



**HAL**  
open science

# Machine learning and deep learning prediction of patient specific quality assurance in breast IMRT radiotherapy plans using Halcyon specific complexity indices

Christine Boutry, Noémie Moreau, Cyril Jaudet, Laetitia Lechippey, Aurélien Corroyer-Dulmont

## ► To cite this version:

Christine Boutry, Noémie Moreau, Cyril Jaudet, Laetitia Lechippey, Aurélien Corroyer-Dulmont. Machine learning and deep learning prediction of patient specific quality assurance in breast IMRT radiotherapy plans using Halcyon specific complexity indices. *Radiotherapy & Oncology*, 2024, 200, pp.110483. 10.1016/j.radonc.2024.110483. hal-04676118

**HAL Id: hal-04676118**

**<https://normandie-univ.hal.science/hal-04676118v1>**

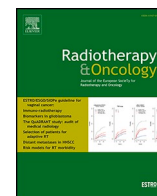
Submitted on 23 Aug 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License



## Original Article

# Machine learning and deep learning prediction of patient specific quality assurance in breast IMRT radiotherapy plans using Halcyon specific complexity indices

Christine Boutry<sup>a</sup>, Noémie N Moreau<sup>a,b</sup>, Cyril Jaudet<sup>a</sup>, Laetitia Lechippey<sup>a</sup>, Aurélien Corroyer-Dulmont<sup>a,b,\*</sup>

<sup>a</sup> Medical Physics Department, Centre François Baclesse, 14000 Caen, France

<sup>b</sup> Université de Caen Normandie, CNRS, Normandie Université, ISTCT UMR6030, GIP CYCERON, F-14000 Caen, France



## ARTICLE INFO

## Keywords:

Dual-layered MLC  
Machine Learning  
Deep Learning  
Radiotherapy  
Halcyon  
Quality assurance

## ABSTRACT

**Introduction:** New radiotherapy machines such as Halcyon are capable of delivering dose-rate of 600 monitor-units per minute, allowing large numbers of patients treated per day. However, patient-specific quality assurance (QA) is still required, which dramatically decrease machine availability. Innovative artificial intelligence (AI) algorithms could predict QA result based on complexity metrics. However, no AI solution exists for Halcyon machines and the complexity metrics to be used have not been definitively determined. The aim of this study was to develop an AI solution capable of firstly determining the complexity indices to be obtained and secondly predicting patient-specific QA in a routine clinical setting.

**Methods:** Three hundred and eighteen beams from 56 patients with breast cancer were used. The seven complexity indices named Modulation-Complexity-Score (MCS), Small-Aperture-Score (SAS10), Beam-Area (BA), Beam-Irregularity (BI), Beam-Modulation (BM), Gantry and Collimator angles were used as input to the AI model. Machine learning (ML) and deep learning (DL) models using tensorflow were set up to predict DreamDose QA conformance.

**Results:** MCS, BI, gantry and collimator angle are not correlated with QA compliance. Therefore, ML and DL models were trained using SAS10, BA and BM complexity indices. ROC analyses enabled to find best predicted probability threshold to increase specificity and sensitivity. ML models did not show satisfactory performance with an area under-the-curve (AUC) of 0.75 and specificity and sensitivity of 0.88 and 0.86. However, optimised DL model showed better performance with an AUC of 0.95 and specificity and sensitivity of 0.98 and 0.97.

**Conclusion:** The DL model demonstrated a high degree of accuracy in its predictions of the quality assurance (QA) results. Our online predictive QA-platform offers significant time savings in terms of accelerator occupancy and working time.

## Introduction

Over the last decade, various technological advancements have led to improved radiotherapy techniques and methods. In the case of breast cancer radiotherapy, Intensity-Modulated-Radiation-Therapy (IMRT) and Volumetric-Modulated-Arc-Therapy (VMAT) were common treatment techniques. IMRT has significantly improved target coverage while sparing normal tissue compared to 3D conformal radiotherapy. The IMRT treatment plan is very complex due to numerous variables including the position of the multi-leaf-collimator (MLC), speeds, gantry

rotation and beam stability [1]. As treatment plans become more complex, the challenge is to take into account patient anatomy, dosimetry constraints, optimization algorithms, and machine capabilities to ensure that the treatment delivered is consistent with the parameters originally planned in the treatment-planning-system (TPS). The introduction of Halcyon machines to the radiotherapy market has provided robust access to advanced IMRT techniques for patient treatment. It is necessary to conduct patient-specific quality assurance (QA) for each treatment plan before the administering of actual patient treatment takes place. However, although these machines allow for technically sophisticated

\* Corresponding author at: Medical Physics Department, CLCC François Baclesse, 14000 Caen, France.

E-mail address: [a.corroyer-dulmont@baclesse.unicancer.fr](mailto:a.corroyer-dulmont@baclesse.unicancer.fr) (A. Corroyer-Dulmont).

<https://doi.org/10.1016/j.radonc.2024.110483>

Received 15 April 2024; Received in revised form 5 July 2024; Accepted 14 August 2024

Available online 17 August 2024

0167-8140/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

treatments to be delivered quickly, the mandatory pre-treatment checks can quickly become a constraint in terms of machine downtime [2]. Nowadays, research groups are using artificial intelligence (AI) to delineate tumours and automate dosimetry. This will allow to reduce time before treatment of the disease with less anatomical change, less tumor progression, and fewer repeated treatment plans. The prediction of the QA is a very important aspect in the revision of the treatment plans. We previously published [3] an AI solution aiming to predict QA of VMAT treatment using complexity indices. However, due to the novelty of Halcyon machine, new type of complexity indices must be developed to base AI model on them. Tamura and colleagues [4] have proposed new complexity indices but, to the best of our knowledge, no AI model using these indices have been developed.

Furthermore, in our center as others, for Halcyon's QA, zero percent of the treatment plans are non-compliant. It is therefore crucial to implement a patient quality control strategy that is adapted to these new machines in order to validate the treatment process.

Nowadays, the clinically available solution for patient QA is the one offered by Varian with its Portal Dosimetry option. An MV image is acquired as each beam is delivered and compared to a predicted image calculated by the TPS. While this integrated solution may seem very attractive in terms of efficiency, solutions based on dose prediction are not considered to be fully reliable. For this study, pre-processing checks using a new DreamDose application based on a patented technology to convert the acquired image into a planar dose in water under the reference calibration conditions of the machine (SSD and depth) have been developed [5 6]. In this study, DreamDose measurement was considered as ground truth as it was shown to be more sensitive than portal dosimetry in the literature [7,8].

The objectives of this study were firstly to evaluate the pertinence of new complexity indices in a Halcyon patient-specific QA prediction purpose and secondly to develop AI solutions based on these selected complexity indices able to predict patient-specific QA conformance in Halcyon machine.

## Materials and methods

### Patients

This retrospective study was approved by the local institutional review board. Fifty-six patients with breast cancer referred to our oncology centre between June-2023 and September-2023 were included. The Declaration of Helsinki, which defines guidelines for the conduct of health research, was used for this study. The characteristics of the study population are described in [Supplementary Table 1](#).

### Radiotherapy plans

IMRT was delivered using a Halcyon (Varian®) machine. Patient received breast alone treatment. The controlled beams are derived from treatment plans based mainly on four IMRT – sliding window beams with incidences close to the tangential ballistics defined in conformal radiotherapy. Optimisation in the TPS is carried out on several sets of CT images simulating movements of the patient's external anatomy, and the IMRT technique allows a default margin of 5 mm to be applied to all the MLC leaves when the final dose distribution is calculated. IMRT plans were calculated using Raystation®-TPS. Each IMRT plan contained multiple beams, resulting in a total of 318 beams used to calculate the prediction model. Complexity indices were obtained from the RT plans using Raystation® TPS and an in-house Python code available at: [https://github.com/AurelienCD/DeepLearning\\_Patient-HalcyonQA\\_Prediction](https://github.com/AurelienCD/DeepLearning_Patient-HalcyonQA_Prediction) [9] and “*Get\_MCSV\_index.py*”, “*Get\_SAS10\_index.py*” or “*Get\_PA\_PI\_PM\_index.py*”. In order to achieve a simple solution applicable to routine clinical settings, only seven complexity indices were used: Modulation-complexity-score (MCS), Small-aperture-score (SAS10) and Beam-Area (BA), Beam-Irregularity (BI) and Beam-Modulation (BM),

which are the averaged field area, the averaged degree of difference from the circular field and the averaged proportion of the field area to the union area at the beam level. These complexity indices were obtained from the application of Tamura and colleagues founding [4]. Finally, gantry and collimator angles were also used in this study. Complexity indices were extracted from each beam and not from each plan.

### Patient specific quality assurance

In this study, 2D controls were performed on the basis of integrated images automatically acquired by the EPID (Varian-AS1200). These images were acquired for each irradiation beam of the treatment plans and were exported in DICOM format. The DreamDose application uses a patented model to convert the grey level into a calculated dose to water under the machine's reference conditions.

The application then compares this measured planar dose to the calculated planar dose exported from the TPS using a gamma index analysis. In the literature, thresholds of gamma index metrics with dose-difference (%) and distance-to-agreement (mm) are used to establish QA compliance. However, as mentioned by the Task-Group-218 of the American Association of Physicist in Medicine, even if 3%/3mm is most commonly used, there is no real consensus on the values of the thresholds for each variable to be used [10]. For this reason, in this study, we choose to develop an AI model capable of predicting QA conformity in several cases of dose-difference/distance-to-agreement: 2.5%/2.5 mm, 3%/3mm, 3%/2mm, 2%/3mm, 2%/2.5 mm. As presented in the results section, firstly, ML and DL models were developed to predict QA conformity of the balanced 2%/2.5 mm class. Specific class of 2%/2.5 mm because of the balanced proportion of this class and because one of the aims of the study was to determine the minimum gamma-index criterion enabling more than 95 % of favourable points to be obtained. With this in mind, we chose 2%/2.5 mm as the minimum value for this criterion, as a criterion of 2%/2mm did not allow us to differentiate the DoseDream methodology from the Portal Dosimetry methodology. Subsequently, a second DL model was developed to provide the probabilities of QA conformity for the 2.5%/2.5 mm, 3%/3mm, 3%/2mm, 2%/3mm and 2%/2.5 mm classes in a single instance (called “Multiclass prediction” in the results section). Finally, QA was defined as non-compliant if the local gamma index is below than 0.95 %. Proportion of compliant and non-compliant QA for each dose-difference/distance-to-agreement was details in the [Supplementary Table 2](#).

### Artificial intelligence algorithms

#### Machine learning (ML)

The input data of the ML models were seven quantitative items obtained from the RT beam representing the complexity of the radiotherapy treatment. SAS10, MCSV, BA, BI, BM, and gantry and collimator angles (Gantry- $\alpha$  and Colli- $\alpha$ , respectively). The goal of the ML models was to predict the two classes of compliant and non-compliant QA (0 and 1, respectively) from these seven complexity indices as inputs. The models were developed with hyperparameter optimization and cross-validation using the Sklearn Python library [11]. Training and testing were allocated proportionally to 80 % and 20 %, respectively. Several ML models were tested and compared by evaluating their prediction scores: Linear Discriminant, Ridge, KNeighbors, Gaussian-NB, Decision-Tree, Support-Vector-Classifer, Stochastic-Gradient-Descent and Random-Forest-Classifer. ML model hyperparameters were tuned to obtain higher validation score and area under-the-curve (AUC). After model optimization, receiver-operating-characteristic (ROC) curves and confusion matrices were computed for the best models. This was done to find the optimal probability threshold to minimize false positives. All ML codes are available at [https://github.com/AurelienCD/DeepLearning\\_Patient-HalcyonQA\\_Prediction](https://github.com/AurelienCD/DeepLearning_Patient-HalcyonQA_Prediction) and “*Machine\_Learning.ipynb*”.

**Deep learning (DL)**

Deep learning (DL) methods enable more complex algorithms that can improve performance in certain cases [12]. For this reason, we also develop DL algorithms that use the same data input and output as the ML model. The DL model consists of a multilayer-deep-learning (artificial-neural-network) model to obtain probabilities of compliant and non-compliant QA. The model is computed using Tensorflow/Keras library [13]. Optimization of hyperparameters such as activation-function, number of layers, number of neurons in each layer, loss-function, regularizer-function, learning-rate, metrics, and percentage of dropout was performed using KerasTuner.

This optimization resulted in three dense layers of 2048, 4096, and 512 neurons, respectively, and an output layer of five neurons (for the five dose difference/distance classes) with the Adam optimizer at a learning rate of 0.001. Activation functions were relu, elu, and gelu for the three dense layers. The sigmoid final function was used to obtain the prediction of QA compliance for the five classes. The model was built with a binary cross-entropy loss function and validation binary accuracy optimization metrics on 4000 epochs.

The final version of the DL model predicting the five classes of dose difference/distance to agreement is shown in Fig. 1.

As with the majority of medical datasets, our data were imbalanced for certain classes of dose difference/distance to agreement (mainly classes 2%/3mm and 3%/3mm with almost all cases in agreement). To avoid the negative impact of this characteristic, Imblearn-Python-library and Keras-class-weighting were used [14]. In addition, the weights of the model were adjusted with respect to the proportion of classes to handle the unbalanced data. Performance of the DL model was evaluated considering the accuracy score of the test dataset, the AUC of the ROC curves, specificity, sensitivity and finally the number of false positives. ML and DL models have been trained using A6000 GPU obtained with the "Booster IA" grant of the Région Normandie.

DL codes are available at [https://github.com/AurelienCD/DeepLearning\\_Patient-HalcyonQA\\_Prediction](https://github.com/AurelienCD/DeepLearning_Patient-HalcyonQA_Prediction) and "Deep\_Learning.ipynb".

**Statistical analyses**

All data are expressed as mean ± SD. Student's *t*-test was used to compare the different quantitative metrics in the 2%/2.5 mm class

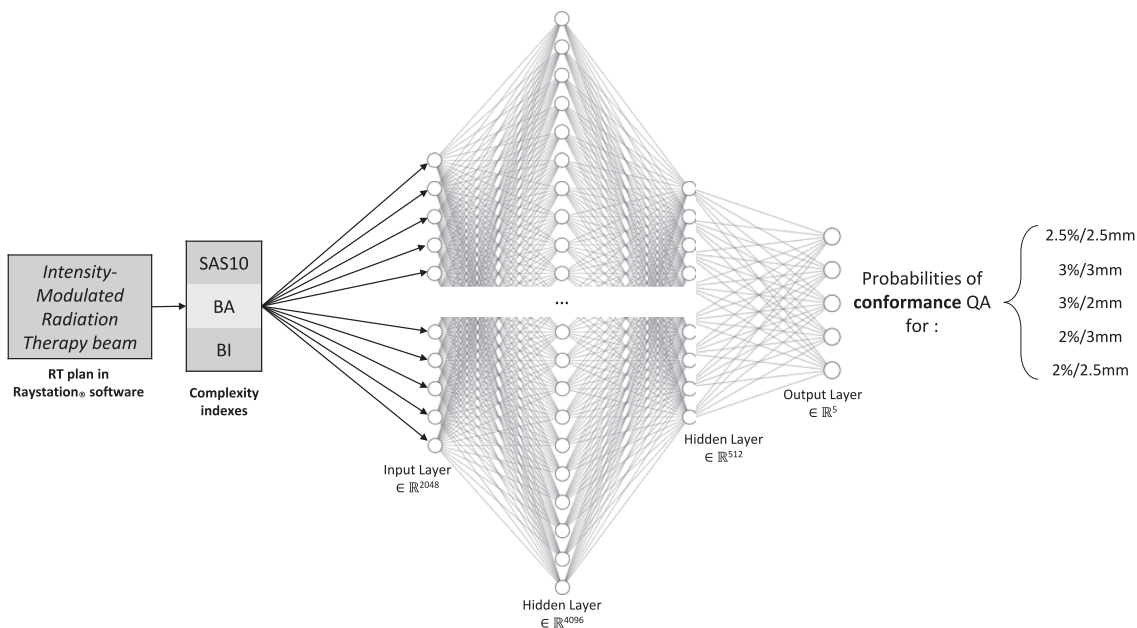
which was a balanced class allowing to performed *t*-test analysis. All statistical analyses were performed using SciPy-Python-library [15]. Data visualization used Seaborn library [16,17]. All Python code used in the analysis is available at [https://github.com/AurelienCD/DeepLearning\\_Patient-HalcyonQA\\_Prediction](https://github.com/AurelienCD/DeepLearning_Patient-HalcyonQA_Prediction) and "Statistical\_Analysis.ipynb".

**Results**

The validity of SAS10, MCSV, BA, BI, BM, and gantry and collimator angles (Gantry- $\alpha$  and Colli- $\alpha$ , respectively) for predicting QA compliance was first evaluated. As shown in Fig. 2, only the SAS10, BA and Bi variables showed significant differences between the compliant and non-compliant groups ( $p < 0.001$ ,  $p < 0.001$  and  $p < 0.001$  for SAS10, BA and BI, respectively and  $p = 0.07$ ,  $p = 0.26$ ,  $0.48$  and  $p = 0.90$  for MCSV, BM, Gantry- $\alpha$  and Colli- $\alpha$ , respectively). The other complexity indices did not appear to be correlated with the conformance status of the beam, so we decided to develop ML and DL models using only the SAS10, BA and BI complexity indices as input data.

We first developed an ML model capable of predicting only one class of dose difference/distance (2%/2.5 mm). As shown in Fig. 3, the performance of the different models is around 0.7 for training and 0.6 for validation. The Random-Forest-Classifer (RFC) seems to be the most promising with a training score of 0.94 and a validation score of 0.72. The RFC model gave a percentage of predictions of conformity QA between zero and one. The optimal threshold to apply to this percentage to allow higher true positive (TP) and true negative (TN), and lower false positive (FP) and false negative (FN) was determined using the ROC curve as shown in Fig. 3C. The performance of the RFC model resulted in the confusion matrix shown in Fig. 3D. Among the 318 beams, 120 and 155 were correctly predicted as TP and TN, respectively (Fig. 3D). However, 20 and 23 QA results were incorrectly identified as compliant and non-compliant QA, respectively which resulted in an AUC of 0.87. The sensitivity and specificity of the ML model were 0.84 and 0.89, respectively. These results were not sufficient enough, so a more complex AI model was developed using deep learning approaches.

As for the ML model, the DL model was first developed for the 2%/2.5 mm class. Better results were observed compared to the ML model with a validation binary accuracy of 0.86. As shown in Fig. 4A, the optimal probabilistic threshold was determined using the ROC curve,



**Fig. 1.** Deep Learning model for Halcyon QA prediction. Complexity indices were obtained from RT plan and used as input of a deep learning model which predict five probabilities of conformance QA for different dose-difference (%) and distance-to-agreement (mm) cases.

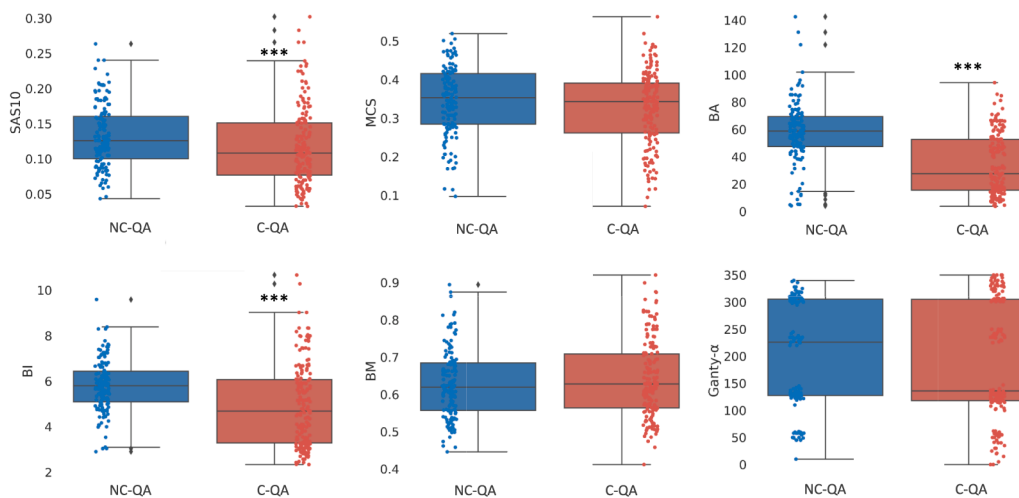


Fig. 2. Complexity indices differences in conformance (C-QA) and non-conformance QC (NC-QA) in the 2%/2.5 mm class. \*\*\*p < 0.001.

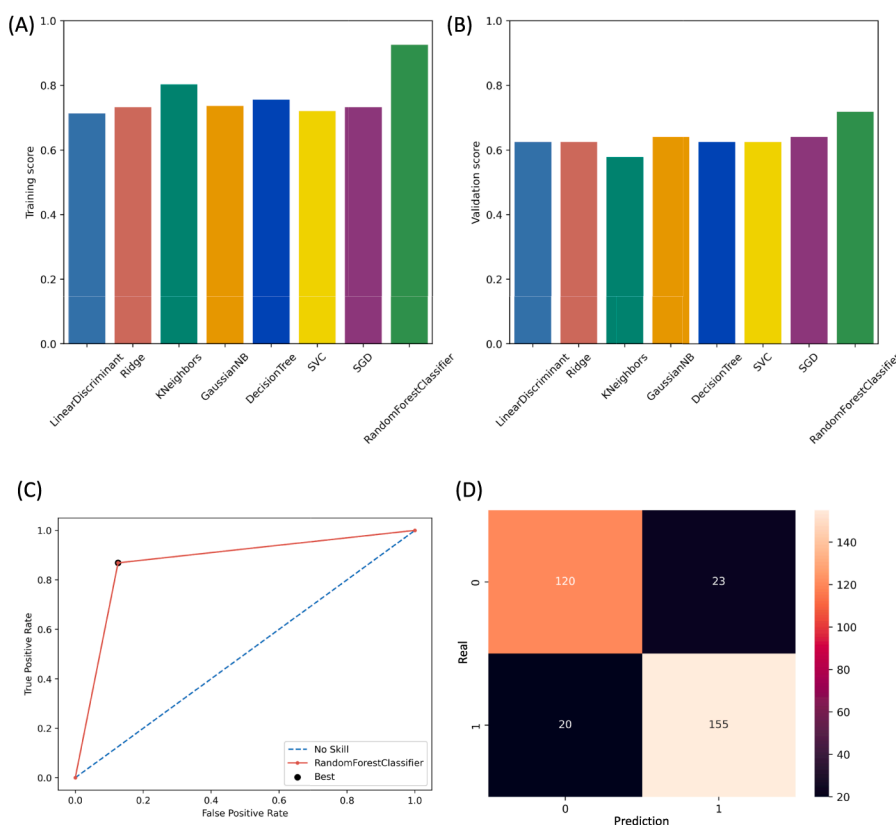


Fig. 3. ML models performance for single QA conformity class prediction. Several ML algorithms were evaluate and the training score and validation score are report in (A) and (B). (C) and (D) show the ROC curve and confusion matrix of the best ML algorithms (RandomForestClassifier) to predict single QA conformity class.

which resulted in an AUC of 0.97. Using this threshold, the number of FN and FP was low with only 18 among the 318 beams (TP and TN were 134 and 166, respectively, Fig. 4B). Sensitivity and specificity of the unique class DL model were 0.94 and 0.95, respectively. In the literature, there is no consensus for dose difference/distance-to-agreement class that can be used. For this reason, we decided to develop a multiclass DL prediction model that is capable of predicting five different classes of dose difference/distance-to-agreement.

The multiclass DL model took as input data the three best complexity

indices (SAS10, BA and Bi) and gave as output data five values between zero and one as percentage of QA agreement for the five different dose difference/distance to agreement classes. The multiclass DL model gave an interesting performance with a validation binary accuracy of 0.84. As for the previous models, the ROC curve gave the optimal threshold to apply to the different classes (Fig. 4C). The predictive performance of the final model is shown in the confusion matrix in Fig. 4D. Of the 318 beams in the five different classes, only 26 and 31 QAs were misclassified, but 335 and 1198 control results were correctly predicted.



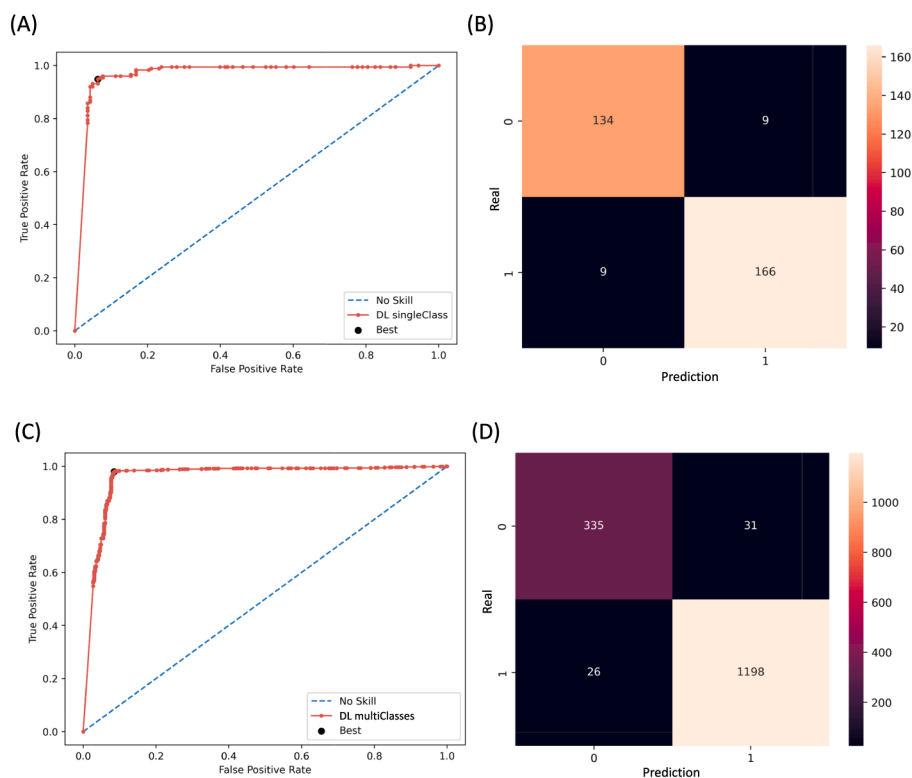


Fig. 4. ROC curve and Confusion matrix show the accuracy of the DL models with low number of false positive and negative to predict single QA conformity class (A and B) and multiclass QA conformity class (C and D).

These results presented a sensitivity and specificity of 0.92 and 0.98, respectively, which are similar to the DL single class model, but with the direct prediction of five different dose difference/distance-to-agreement classes.

As shown in Table 1, the DL models show better results than the ML model and allow the prediction of five different dose difference/distance to agreement classes.

One important requirement of this study was to develop a simple approach that could be easily implemented in clinical practice. To achieve this, an application programming interface (API) was developed that can be used in a clinical setting. The interface requires the input of three selected complexity indices (SAS10, BA, BM) and then the DL model predicts patient-specific QA compliance. The API can be freely accessed at <https://aureliencd-halcyonqualitycontrolapi-ba47xul4uka-xjrw4vxp5s.streamlit.app>. On average, it takes less than two minutes to predict conformance with the platform per beam. By performing the QA measuring only for the plans with a non-conformance prediction, we can save approximately 65 min per week compared to the usual 210 min required for patient-specific QA for Halcyon. This allows us to treat an additional four patients. Please refer to Supplementary Fig. 1 for the complete workflow we have proposed for implementing this solution in our center.

**Discussion**

Patient-specific QA is mandatory in RT, but it can be time-consuming

and affect the availability of RT machines to treat patients [18]. To address this issue, it has been proposed to use models that use RT plan complexity indices as input to predict QA outcomes while ensuring treatment safety [19]. In this context, we previously proposed an AI solution able to predict QA outcomes for VMAT single-layered MLC [3]. However, even if single-layered MLC still represent the majority of RT machine, new dual-layered MLC linac are not proposed and complexity indices as well as AI prediction models are not valid in that context. Tamura and colleagues [4], have proposed new complexity indices adapted to dual-layered MLC linac and shown that eight indices (MSC, total number of monitor-units (MU), averaged-field-area (PA), average-degree-of-difference-from-the-circular (PI), average-proportion-of-the-field-area-against-the-union-area (PM), leaf-sequence-variability (LSV), SAS, effective-distal-MLC-score (EDS)) are able to characterize the complexity of Halcyon (a dual-layered MLC linac) treatment plans. However, based on our analysis of the relevance of the compliancy indices for QA prediction purpose, only three variables (SAS10, BA and BI) were interesting with significant differences between the compliant and non-compliant groups.

A recent study by Zhu and colleagues [20] proposed ML model for QA outcome prediction in dual-layered MLC linac using 213 IMRT treatment-plans, including 1383 beams from head and neck, chest, abdominal and pelvic cancers. Criteria of 1 %/1mm, 2 %/2mm and 3 %/2mm with a 10 % threshold were used for gamma analysis. As input data for the training models, 33 complexity indices were extracted for each beam. The gradient-boosting-decision-tree was the model with the

**Table 1**  
Machine learning and deep learning performances.

	AUC	Sensitivity	Specificity	Accuracy	TP	TN	FP	FN
ML single class (2.5 %/2.5 mm)	0.87	0.84	0.89	0.86	120	155	20	23
DL single class (2.5 %/2.5 mm)	0.97	0.94	0.95	0.94	134	166	9	9
DL multiclass	0.96	0.92	0.98	0.96	335	1198	26	31

best performance. In validation set, the minimal mean prediction error was 1.93 % for 1 %/mm, 1.16 % for 2 %/2mm and 0.78 % for 3 %/3mm. In our study, ML model resulted in an AUC of 0.87 and sensitivity and specificity of 0.84 and 0.89, respectively. These results were not sufficient enough, so a more complex AI model was developed using deep learning approaches. Concerning the DL model, firstly, in comparison to Zhu and colleagues study, our study showed an AUC of 0.95 and specificity and sensitivity of 0.98 and 0.97 with the DL models, which showed better performance with only three complexity indices. Secondly, authors use portal dosimetry which could be not sensitive enough to assess plan conformance. In our clinical routine, using portal dosimetry means 100 % of the beams which pass the criteria 2.5 %/2.5 mm. This result question the utility of this type of control in clinical routine. In comparison, for the 100 % beams conform using portal dosimetry, 31 % were no conform with the DreamDose solution. Finally, our study proposed QA prediction for breast cancer which is not proposed by the precedent study whereas breast cancer treatment plans are known to be most complicated one [21]. Concerning the DreamDose solution: the prediction solutions are based on a complete modelling of the measurement chain (accelerator and EPID) combined with a calculation including all the mechanical and dosimetric parameters of the treatment plans. This leads to a new dose calculation algorithm in the EPID, not a dose measurement solution. Our solution is based on a calibration of delivered dose by framing the detector response as a function of beam energy, independently of other irradiation conditions. The measurement of the interactions taking place in the detector is therefore directly and simply converted into a dose in water, in a similar way to a measurement carried out with an ionisation chamber or a solid detector. Although 2D patient QA solutions, unlike 3D solutions, make it easier to consider pre-treatment checks for each patient, the large number of patients treated on a Halcyon-type machine means that the time spent on these checks is too high. This study shows that it is possible to consider reducing the number of beams checked and thus saving machine time for patient treatment but also allow physicists to focus their attention and reflection on the beams to be inspected. It would then be possible to carry out a more in-depth analysis of these beams and possibly define new instructions to be followed when optimizing treatment plans, which would generate beams that no longer require QA. This predictive process can therefore be used to optimize patient management. In medicine it is necessary to have as low as possible false negatives. In our case if the wrong estimation is a false positive it means that we will performed manually the QA according to the **Supplementary Fig. 1** and then discover that the QA was correct. False negative case means that the prediction will be “conformed” but reality will be non-conformed which is an important problem for clinical routine. For this reason we have optimised the algorithms to maintain as low as possible the false negatives accepting some false positives.

To speed up the process of patient QA in RT, a significant number of studies have developed AI models. However, the number of studies with clinical implementation of the models has been very small. In one study, a clinical implementation was performed, but the results obtained for the AUC were at best 0.869 [22]. Our first goal in this study was to develop an AI solution for QA in IMRT that can be easy to use for everyone and can be implemented in clinical routine. For this reason, we used only three complexity indices in the modelling, which can be easily obtained during the RT plan preparation step. At the same time, we packaged the solution in an API to make it easy for everyone to use. The reproducibility of the AI models in different clinical settings and at different centers is a key factor for their implementation into clinical practice. Using centralized (due to the fact that complexity indices are non-patient specific) and then a federated learning approach can lead to the development of a global model based on data from different centers [23]. The next step is to use federated learning with volunteer centers to improve our model and make it more relevant for other centers.

## Conclusion

Patient-specific QA requires time on the treatment machine, during which time it is impossible to treat patients. Therefore, the objective of this study was to develop an alternative solution for patient-specific QA that would make treatment machines more available to patients. To predict patient-specific QA compliance for treatments, we developed AI models based on complexity indices. In a busy RT department, the AI solution was used to automate patient-specific assurance. The results were conclusive for the IMRT in Halcyon machine. To implement this study in clinical routine, API was developed.

## Funding statement

This study was funded by the Région Normandie thought “Booster IA” grant. NN.M. was supported by the Région Normandie.

## CRedit authorship contribution statement

**Christine Boutry:** Writing – review & editing, Methodology, Data curation, Conceptualization. **Noémie N Moreau:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation. **Cyril Jaudet:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Conceptualization. **Laetitia Lechippey:** Writing – review & editing, Conceptualization. **Aurélien Corroyer-Dulmont:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability Statement:

The data presented in this study can be sent upon reasonable request and the python code used are openly available at [https://github.com/AurelienCD/DeepLearning\\_Patient-HalcyonQA\\_Prediction](https://github.com/AurelienCD/DeepLearning_Patient-HalcyonQA_Prediction).

## Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.radonc.2024.110483>.

## References

- [1] Duan L, Qi W, Chen Y, Cao L, Chen J, Zhang Y, et al. Evaluation of complexity and deliverability of IMRT treatment plans for breast cancer. *Sci Rep* 2023 Dec 6;13: 21474.
- [2] Agazaryan N, Solberg TD, DeMarco JJ. Patient specific quality assurance for the delivery of intensity modulated radiotherapy. *J Appl Clin Med Phys* 2003;4:40–50.
- [3] Moreau N, Bonnor L, Jaudet C, Lechippey L, Falzone N, Batalla A, et al. Deep hybrid learning prediction of patient-specific quality assurance in radiotherapy: implementation in clinical routine. *Diagnostics (Basel)*. 2023 Mar 2;13(5).
- [4] Tamura M, Matsumoto K, Otsuka M, Monzen H. Plan complexity quantification of dual-layer multi-leaf collimator for volumetric modulated arc therapy with Halcyon linac. *Phys Eng Sci Med* 2020 Sep;43:947–57.
- [5] Christine Boutry-Duthil. Method for estimating the dose administered by an external radiotherapy system.
- [6] Boutry C, Sors A, Fontaine J, Delaby N, Delpont G. Technical Note: A simple algorithm to convert EPID gray values into absorbed dose to water without prior knowledge. *Med Phys* 2017;44.
- [7] Dogan N, Mijneer BJ, Padgett K, Nalichowski A, Wu C, Nyflot MJ, et al. AAPM task group report 307: use of EPIDs for patient-specific IMRT and VMAT QA. *Med Phys* 2023;50.
- [8] Razinskas G, Schindhelm R, Sauer OA, Wegener S. Sensitivity and specificity of Varian Halcyon’s portal dosimetry for plan-specific pre-treatment QA. *J Appl Clin Med Phys* 2023;24.

- [9] Corroyer-Dulmont A. DeepLearning\_Patient-HalcyonQA\_Prediction [Internet]. [cited 2023 Dec 29]. Available from: [https://github.com/AurelienCD/DeepLearning\\_Patient-HalcyonQA\\_Prediction](https://github.com/AurelienCD/DeepLearning_Patient-HalcyonQA_Prediction).
- [10] Miften M, Olch A, Mihailidis D, Moran J, Pawlicki T, Molineu A, et al. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM Task Group No. 218. *Med Phys*. 2018 Apr;45(4):e53–83.
- [11] Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in python. *J Mach Learn Res* 2011;12:2825–30.
- [12] Jan B, Farman H, Khan M, Imran M, Islam IU, Ahmad A, et al. Deep learning in big data analytics: a comparative study. *Comput Electr Eng* 2019 May;75:275–87.
- [13] Chollet, F. & others. Keras [Internet]. 2015. Available from: <https://docs.anacanda.com>.
- [14] Lemaitre G, Fernando N, Aridas C. Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. Available from *J Mach Learn Res* [Internet] 2017;18:1–5. <http://arxiv.org/abs/1512.03502>.
- [15] Anaconda Software Distribution [Internet]. Anaconda Inc.; 2020. Available from: <https://docs.anacanda.com>.
- [16] Waskom M. Seaborn: statistical data visualization. *J Open Source Softw* 2021;6:3021.
- [17] Hunter JD. Matplotlib: a 2D graphics environment. Available from *Comput Sci Eng* [Internet] 2007;9:90–5. <http://ieeexplore.ieee.org/document/4160265/>.
- [18] Agazaryan N. Patient specific quality assurance for the delivery of intensity modulated radiotherapy. *J Appl Clin Med Phys* 2003 Jan;4:40.
- [19] Antoine M, Ralite F, Soustiel C, Marsac T, Sargos P, Cugny A, et al. Use of metrics to quantify IMRT and VMAT treatment plan complexity: a systematic review and perspectives. *Phys Med* 2019 Aug;64:98–108.
- [20] Zhu H, Zhu Q, Wang Z, Yang B, Zhang W, Qiu J. Patient-specific quality assurance prediction models based on machine learning for novel dual-layered MLC linac. *Med Phys* 2023 Feb;50:1205–14.
- [21] Jazouli Z, Muraro S, Julian D. Patient-QA prediction: A new approach of complexity indexes. *Phys Med Europ J Med Phys* 2021.
- [22] Yang R, Yang X, Wang L, Li D, Guo Y, Li Y, et al. Commissioning and clinical implementation of an Autoencoder based Classification-Regression model for VMAT patient-specific QA in a multi-institution scenario. *Radiother Oncol* 2021 Aug;161:230–40.
- [23] Pati S, Baid U, Edwards B, Sheller M, Wang SH, Reina GA, et al. Federated learning enables big data for rare cancer boundary detection. *Nat Commun* 2022 Dec 5;13:7346.