

Using KMeans Clustering to Evaluate and Alert for Deviations of Linac Photon Beam Parameters

Narmada Chinnakannan, Punithavelan Nallamuthu*

Abstract

Objective: To analyse the daily measured Dosimetric Quality Assurance (QA) parameters of linear accelerator (linac) using Unsupervised Machine Learning (ML) Algorithm thereby evaluating the current machine status and to highlight the probable cause of the 'out-of-range' measured parameter. **Methods:** Five parameters measured using PTW QuickCheckweblin device in a linac is subjected to KMeans clustering technique. The measured parameters comprise of Central Axis Dose (CAX), Beam Flatness, SymmetryLR, SymmetryGT and Beam Quality (BQF). Data from Varian with 55- and 107-day's measurements and from Elekta with 75 days measurements from 2 beam matched linacs were used in this clustering technique. **Results:** This evaluation is used to review the current linac status and obtain 1) upper and lower limits of each parameter (CAX, Flatness, Symmetry, Beam Quality), 2) Frequency of the days when the linac parameters are closer to the target value and when they deviate from the target value. 3) The date when these parameters deviate from the estimated limits. 4) The probable reason for the deviation and 5) Finally if the machine requires maintenance. This methodology ensures that the machine is always closest to the target value, thus providing quality radiation treatment for the cancer patients. Moreover, the performance of the linac is studied meticulously and the need for maintenance is alerted before the linac beam shows marked deviation from the base value. **Conclusion:** KMeans clustering is a very simple and easy to use ML tool. With quick computation time and with lesser data it can arrive at the actual limits of the linac parameters and help to determine if the linac needs maintenance well in advance.

Keywords: Daily QA of Linac- Quick Check Weblin- Radiotherapy- Clustering, Quality Assurance- Dosimetry

Asian Pac J Cancer Prev, 25 (1), 305-315

Introduction

Treating cancer patients with radiation is one of the prime modalities in Oncology department. External beam radiotherapy is achieved using C or O type Linear accelerators (Linac) that can produce photons and electrons. These linac should deliver the photon and electron beams daily within certain limits as recommended by Hanley J et al., (2009) [1] from the base line values measured during commissioning. This is to ensure that the measured base line values input into the Treatment Planning System (TPS), with which every patient is planned, is also delivered every day to get the desired result. Linac, deviating from the baseline values will not give optimal treatment. To prevent this, it is important to have the beam parameters as close as possible to the TPS commissioned model with some bounds. Minimising this boundary ensures quality and precision treatment to the patient. Therefore, the efficiency of every linac should be assessed individually and continue to maintain it.

As Binny et al., (2016) [2] correctly stated, every linac require verification to account for uncertainties in

linac's mechanical positions, focal spot position etc, in this study we can see how two beam matched units show varying parameters for the daily QA. D.Jiang et al., (2020) [3] could observe the drift in the absolute output for all energies but unable to conclude that the issue was with the monitor chamber until after 200 days of measurements and after intense analysis. A tool is hence required to not only measure the daily QA parameters but also to evaluate them and help to identify the issue if the measurement shows marked deviation. Many studies were made using Statistical Process Control like cumulative sum (CUSUM) chart, Exponentially Weighted Moving Average (EWMA) chart, Ishikawa diagrams and Shewhart charts [4-6]. CUSUM charts and Shewhart charts respond very well for small shifts but are very slow in recognising large shifts. It also gives importance to the most recent data. Weighted Moving Average (WMA) charts work fine with normal data but fails with out-of-range data. Ishikawa diagrams does not show the development of a problem and the process needs to be repeated for every situation following the workflow each time. Li and Chan (2016) [7] have studied 5 years of daily QA measurements using data-

driven Artificial Neural Network and have discussed that overfitting is the major issue requiring a very large number of data, but a large data affects the splitting up of data for training, testing and validation. Hence, a tool which can analyse short term data and a large set of data, alerting when there are out-of-range values with a reasoning will help to deliver precise treatment to patients as planned.

In this article the past records of the linac are analysed using clustering technique and the following are determined: 1) Upper and lower limits of each parameter (CAX, Flatness, Symmetry, Beam Quality), 2) Frequency of the days when the linac was closest to the target value and when it was deviating from the target 3) The date when these parameters are deviating from the estimated limits. 4) The probable reason for the deviation and 5) Finally if the machine requires a maintenance.

Materials and Methods

Linac Daily QA

PTW Quickcheckwebline is a wireless device that records the radiation automatically and displays the parameters involved. There are 13 ionisation chambers in this device to capture the radiation and display the CAX, Flatness, Symmetry GT, SymmetryLR and BeamQuality for photons and electrons.

CAX

This gives the central axis dose measured at isocentre as absorbed dose to water.

Flatness

For a flattened photon beam the flatness defines how flat the profiles are and as per IEC 60976 it is the percentage dose ratio of the maximum to minimum value within the flattened region.

Flatness = $(D_{\max} / D_{\min}) * 100$ within the flattened region (IEC 60976)

Symmetry

Symmetry is the percentage of maximum deviation of the left-side dose from the right-side dose within the flattened region. Symmetry from left to right of the source is termed as Symmetry LT and from Gun to Target is referred as Symmetry GT.

Symmetry = $(D_{(x)} / D_{(-x)}) * 100$ within flattened region (IEC 60976)

Beam quality

This remains fixed for every energy type representing the penetration and attenuation of the beam. The Half Value Layer (HVL), Nominal acceleration potential (NAP) and Tissue Phantom ratio of 20 cm depth to that of 10 cm depth ($TPR_{20/10}$) are some of the methods of evaluating the beam quality.

Nyaichyai et al., (2022) [8] verified the suitability of PTW Quickcheck device for routine quality assurance of the linac for output, energy, flatness and symmetry. Nicewonger et al., (2019) [9] found the PTW Quick check device to be a suitable tool for daily testing quantitatively and efficient solution qualitatively. D.Jiang et al., (2020)

[3] observed that PTW Quick check device produced fitting linearity and reproducibility when compared with Farmer chamber. Dhoju et al., (2023) [10] concluded Beam monitorization following quality assurance protocol improves quality of the exposed beam during treatment procedure of patients. These works suggest that PTW QuickCheck webline is a reliable tool and the measurements can be effectively and efficiently utilised to automatically decide on the status of the linac

Desired protocol (IEC, Varian, Elekta, AAPM TG45 etc.) to arrive at the linac parameters, can be chosen from the available list of various international protocols in the Quick Check Webline software. When the data for TPS is collected, the base data for the Quickcheck is also collected and normalised. Subsequent measurements are performed daily and compared with the base value. A standard value of 2% or 3% is given for the upper and lower limits from the target value as per the clinical protocol. If the linac's performance can be evaluated, then the limits can be set uniquely for every linac.

Datasets

Daily QA measurements from different make and model of linac was collected and subjected to this analysis successfully. In this paper Varian Truebeam with 55 & 107 datasets and two beam matched Elekta Synergy machines with 75 datasets are analysed.

KMeans Clustering technique

Machine Learning (ML) is a subset of Artificial Intelligence and has found immense usage in different areas of Radiotherapy like imaging, classification, and prediction [11-13]. Unsupervised ML is a technique used for classification based on the data without any manual intervention [7]. Clustering or grouping is one of the Unsupervised ML algorithms [14, 15] and amongst the different methods KMeans clustering is chosen for this study using Python. [16-18].

KMeans involves grouping the data into clusters depending on how close they are to each other. Centroid in KMeans is the centre of a cluster such that it is the mean of all the points in that cluster. In KMeans, first a centroid is randomly set within the datapoints and then the Euclidean distance between this centroid and each of the available points is calculated. The mean distance from all the points is set as the centroid for that cluster. If a data shows very large distance, then it is considered as centroid of the next cluster. Through iterative process the centroids are determined, and the process stops when there is no change in the position of the centroid or when the number of given iterations is reached. The number of clusters need to be defined prior in KMeans. Centroids which are the centre of the clusters are formed until the defined number of clusters are achieved. Finally, we get the classification of all the data points into K clusters. The data points within a cluster should be closer to that centroid and away from the other centroids.

Reasons for applying KMeans for the daily dosimetry parameters

To ensure that the datasets are eligible for clustering,

Hopkin Statistic that tests the spatial randomness of the data is applied individually to the CAX, Flatness, Symmetry and BeamQuality parameters. It assesses the clustering tendency of a data set by measuring the probability that a given data set is generated by a uniform data distribution. This factor was 0.8 for CAX, 0.9 for Flatness and Symmetry and 0.95 for Beam Quality, implying that the daily QA parameters are perfect candidates for clustering. Of the different clustering techniques, KMeans works well for non-linear dataset which is true with each of the daily QA parameters (CAX, Flatness, Sym GT, Sym LR & BQF). The target value for each of these parameters is known for a commissioned linac. We can expect the clusters to be distributed above and below the target value with limited bounds. The datasets are 1Dimensional with only the dose value (for CAX) and so standardisation can be ignored. Additionally, the measurements can be directly input into the code to form clusters without any cleansing or preparation. This is of advantage as usually any data to be subjected to ML needs to be cleansed, prepared, and normalised. Another fact is a minimum of 30 number of data can also be analysed. This helps to understand how the linac behaved in those 30 days.

Parameters used for the clustering

KMeans requires initialisation method, number of clusters along with other parameters to form the clusters. KMeans++ was chosen for the initialisation method and 5 was chosen for the number of clusters. Out of the different initialisation methods that chooses the centroids, KMeans++ is found to be effective as it selects only one centroid initially and then moves on in selecting the other centroids. By using the ‘Elbow’ method, the number of clusters which will yield better result is chosen. For the number of data used in this article 5 clusters gave optimal results with KMeans_Inertia <1 and so 5 was chosen.

To maintain the reproducibility the random state parameter of KMeans is kept fixed (42 in this case). As this value remains the same, we can get the same consistent

results for multiple trials.

Applying KMeans to CAX values of Varian Truebeam

On subjecting 55 days CAX values of a Varian Truebeam to KMeans clustering we get the group of data as in Figure 1. The CAX dose is along the horizontal axis and the clusters, usually 0, 1, 2 etc is along the vertical axis.

Figure 1 shows five different clusters with 0.56 as Silhouette score which justifies the clustering. The five clusters are named as 0,1,2,3,4. As the target value is fixed as 100, cluster 1 is the best group (~100 to ~100.25) of data and it holds 13 out of 55 data. The next range of good cluster is 4, (~100.25 to ~100.5) and this holds 20 out of 55 data. The next range is cluster 3 (~99.75 to ~100) but with only 5 out of 55. Most of the output values are greater than 100 (from the plot), implying that this energy of this Truebeam machine has the tendency to deliver slightly higher output (>100) The lower limit can be set as 99.5 and the upper limit as 101.5. Clusters 4 & 1 can be marked as the ones closest to the target and for 33 days (20+13) the linac was within these clusters. Cluster 3 and Cluster 0 can be considered as “out-of-range” data. Though this customer had given 3% as the acceptable CAX deviation, this linac shows <1.0% deviation on most of the days and the target is 100.5.

To show that using a smaller dataset or larger dataset give the same prediction, the above clustering was repeated with 107 datasets. In Figure 2 107 datasets of the same linac is clustered. The cluster number (0 to 4) may differ, but the grouping of the data almost remains same excepting for some outliers (102.5). Here cluster 4 represents the best group (~99.75 to ~100.25) with 23 out of 107 in this group. Also, it is noted that the output is more than 100 in most of the days (0+4+3 clusters) 64 out of 107 when compared to the lesser than 100 values which is 20 out of 107. This set of data also shows 100.5 as the target value and 1% can be the acceptable deviation.

The above two sets of process with 55- and 107-days measurements show that with minimal set of data it is

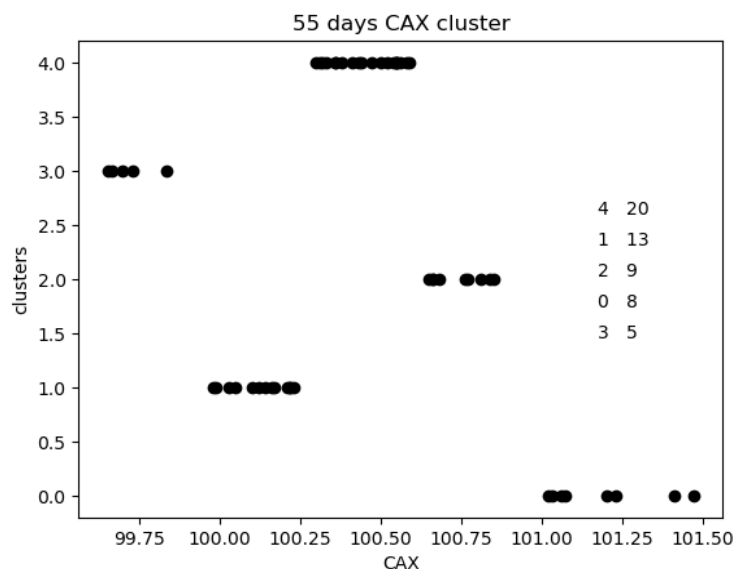


Figure 1. CAX Clusters with 55 Days Data

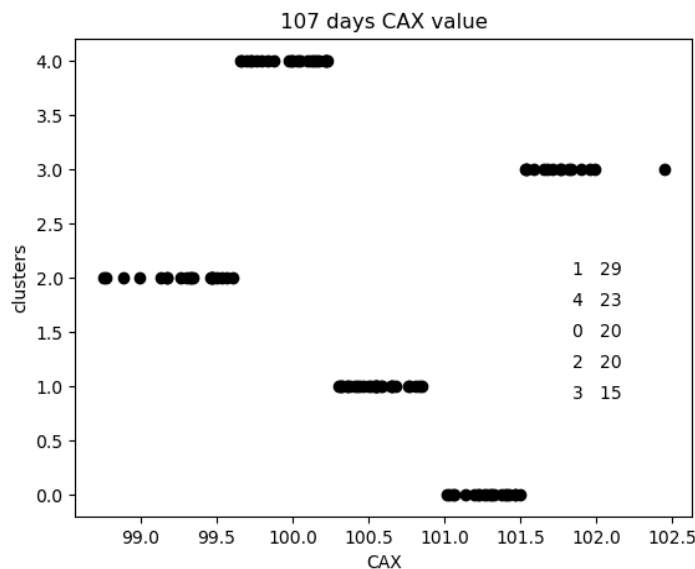


Figure 2. CAX Clusters with 107 Days Data

possible to estimate the limits within which the linac functions and frequency of best days. The clusters when transferred to the respective dates of measurements can predict the performance of the linac. Applying these clusters to the regular measurements helps to decide on which category they fall and alerts if the values are out-of-limits as defined by the training dataset. This is explained in detail in the next section.

Thus, with the existing data a model can be trained for KMeans cluster to analyse the data and know the customised limits of every individual linac parameter. Using this trained model the subsequent measurements can be predicted to be either “within tolerance” or “out of tolerance”. Here the tolerance limit is more specific to the that linac.

In the same context, models can be trained for Flatness, Symmetry GT, Symmetry LR and Beam Quality.

Detailed study of Beam Matched Elekta Linacs Training dataset

Elekta Infinity and Elekta Synergy are two beam matched units whose data were used for this study. Randomly 75 days measurements (from 03FEB2020 to 15MAY2020) were used as training datasets and the clusters obtained were mapped to the date of measurements scoring only the group that is out-of-tolerance or the group with lesser number of candidates. This is then compared to the date when the re-normalisation was done on the device. Renormalisation is usually done when there is continuous gross deviation of the measured parameters from the set tolerance limit (indicating a change in the linac’s behaviour) or when the linac is independently tuned. The datasets of this study had 2% limits from target value (100) for the CAX and Symmetry, 3% for the FLATNESS and 1% for the BeamQuality for which the target is 6.

The clusters of all the parameters are shown in Figure 3 for the Infinity. The outliers can be discarded if present, but this dataset does not have any outliers, so all

the data are included for assessment. From the clusters the upper and the lower limits for each parameter can be arrived as discussed under “Applying KMeans to CAX values of Varian Truebeam”. The frequency of clusters closest to the target and the ones away from the target can also be observed from the legend of each parameter.

To correlate these with the dates of measurement, Table 1 is generated using MS Excel. The clusters generated from python are written into the spreadsheet and the actual measurements are given along with the clusters for the respective dates. Only those clusters that showed marked deviation from the target value or clusters with a smaller number of data are marked in the Table. For example, in Figure 3, CAX_Infinity plot has a legend of 3 & 4 with a smaller number of clusters and can be considered as out-of-limits. Hence the numbers 3 or 4 is used for corresponding date in Table 1.

Between 16th and 20th March, while the actual measurements do not show deviation (2%,3% and 1% respectively from the target value), the clustering indicates that on these days the values are away from the target value. Customer had re-normalised the values in the Quickcheck device on 21st march after when, the clusters fall closest to the target value.

Similarly, clusters for Synergy can be seen in Figure 4 and the corresponding Table 2 gives the analysis with date of measurement. In Table 2 we can observe the similar behaviour for the other beam matched linac Synergy but at different date (before 1st March). In this case there is a continuous indication that “SYMGMT” is not within the expected clusters. While the actual measured value does not indicate this. Further for this linac the normalisation done on 1st March after which the clusters fall closer to the target value.

Accuracy of the Clusters

The Silhouette factor for each parameter of the linac are listed in Table 3. This is a good indication that the formed clusters closely represent the linac. This factor

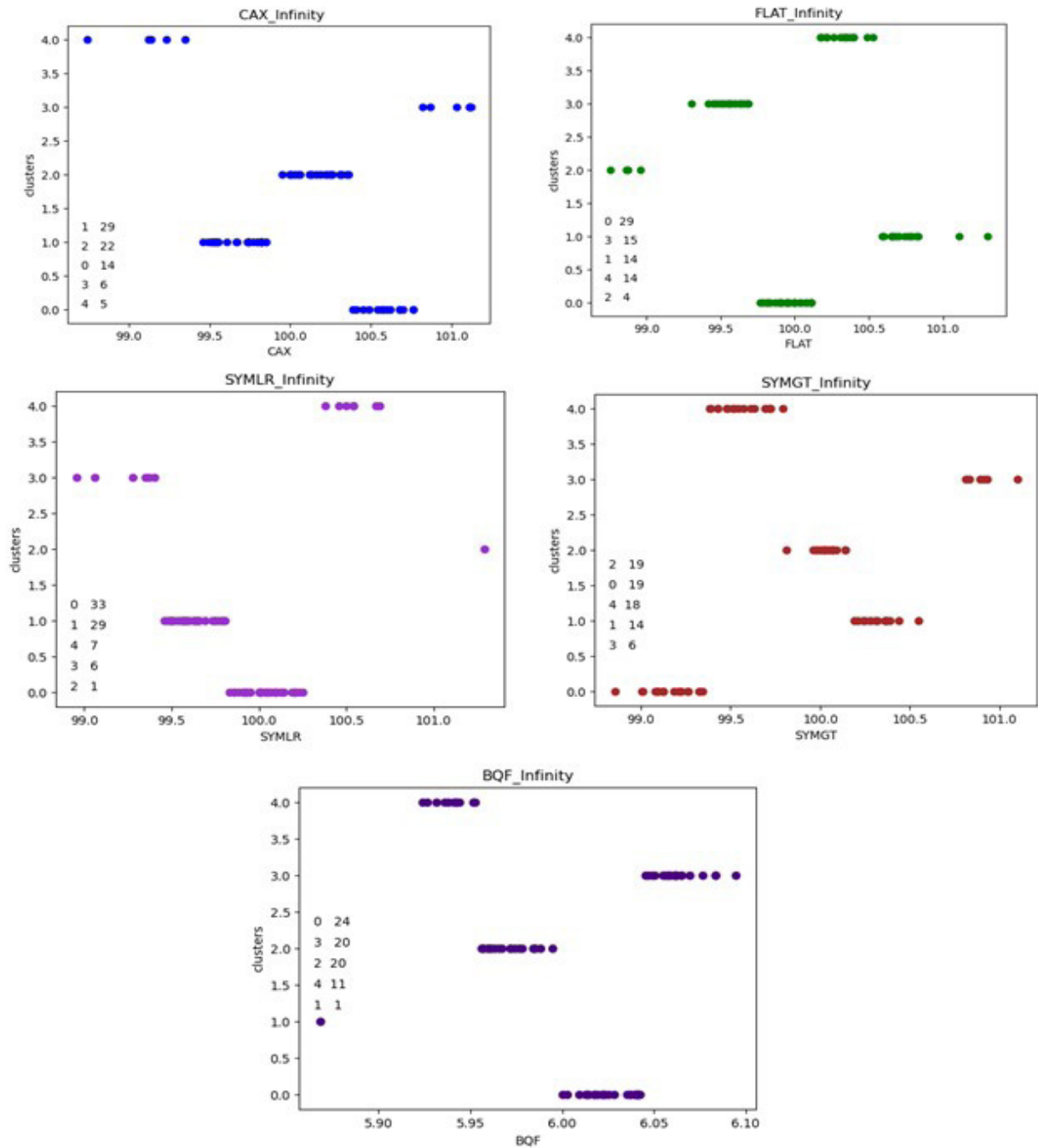


Figure 3. Clusters of Measured Parameters of Infinity-Training Dataset

is a measure of how alike an object is to its own cluster in comparison to other clusters. The Silhouette factor, calculated using any distance metric like Euclidean or Manhattan distance, for each data point provides a graphical representation of how well each object has been classified. The formula for calculating the Silhouette coefficient is as follows:

$$\text{silhouette factor} = \frac{(\text{separation} - \text{cohesion})}{\max(\text{separation}, \text{cohesion})}$$

where separation is the distance between a data point and the nearest cluster that the data point is not a part of, and cohesion is the average distance between a data point and all other data points in the same cluster.

The Silhouette coefficient ranges from -1 to +1, where a high value indicates that the object is well matched to its own cluster and poorly matched to neighbouring clusters. If most objects have a high value, then the clustering configuration is appropriate. If many points have a low or negative value, then the clustering configuration may have too many or too few clusters.

Model for errors

On occasions when the daily check device shows an error for a particular parameter, it is difficult to pinpoint to the apt reason. To train the model that can predict the probable errors, measurements were done with purposely made handpicked few errors and the data were used

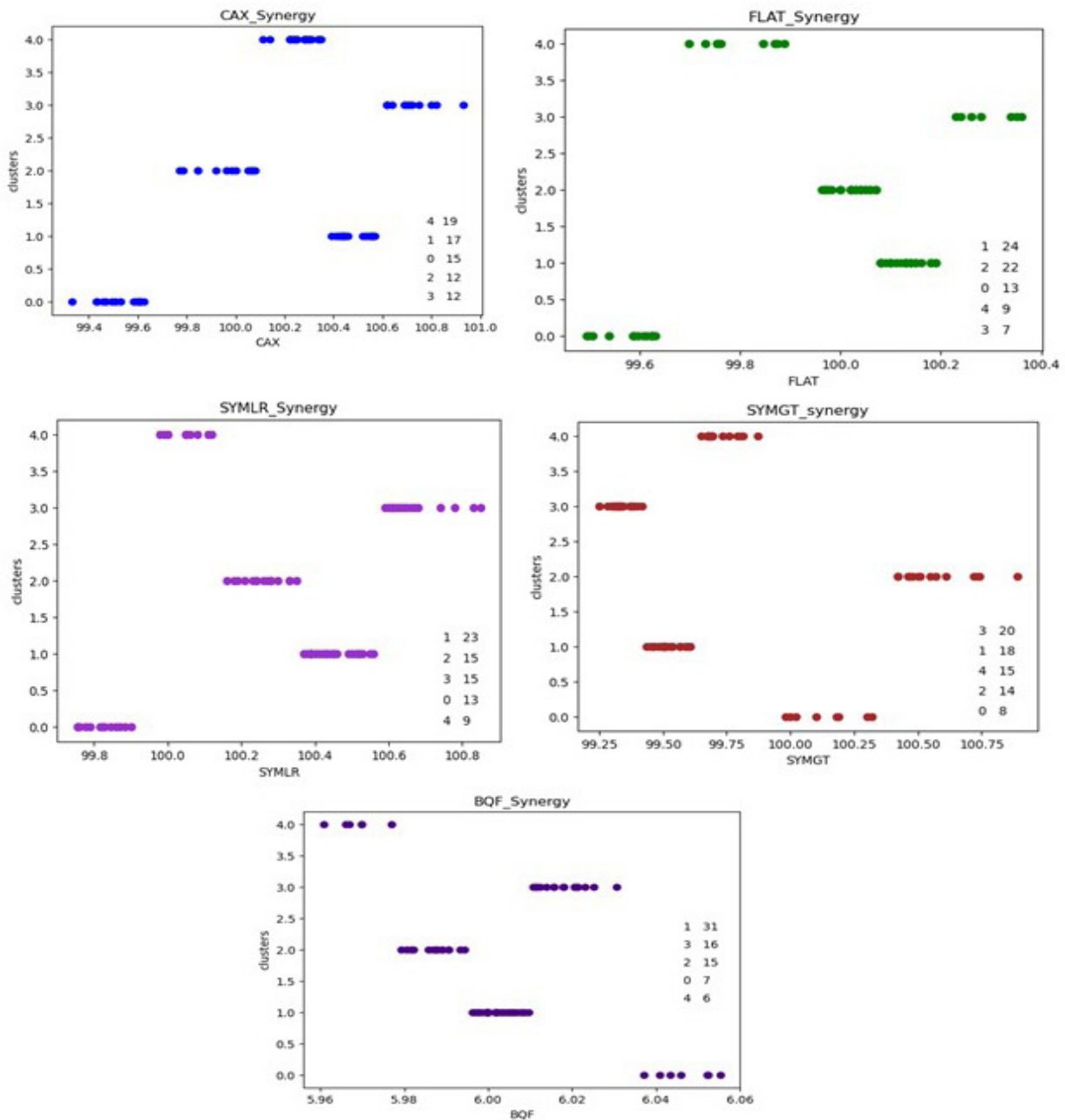


Figure 4. Clusters of Measured Parameters of Synergy-Training Dataset

to create a model. In the case of CAX dose error, the following are some of the errors encountered: Lesser Monitor units (MU) delivered than the baseline, Greater MU delivered than the baseline, Linac output variation, Field size is different, Energy is different, Set up error. This model has 7 clusters, one for each error. A measurement stated as “out-of-tolerance” in the analysis model can be subjected to this “foreseeing model” which can pinpoint to the error that has caused the deviation in the data (Figure 5A, 5B).

In case of beam quality, three erratic situations were reproduced as in Figure 6A, 6B. Similarly, the probable errors for Flatness and symmetry can also be generated. Also, one error can cause two or three parameters to deviate from baseline. For example, when a larger field size, in comparison to baseline data, is used both Flatness

and BQF can fail. Thus, the errors generated can also be unique to every linac. This database can be built by adding the forced errors and actual errors which will help in the long run for that linac. The different type of errors classified as setup errors, method and measurement errors, machine errors and environmental errors [6] can be incorporated with clusters to get the model for errors.

Test dataset

The clusters obtained with the 75 training datasets gives a good picture of the existing linac condition. The same model can be applied to another set of data and check the linac behaviour. The log registry of the linac was analysed to check the dates when the linac was subjected to maintenance. As there were many entries of engineer visit, those visits related to the tuning of the beam like

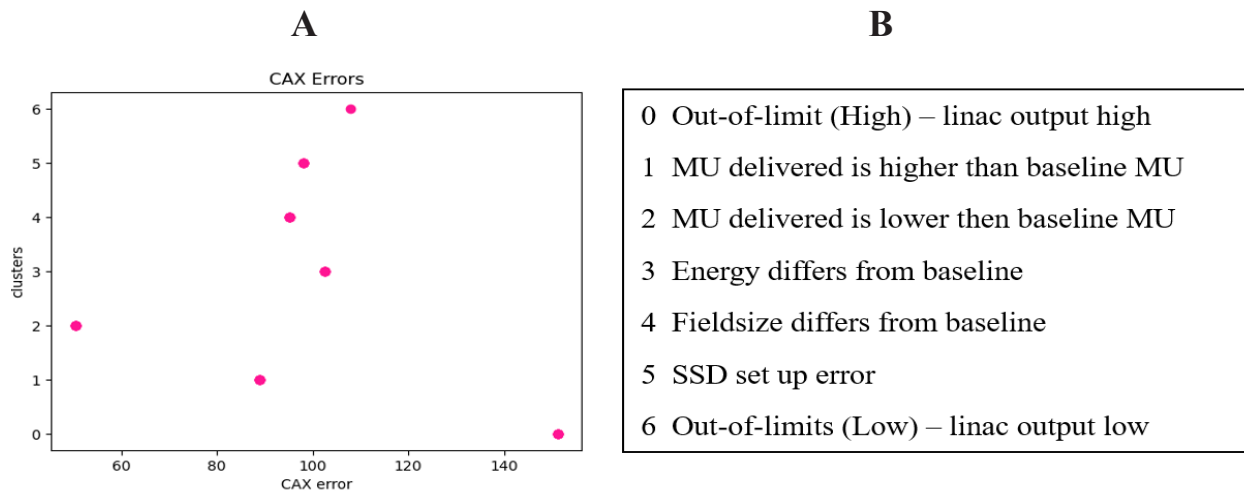


Figure 5. A, Clusters based on CAX errors; B, List of CAX Errors

dose rate error, beam timer error, beam mu ch2, Preventive maintenance were taken as reference. Few days before and after these visits were analysed to check if the clustering can help to identify the issue. Following Table 4, Table 5 show these details where days marked in bold with bigger font represent the days of beam tuning along with the original measured value and cluster group. The measured values that do not fall within the cluster limits are marked as OLH for out-of-limit on the upper limit and OLL for out-of-limit on the lower limit. The cells highlighted in green indicate the days when the parameter was closest to the target value. In most cases we can observe that after the machine underwent a maintenance, the days following it showed results that are closer to the target value or do not fall out-of-range.

Results

The clusters of the trained datasets help to visualise the behaviour of the linac. The clusters arrived using the training datasets helps to set linac specific upper and lower limit for each parameter. The frequency of days when the linac was close to the target can be obtained that helps to understand the stability of the machine. Tabulation of the

cluster groups with the date of measurement enables to assess the status of the beam and check if any tuning of the parameters or renormalisation of the Quick check device is required. If there is gross deviation the reason can be determined from the Model to reason errors. Comparing the dates of maintenance with the cluster group shows that after the beam maintenance the clusters fall closer to the target group as indicated by the green colour cells. Conversely, a maintenance visit can be planned if any one cluster is continuously out-of-limit or more than two clusters are not under tolerance. As the limits used here are very tight (about 0.5 %), the beam is always under check and prevents it from grossly deviating (say beyond 2%). This ensures that even very high dose treatments like SRS can be accomplished with excellent results.

Discussion

KMeans clustering can be considered as an expressive tool to evaluate the daily dosimetry parameters. Knowing the range within which the linac usually behaves helps to have a good control over the patient specific quality assurance. If the linac deviates continuously from the usual range, then immediate action can be taken before

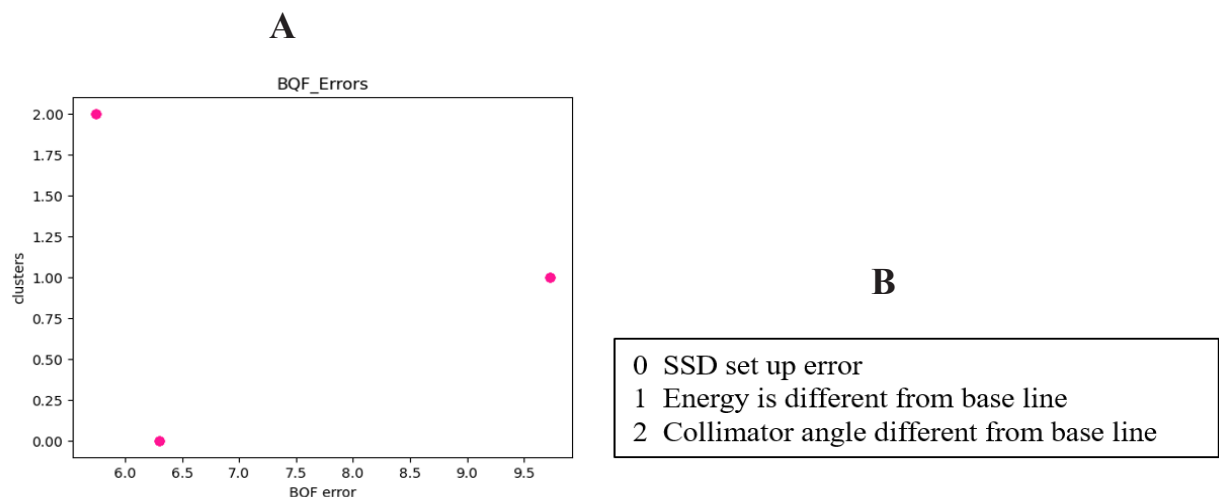


Figure 6. A, Clusters based on BQF errors; B, List of BQF Errors

Table 1. Table Showing the Out-of-Tolerance Clusters and the Actual Measurements for Infinity Machine

DATE	CAX	FLAT	SYMLR	SYMGT	BQF	CAX	FLAT	SYMLR	SYMGT	BQF
04/Mar/2020	3					101.03	99.546	99.505	100.07	5.9718
05/Mar/2020						100.32	100.33	99.91	99.975	5.9739
06/Mar/2020					4	100.87	99.63	99.564	100.14	5.9632
09/Mar/2020						99.817	100.79	99.856	100.04	5.9669
11/Mar/2020					4	99.851	100.66	99.64	100.21	5.9407
12/Mar/2020					4	100.68	99.791	99.491	100.39	5.9421
13/Mar/2020					4	100.41	99.963	99.572	100.02	5.9595
16/Mar/2020		2	3	3	1	99.743	98.756	98.957	101.1	5.8681
17/Mar/2020		2		3	4	99.51	98.872	99.628	100.93	5.9357
18/Mar/2020		2	3	3	4	99.551	98.864	99.277	100.91	5.9238
19/Mar/2020	4					98.743	99.949	100.14	100.55	5.984
20/Mar/2020	4	2		3	4	99.235	98.958	99.658	100.81	5.9516
21/Mar/2020						100	100	100	100	6
23/Mar/2020						100.13	99.556	99.598	99.812	6.0029
24/Mar/2020						99.516	100.49	100.09	99.21	6.0574
25/Mar/2020						99.525	100.35	100.01	99.428	6.0395
26/Mar/2020						100.26	99.472	99.493	99.722	6.0141
27/Mar/2020	4					99.347	100.77	100.07	98.855	6.0766
28/Mar/2020	4					99.141	100.66	100.05	99.094	6.0581
30/Mar/2020						99.825	100.11	100.02	99.516	6.0406
31/Mar/2020						99.85	99.868	99.788	99.698	6.0255
01/Apr/2020						99.798	100.04	99.832	99.478	6.0469
02/Apr/2020						99.545	100.17	100.06	99.345	6.056

Table 2. Table Showing the Out-of-Tolerance Cluster Group and the Actual Measurements for Synergy Machine

DATE	CAX	FLAT	SYMLR	SYMGT	BQF	CAX	FLAT	SYMLR	SYMGT	BQF
21-Feb-20				0	0	99.61	99.61	99.85	99.98	6.06
24-Feb-20				0	0	99.33	99.62	99.87	100.19	6.04
25-Feb-20		4	4	0	0	99.44	99.76	100.06	100.30	6.05
26-Feb-20		4		2		99.60	99.70	99.86	100.47	6.03
27-Feb-20		4	4	0	0	99.58	99.73	99.98	100.10	6.05
28-Feb-20		4	4	2		99.63	99.88	99.99	100.89	6.00
1-Mar-20			4	0		100.00	100.00	100.00	100.00	6.00
2-Mar-20			4			100.62	100.13	100.08	99.76	6.00
3-Mar-20						100.06	100.08	100.19	99.79	5.99
4-Mar-20		4	4			100.75	99.87	100.11	99.57	6.01
5-Mar-20						100.56	100.04	100.40	99.60	5.99
6-Mar-20						100.52	99.96	100.38	99.39	6.00
9-Mar-20						100.31	100.09	100.49	99.32	6.00

Table 3. Silhouette Factors Showing the Accuracy of the Clustering

	Infinity	Synergy
Cax	0.557	0.596
FLAT	0.587	0.585
SYMLR	0.522	585
SYMGT	0.622	0.619
BQF	0.528	0.59

the beam characteristics fall well below the norms. By accumulating the errors, a model can also be easily trained with which the reason for the failure can be got instantaneously which in turn helps to keep a check on the linac. Above all the KMeans clustering is a very simple and easy to use tool with quick computation time and with lesser data. As more advanced treatment technique like Stereotactic radiosurgery, stereotactic radiotherapy etc involves very large dose, the important beam parameter's

Table 4. Comparison of Clusters and Maintenance Visit for Infinity Machine

DATE	CAX	FLAT	SYMLR	SYMGT	BQF					
30-Dec-19	100.73	100.18	100.43	100.2	6.0179	0	4	4	0	2
31-Dec-19	100.03	100.61	100.99	100.06	6.0517	2	1	<OLH>	2	0
1-Jan-20	100.71	100.47	100.71	100.15	6.0268	0	4	4	2	2
2-Jan-20	100.4	100.69	100.94	100.09	6.0398	0	1	<OLH>	2	2
3-Jan-20	100.11	101.37	101.37	100.01	6.0531	2	<OLL>	<OLH>	2	0
6-Jan-20	100.48	100.61	100.84	99.94	6.0451	0	1	<OLH>	2	2
7-Jan-20	100.7	100.66	100.87	100.02	6.0372	0	1	<OLH>	2	2
8-Jan-20	101.47	99.398	99.993	100.45	5.9985	<OLH>	3	0	0	3
10-Jan-20	100.68	100.05	100.32	100.26	6.0094	0	0	4	0	2
13-Jan-20	100.69	100.09	100.57	100.45	6.0151	0	0	4	0	2
14-Jan-20	100.88	99.948	100.62	100.24	6.0266	3	0	4	0	2
15-Jan-20	100.59	99.78	100.21	100.12	6.0108	0	0	0	2	2
23-Jan-20	100.43	100.04	100.43	100.35	6.0056	0	0	4	0	3
24-Jan-20	100.31	100.28	100.64	100.22	6.0235	2	4	4	0	2
28-Jan-20	100.59	100.38	100.6	100.23	6.0194	0	4	4	0	2
29-Jan-20	101.01	100.01	100.42	100.04	6.0247	<OLH>	0	4	2	2
30-Jan-20	100.93	99.946	100.5	100.02	6.0361	3	0	4	2	2
31-Jan-20	100.48	100.24	100.5	100.15	6.0105	0	4	4	2	2
3-Feb-20	99.823	101.3	101.29	100.07	6.0488	1	<OLH>	<OLH>	2	0
4-Feb-20	100.76	100.03	100.46	100.05	6.0282	0	0	4	2	2
5-Feb-20	101.11	99.57	100.09	100.83	5.9655	<OLH>	3	0		1
6-Feb-20	100.7	99.813	100.5	100.24	6.0221	0	0	4	0	2
7-Feb-20	101.11	99.898	100.54	100.14	6.0216	<OLH>	0	4	2	2
8-Feb-20	100.76	99.897	100.1	100.89	5.9437	0	0	0		1
10-Feb-20	100.82	99.913	100.54	100.37	6.0171	3	0	4	0	2
11-Feb-20	99.823	100.18	99.571	99.961	5.9594	1	4	1	2	1
12-Feb-20	100.06	100.08	99.79	100.07	5.9721	2	0	1	2	3
13-Feb-20	100.31	100.11	99.771	100.31	5.9567	2	0	1	0	1
14-Feb-20	99.952	100.74	100.38	99.982	5.9947	2	1	4	2	3
17-Feb-20	100.03	100.83	100.2	100.03	5.9845	2	1	0	2	3
18-Feb-20	100.23	99.823	99.473	99.999	5.9562	2	0	1	2	1
19-Feb-20	100.58	99.998	99.629	100.02	5.9612	0	0	1	2	1
20-Feb-20	99.998	100.21	99.949	100.06	5.9764	2	4	0	2	3
21-Feb-20	100.6	99.908	99.656	100.25	5.957	0	0	1	0	1
24-Feb-20	100.54	99.63	99.402	100.36	5.9379	0	3	3	0	1
25-Feb-20	100.49	99.689	99.538	100.37	5.9526	0	3	1	0	1
3-Aug-20	99.463	100.85	100.98	99.223	6.1038	1	1	<OLH>	0	0
4-Aug-20	99.106	101.32	101.49	99.104	6.1481	<OLL>	<OLH>	<OLH>	0	0
5-Aug-20	99.222	101.02	100.9	98.946	6.1165	<OLL>	<OLH>	<OLH>	<OLL>	0
6-Aug-20	99.683	100.17	100.64	99.478	6.1094	1	4	4	2	0
7-Aug-20	99.465	100.51	100.85	99.15	6.1189	1	4	<OLH>	0	0
10-Aug-20	99.376	100.32	100.63	99.321	6.1068	4	4	4	0	0
11-Aug-20	99.721	100.47	101	99.396	6.1278	1	4	<OLH>	4	0
13-Aug-20	99.881	100.25	100.92	99.421	6.1084	1	4	<OLH>	4	0
14-Aug-20	100.15	99.743	100.28	100.03	6.0695	2	3	0	2	0
17-Aug-20	99.803	99.719	100.09	99.651	6.0735	1	3	0	4	0
18-Aug-20	100.17	99.624	100.31	99.752	6.0731	2	3	4	4	0
19-Aug-20	99.702	100.06	100.5	99.651	6.0761	1	0	4	4	0
20-Aug-20	99.745	99.653	100.31	99.56	6.0794	1	3	4	4	0

Table 4. Continued

DATE	CAX	FLAT	SYMLR	SYMGT	BQF					
21-Aug-20	99.098	100.48	100.66	99.662	6.0737	<OLL>	4	4	4	0
24-Aug-20	99.852	99.735	100.15	99.612	6.0632	1	3	0	4	0
25-Aug-20	99.835	100.16	100.78	99.577	6.1101	1	4	4	4	0
20-May-21	99.765	99.444	99.881	99.293	6.0302	1	3	0	0	0
21-May-21	100.02	99.351	99.982	99.192	6.0465	2	3	0	0	3
24-May-21	101.34	100.22	100.46	99.743	5.9246	<OLH>	4	4	4	
25-May-21	99.616	100.02	100.35	99.2	6.0529	1	0	4	0	3
26-May-21	99.171	101.19	101.05	99.075	6.0613	<OLL>	<OLH>	<OLH>	0	3
27-May-21	99.568	99.53	99.868	99.224	6.034	1	3	1	0	0
28-May-21	99.535	99.448	99.778	98.987	6.0286	1	3	1	<OLL>	0
29-May-21	98.744	99.876	99.654	100.12	5.9522	4	0	1	2	4
31-May-21	99.174	99.305	99.572	99.099	6.0178	4	3	1	0	0

Table 5. Comparison of Clusters and Maintenance visit for Synergy Machine

DATE	CAX	FLAT	SYMLR	SYMGT	BQF					
11-Feb-20	99.507	99.755	100.18	100.32	6.0252	0	4	2	3	3
12-Feb-20	99.494	99.761	100.24	100.74	6.0018	0	4	2	<OLH>	1
13-Feb-20	99.591	99.632	99.815	100.51	6.0036	0	0	0	<OLH>	1
14-Feb-20	99.606	99.626	100.05	100.61	6.0113	0	0	4	<OLH>	3
15-Feb-20	99.613	99.846	100.12	100.73	5.9998	0	4	4	<OLH>	1
17-Feb-20	99.921	99.596	99.828	100.5	6.0156	2	<OLL>	0	<OLH>	3
18-Feb-20	99.785	99.626	99.824	100.46	6.0097	2	0	0	2	1
19-Feb-20	99.53	99.889	100.05	100.72	5.9981	0	4	4	<OLH>	1
20-Feb-20	99.459	99.587	99.755	100.02	6.0523	0	<OLL>	<OLL>	3	<OLH>
21-Feb-20	99.613	99.609	99.846	99.983	6.0553	0	0	0	3	<OLH>
24-Feb-20	99.331	99.623	99.87	100.19	6.0433	<OLL>	0	0	0	0
25-Feb-20	99.435	99.758	100.06	100.3	6.0458	0	4	4	3	0
26-Feb-20	99.602	99.699	99.859	100.47	6.0307	0	4	0	2	3
2/12/2022	101.34	100.19	99.683	100.22	5.9577	<OLH>	1	<OLL>	0	<OLL>
2/14/2022	99.993	100.66	99.888	100.8	5.9478	2	<OLH>	0	<OLH>	<OLL>
2/15/2022	99.579	100.59	99.706	100.69	5.9274	0	3	<OLL>	<OLH>	<OLL>
2/16/2022	99.356	100.26	100.28	100.38	5.9803	<OLL>	3	2	2	2
2/17/2022	100.02	100.36	100.33	100.47	5.9998	2	3	2	2	1
2/18/2022	99.97	100.47	100.57	100.48	6.0034	2	3	3	2	1
2/19/2022	99.667	100.71	100.73	100.72	6.0011	0	3	3	<OLH>	1
3/16/2022	100.53	100.21	100.15	100.21	6.0174	1	1	4	0	3
3/17/2022	100.21	100.17	100.43	100.18	6.0236	4	1	1	0	3
3/21/2022	100.21	99.916	99.218	99.798	5.9558	4	2	<OLL>	4	<OLL>
3/22/2022	100.27	100.43	100.25	100.64	5.9972	4	3	2	<OLH>	1
3/23/2022	100.07	100.21	100.44	100.16	6.0246	2	1	1	0	3
3/24/2022	99.848	100.1	100.35	100.04	6.0263	2	1	2	0	3
3/25/2022	99.853	100.15	99.183	100.4	5.9481	2	1	<OLL>	2	<OLL>
3/26/2022	100.23	100.38	100.62	100.2	6.0433	4	3	3	0	0
4/18/2022	99.965	100.3	100.7	100.24	6.0348	2	3	<OLH>	0	0
4/19/2022	100.32	100.24	100.56	100.29	6.0156	4	3	1	0	3
4/20/2022	100.24	100.42	100.49	100.57	6.009	4	3	1	<OLH>	1
4/21/2022	99.937	100.31	100.49	100.36	6.0084	2	3	1	2	1
4/22/2022	100.05	100.25	100.63	100.13	6.0392	2	3	3	0	0
4/25/2022	100.04	100.19	100.36	100.34	6.0155	2	1	1	0	3
4/26/2022	99.772	100.37	100.4	100.56	5.9885	2	3	1	<OLH>	2
4/27/2022	99.733	100.44	100.17	100.74	5.9631	0	3	2	2	4

limits can be made more stringent and unique using the KMeans trained dataset.

Author Contribution Statement

Narmada Chinnakannan is responsible for conceptualization, Methodology, Software and Validation. Punithavelan Nallamuthu Supervised the work.

Acknowledgements

The authors are thankful for the support extended by Artemis hospitals, Haryana, India for sharing the measured data from their Quickcheckweblne and the log registry related to the machine maintenance.

Future work

In this study the model for the error was created only for few data. This can be extended to incorporate all possible errors.

Funding

This research did not receive any specific grant from funding agencies in the pubic, commercial or not-for-profit sectors.

Declaration of Competing interests:

The authors declare that they have no competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Narmada Chinnakannan is an employee of PTW India.

Data availability

The data used for this study are available from the corresponding author on request.

References

1. Hanley J, Dresser S, Simon W, Flynn R, E Klein E, Letourneau D, et al. AAPM task group 198 report: An implementation guide for tg 142 quality assurance of medical accelerators. *Med Phys.* 2021;48:e830-85.
2. Binny D, Aland T, Archibald-Heeren BR, Trapp JV, Kairn T, Crowe SB. A multi-institutional evaluation of machine performance check system on treatment beam output and symmetry using statistical process control. *J Appl Clin Med Phys.* 2019;20(3):71-80. <https://doi.org/10.1002/acm2.12547>.
3. Jiang D, Wang X, Dai Z, Shen J, Wang D, Bao Z, et al. Systematic and comprehensive analysis of the dose-response characteristics of a morning quality check of a linear accelerator and an important application of accelerator performance prediction. *Int J Radiat Res.* 2020;18(4):841-51. <https://doi.org/10.52547/ijrr.18.4.841>.
4. Pawlicki T, Whitaker M, Boyer AL. Statistical process control for radiotherapy quality assurance. *Med Phys.* 2005;32(9):2777-86. <https://doi.org/10.1118/1.2001209>.
5. Sanghangthum T, Suriyapee S, Srisatit S, Pawlicki T. Retrospective analysis of linear accelerator output constancy checks using process control techniques. *J Appl Clin Med Phys.* 2013;14(1):4032. <https://doi.org/10.1120/jacmp.v14i1.4032>.
6. Pal B, Pal A, Das S, Palit S, Sarkar P, Mondal S, et al.

Retrospective study on performance of constancy check device in linac beam monitoring using statistical process control. *Rep Pract Oncol Radiother.* 2020;25(1):91-9. <https://doi.org/10.1016/j.rpor.2019.12.004>.

7. Li Q, Chan M, Wang B, Shi C. Clustering breathing curves in 4d radiotherapy by using multiple machine learning tools: K-means and hierarchical clustering algorithms. in proceedings of the 11th annual machine learning symposium (new york, ny);2017:28-9.
8. Nyaichyai KS, Jha D, Adhikari KP. Monitoring linear accelerator output constancy and overall performacne using the ptw quickcheck weblne. *JNPS.* 2022;8:66-74.
9. Nicewonger D, Myers P, Saenz D, Kirby N, Rasmussen K, Papanikolaou N, et al. Ptw quickcheck weblne: Daily quality assurance phantom comparison and overall performance. *J BUON.* 2019;24:1727-34.
10. Dhoju N, Pudasainee A, Jha B, Pudasainee A, Yadav PK, Pokharel A, et al. Monitoring linear accelerator beam with daily quality assurance phantom. *Sci World J.* 2023;16:5-11.
11. Issam El Naqa, Ruijiang Li, Martin J. Murphy. Machine learning in radiation oncology: Theory and applications. 2015th Edition. Springer International Publishing Switzerland: Springer;2015.
12. Weidlich V, Weidlich GA. Artificial intelligence in medicine and radiation oncology. *Cureus.* 2018;10(4):e2475. <https://doi.org/10.7759/cureus.2475>.
13. Kourou K, Exarchos TP, Exarchos KP, Karamouzis MV, Fotiadis DI. Machine learning applications in cancer prognosis and prediction. *Comput Struct Biotechnol J.* 2015;13:8-17. <https://doi.org/10.1016/j.csbj.2014.11.005>.
14. Saxena AK, Prasad M, Gupta A, Bharill N, Patel OP, Tiwari A, et al. A review of clustering techniques and developments. *Neurocomputing.* 2017;267:664-81.
15. Li H, Galperin-Aizenberg M, Pryma D, Simone CB, 2nd, Fan Y. Unsupervised machine learning of radiomic features for predicting treatment response and overall survival of early stage non-small cell lung cancer patients treated with stereotactic body radiation therapy. *Radiother Oncol.* 2018;129(2):218-26. <https://doi.org/10.1016/j.radonc.2018.06.025>.
16. Fabian Pedregosa, Gael Varoquaux, Gramfort A, et al. Scikit-learn: Machine learning in python. *JMLR.* 2011;12:2825-30.
17. Yedla M, Rao S, Pathakota, Srinivasa TM. Enhancing k-means clustering algorithm with improved initial center. *IJCSIT.* 2010;1(2):121-125
18. Aristidis Likas NV, Jakob J Verbeek. The global k-means clustering algorithm. *Pattern Recogn.* 2003;36:451-61.



This work is licensed under a Creative Commons Attribution-Non Commercial 4.0 International License.