# Statistical approaches to analyze the log file of Elekta versa HD and the DICOM file of RT plan for machine learning

Pilsu Kim · Manjin Ha · Byung Jun Min †
Department of Radiation Oncology, Chungbuk National University Hospital,
776, 1 Sunhwan-ro, Seowon-gu, Cheongju, Korea
† E-mail: bjmin.kr@gmail.com

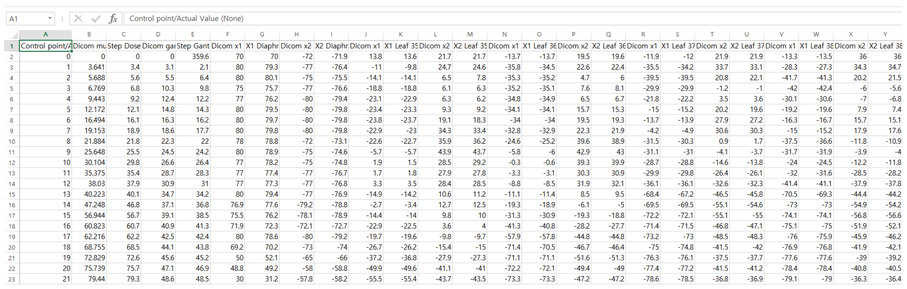
## Introduction

- Recently, many studies have been conducted to apply machine learning or deep learning for big data, which has been in the spotlight, to the department of radiation oncology [1][2].
- However, there is a difficulty to choose a proper model for data before doing a machine learning [3].
- The aim of this research is that data extracted from DICOM files of RT plan and log files are statistically analyzed to select a model of machine learning.

## Materials and Methods

- Data were extracted from the log files of Elekta versa HD and from the DICOM files of the RT Plan.
- The log files were recorded every 0.04 seconds, and the data in the DICOM file were recorded only at control points, so the data in the log files were aligned based on the DICOM files.

Fig 1. Extracted data from the log and the DICOM file



| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Control point/A | Dicom mu | Step Dose | Dicom ga | Step Gant | Dicom x1 | X1 Diaphr | Dicom x2 | X2 Diaphr | Dicom x1 | X1 Leaf 35 | Dicom x2 | X2 Leaf 35 | Dicom x1 | X1 Leaf 36 | Dicom x2 | X2 Leaf 36 | Dicom x1 | X1 Leaf 37 | Dicom x2 | X2 Leaf 37 | Dicom x1 | X1 Leaf 38 | Dicom x2 | X2 Leaf 38 |
| 2 | 0 | 0 | 0 | 0 | 359.6 | 70 | 70 | -72 | -71.9 | 13.8 | 13.6 | 21.7 | 21.7 | -13.7 | -13.7 | 19.5 | 19.6 | -11.9 | -12 | 21.9 | 21.9 | -13.3 | -13.5 | 36 | 36 |
| 3 | 1 | 3.641 | 3.4 | 3.1 | 2.1 | 80 | 79.3 | -77 | -76.4 | -11 | -9.8 | 24.7 | 24.6 | -35.8 | -34.5 | 22.6 | 22.4 | -35.5 | -34.2 | 33.7 | 33.1 | -28.3 | -27.3 | 34.3 | 34.7 |
| 4 | 2 | 5.688 | 5.6 | 5.5 | 6.4 | 80 | 80.1 | -75 | -75.5 | -14.1 | -14.1 | 6.5 | 7.8 | -35.3 | -35.2 | 4.7 | 6 | -39.5 | -39.5 | 20.8 | 22.1 | -41.7 | -41.3 | 20.2 | 21.5 |
| 5 | 3 | 6.769 | 6.8 | 10.3 | 9.8 | 75 | 75.7 | -77 | -76.6 | -18.8 | -18.8 | 6.1 | 6.3 | -35.2 | -35.1 | 7.6 | 8.1 | -29.9 | -29.9 | -1.2 | -1 | -42 | -42.4 | -6 | -5.6 |
| 6 | 4 | 9.443 | 9.2 | 12.4 | 12.2 | 77 | 76.2 | -80 | -79.4 | -23.1 | -22.9 | 6.3 | 6.2 | -34.8 | -34.9 | 6.5 | 6.7 | -21.8 | -22.2 | 3.5 | 3.6 | -30.1 | -30.6 | -7 | -6.8 |
| 7 | 5 | 12.172 | 12.1 | 14.8 | 14.3 | 80 | 79.5 | -80 | -79.8 | -23.4 | -23.3 | 9.3 | 9.2 | -34.1 | -34.1 | 15.7 | 15.3 | -15 | -15.2 | 20.2 | 19.6 | -19.2 | -19.6 | 7.9 | 7.4 |
| 8 | 6 | 16.494 | 16.1 | 16.3 | 16.2 | 80 | 79.7 | -80 | -79.8 | -23.8 | -23.7 | 19.1 | 18.3 | -34 | -34 | 19.5 | 19.3 | -13.7 | -13.9 | 27.9 | 27.2 | -16.3 | -16.7 | 15.7 | 15.1 |
| 9 | 7 | 19.153 | 18.9 | 18.6 | 17.7 | 80 | 79.8 | -80 | -79.8 | -22.9 | -23 | 34.3 | 33.4 | -32.8 | -32.9 | 22.3 | 21.9 | -4.2 | -4.9 | 30.6 | 30.3 | -15 | -15.2 | 17.9 | 17.6 |
| 10 | 8 | 21.884 | 21.8 | 22.3 | 22 | 78 | 78.8 | -72 | -73.1 | -22.6 | -22.7 | 35.9 | 36.2 | -24.6 | -25.2 | 39.6 | 38.9 | -31.5 | -30.3 | 0.9 | 1.7 | -37.5 | -36.6 | -11.8 | -10.9 |
| 11 | 9 | 25.648 | 25.5 | 24.5 | 24.2 | 80 | 78.9 | -75 | -74.6 | -5.7 | -5.7 | 43.9 | 43.7 | -5.8 | -6 | 42.9 | 43 | -31.1 | -31 | -4.1 | -3.7 | -31.7 | -31.9 | -3.9 | -4 |
| 12 | 10 | 30.104 | 29.8 | 26.6 | 26.4 | 77 | 78.2 | -75 | -74.8 | 1.9 | 1.5 | 28.5 | 29.2 | -0.3 | -0.6 | 39.3 | 39.9 | -28.7 | -28.8 | -14.6 | -13.8 | -24 | -24.5 | -12.2 | -11.8 |
| 13 | 11 | 35.375 | 35.4 | 28.7 | 28.3 | 77 | 77.4 | -77 | -76.7 | 1.7 | 1.8 | 27.9 | 27.8 | -3.3 | -3.1 | 30.3 | 30.9 | -29.9 | -29.8 | -26.4 | -26.1 | -32 | -31.6 | -28.5 | -28.2 |
| 14 | 12 | 38.03 | 37.9 | 30.9 | 31 | 77 | 77.3 | -77 | -76.8 | 3.3 | 3.5 | 28.4 | 28.5 | -8.8 | -8.5 | 31.9 | 32.1 | -36.1 | -36.1 | -32.6 | -32.3 | -41.4 | -41.1 | -37.9 | -37.8 |
| 15 | 13 | 40.223 | 40.1 | 34.7 | 34.2 | 80 | 79.4 | -77 | -76.9 | -14.9 | -14.2 | 10.6 | 11.2 | -11.1 | -11.4 | 8.5 | 9.5 | -68.4 | -67.2 | -46.5 | -45.8 | -70.5 | -69.3 | -44.4 | -44.2 |
| 16 | 14 | 47.248 | 46.8 | 37.1 | 36.8 | 76.9 | 77.6 | -79.2 | -78.8 | -2.7 | -3.4 | 12.7 | 12.5 | -19.3 | -18.9 | -6.1 | -5 | -69.5 | -69.5 | -55.1 | -54.6 | -73 | -73 | -54.9 | -54.2 |
| 17 | 15 | 56.944 | 56.7 | 39.1 | 38.5 | 75.5 | 76.2 | -78.1 | -78.9 | -14.4 | -14 | 9.8 | 10 | -31.3 | -30.9 | -19.3 | -18.8 | -72.2 | -72.1 | -55.1 | -55 | -74.1 | -74.1 | -56.8 | -56.6 |
| 18 | 16 | 60.823 | 60.7 | 40.9 | 41.3 | 71.9 | 72.3 | -72.1 | -72.7 | -22.9 | -22.5 | 3.6 | 4 | -41.3 | -40.8 | -28.2 | -27.7 | -71.4 | -71.5 | -46.8 | -47.1 | -75.1 | -75 | -51.9 | -52.1 |
| 19 | 17 | 62.216 | 62.2 | 42.5 | 42.4 | 80 | 78.6 | -80 | -79.2 | -19.7 | -19.6 | -9.8 | -9.7 | -57.9 | -57.8 | -44.8 | -44.8 | -73.2 | -73 | -48.5 | -48.3 | -76 | -75.9 | -45.9 | -46.2 |
| 20 | 18 | 68.755 | 68.5 | 44.1 | 43.8 | 69.2 | 70.2 | -73 | -74 | -26.7 | -26.2 | -15.4 | -15 | -71.4 | -70.5 | -46.7 | -46.4 | -75 | -74.8 | -41.5 | -42 | -76.9 | -76.8 | -41.9 | -42.1 |
| 21 | 19 | 72.829 | 72.6 | 45.6 | 45.2 | 50 | 52.1 | -65 | -66 | -37.2 | -36.8 | -27.9 | -27.3 | -71.1 | -71.1 | -51.6 | -51.3 | -76.3 | -76.1 | -37.5 | -37.7 | -77.6 | -77.6 | -39 | -39.2 |
| 22 | 20 | 75.739 | 75.7 | 47.1 | 46.9 | 48.8 | 49.2 | -58 | -58.8 | -49.9 | -49.6 | -41.1 | -41 | -72.2 | -72.1 | -49.4 | -49 | -77.4 | -77.2 | -41.5 | -41.2 | -78.4 | -78.4 | -40.8 | -40.5 |
| 23 | 21 | 79.44 | 79.3 | 48.6 | 48.5 | 30 | 31.2 | -57.8 | -58.2 | -55.5 | -55.4 | -43.7 | -43.5 | -73.3 | -73.3 | -47.2 | -47.2 | -78.6 | -78.5 | -36.8 | -36.9 | -79.1 | -79 | -36.3 | -36.4 |

- The correlation coefficient of each variables was calculated using the free statistical program R (version x64 4.0.4). In addition, the model of linear regression was established by extracting highly correlated variables whose correlation coefficients are more than about 0.9 and t-testing was performed for significance tests [4].

## Conclusion

- It is expected to apply a linear regression for one of the machine learning models, which has various models.
- The MLC variables are associated with MU variable, so it will be analyzed more with other statistical methods.
- Furthermore, it will be helpful to predict the data which will be written in a log file because the linear regression is a statistical method to calculate the next values.

## Acknowledgement

This research was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2020R1A2C4001910 and 2020M2D9A1094075).

## Results and Discussion

- The correlation coefficients which represents the relation of variables are shown in Table Ⅰ, which omitted the remainder of MLC variables, because they are similar to the MLC 35 variable.

**Table Ⅰ.** The correlation coefficients of variables in DICOM and log files (MLC Leaf 35, D : DICOM, L : log file)

| | D.Mu | L.Mu | D.Gan. Angle | L.Gan. Angle | D.X1. Jaw | L.X1. Jaw | D.X2. Jaw | L.X2. Jaw | D.X1. Lf35 | L.X1. Lf35 | D.X2. Lf35 | L.X2. Lf35 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| D.Mu | 1 | 0.9999 | -0.006 | -0.008 | -0.013 | -0.013 | -0.055 | -0.055 | 0.5463 | 0.5458 | 0.4816 | 0.4833 |
| L.Mu | 0.9999 | 1 | -0.006 | -0.008 | -0.013 | -0.013 | -0.054 | -0.054 | 0.5462 | 0.5457 | 0.4816 | 0.4832 |
| D.Gan. Angle | -0.006 | -0.006 | 1 | 0.9893 | -0.127 | -0.127 | 0.1149 | 0.1148 | 0.1176 | 0.1139 | 0.1007 | 0.0993 |
| L.Gan. Angle | -0.008 | -0.008 | 0.9893 | 1 | -0.129 | -0.129 | 0.1175 | 0.1173 | 0.1161 | 0.1124 | 0.0996 | 0.0981 |
| D.X1. Jaw | -0.013 | -0.013 | -0.127 | -0.129 | 1 | 0.9992 | -0.704 | -0.705 | -0.078 | -0.068 | 0.0693 | 0.0678 |
| L.X1. Jaw | -0.013 | -0.013 | -0.127 | -0.129 | 0.9992 | 1 | -0.705 | -0.706 | -0.078 | -0.067 | 0.0696 | 0.0680 |
| D.X2. Jaw | -0.055 | -0.054 | 0.1149 | 0.1175 | -0.704 | -0.705 | 1 | 0.9996 | 0.0188 | 0.0191 | -0.099 | -0.099 |
| L.X2. Jaw | -0.055 | -0.054 | 0.1148 | 0.1173 | -0.705 | -0.706 | 0.9996 | 1 | 0.0187 | 0.0191 | -0.099 | -0.099 |
| D.X1. Lf35 | 0.5463 | 0.5462 | 0.1176 | 0.1161 | -0.078 | -0.078 | 0.0188 | 0.0187 | 1 | 0.9990 | 0.8976 | 0.8975 |
| L.X1. Lf35 | 0.5458 | 0.5457 | 0.1139 | 0.1124 | -0.068 | -0.067 | 0.0191 | 0.0191 | 0.9990 | 1 | 0.8988 | 0.8995 |
| D.X2. Lf35 | 0.4816 | 0.4816 | 0.1007 | 0.0996 | 0.0693 | 0.0696 | -0.099 | -0.099 | 0.8976 | 0.8988 | 1 | 0.9995 |
| L.X2. Lf35 | 0.4833 | 0.4832 | 0.0993 | 0.0981 | 0.0678 | 0.0680 | -0.099 | -0.099 | 0.8975 | 0.8995 | 0.9995 | 1 |

- It is estimated that some variables have a linear relationship with other variables because of the result of t-test in Table Ⅱ.

**Table Ⅱ.** The t-test's p-value of correlation coefficients containing MU, gantry angle, jaw positions, X1 leaf 35.

| | D.MU | D.Gan. angle | D.Jaw X1 | D.Jaw X2 | D,MLC X1 35 |
|---|---|---|---|---|---|
| D.MU | | | | | |
| D.Gan. angle | 0.3377 | | | | |
| D.Jaw X1 | 0.05322 | $2.0\times10^{-16}$ | | | |
| D.Jaw X2 | $5.0\times10^{-15}$ | $2.0\times10^{-16}$ | $2.0\times10^{-16}$ | | |
| D.MLC X1 35 | $2.0\times10^{-16}$ | $2.0\times10^{-16}$ | $2.0\times10^{-16}$ | 0.00744 | |

- The result of linear regression is illustrated in Table Ⅲ. The coefficient is almost one, and the intercept is approximately less than 0.2.

**Table Ⅲ.** The result of linear regression containing MU, gantry angle, jaw positions, X1 leaf 36-38 and X2 leaf 36-38.

| | coefficient | intercept | p-value |
|---|---|---|---|
| MU | 0.999 | -0.146 | $1.58\times10^{-8}$ |
| Gantry angle | 0.989 | 0.1861 | $5.37\times10^{-16}$ |
| Jaw X1 position | 0.997 | 0.214 | $2.00\times10^{-16}$ |
| Jaw X2 position | 0.998 | -0.157 | $2.00\times10^{-16}$ |
| MLC X1 36 | 0.993 | -0.167 | $2.00\times10^{-16}$ |
| MLC X2 36 | 0.996 | 0.066 | $2.00\times10^{-16}$ |
| MLC X1 37 | 0.993 | -0.164 | $2.00\times10^{-16}$ |
| MLC X2 37 | 0.995 | 0.066 | $2.00\times10^{-16}$ |
| MLC X1 38 | 0.994 | -0.141 | $2.00\times10^{-16}$ |
| MLC X2 38 | 0.996 | 0.056 | $2.00\times10^{-16}$ |

## References

1. Naqa I, Irrer J, Ritter TA, et al. "Machine learning for automated qual-ity assurance in radiotherapy: a proof of principle using EPID datadescription", Med Phys. 2019;46:1914–1921
2. D, Zhang X, Li H, et al. "Predicting gamma passing rates for portaldosimetry-based IMRT QA using machine learning." Med Phys. 2019;46:4666–4675.
3. Geron, Aurelien, "Hands-on Machine Learning With Scikit-learn, Keras, and Tensorflow (Concepts, Tools, and Techniques to Build Intelligent Systems)", Oreilly&AssociatesInc ,2018
4. Robert Hogg, "Introduction to Mathematical Statistics", PearsonHigherEducation, 2013