Machine and Deep Learning for Patient-Specific Quality Assurance in Intensity Modulated Radiotherapy Using Log Files: Current Techniques and Emerging Directions

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*—This paper emphasizes emerging strategies in Patient-Specific Quality Assurance (PSQA) for Intensity Modulated Radiotherapy, with particular focus on the use of trajectory log files to enhance computational efficiency and clinical throughput. These log files passively record machine parameters throughout treatment, offering a compelling alternative to conventional phantom-based verification methods, which are resource-intensive and limited in their ability to capture patient-specific variability. Recent advancements have demonstrated the potential of algorithms such as Support Vector Machines, tree-based algorithms, and Artificial Neural Networks to improve the predictive accuracy and robustness of PSQA systems. While current best practices remain essential for ensuring baseline treatment safety, new models should meet additional demands. To maintain high standards of patient care, these models must be explainable, adaptable to evolving clinical workflows, and capable of continuous updates as treatment techniques advance. These attributes are key to enabling clinical integration and establishing a scalable, data-driven framework for personalized, real-time quality assurance in radiation oncology. They are the keystone in turning proof of concept into clinical reality.**

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

I. Introduction

This is an extended version of the review paper [1] published at AI Health 2025.

The American Cancer Society has estimated over 2 million new cases of cancer in 2024 [2]. About 50% of all cancer patients are expected to receive radiotherapy at some point during treatment [3]. 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 [4]. 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 [5]–[7]. 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 [5][8][9]. 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 healthcare [10]–[12]. 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 [13]. This is specifically advantageous

in fields, such as radiation therapy, where advancements are rapid, and techniques are constantly changing [14]–[17].

This 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 II provides the theoretical context for log-file based PSQA. Section III explores the various applications of machine learning and deep learning models for PSQA. Section IV contains the discussion. Section V details future directions and concludes with final remarks.

II. 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

As external beam radiation therapy has evolved, increasing complexity in both treatment planning and delivery has driven the need for more sophisticated quality assurance (QA) approaches. Intensity-modulated radiation therapy (IMRT), together with its rotational counterpart volumetric modulated arc therapy (VMAT), exemplifies this complexity. To understand how this progression impacts QA requirements, it is helpful to begin with three-dimensional conformal radiation therapy (3D-CRT), a foundational technique that remains in clinical use and provides a baseline for comparison with more advanced delivery methods.

3D-CRT uses static radiation beams delivered from multiple angles, shaped to match the target volume using manually designed beam apertures. These treatment plans are relatively simple, often allowing for manual dose calculations and straightforward verification. Due to their efficiency and reliability, 3D-CRT techniques are still employed in both curative and palliative settings, particularly when target volumes are geometrically uncomplicated and located away from critical structures.

In contrast, IMRT delivers beams with variable intensity across each field, enabling greater dose conformity to complex or concave target volumes. This is particularly beneficial for treating tumors located adjacent to radiosensitive organs. IMRT may be delivered using a fixed gantry through segmental (step-and-shoot) or dynamic (sliding window) techniques, and may also utilize compensator-based systems. VMAT further expands upon IMRT by delivering radiation in a continuous arc, dynamically modulating beam intensity, multileaf collimator (MLC) positions, and gantry speed throughout the rotation.

These treatments are administered using a computer-controlled linear accelerator (linac), which generates high-energy x-rays by accelerating electrons and directing them toward a metal target. The resulting radiation beam is shaped by the MLCs and guided according to complex instructions generated through inverse planning, typically based on computed tomography (CT) imaging [18].

While these advancements allow for highly conformal dose delivery and improved normal tissue sparing, they also introduce significant complexity. Unlike 3D-CRT, IMRT and VMAT plans are dynamic, with beam parameters changing continuously during treatment. This makes manual verification impractical and necessitates the use of advanced computational QA systems. Consequently, modern QA must now account for machine mechanics, patient positioning, beam geometry, dose rate modulation, and the real-time behavior of delivery components.

To ensure that each patient's plan can be delivered safely and accurately, patient-specific quality assurance (PSQA) is used. Given the individualized and time-varying nature of modulated treatments, PSQA plays a critical role in identifying errors and maintaining treatment fidelity in clinical settings [19][20]. 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 [21].

Currently, IMRT and VMAT treatment plans undergo physical measurements prior to delivery to confirm accurate dose output. These pre-treatment verifications are typically performed on a phantom and measured using devices such as ion chambers, diode arrays, film, or electronic portal imaging devices (EPIDs). However, these measurements are conducted in advance and may not capture real-time deviations during treatment. This introduces the risk of mechanical or dosimetric discrepancies between pre-treatment QA and actual delivery.

The gamma passing rate (GPR) is the most widely used metric for comparing planned and measured dose distributions. It is based on the gamma index, a composite metric introduced in 1998 that combines dose difference (DD) and spatial discrepancy-referred to as distance to agreement (DTA)—into a single score [22]. The GPR method computes the gamma index for each voxel and classifies it as pass or fail depending on whether the selected acceptance criteria are met [23].

Despite its routine use in clinical workflows, GPR has several well-documented limitations [24]–[26]. It is sensitive to dose grid resolution and often exhibits weak correlation with clinically significant dose errors. Furthermore, it has poor specificity and sensitivity in detecting subtle delivery inaccuracies that could affect patient outcomes. Although alternative metrics have been introduced over the past two decades, GPR remains the de facto clinical standard.

The most common PSQA workflow involves recalculating the dose distribution onto a phantom geometry. The treatment plan is then delivered and measured using QA devices. Differences between the measured and planned dose distributions are evaluated using gamma analysis, as outlined in American Association of Physicists in Medicine (AAPM) Task Group reports 119 and 218 [27][28]. These guidelines recommend that at least 90% of measured points meet the defined dose difference and DTA criteria, typically set at 3% and 2 mm, respectively. However, this process is resource-intensive and time-consuming, often requiring after-hours use of clinical equipment to avoid disrupting treatment schedules. Additionally,

the robustness of these methods and their ability to detect certain failure modes remains under debate [7][29][30].

Log file-based PSQA has emerged as a promising alternative to traditional measurement based verification. Rather than relying on physical detectors, it uses automatically generated machine log files to assess treatment delivery accuracy. These files capture key delivery variables, including radiation output, MLC positions, gantry and couch angles, beam status, and timing at fixed intervals (typically every 20 milliseconds) [31]. The recorded values represent the minimal data necessary to validate and troubleshoot the treatment process. Since the dataset is time resolved, it enables frame by frame reconstruction of the actual delivery, which can then be compared directly to the planned settings for error detection.

While promising, log file–based PSQA is not without limitations. Because log files are generated by the linac itself, they cannot detect hardware miscalibrations—such as incorrect MLC leaf positioning [8] - or software-related errors introduced during treatment planning. This may lead to discrepancies between recorded machine behavior and the actual delivered dose. In the case of MLC discrepencies, any difference above 1mm can lead to field edge misalignments that risk radiation exposure to nearby organs [32].To address these issues, enhanced QA protocols for the linac and more sensitive machine QA tools are recommended, particularly for verifying MLC performance [33]–[35].

The structure and resolution of these log files are critical for their integration into machine learning workflows. Variations in data format, parameter naming, and sampling frequency between vendors and machine models can significantly impact feature extraction and model performance. In recent studies, log files, treatment planning system (TPS) data, and modulation complexity scores (MCS) [34] have been used to develop machine learning models that predict GPR as a surrogate for plan deliverability.

Several early studies have evaluated the feasibility of log file–based PSQA and reported encouraging results. Most of this research has focused on specific disease sites such as head and neck, prostate, and lung cancer, often in small patient cohorts. In addition, the majority of published studies have used Varian linacs, with relatively limited evaluation of Elekta systems or other delivery platforms.

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 [36]. 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 [37].

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 [38]–[40].

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 [38].

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 [39]. XGBoost builds upon the gradient boosting algorithm by including L1 (Lasso) and L2 (Ridge) regularization to prevent overfitting [41][42]. 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 [40].

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 [43]. With the advent of backpropagation and activation functions -such as the Rectified Linear Unit (ReLU) [44]- 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 [45][46]. 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 [47]. CNNs' 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. CNNs' decision making can be extremely difficult to interpret [48].

Despite the demonstrated utility of SVMs, tree-based models, and CNNs in log file–based PSQA, a common limitation across these approaches is their lack of adaptability and scalability. Most models are trained on static datasets and evaluated under fixed conditions, which presents a challenge in clinical environments where treatment techniques, machine behavior, and planning protocols are continuously evolving.

Once trained, these models typically require complete retraining or manual fine-tuning to incorporate new data or adapt to changes in treatment delivery. This static approach limits their long-term clinical utility and increases the risk of model degradation in the face of equipment updates, workflow modifications, or shifts in patient population characteristics. Notably, few studies have systematically evaluated how quickly treatment plans evolve in practice or how these changes may impact machine learning model performance.

In addition, many models struggle to scale effectively across institutions or linear accelerator (linac) vendors due to differences in log file formatting, planning conventions, and QA workflows. These inconsistencies can significantly hinder model generalizability and limit cross-site implementation.

III. 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 of machine learning models in IMRT and VMAT PSQA fall into two main categories: parameter prediction and error detection studies (see Table I). Across the 20 studies summarized in Table I, several clear trends emerge. Tree-based models were the most commonly used machine learning approach, appearing in 50% of studies (10/20), followed by convolutional or artificial neural networks (CNNs/ANNs) in 45% (9/20), and support vector machines (SVMs) in 35% (7/20). Seven studies also explored other model types such as k-nearest neighbors or ensemble hybrids. The majority (12/20) employed a parameter prediction approach, while 8 focused on error detection. Gamma passing rate (GPR) remained the most frequent outcome metric, though some studies attempted direct dosimetric prediction or error classification. While both IMRT and VMAT were well represented—appearing in 10 and 11 studies, respectively—only 4 studies were limited to a single anatomical site, suggesting growing efforts to develop more generalizable models across varied treatment contexts. Nevertheless, most studies still relied on single-institution datasets, and few incorporated data from multiple vendors. These trends emphasize the need for multi-center collaborations and broader clinical diversity to support scalable, real-world PSQA tools.

Models were evaluated using both error-based and classification-based metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Spearman's Coefficient (SC), and Area Under the Curve (AUC), all of which are standard metrics for regression and classification performance. The choice of evaluation metric often depends on the model's output format and the specific goals of the study—whether continuous value prediction or binary classification. As a result, direct comparisons between studies can be challenging, particularly when different endpoints or performance criteria are reported. This variability underscores the need for standardized evaluation frameworks in future work to better assess and compare model effectiveness.

B. Drawbacks and Limitations

Tomori et al. [49], Lam et al. [50], Ono et al. [51], Huang et al. [52], Wang et al. [53], and Song et al. [54] 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 I. 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. [55], Sakai et al. [56], and Nyflot et al. [57] 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 [50]. 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 [55]. Sakai et al. included radiomic data which resulted in higher sensitivity and specificity for MLC position and MLC modeling errors [56]. Hirashima et al. utilized a combination of 3D dosiomic features and plan complexity in a tree-based model [58]. Tomori et al.'s GPR prediction-based CNN model struggled with overestimating low GPR values and underestimating GPR in the test set [49][59]. Song et al. developed a novel model that weighed the MSE loss function to mitigate this class imbalance with promising results [54]. 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 [27]. Although Varian machines are widely used in the US, Elekta machines are also used.

IV. 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 [60]. 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 [61][62]. Additionally, encompassing both Varian and Elekta machines is essential to ensure this PSQA strategy is accurate on both platforms [63].

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 [64]. The protocol leveraged plan complexity to designate plans as acceptable (green light), requires further verifcation (orange light), or unacceptable (red light) [64]. Fuzzy logic follows similar reasoning and has been successfully applied to radiation control systems and treatment plan optimization [65][66]. 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 precesssors, achieving better performance with fewer connections [67].

While fuzzy logic offers a way to model uncertainty in PSQA, it also highlights a broader need: the development of systems that can continuously adapt to changing clinical conditions. Future models must be not only accurate, but also adaptable, scalable, and self-evolving. Instead of relying on retraining static models each time conditions change, future systems should be capable of continuous learning and modular updates.

Incremental learning is one potential strategy where models are can be updated gradually as new data is introduced [68]. This avoids the need for complete retraining by using an adaptable models that can change in real-time. It is particularly well suited for large datasets that requires stability. However, these approaches are vulnerable to catastrophic forgetting where older knowledge can be lost when incorporating new data [69]. Techniques such as elastic weight consolidation can help address this issue by preserving important parameters, though they introduce new challenges in implementation and tuning. Other mitigation techniques include replay, template-based classification, and context dependent processing[70]–[72].

Fine-tuning offers another solution with further training on related datasets [73]. This could be particularly effective when adapting models to new machines, clinics, or delivery methods. It is less computationally demanding than full retraining, but fine-tuning must be handled carefully to avoid overfitting, especially in settings with limited or imbalanced data which are already inherent issues with log-file based PSQA models.

Genetic algorithms are another potential method for model evolution [74][75]. By using population-based search strategies, these algorithms can explore different architectures, hyperparameters, and feature selections over time. Genetic algorithms are computationally intensive which can make them difficult to implement. However, they are well suited for continuous optimization in non-critical processing environments and are not as prone to overfitting.

Figure 1 depicts a potential workflow that incorproates suggested improvments to prior models. These To address the limitations of prior log file-based IMRT PSQA modeling strategies, a revised workflow is proposed, as illustrated in Figure 1 Previous machine learning and deep learning models often relied on data from a single institution, which greatly limited model generalizability and increased the risk of overfitting due to site-specific bias. In the proposed framework, the input data is expanded to incorporate contributions from multiple institutions, enhancing the diversity of the training dataset and improving the model's capacity for generalizability. Additionally, the integration of a genetic algorithm with fuzzy logic is expected to improve feature selection and model optimization. The genetic

algorithm enables a more robust exploration of potential feature sets, while fuzzy logic facilitates flexible decision-making in response to variability both within and across institutions. Together, these modifications are intended to yield models with not only increased predictive accuracy but also significantly enhanced clinical applicability by allowing the system to evolve alongside future clinical shifts.

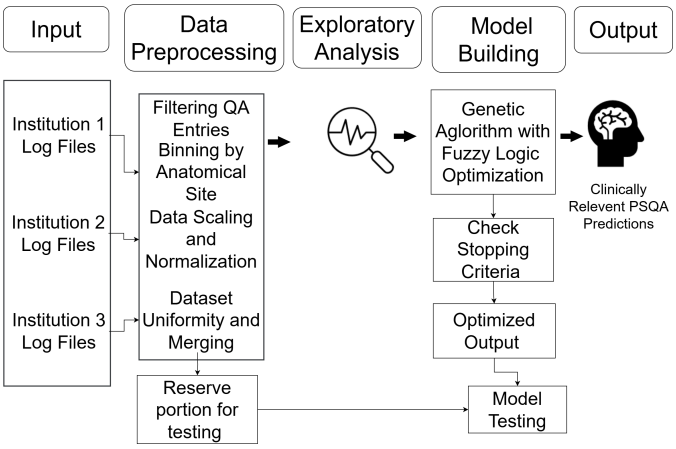

Figure 1. Suggested workflow for log file-based IMRT PSQA modeling. Enhancements include the incorporation of multi-institutional input data to improve model generalizability. The integration of a genetic algorithm with fuzzy logic is expected to enhance optimization, feature selection, and adaptability.

As a supplemental technique to the aforementioned models, federated learning can provide a network-preserving framework for scaling PSQA across multiple clinics [76]. Rather than sharing patient data, each institution trains a local model on its own dataset and transmits only model updates (e.g., gradients or weights) to a central server. The server then aggregates the updates into a global model and redistributes it. While federated learning effectively addresses data-sharing restrictions, its implementation presents challenges such as non-identically distributed data across sites, communication inefficiencies, and synchronization issues. Nevertheless, it remains a promising direction for building generalizable, robust QA models at scale.

V. Conclusions and Future Directions

Log file–based PSQA has emerged as a viable and efficient alternative to traditional phantom-based methods, particularly for IMRT and VMAT treatments. By leveraging machine learning models such as SVMs, tree-based algorithms, and CNNs, recent studies have demonstrated the ability to predict GPR outcomes and detect delivery errors directly from log file data. These methods offer a promising path toward scalable, real-time QA workflows that reduce clinical burden while maintaining—or even enhancing—treatment safety.

However, several challenges must be addressed before these tools can be widely implemented in clinical practice. Most current models rely on machine-reported parameters, limiting their ability to detect mechanical miscalibrations or TPS-related errors. In addition, the majority are trained on single-institution datasets, with limited anatomical and vendor diversity, which restricts their generalizability. These limitations highlight the need for multi-institutional datasets that reflect a broader spectrum of treatment techniques, patient populations, and machine types [54][64][77]–[79].

Future research must prioritize both generalizability and clinical adaptability. Multi-center collaborations that incorporate diverse planning protocols and hardware systems will be critical to developing robust, transferable models. Models must also be designed to remain effective in evolving clinical environments. Techniques such as incremental learning, transfer learning, and modular architectures can enable continuous model improvement without requiring full retraining. Federated learning also offers a promising privacy-preserving strategy for distributed model development across institutions.

Equally important, especially in the clinical setting, is the need for transparency and interpretability. Integrating explainable machine learning and deep learning tools can help clinicians understand how models generate predictions and identify which features contribute to error detection. This not only fosters trust and accountability, but also facilitates earlier, more targeted interventions.

Altogether, these advancements represent a shift toward faster, more efficient, and responsive QA systems. Future PSQA workflows should ideally evolve in step with technological innovation, while enhancing precision and safety in modern radiation oncology.

References

[1] K. De Jesus, L. Dunn, D. Thomas, and L. Sztandera, "Recent advances in machine learning for log file-based psqa for imrt and vmat", in *Proceedings of the 2025 Second International Conference on AI-Health*, Mar. 2025, ISBN: 978-1-68558-247-0.

[2] 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. Accessed: Nov. 10, 2024. [Online]. Available: http://dx.doi.org/10.3322/caac.21820.

[3] A. A. for Cancer Research, *Aacr Cancer Progress Report 2024*. AACR, 2024, ISBN: 979-8-9857852-7-2. Accessed: Nov. 10, 2024.

[4] R. J. Hutten et al., "Worsening racial disparities in utilization of intensity modulated radiation therapy", *Advances in Radiation Oncology*, vol. 7, no. 3, p. 100 887, Jan. 2022, ISSN: 24521094. DOI: 10.1016/j.adro.2021.100887. Accessed: Nov. 10, 2024. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2452109421002451.

[5] 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. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1088/0031-9155/59/9/N49.

[6] 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, Sep. 2014. DOI: 10.1120/jacmp.v15i6.4994. Accessed: Oct. 24, 2024. [Online]. Available: http://dx.doi.org/10.1120/jacmp.v15i6.4994.

[7] 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. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1120/jacmp.v16i1.5385.

[8] 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. Accessed: Oct. 24, 2024. [Online]. Available: http://dx.doi.org/10.1118/1.4949002.

[9] 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. Accessed: Oct. 24, 2024. [Online]. Available: http://dx.doi.org/10.1002/acm2.13660.

[10] 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. 101 822, Apr. 2020, ISSN: 09333657. DOI: 10.1016/j.artmed.2020.101822. Accessed: Jan. 22, 2025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0933365719310437.

[11] 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. Accessed: Oct. 20, 2021. [Online]. Available: http://www.nature.com/articles/s41591-018-0316-z.

[12] K. Rasheed et al., "Explainable, trustworthy, and ethical machine learning for healthcare: A survey", *Computers in Biology and Medicine*, vol. 149, p. 106 043, Oct. 2022. DOI: 10.1016/j.compbiomed.2022.106043. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/j.compbiomed.2022.106043.

[13] 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. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/S1470-2045(19)30149-4.

[14] 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. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.4143/crt.2021.706.

[15] 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. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/S0140-6736(21)00233-6.

[16] 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. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/j.ijrobp.2021.02.047.

[17] 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. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1088/1361-6560/ab5507.

[18] 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, Sep. 2008. DOI: 10.1118/1.3002305. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1118/1.3002305.

[19] 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. Accessed: Oct. 20, 2024. [Online]. Available: http://dx.doi.org/10.1002/mp.14014.

[20] 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, Sep. 2020. DOI: 10.1002/acm2.13051. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1002/acm2.13051.

[21] 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. Accessed: Jul. 1, 2020. [Online]. Available: http://dx.doi.org/10.1118/1.3190392.

[22] D. A. Low, W. B. Harms, S. Mutic, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions", *Medical Physics*, vol. 25, no. 5, pp. 656–661, May 1998. DOI: 10.1118/1.598248.

[23] D. A. Low and J. F. Dempsey, "Evaluation of the gamma dose distribution comparison method", *Medical Physics*, vol. 30, no. 9, pp. 2455–2464, Sep. 2003. DOI: 10.1118/1.1598711.

[24] B. E. Nelms, H. Zhen, and W. A. Tomé, "Per-beam, planar imrt qa passing rates do not predict clinically relevant patient dose errors", *Medical Physics*, vol. 38, no. 2, pp. 1037–1044, Feb. 2011. DOI: 10.1118/1.3544657.

[25] B. E. Nelms et al., "Evaluating imrt and vmat dose accuracy: Practical examples of failure to detect systematic errors when applying a commonly used metric and action levels", *Medical Physics*, vol. 40, no. 11, p. 111 722, Sep. 2013. DOI: 10.1118/1.4826166.

[26] T. Depuydt, A. Van Esch, and D. P. Huyskens, "A quantitative evaluation of imrt dose distributions: Refinement and clinical assessment of the gamma evaluation", *Radiotherapy and Oncology*, vol. 62, no. 3, pp. 309–319, Mar. 2002. DOI: 10.1016/s0167-8140(01)00497-2.

[27] 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, Sep. 2009. DOI: 10.1118/1.3238104. Accessed: Oct. 20, 2024. [Online]. Available: http://dx.doi.org/10.1118/1.3238104.

[28] 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. Accessed: Oct. 20, 2024. [Online]. Available: http://dx.doi.org/10.1002/mp.12810.

[29] 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. 070 601, Jul. 2013. DOI: 10.1118/1.4794929. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1118/1.4794929.

[30] 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. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1118/1.3512801.

[31] 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. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1120/jacmp.v3i2.2578.

[32] N. L. Childress, C. Bloch, R. A. White, M. Salehpour, and I. I. Rosen, "Detection of imrt delivery errors using a quantitative 2d dosimetric verification system", *Medical Physics*, vol. 32, no. 1, pp. 153–162, Jan. 2005. DOI: 10.1118/1.1829171.

[33] 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. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1016/j.prro.2011.04.008.

[34] 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. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1118/1.3276775.

[35] 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. 071 718, Jul. 2013. DOI: 10.1118/1.4810969. Accessed: Oct. 10, 2024. [Online]. Available: http://dx.doi.org/10.1118/1.4810969.

[36] B. Sun et al., "Initial experience with TrueBeam trajectory log files for radiation therapy delivery verification", *Practical Rdiation Oncology*, vol. 3, no. 4, e199–208, Dec. 2013. DOI: 10.1016/j.prro.2012.11.013. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/j.prro.2012.11.013.

[37] C. Cortes and V. Vapnik, "Support-vector networks", *Machine Learning*, vol. 20, no. 3, pp. 273–297, Sep. 1995. DOI: 10.1007/BF00994018.

[38] 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.

[39] 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. Accessed: Jul. 5, 2020. [Online]. Available: http://projecteuclid.org/euclid.aos/1013203451.

[40] 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. Accessed: Jan. 22, 2025. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2939672.2939785.

[41] 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. Accessed: Jun. 10, 2016. [Online]. Available: http://doi.wiley.com/10.1111/j.2517-6161.1996.tb02080.x.

[42] 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. Accessed: Jan. 22, 2025. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1080/00401706.1970.10488634.

[43] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity", *The Bulletin of Mathematical Biophysics*, vol. 5, pp. 115–133, 1943. DOI: 10.1007/BF02478259.

[44] K. Fukushima, "Visual feature extraction by a multilayered network of analog threshold elements", *IEEE Transactions on Systems Science and Cybernetics*, vol. 5, no. 4, pp. 322–333, 1969. DOI: 10.1109/TSSC.1969.300225.

[45] A. G. Ivakhnenko, "The group method of data handling—a rival of the method of stochastic approximation", *Soviet Automatic Control*, vol. 1, no. 3, pp. 43–55, 1968.

[46] J. Schmidhuber, "Deep learning in neural networks: An overview", *Neural Networks*, vol. 61, pp. 85–117, Jan. 2015. DOI: 10.1016/j.neunet.2014.09.003.

[47] 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.

[48] 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, 2022, ISSN: 0020-0255. DOI: https://doi.org/10.1016/j.ins.2022.10.013. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S002002552201132X.

[49] 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. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.13112.

[50] 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. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.13752.

[51] 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. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.13669.

[52] 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. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/10.1177/15330338221104881.

[53] 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, Sep. 2020. DOI: 10.1088/1361-6560/abb31c. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1088/1361-6560/abb31c.

[54] 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. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/10.1186/s13014-024-02496-5.

[55] 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. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1016/j.ejmp.2020.03.022.

[56] 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. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.14699.

[57] 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. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.13338.

[58] 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. Accessed: Jan. 17, 2025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0167814020306769.

[59] 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. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.14682.

[60] 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. Accessed: Sep. 23, 2024. [Online]. Available: http://dx.doi.org/10.1002/acm2.13375.

[61] 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. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.12890.

[62] 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, Sep. 2021. DOI: 10.1016/j.ijrobp.2021.07.515. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1016/j.ijrobp.2021.07.515.

[63] 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. Accessed: Jan. 22, 2025. [Online]. Available: http://dx.doi.org/10.1118/1.2969070.

[64] 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, Sep. 2022. DOI: 10.1016/j.ejmp.2022.10.004. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/10.1016/j.ejmp.2022.10.004.

[65] 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.

[66] 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. Accessed: Jan. 22, 2025. [Online]. Available: https://link.springer.com/10.1007/s00500-020-05126-w.

[67] 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. Accessed: Jan. 18, 2025.

[68] G. M. van de Ven and A. S. Tolias, "Three types of incremental learning", *Nature Machine Intelligence*, vol. 4, no. 12, pp. 1185–1197, Dec. 2022. DOI: 10.1038/s42256-022-00568-3.

[69] J. Kirkpatrick et al., "Overcoming catastrophic forgetting in neural networks", *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 13, pp. 3521–3526, Mar. 2017. DOI: 10.1073/pnas.1611835114.

[70] H. Shin, J. K. Lee, J. Kim, and J. Kim, "Continual learning with deep generative replay", in *Advances in Neural Information Processing Systems*, I. Guyon et al., Eds., vol. 30, Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/0efbe98067c6c73dba1250d2beaa81f9-Paper.pdf.

[71] S.-A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "Icarl: Incremental classifier and representation learning", in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5533–5542. DOI: 10.1109/CVPR.2017.587.

[72] N. Y. Masse, G. D. Grant, and D. J. Freedman, "Alleviating catastrophic forgetting using context-dependent gating and synaptic stabilization", *Proceedings of the National Academy of Sciences*, vol. 115, no. 44, E10467–E10475, 2018. DOI: 10.1073/pnas.1803839115. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1803839115. [Online]. Available: https://www.pnas.org/doi/abs/10.1073/pnas.1803839115.

[73] S. J. Pan and Q. Yang, "A survey on transfer learning", *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010. DOI: 10.1109/TKDE.2009.191.

[74] "Genetic algorithms and machine learning", *Machine Learning*, vol. 3, no. 2–3, pp. 95–99, Oct. 1988. DOI: 10.1007/BF00113892.

[75] S. Katoch, S. S. Chauhan, and V. Kumar, "A review on genetic algorithm: Past, present, and future", *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 8091–8126, 2021. DOI: 10.1007/s11042-020-10139-6.

[76] P. Kairouz et al., "Advances and open problems in federated learning", *Found. Trends Mach. Learn.*, vol. 14, no. 1–2, pp. 1–210, Jun. 2021, ISSN: 1935-8237. DOI: 10.1561/2200000083. [Online]. Available: https://doi.org/10.1561/2200000083.

[77] 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. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/10.1088/2057-1976/ac82c6.

[78] 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. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/10.3233/%7BXST%7D-230412.

[79] 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. Accessed: Sep. 23, 2024. [Online]. Available: http://dx.doi.org/10.1093/jrr/rrae033.

[80] 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. Accessed:

Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1088/0031-9155/61/6/2514.

[81] 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. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1088/1361-6560/ab142e.

[82] 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. Accessed: Jan. 17, 2025. [Online]. Available: http://dx.doi.org/10.1002/mp.14559.

[83] 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. Accessed: Jan. 17, 2025. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2352914819303661.

[84] 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. Accessed: Oct. 20, 2024. [Online]. Available: http://dx.doi.org/10.1002/mp.14670.

[85] 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. Accessed: Nov. 3, 2024. [Online]. Available: http://dx.doi.org/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 [80]	VMAT	74 plans (3,161,280 data points)	Multiple	RF	Error detection	6	RMSE= 0.193mm (linear regression)
Tomori et al. 2018 [49]	IMRT	60 plans	Prostate	CNN	Parameter prediction	N/A	Errors within 1.10% at 3%/3mm criteria
Interian et al. 2018 [61]	IMRT	498 plans	Multiple	CNN	Parameter prediction	N/A	MAE= 0.70% at 3%/3mm criteria
Lam et al. 2019 [50]	IMRT	1497 beams	Multiple	Tree-based	Parameter prediction	31	Errors within 3% for 98% of predictions at 2%/2mm criteria
Ono et al. 2019 [51]	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 [81]	VMAT	1,620 beams	Multiple	SVM	Error detection	60	AUC=0.88 (macro-averaged)
Nyflot et al. 2019 [57]	IMRT	186 beams (558 images)	Multiple	SVM, Decision Tree, Other	Error detection	145	Accuracy= 64.3% for SVM
Ma et al. 2020 [82]	IMRT	180 beams (1,620 images)	Multiple	SVM, RF, Other	Error detection	276	AUC=0.86 for linear SVM
Osman et al. 2020 [19]	IMRT	10 plans (360,800 datapoints)	Multiple	ANN	Error detection	14	RMSE=0.0096mm
Wall and Fontenot 2020 [83]	VMAT	500 plans	Multiple	SVM, Tree-Based, ANN	Parameter prediction	241	MAE=3.75% at 3%/3mm criteria (SVM)
Hirashima et al. 2020 [58]	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 [53]	VMAT	276 Plans	Multiple	ANN	Parameter prediction	N/A	Absolute prediction error=1.76% at 3%/3mm criteria
Kimura et al. 2020 [55]	VMAT	161 Beams	Prostate	CNN	Error detection	54	Accuracy=0.94
Tomori et al. 2020 [59]	VMAT	147 plans	Multiple	CNN	Parameter prediction	N/A	MAE=0.63% at 3%/3mm criteria
Sakai et al. 2021 [56]	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 [84]	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 [52]	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 [64]	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 [85]	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 [54]	IMRT	204 plans/2,348 fields	Nasopharyngeal Carcinoma	CNN	Parameter prediction	1-8	AUC= 0.92 with 0.77 sensitivity and 0.89 specificity