

Recent Advances in Machine Learning for Log File-Based PSQA for IMRT and VMAT

Kellin M. DeJesus

Cell Biology and Regenerative Medicine

Thomas Jefferson University

Philadelphia, Pennsylvania, USA

e-mail: kmd119@students.jefferson.edu

Leon Dunn

Medical Physics

Genesis Care

Victoria, Australia

e-mail: info@efilmqa.com

David Thomas

Medical Physics

Thomas Jefferson University

Philadelphia, Pennsylvania, USA

e-mail: david.thomas2@jefferson.edu

Les Sztandera

Computer Science

Thomas Jefferson University

Philadelphia, Pennsylvania, USA

e-mail: les.sztandera@jefferson.edu

Abstract—The paper addresses the critical need for a faster and more efficient approach to Patient-Specific Quality Assurance (PSQA) in radiation therapy. The accuracy of PSQA is crucial for the safety of radiation therapy, particularly with complex procedures like Intensity-Modulated Radiation Therapy (IMRT) and Volumetric-Arc Radiation Therapy (VMAT). Traditional phantom-based methods, while effective, are time-consuming and fail to account for patient-specific variability and real-time treatment adjustment. To address these limitations, alternative strategies leveraging trajectory log files—automatically recorded during treatment—have emerged as promising tools for PSQA. In recent years, the application of machine learning and deep learning algorithms to trajectory log files has been increasingly studied in literature. These algorithms have shown notable progress in predicting PSQA outcomes and detecting errors, though further development is required before they can be fully integrated into clinical practice. By surveying key studies, the paper highlights the potential of algorithms such as support vector machines, tree-based methods, and convolutional neural networks to enhance the efficiency and accuracy of log file-based PSQA. The findings underscore the promise of these techniques in replacing traditional methods while addressing current challenges to pave the way for clinical integration.

Keywords—*deep learning; machine learning; quality assurance; volumetric-arc radiation therapy; intensity-modulated radiation therapy.*

The American Cancer Society has estimated over 2 million new cases of cancer in 2024 [1]. About 50% of all cancer patients are expected to receive radiotherapy at some point during treatment [2]. The proportion of radiotherapy patients receiving Intensity-Modulated Radiotherapy (IMRT) and Volumetric Modulated Arc Therapy (VMAT) has steadily increased over time from 22% in 2004 to 57.8% in 2017 [3]. IMRT and VMAT are routine but complex cancer treatment modalities

that require time-consuming Quality Assurance (QA) measures. Log file-based Patient-Specific Quality Assurance (PSQA) has been proposed as an alternative method that can be performed in real-time on a fraction-by-fraction basis [4][5][6]. Studies comparing log file-based PSQA have identified differences between log file recordings and actual behavior of machines during treatment, however, several mitigation strategies have been proposed [4][7][8]. These studies have given new insights into the potential for more efficient PSQA, however, they have been limited by small cohort size.

Machine learning, and by extension deep learning, have rapidly gained traction as essential tools for advancing health-care [9][10][11]. Machine learning can process and analyze large, complex datasets to identify patterns and make predictions that can be implemented to improve patient outcomes, increase treatment efficiency, and aid in clinical decision-making. Machine learning algorithms can automate time-consuming tasks. This can reduce the workload on medical professionals, reduce waiting times, and mitigate the risks of human error. Unlike traditional strategies for automation that are static after their implementation, these algorithms can evolve over time with additional data. Updates are made constantly to maintain or improve accuracy [12]. This is specifically advantageous in fields, such as radiation therapy, where advancements are rapid, and techniques are constantly changing [13][14][15][16].

The following paper thus endeavors to give a brief but comprehensive overview of the current status of machine learning for log-file based PSQA measures. This paper is structured as follows: Section 2 provides the theoretical context

for log-file based PSQA. Section 3 explores the various applications of machine learning and deep learning models for PSQA. Section 4 discusses future directions and concludes with final remarks.

I. BACKGROUND

We will provide an overview of the theories behind the use of log files for PSQA and the theory for the most successful machine learning algorithms to date.

A. Log File-Based PSQA

IMRT is of particular value when treating tumors with complex or concave shapes, especially those located near radiosensitive normal tissues. It uses a computer-controlled linear accelerator (linac) that can rotate around the patient on a gantry. This process excites electrons via microwave technology, which then collide with a heavy metal target to produce high-energy x-rays. These beams are shaped by the Multileaf Collimator (MLC) as they exit the machine. The intensity of each beam segment, MLC shaping, and gantry rotation are all determined using 3D imaging prior to treatment. The precision in dose delivery provided by IMRT allows for the irradiation of the tumor while sparing nearby healthy tissues, making it especially useful for tumors located near critical organs [17].

Due to the complexity of IMRT and VMAT treatment plans, each patient's treatment plan requires inversely optimized planning. Before treatment begins, these plans are often measured on the linac using detector arrays. The added complexity necessitates additional QA measures to ensure patient safety in clinical settings [18][19]. Confirmations of machine performance and patient treatment plan accuracy are essential. These verifications include assessing patient positioning, machine mechanical accuracy, dose distribution, and beam geometry. Given the complex and highly variable nature of each treatment plan, PSQA is required [20].

Currently, IMRT and VMAT treatment plans undergo physical measurements of the plan parameters before treatment begins to ensure the machine delivers the intended dose. However, these measurements are often done in advance and may not account for real-time deviations that may occur before the treatment begins. As such, the potential for mechanical changes resulting in dose discrepancies between PSQA and actual treatment remains a concern. The most common procedure involves recalculating the dose distribution of a patient's treatment plan onto a suitable phantom. The dose distribution is then measured using various devices, such as film, ion chambers, diode arrays, or Electronic Portal Imaging Devices (EPIDs). Differences between the measured and planned dose distributions are quantified using gamma analysis, as outlined in American Association of Physicists in Medicine Task Groups (AAPM TGs) 119 and 218. These guidelines recommend that over 90% of measured points should fall within a 3% dose difference and a 2mm distance-to-agreement (DTA) [21][22]. However, this process is time and labor-intensive, often requiring after-hours work to avoid

interrupting treatment schedules. Additionally, there is ongoing debate over the efficacy of these methods, particularly regarding their robustness and ability to detect potential failure modes [6][23][24].

Log file-based PSQA offers an alternative to traditional methods by utilizing automatically generated log files from radiation treatment machines to verify the accuracy of treatment plans. These log files capture data such as radiation output, MLC positions, gantry and couch positions, beam angles, and timing information. This data can then be compared to the treatment plan to identify potential errors [25]. Log files, TPS files, and Mean Complexity Scores (MCS) have been used to develop prediction models for Gamma Passing Rate (GPR), a key metric in PSQA [26]. Recent studies have indicated discrepancies between the recorded data and the actual performance of the machine, particularly in terms of MLC positioning [26][27][28]. However, since log files are generated by the linac, they do not detect mechanical miscalibrations, such as incorrect leaf positioning. Moreover, they cannot account for low plan quality or errors originating from the treatment planning system (TPS). To mitigate these limitations, enhanced QA procedures specifically for the linac, combined with more sensitive machine QA tools, are recommended to ensure MLC accuracy [26].

B. Machine Learning and IMRT/VMAT

Treatment log files record various parameters of radiation delivery, such as MLC position, dose rates, beam angles, and gantry positions in real-time during the course with recordings taken every few milliseconds [29]. As highly structured, real-time, and extensive data capture, these files would be particularly difficult to analyze manually. Log files are thus particularly well-suited to machine learning algorithms for pattern recognition and error prediction. Models range from simple classification techniques to complex deep learning algorithms. The most successful models in the literature include Support Vector Machines (SVMs), tree-based algorithms, and Artificial Neural Networks (ANNs).

SVMs are effective for classification tasks for log file-based PSQA. They can distinguish between compliant and non-compliant treatment sessions by setting predefined acceptable ranges for discrepancies between planned and delivered values for parameters within the log file, such as dose rate, MLC positions, and beam angles. This allows for quick identification of errors as they occur so that a clinician can be alerted. However, SVM is limited to cases where there are clear distinctions between compliant and non-compliant values. SVM is also sensitive to noise and outliers and is not well suited for multi-class tasks [30].

Tree-based algorithms are non-parametric and based on hierarchical, tree-like structures. Each tree is made up of nodes that represent decisions based on feature values. The branches represent possible outcomes or decisions. They are well-suited for non-linear relationships between features and can partition the feature space in more complex ways than linear models. Tree-based machine learning models include

Random Forest (RF), Gradient Boosting, and Extreme Gradient Boosting (XGBoost) algorithms [31][32][33].

RF models can leverage many decision trees to map the involvement of multiple interacting features to identify more subtle discrepancies between expected and delivered values. It can detect complex relationships within the treatment data that would not be as apparent with simpler methods such as SVM. Due to the ensemble nature of the algorithm, RFs are difficult to interpret and feature importance scores are only rough approximations. They can show bias toward categorical features with many levels. RFs also require a lot of optimizations for hyperparameter tuning [31].

Gradient Boosting uses decision trees as its base and adjusts instance weights with each iteration by fitting new predictors to errors in the preceding iteration. Individual decision trees are differentiated by a different subset of features to select the best split. Each new tree accounts for the errors of the preceding ones. This approach can be slow to train and is prone to overfitting [32]. XGBoost builds upon the gradient boosting algorithm by including L1 (Lasso) and L2 (Ridge) regularization to prevent overfitting [34][35]. It also grows trees with a depth-first approach and can train trees in parallel, which increases the speed of training. Although these two models are less prone to overfitting than RF, they do still pose some risk of overfitting. They also exhibit hyperparameter sensitivity and require careful tuning, especially for large datasets. Like other tree-based models, they both struggle with extrapolation beyond the training dataset [33].

ANNs are based on the McCulloch-Pitts artificial neuron model. The model represents a neuron as a binary threshold unit and inputs are assigned weights before being summed, and compared against a specific threshold to determine the neuron's output. This effectively enables the representation of logical functions [36]. With the advent of backpropagation and activation functions -such as the Rectified Linear Unit (ReLU)- Deep Neural Networks (DNNs) further built upon the ANN model by increasing the number of hidden layers which enabled more complex patterns and representations to be modeled [37][38]. Deep learning models such as convolutional neural networks (CNNs), have more recently been applied to log file-based PSQA. CNNs are well-suited to image classification, making them ideal for use with fluence maps that can be generated by log file data. CNNs apply filters to detect desired features, reduce spatial dimensions to retain the most important features, and then perform final classification or predictions. They circumvent the need for manual feature selection. They are highly scalable for large datasets and have improved computational efficiency [39]. CNN's capabilities for detecting highly complex and time-dependent errors make them ideal for log file-based PSQA applications. They can identify small misalignments in MLC positions, irregular dose rate fluctuations, as well as other more subtle anomalies that may be missed by more traditional machine learning models. To prevent overfitting, large, labeled datasets are required and can be vulnerable to being misled by small input changes. CNN's decision making can be extremely difficult to interpret

[40].

II. EXAMPLES OF RECENT APPLICATIONS

This section will summarize the current machine learning applications for IMRT/VMAT PSQA within literature, including both drawbacks and advantages.

A. Recent Models for IMRT/VMAT PSQA

Most current applications for these models in IMRT and VMAT PSQA can be classified as either parameter prediction studies or error detection studies (see Table 1). Most parameter prediction studies are structured to predict GPR, with error detection studies predominantly performed on induced error data input. Either approach appears to struggle with similar limitations.

B. Drawbacks and Limitations

Tomori et al.[41], Lam et al.[43], Ono et al.[44], Huang et al.[45], Wang et al.[46], and Song et al.[47] used the parameter prediction approach. Using a prediction approach, all studies indicated that machine learning models could be effectively trained using log files to predict machine parameters at the time of treatment delivery for new treatment plans. These studies vary in the models explored, including SVM, RF, CNNs, and others. All models have relatively promising accuracy as seen in Table 1. However, Tomori et al.'s scope was limited to prostate IMRT plans, Huang et al. was limited to chest IMRT plans, and Song et al. was similarly limited to nasopharyngeal carcinoma and only used static gantry IMRT plans. Lam et al. included plans for multiple anatomical sites but were still specific to IMRT. Ono et al. and Wang et al. were specific to VMAT plans. Ono et al. and Lam et al. both performed their studies on multiple linear accelerators, but only Lam et al. used data from more than one institution. All six studies acknowledge that by using trajectory files, which are dependent on the linear accelerator itself, there is some vulnerability to machine-based error. As such, most log file-based PSQA is considered an enhancement to other QA measures that ensure the machine is calibrated appropriately, either with separate protocols or by incorporating additional sources of data into future models.

Error detection studies such as those by Kimura et al.[48], Sakai et al.[49], and Nyflot et al.[50] were similarly limited to one treatment plan type from a single institution. The only study that incorporated both VMAT and IMRT plans into a single study was an error detection study by Chuang et al. However, the study was only focused on MLC errors.

C. Positive Developments

These preliminary studies have gleaned significant insights into creating a holistic model for automating PSQA using log file data with a clear improvement upon methods over time. Lam et al. trained their model for predicting dosimetric effects in lieu of GPR to overcome any discrepancies between gamma index and errors that are clinically relevant [43]. Kimura et al. directly compared gamma map-based CNN models with dose

difference map-based CNN models and found dose difference maps were more accurate [48]. Sakai et al. included radiomic data which resulted in higher sensitivity and specificity for MLC position and MLC modeling errors [49]. Hirashima et al. utilized a combination of 3D dosiomic features and plan complexity in a tree-based model [52]. Tomori et al.'s GPR prediction-based CNN model struggled with overestimating low GPR values and underestimating GPR in the test set [41][42]. Song et al. developed a novel model that weighed the MSE loss function to mitigate this class imbalance with promising results [47]. However, as all these studies have been limited to relatively small, single, or double institution datasets, their results are difficult to directly compare to one another. Additionally, most of the literature has been performed using Varian machines [21]. Although Varian machines are widely used in the US, Elekta machines are also used.

III. DISCUSSION

Literature has broadly indicated that CNNs and other Deep Learning models appear to be the most successful at creating a model that is robust against certain biases seen in SVM and tree-based algorithms [53]. Although some studies have utilized data augmentation, most studies have agreed that to bring these findings to a clinically relevant standpoint, sufficient data must be collected from multiple institutions, techniques, treatment machines, and anatomical sites [54][55]. Additionally, encompassing both Varian and Elekta machines is essential to ensure this PSQA strategy is accurate on both platforms [56].

Furthermore, past work has predominantly focused on deterministic methods, which are ideal for providing direct, quantitative evaluations of dose delivery accuracy. While these are incredibly important in the overall application of the model, there are many aspects of treatment that carry uncertainty. Error tolerance, dose assessments, and multi-criteria evaluations are all subject to imprecision. Cilla et al. approached these aspects by using a "traffic light" protocol [57]. The protocol leveraged plan complexity to designate plans as acceptable (green light), requires further verification (orange light), or unacceptable (red light). Fuzzy logic follows similar reasoning and has been successfully applied to radiation control systems and treatment plan optimization [58][59]. Fuzzy logic uses fuzzy sets and linguistic variables to model uncertain or imprecise information. Desired variables can be assigned degrees of truth rather than a yes/no value. When applied to complex systems, this mathematical system eliminates the restriction of binary values to create more human-like decision making. The Fuzzy-CID3 (F-CID3) algorithm is a tree-based, hybrid method that combines neural networks and fuzzy sets, generating its own topology. Using a neural fuzzy number tree with a class separation method, the F-CID3 algorithm simplifies architecture compared to preprocessors, achieving better performance with fewer connections [60].

IV. CONCLUSIONS AND FUTURE DIRECTIONS

Recent work has proposed log file-based PSQA as a promising solution to the limitations of traditional phantom-

based QA methods by leveraging Machine Learning algorithms to predict IMRT/VMAT QA outcomes and detect errors [6]. These algorithms, including SVMs, tree-based models, and CNNs, have demonstrated substantial progress in using log files for treatment plan verification.

While studies show the potential of log file-based PSQA, they also highlight key limitations. These include the inability to detect mechanical miscalibration or treatment planning errors, and the restricted scope of available data [45][47][61]. Mechanical errors can be mitigated through enhanced QA protocols for linacs and the incorporation of more sensitive machine tools. Moreover, issues such as insufficient training data for cancer site stratification and the lack of multi-institutional studies with diverse machine types remain significant barriers to widespread implementation [47][57][61][62][63].

Given the time-consuming nature of current PSQA protocols, log file-based PSQA, combined with AI model predictions, offers an efficient alternative. Future studies should focus on creating larger, multi-institutional datasets and exploring features within machine learning models that identify key factors in treatment failure. As machine learning and deep learning models evolve, their integration into clinical practice could lead to more efficient, accurate, and real-time quality assurance for radiation therapy.

REFERENCES

- [1] R. L. Siegel, A. N. Giaquinto, and A. Jemal, "Cancer statistics, 2024.", *CA: A Cancer Journal for Clinicians*, vol. 74, no. 1, pp. 12–49, Jan. 2024. DOI: 10.3322/caac.21820.
- [2] A. A. for Cancer Research, *Aacr Cancer Progress Report 2024*. AACR, 2024, ISBN: 979-8-9857852-7-2.
- [3] R. J. Hutten et al., "Worsening racial disparities in utilization of intensity modulated radiation therapy.", *Advances in radiation oncology*, vol. 7, no. 3, p. 100887, Jan. 2022, ISSN: 24521094. DOI: 10.1016/j.adro.2021.100887.
- [4] A. Agnew, C. E. Agnew, M. W. D. Grattan, A. R. Hounsell, and C. K. McGarry, "Monitoring daily MLC positional errors using trajectory log files and EPID measurements for IMRT and VMAT deliveries.", *Physics in Medicine and Biology*, vol. 59, no. 9, N49–63, May 2014. DOI: 10.1088/0031-9155/59/9/N49.
- [5] C. E. Agnew, D. M. Irvine, and C. K. McGarry, "Correlation of phantom-based and log file patient-specific QA with complexity scores for VMAT.", *Journal of applied clinical medical physics / American College of Medical Physics*, vol. 15, no. 6, p. 4994, Nov. 2014. DOI: 10.1120/jacmp.v15i6.4994.
- [6] N. Childress, Q. Chen, and Y. Rong, "Parallel/opposed: IMRT QA using treatment log files is superior to conventional measurement-based method.", *Journal of applied clinical medical physics / American College of Medical Physics*, vol. 16, no. 1, p. 5385, Jan. 2015. DOI: 10.1120/jacmp.v16i1.5385.
- [7] B. Neal et al., "A clinically observed discrepancy between image-based and log-based MLC positions.", *Medical Physics*, vol. 43, no. 6, pp. 2933–2935, Jun. 2016. DOI: 10.1118/1.4949002.
- [8] M. Barnes et al., "Insensitivity of machine log files to MLC leaf backlash and effect of MLC backlash on clinical dynamic MLC motion: An experimental investigation.", *Journal of applied clinical medical physics / American College of Medical Physics*, vol. 23, no. 9, e13660, Sep. 2022. DOI: 10.1002/acm2.13660.

- [9] J. Waring, C. Lindvall, and R. Umeton, "Automated machine learning: Review of the state-of-the-art and opportunities for healthcare.", *Artificial Intelligence in Medicine*, vol. 104, p. 101822, Apr. 2020, ISSN: 09333657. DOI: 10.1016/j.artmed.2020.101822.
- [10] A. Esteva et al., "A guide to deep learning in healthcare.", *Nature Medicine*, vol. 25, no. 1, pp. 24–29, Jan. 2019, ISSN: 1078-8956. DOI: 10.1038/s41591-018-0316-z.
- [11] K. Rasheed et al., "Explainable, trustworthy, and ethical machine learning for healthcare: A survey.", *Computers in biology and medicine*, vol. 149, p. 106043, Oct. 2022. DOI: 10.1016/j.combiomed.2022.106043.
- [12] K. Y. Ngiam and I. W. Khor, "Big data and machine learning algorithms for health-care delivery.", *The Lancet Oncology*, vol. 20, no. 5, e262–e273, May 2019. DOI: 10.1016/S1470-2045(19)30149-4.
- [13] J. M. Park, J.-I. Kim, and H.-G. Wu, "Technological advances in charged-particle therapy.", *Cancer research and treatment : official journal of Korean Cancer Association*, vol. 53, no. 3, pp. 635–640, Jul. 2021. DOI: 10.4143/crt.2021.706.
- [14] R. A. Chandra, F. K. Keane, F. E. M. Voncken, and C. R. Thomas, "Contemporary radiotherapy: Present and future.", *The Lancet*, vol. 398, no. 10295, pp. 171–184, Jul. 2021. DOI: 10.1016/S0140-6736(21)00233-6.
- [15] J. Bertholet, Y. Vinogradskiy, Y. Hu, and D. J. Carlson, "Advances in image-guided adaptive radiation therapy.", *International Journal of Radiation Oncology, Biology, Physics*, vol. 110, no. 3, pp. 625–628, Jul. 2021. DOI: 10.1016/j.ijrobp.2021.02.047.
- [16] S. Bartzsch et al., "Technical advances in x-ray microbeam radiation therapy.", *Physics in Medicine and Biology*, vol. 65, no. 2, 02TR01, Jan. 2020. DOI: 10.1088/1361-6560/ab5507.
- [17] C. X. Yu, C. J. Amies, and M. Svatos, "Planning and delivery of intensity-modulated radiation therapy.", *Medical Physics*, vol. 35, no. 12, pp. 5233–5241, Dec. 2008. DOI: 10.1118/1.3002305.
- [18] A. F. I. Osman, N. M. Maalej, and K. Jayesh, "Prediction of the individual multileaf collimator positional deviations during dynamic IMRT delivery priori with artificial neural network.", *Medical Physics*, vol. 47, no. 4, pp. 1421–1430, Apr. 2020. DOI: 10.1002/mp.14014.
- [19] P. Szeverinski et al., "Error sensitivity of a log file analysis tool compared with a helical diode array dosimeter for VMAT delivery quality assurance.", *Journal of applied clinical medical physics / American College of Medical Physics*, vol. 21, no. 11, pp. 163–171, Nov. 2020. DOI: 10.1002/acm2.13051.
- [20] E. E. Klein et al., "Task group 142 report: Quality assurance of medical accelerators.", *Medical Physics*, vol. 36, no. 9, pp. 4197–4212, Sep. 2009. DOI: 10.1118/1.3190392.
- [21] G. A. Ezzell et al., "IMRT commissioning: Multiple institution planning and dosimetry comparisons, a report from AAPM task group 119.", *Medical Physics*, vol. 36, no. 11, pp. 5359–5373, Nov. 2009. DOI: 10.1118/1.3238104.
- [22] M. Miften et al., "Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM task group no. 218.", *Medical Physics*, vol. 45, no. 4, e53–e83, Apr. 2018. DOI: 10.1002/mp.12810.
- [23] R. A. C. Siochi, A. Molineu, and C. G. Orton, "Point/counterpoint. patient-specific QA for IMRT should be performed using software rather than hardware methods.", *Medical Physics*, vol. 40, no. 7, p. 070601, Jul. 2013. DOI: 10.1118/1.4794929.
- [24] J. C. Smith, S. Dieterich, and C. G. Orton, "Point/counterpoint. it is still necessary to validate each individual IMRT treatment plan with dosimetric measurements before delivery.", *Medical Physics*, vol. 38, no. 2, pp. 553–555, Feb. 2011. DOI: 10.1118/1.3512801.
- [25] D. W. Litzenberg, J. M. Moran, and B. A. Fraass, "Verification of dynamic and segmental IMRT delivery by dynamic log file analysis.", *Journal of applied clinical medical physics / American College of Medical Physics*, vol. 3, no. 2, pp. 63–72, 2002. DOI: 10.1120/jacmp.v3i2.2578.
- [26] A. L. McNiven, M. B. Sharpe, and T. G. Purdie, "A new metric for assessing IMRT modulation complexity and plan deliverability.", *Medical Physics*, vol. 37, no. 2, pp. 505–515, Feb. 2010. DOI: 10.1118/1.3276775.
- [27] J. M. Moran et al., "Safety considerations for IMRT: Executive summary.", *Practical radiation oncology*, vol. 1, no. 3, pp. 190–195, Sep. 2011. DOI: 10.1016/j.prro.2011.04.008.
- [28] L. Masi, R. Doro, V. Favuzza, S. Cipressi, and L. Livi, "Impact of plan parameters on the dosimetric accuracy of volumetric modulated arc therapy.", *Medical Physics*, vol. 40, no. 7, p. 071718, Jul. 2013. DOI: 10.1118/1.4810969.
- [29] B. Sun et al., "Initial experience with TrueBeam trajectory log files for radiation therapy delivery verification.", *Practical radiation oncology*, vol. 3, no. 4, e199–208, Dec. 2013. DOI: 10.1016/j.prro.2012.11.013.
- [30] C. Cortes and V. Vapnik, "Support-vector networks", *Machine learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995, ISSN: 0885-6125. DOI: 10.1007/BF00994018.
- [31] T. K. Ho, "Random decision forests", in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, 1995, 278–282 vol.1. DOI: 10.1109/ICDAR.1995.598994.
- [32] J. H. Friedman, "Greedy function approximation: A gradient boosting machine", *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, Oct. 2001, ISSN: 0090-5364. DOI: 10.1214/aos/1013203451.
- [33] T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system", in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*, New York, New York, USA: ACM Press, Aug. 2016, pp. 785–794, ISBN: 9781450342322. DOI: 10.1145/2939672.2939785.
- [34] R. Tibshirani, "Regression shrinkage and selection via the lasso", *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, Jan. 1996, ISSN: 00359246. DOI: 10.1111/j.2517-6161.1996.tb02080.x.
- [35] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems", *Technometrics : a journal of statistics for the physical, chemical, and engineering sciences*, vol. 12, no. 1, pp. 55–67, Feb. 1970, ISSN: 0040-1706. DOI: 10.1080/00401706.1970.10488634.
- [36] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, Dec. 1943, ISSN: 0007-4985. DOI: 10.1007/BF02478259.
- [37] A. G. Ivakhnenko, "The group method of data handling; a rival of the method of stochastic approximation", 1968.
- [38] R. H. Hahnloser, R. Sarpeshkar, M. A. Mahowald, R. J. Douglas, and H. S. Seung, "Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit.", *Nature*, vol. 405, no. 6789, pp. 947–951, Jun. 2000. DOI: 10.1038/35016072.
- [39] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998. DOI: 10.1109/5.726791.
- [40] W. Ding, M. Abdel-Basset, H. Hawash, and A. M. Ali, "Explainability of artificial intelligence methods, applications and challenges: A comprehensive survey", *Information sciences*, vol. 615, pp. 238–292, Nov. 2022, ISSN: 00200255. DOI: 10.1016/j.ins.2022.10.013.

- [41] S. Tomori et al., “A deep learning-based prediction model for gamma evaluation in patient-specific quality assurance.”, *Medical Physics*, Jul. 2018. DOI: 10.1002/mp.13112.
- [42] S. Tomori et al., “Systematic method for a deep learning-based prediction model for gamma evaluation in patient-specific quality assurance of volumetric modulated arc therapy.”, *Medical Physics*, vol. 48, no. 3, pp. 1003–1018, Mar. 2021. DOI: 10.1002/mp.14682.
- [43] D. Lam et al., “Predicting gamma passing rates for portal dosimetry-based IMRT QA using machine learning.”, *Medical Physics*, vol. 46, no. 10, pp. 4666–4675, Oct. 2019. DOI: 10.1002/mp.13752.
- [44] T. Ono et al., “Prediction of dosimetric accuracy for VMAT plans using plan complexity parameters via machine learning.”, *Medical Physics*, vol. 46, no. 9, pp. 3823–3832, Sep. 2019. DOI: 10.1002/mp.13669.
- [45] Y. Huang et al., “Deep learning for patient-specific quality assurance: Predicting gamma passing rates for IMRT based on delivery fluence informed by log files.”, *Technology in Cancer Research & Treatment*, vol. 21, p. 15 330 338 221 104 881, 2022. DOI: 10.1177/15330338221104881.
- [46] L. Wang et al., “Multi-task autoencoder based classification-regression model for patient-specific VMAT QA.”, *Physics in Medicine and Biology*, vol. 65, no. 23, p. 235 023, Nov. 2020. DOI: 10.1088/1361-6560/abb31c.
- [47] W. Song et al., “Improving the performance of deep learning models in predicting and classifying gamma passing rates with discriminative features and a class balancing technique: A retrospective cohort study.”, *Radiation Oncology*, vol. 19, no. 1, p. 98, Jul. 2024. DOI: 10.1186/s13014-024-02496-5.
- [48] Y. Kimura, N. Kadoya, S. Tomori, Y. Oku, and K. Jingu, “Error detection using a convolutional neural network with dose difference maps in patient-specific quality assurance for volumetric modulated arc therapy.”, *Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics (AIFB)*, vol. 73, pp. 57–64, May 2020. DOI: 10.1016/j.ejmp.2020.03.022.
- [49] M. Sakai et al., “Detecting MLC modeling errors using radiomics-based machine learning in patient-specific QA with an EPID for intensity-modulated radiation therapy.”, *Medical Physics*, vol. 48, no. 3, pp. 991–1002, Mar. 2021. DOI: 10.1002/mp.14699.
- [50] M. J. Nyflot, P. Thammasorn, L. S. Wootton, E. C. Ford, and W. A. Chaovalitwongse, “Deep learning for patient-specific quality assurance: Identifying errors in radiotherapy delivery by radiomic analysis of gamma images with convolutional neural networks.”, *Medical Physics*, vol. 46, no. 2, pp. 456–464, Feb. 2019. DOI: 10.1002/mp.13338.
- [51] K.-C. Chuang, W. Giles, and J. Adamson, “A tool for patient-specific prediction of delivery discrepancies in machine parameters using trajectory log files.”, *Medical Physics*, vol. 48, no. 3, pp. 978–990, Mar. 2021. DOI: 10.1002/mp.14670.
- [52] H. Hirashima et al., “Improvement of prediction and classification performance for gamma passing rate by using plan complexity and dosiomics features.”, *Radiotherapy and Oncology*, vol. 153, pp. 250–257, Dec. 2020, ISSN: 01678140. DOI: 10.1016/j.radonc.2020.07.031.
- [53] A. F. I. Osman and N. M. Maalej, “Applications of machine and deep learning to patient-specific IMRT/VMAT quality assurance.”, *Journal of applied clinical medical physics / American College of Medical Physics*, vol. 22, no. 9, pp. 20–36, Sep. 2021. DOI: 10.1002/acm2.13375.
- [54] Y. Interian et al., “Deep nets vs expert designed features in medical physics: An IMRT QA case study.”, *Medical Physics*, vol. 45, no. 6, pp. 2672–2680, Jun. 2018. DOI: 10.1002/mp.12890.
- [55] D. M. Fondevila, P. J. Rios, D. M. D. Peñalva, and S. Arbiser, “Predicting gamma passing rates for portal dosimetry-based IMRT QA using deep learning.”, *International Journal of Radiation Oncology, Biology, Physics*, vol. 111, no. 3S, e110–e111, Nov. 2021. DOI: 10.1016/j.ijrobp.2021.07.515.
- [56] I. J. Das et al., “Accelerator beam data commissioning equipment and procedures: Report of the TG-106 of the therapy physics committee of the AAPM.”, *Medical Physics*, vol. 35, no. 9, pp. 4186–4215, Sep. 2008. DOI: 10.1118/1.2969070.
- [57] S. Cilla et al., “Prediction and classification of VMAT dosimetric accuracy using plan complexity and log-files analysis.”, *Physica medica : PM : an international journal devoted to the applications of physics to medicine and biology : official journal of the Italian Association of Biomedical Physics (AIFB)*, vol. 103, pp. 76–88, Nov. 2022. DOI: 10.1016/j.ejmp.2022.10.004.
- [58] T.-F. Lee et al., “A fuzzy system for evaluating radiation treatment plans of head and neck cancer”, in *2012 9th International Conference on Fuzzy Systems and Knowledge Discovery*, 2012, pp. 510–514. DOI: 10.1109/FSKD.2012.6234043.
- [59] C. Pinter, T. Olding, L. J. Schreiner, and G. Fichtinger, “Using fuzzy logics to determine optimal oversampling factor for voxelizing 3D surfaces in radiation therapy”, *Soft computing*, vol. 24, no. 24, pp. 18 959–18 970, Dec. 2020, ISSN: 1432-7643. DOI: 10.1007/s00500-020-05126-w.
- [60] L. Sztandera, “Computational intelligence foundations”, in *Computational Intelligence in Business Analytics: Concepts, Methods, and Tools for Big Data Applications*, 1st ed., Upper Saddle River, New Jersey: Pearson, 2014, ISBN: 0-13-355208-X.
- [61] Y. Huang et al., “A feasibility study to predict 3D dose delivery accuracy for IMRT using DenseNet with log files.”, *Journal of X-ray science and technology*, vol. 32, no. 4, pp. 1199–1208, 2024. DOI: 10.3233/{XST}-230412.
- [62] P. Viola et al., “Prediction of VMAT delivery accuracy using plan modulation complexity score and log-files analysis.”, *Biomedical physics & engineering express*, vol. 8, no. 5, Aug. 2022. DOI: 10.1088/2057-1976/ac82c6.
- [63] T. Ono et al., “Applications of artificial intelligence for machine- and patient-specific quality assurance in radiation therapy: Current status and future directions.”, *Journal of radiation research*, vol. 65, no. 4, pp. 421–432, Jul. 2024. DOI: 10.1093/jrr/rrae033.
- [64] J. N. K. Carlson et al., “A machine learning approach to the accurate prediction of multi-leaf collimator positional errors.”, *Physics in Medicine and Biology*, vol. 61, no. 6, pp. 2514–2531, Mar. 2016. DOI: 10.1088/0031-9155/61/6/2514.
- [65] D. A. Granville, J. G. Sutherland, J. G. Belec, and D. J. La Russa, “Predicting VMAT patient-specific QA results using a support vector classifier trained on treatment plan characteristics and linac QC metrics.”, *Physics in Medicine and Biology*, vol. 64, no. 9, p. 095 017, Apr. 2019. DOI: 10.1088/1361-6560/ab142e.
- [66] C. Ma et al., “The structural similarity index for IMRT quality assurance: Radiomics-based error classification.”, *Medical Physics*, vol. 48, no. 1, pp. 80–93, Jan. 2021. DOI: 10.1002/mp.14559.
- [67] P. D. Wall and J. D. Fontenot, “Application and comparison of machine learning models for predicting quality assurance outcomes in radiation therapy treatment planning”, *Informatics in Medicine Unlocked*, vol. 18, p. 100 292, 2020, ISSN: 23529148. DOI: 10.1016/j.imu.2020.100292.
- [68] K. S. Lew et al., “Prediction of portal dosimetry quality assurance results using log files-derived errors and machine learning techniques.”, *Frontiers in oncology*, vol. 12, p. 1 096 838, 2022. DOI: 10.3389/fonc.2022.1096838.

TABLE I. SUMMARY OF RECENT STUDIES USING MACHINE LEARNING MODELS FOR IMRT/VMAT PSQA. (AUC= AREA UNDER THE CURVE, MAE= MEAN ABSOLUTE ERROR, RMSE= ROOT MEAN SQUARE ERROR, SR= SPEARMAN’S RANK CORRELATION COEFFICIENT)

Author/Year	Plan Type	Dataset Size	Anatomic Sites	Algorithm	QA Outcome	Feature Count	Key Results
Carlson et al. 2016 [64]	VMAT	74 plans (3,161,280 data points)	Multiple	RF	Error detection	6	RMSE= 0.193mm (linear regression)
Tomori et al. 2018 [41]	IMRT	60 plans	Prostate	CNN	Parameter prediction	N/A	Errors within 1.10% at 3%/3mm criteria
Interian et al. 2018 [54]	IMRT	498 plans	Multiple	CNN	Parameter prediction	N/A	MAE= 0.70% at 3%/3mm criteria
Lam et al. 2019 [43]	IMRT	1497 beams	Multiple	Tree-based	Parameter prediction	31	Errors within 3% for 98% of predictions at 2%/2mm criteria
Ono et al. 2019 [44]	VMAT	600 plans	Multiple	Regression Tree, ANN, Other	Parameter prediction	28	Mean prediction error= -0.2% at 3%/3mm criteria (ANN)
Granville et al. 2019 [65]	VMAT	1,620 beams	Multiple	SVM	Error detection	60	AUC=0.88 (macro-averaged)
Nyflot et al. 2019 [50]	IMRT	186 beams (558 images)	Multiple	SVM, Decision Tree, Other	Error detection	145	Accuracy= 64.3% for SVM
Ma et al. 2020 [66]	IMRT	180 beams (1,620 images)	Multiple	SVM, RF, Other	Error detection	276	AUC=0.86 for linear SVM
Osman et al. 2020 [18]	IMRT	10 plans (360,800 datapoints)	Multiple	ANN	Error detection	14	RMSE=0.0096mm
Wall and Fontenot 2020 [67]	VMAT	500 plans	Multiple	SVM, Tree-Based, ANN	Parameter prediction	241	MAE=3.75% at 3%/3mm criteria (SVM)
Hirashima et al. 2020 [52]	VMAT	1,255 plans	Multiple	Tree-based	Parameter prediction	875	MAE=4.2% and AUC=0.83 at 2%/2mm criteria
Wang et al. 2020 [46]	VMAT	276 Plans	Multiple	ANN	Parameter prediction	N/A	Absolute prediction error=1.76% at 3%/3mm criteria
Kimura et al. 2020 [48]	VMAT	161 Beams	Prostate	CNN	Error detection	54	Accuracy=0.94
Tomori et al. 2020 [42]	VMAT	147 plans	Multiple	CNN	Parameter prediction	N/A	MAE=0.63% at 3%/3mm criteria
Sakai et al. 2021 [49]	IMRT	38 beams (152 error plans)	Multiple	SVM, Tree-based, Other	Error detection	837	AUC=1.00 for leaf transmission factor error, 1.0 for dosimetric leaf gap error, 0.80 for leaf positional error vs. error free (SVM)
Chuang et al. 2021 [51]	IMRT/VMAT	267 IMRT and VMAT plans (10,584,120 data points)	Multiple	Tree-based, Other	Error detection	7	RMSE=0.0085 mm (Boosted Tree Model)
Huang et al. 2022 [45]	IMRT	112 plans	Chest	CNN	Parameter prediction	4	MAE and RMSE decreased with stricter gamma criteria, while SR and R^2 increased as gamma criteria were made stricter (3%/3mm, 3%/2mm, 2%/3mm, and 2%/2mm)
Cilla et al. 2022 [57]	VMAT	651 plans/1,302 arcs	Multiple	SVM, Other	Parameter prediction	3	Precision of 93.1 for gamma % and 92.7% for gamma mean for the testing dataset at 2%/2mm (SVM)
Lew et al. 2022 [68]	VMAT	578 log files	Multiple	RF, SVM, Other	Parameter prediction	13	Average error of less than 2% with 1%/1mm criteria.
Song et al. 2024 [47]	IMRT	204 plans/2,348 fields	Nasopharyngeal Carcinoma	CNN	Parameter prediction	1-8	AUC= 0.92 with 0.77 sensitivity and 0.89 specificity