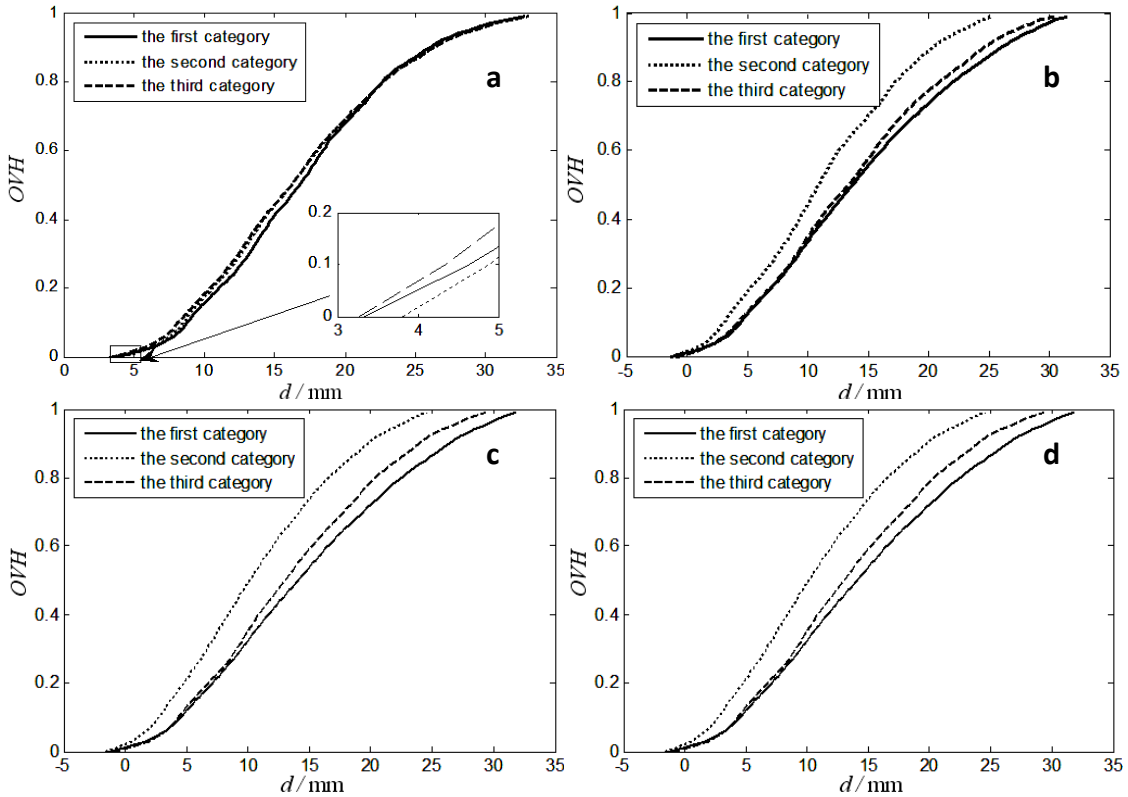


Figure 6. OVH1–OVH4 of the three clustering centers, in which the black curve, the dotted line, and the long-dashed line are the clustering centers of the first, second, and third cluster, respectively **(a)** First feature attribute curve OVH1, the OVH descriptor for the brain stem. **(b)** Second feature attribute curve OVH2, OVH descriptor for the spinal cord. **(c)** Third feature attribute curve OVH3, the OVH descriptor for the left parotid gland. **(d)** Fourth feature attribute curve OVH4, the OVH descriptor for the right parotid gland

Table 3. Fifth and sixth feature attributes of the three clustering centers.

Index of feature attributes	First cluster /mm	Second cluster /mm	Third cluster /mm
5	41.07	28.19	33.99
6	41.37	28.34	34.15

Table 4. Qualitative dose rank prediction of the three clusters.

Index of cluster	Dose of brain stem	Dose of spinal cord	Dose of left parotid gland	Dose of right parotid gland	Dose of left crystal	Dose of right crystal	Volume of target
1st	Medium	Small	Small	Small	Small	Small	Small
2nd	Small	Large	Large	Large	Large	Large	Medium
3rd	Large	Medium	Medium	Medium	Medium	Medium	Large

[Downloaded from ijrr.com on 2026-11-10]
[DOR: 20.1001.1.23223243.2021.19.1.2.3]
[DOI: 10.29252/ijrr.19.1.13]

Table 5. Average doses of OARs for the cases in the three clusters.

Index of Cluster	Brainstem /Gy	Spinal cord/Gy	Left crystal /Gy	Right crystal / Gy	Left parotid gland /Gy	Right parotid gland /Gy
First	48.71	40.57	3.96	4.01	28.84	29.19
Second	52.35	43.28	5.97	5.98	30.73	30.93
Third	50.96	42.72	5.47	5.48	29.62	29.83

DISCUSSION

Currently, IMRT planning is still time consuming and largely depends on the skills and experience of the physician. Rapidly increasing IMRT clinical plans are being generated due to the widespread usage of IMRT in clinical tumor treatments. Employing these large amounts of IMRT plans and effective IMRT planning retrieval technology to guide new cases of IMRT planning can ensure and improve the quality of new IMRT plans. Meanwhile, the cost and time can be reduced and automatic planning can be applied in the near future. However, the IMRT planning retrieval procedure may consume much time from a large-scale IMRT plan dataset, where many cases are relatively different from a new case. This problem could significantly reduce the advantage of such technology. Therefore, the effective retrieval of similar cases in a large-scale IMRT plan dataset is challenging and demanding. The clustering of IMRT plans in the dataset is an effective method to solve the above problem. The k-means algorithm and OVH descriptor have been proven to have good geometric description capabilities. Experimental results showed that IMRT plans can be well clustered and that isolated cases can be well identified by the proposed method, thereby improving retrieval efficiency. The cluster results were also used to predict the dosimetric characteristics of the OARs in each cluster. The results showed that the qualitatively predicted dosimetric characteristics of the OARs in each cluster agreed well with the practical results, and the difficulty of IMRT planning for each cluster can be derived and graded. However, other effective geometric features and their combinations should also be investigated in the future to analyze their benefits to IMRT plan clustering.

Furthermore, with the rapid development of conventional clinical imaging, emerging cell and molecular imaging, computer graphics and computer animations, etc., real 3D scenes including multiple 3D objects are being generated and fast growing. Thus, it is very necessary to develop an automatic and efficient clustering and retrieval system for 3D scenes. Therefore, the proposed clustering method is no doubt promising in such applications.

CONCLUSION

An IMRT plan clustering method based on the k-means algorithm and geometrical features was proposed and evaluated with a collected dataset including 100 cases of NPC IMRT plans. Experimental results showed that the tested IMRT plan dataset can be well clustered into three clusters and the isolated cases can be well identified with the proposed method. By using the geometrical features of each cluster center to predict the dosimetric characteristics of the OARs in each cluster, the dosimetric characteristics of the OARs in each cluster can be qualitatively predicted, which agree well with the practical results of each cluster.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (No. 51575256), the Key research and development plan (Social Development) of Jiangsu Province (No. BE2017730), the Key industrial research and development project of Chongqing (No. cstc2017zdcy-zdzxX0007), and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD). We would also like to thank Wei Song, Mingming Fang and

[Downloaded from ijrr.com on 2026-11-10]

[DOR: 20.1001.1.23223243.2021.19.1.2.3]

[DOI: 10.29252/ijrr.19.1.13]

Dongdong Wang, from Changzhou cancer hospital and Hainan Province cancer hospital of China respectively, for helping us collecting the clinical NPC IMRT plans.

***Conflicts of interest*:** *Declared none.*

REFERENCES

1. Petrovic S, Khussainova G, Jagannathan R (2016) Knowledge-light adaptation approaches in case-based reasoning for radiotherapy treatment planning. *Artif Intell Med,* ***68:*** *17-28.*
2. Schlaefer A, Dieterich S (2011) Feasibility of case-based beam generation for robotic radiosurgery. *Artif Intell Med,* ***52(2):*** *67-75.*
3. Zhou Z, Zhang W, Yu Z (2016) An effective calculation method for an overlap volume histogram descriptor and its application in IMRT plan retrieval. *Phys Medica,* ***32(10):*** *1339-1343.*
4. Wu B, Pang D, Lei S, Gatti J, Tong M, McNutt T, *et al.* (2014) Improved robotic stereotactic body radiation therapy plan quality and planning efficacy for organ-confined prostate cancer utilizing overlap-volume histogram-driven planning methodology. *Radiother Oncol ,****112(2):*** *221-226.*
5. Wu B, Pang D, Lei S, Gatti J, Tong M, McNutt T, *et al.* (2014) Improved robotic stereotactic body radiation therapy plan quality and planning efficacy for organ-confined prostate cancer utilizing overlap-volume histogram-driven planning methodology. *Radiother Oncol,* ***112(2):*** *221-226.*
6. Yang Y, Ford EC, Wu B, Pinkawa M, van Triest B, Campbell P *et al.* (2013) An overlap-volume-histogram based method for rectal dose prediction and automated treatment planning in the external beam prostate radiotherapy following hydrogel injection. *Med Phys,* ***40(1):*** *011709.*
7. Good D, Lo J, Lee WR, Wu QJ, Yin FF, Das SK (2013) A knowledge-based approach to improving and homogenizing intensity modulated radiation therapy planning quality among treatment centers: an example application to prostate cancer planning. *Int J Radiat Oncol Biol Phys,* ***87(1):*** *176-181.*
8. Petit SF, Wu B, Kazhdan M, Dekker A, Simari P, Kumar R, *et al.* (2012) Increased organ sparing using shape-based treatment plan optimization for intensity modulated radiation therapy of pancreatic adenocarcinoma. *Radiother Oncol,* ***102(1):*** *38-44.*
9. Kazhdan M, Simari P, McNutt T, Wu B, Jacques R, Chuang M, *et al.* (2009) A shape relationship descriptor for radiation therapy planning. *International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Berlin Heidelberg. p. 100-108.*
10. Zhang T and Ma F (2017) Improved rough k-means clustering algorithm based on weighted distance measure with Gaussian function. *Int J Comput Math,* ***94(4):*** *663-675.*
11. Jain AK (1998) Data clustering: 50 years beyond K-means. *Pattern Recogn Lett,* ***31(8):****651-66.*

[Downloaded from ijrr.com on 2026-11-10]

[DOR: 20.1001.1.23223243.2021.19.1.2.3]

[DOI: 10.29252/ijrr.19.1.13]

[DOI: 10.29252/ijrr.19.1.13] [DOR: 20.1001.1.23223243.2021.19.1.2.3] [Downloaded from ijrr.com on 2026-11-10]