



Virtual Patient-Specific Quality Assurance of IMRT Using UNet++: Classification, Gamma Passing Rates Prediction, and Dose Difference Prediction

OPEN ACCESS

Edited by:

Jose Eduardo Villarreal Barajas,
Royal Devon and Exeter Hospital,
United Kingdom

Reviewed by:

Maria F. Chan,
Memorial Sloan Kettering Cancer
Center, United States
Yuanzeng Min,
University of Science and Technology
of China, China

*Correspondence:

Weihai Zhuo
whzhuo@fudan.edu.cn
Zhiyong Xu
xzyong12vip@sina.com

[†]These authors have contributed
equally to this work and
share first authorship

Specialty section:

This article was submitted to
Radiation Oncology,
a section of the journal
Frontiers in Oncology

Received: 26 April 2021

Accepted: 30 June 2021

Published: 20 July 2021

Citation:

Huang Y, Pi Y, Ma K, Miao X, Fu S,
Chen H, Wang H, Gu H, Shao Y,
Duan Y, Feng A, Wang J, Cai R,
Zhuo W and Xu Z (2021) Virtual
Patient-Specific Quality Assurance of
IMRT Using UNet++: Classification,
Gamma Passing Rates Prediction, and
Dose Difference Prediction.
Front. Oncol. 11:700343.
doi: 10.3389/fonc.2021.700343

Ying Huang^{1†}, Yifei Pi^{2†}, Kui Ma³, Xiaojuan Miao⁴, Sichao Fu⁴, Hua Chen¹, Hao Wang¹, Hengle Gu¹, Yan Shao¹, Yanhua Duan¹, Aihui Feng¹, Jiyong Wang⁵, Ruxin Cai¹, Weihai Zhuo^{6*} and Zhiyong Xu^{1*}

¹ Shanghai Chest Hospital, Shanghai Jiaotong University, Shanghai, China, ² Department of Radiotherapy, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, China, ³ Varian Medical Systems (China), Beijing, China, ⁴ Department of Hematology, Western Theater General Hospital, Chengdu, China, ⁵ PingAn Health Technology Co. Ltd, Shanghai, China, ⁶ Key Laboratory of Nuclear Physics and Ion Beam Application Ministry of Education, Fudan University, Shanghai, China

The dose verification in radiotherapy quality assurance (QA) is time-consuming and places a heavy workload on medical physicists. To provide a clinical tool to perform patient specific QA accurately, the UNet++ is investigated to classify failed or pass fields (the GPR lower than 85% is considered “failed” while the GPR higher than 85% is considered “pass”), predict gamma passing rates (GPR) for different gamma criteria, and predict dose difference from virtual patient-specific quality assurance in radiotherapy. UNet++ was trained and validated with 473 fields and tested with 95 fields. All plans used Portal Dosimetry for dose verification pre-treatment. Planar dose distribution of each field was used as the input for UNet++, with QA classification results, gamma passing rates of different gamma criteria, and dose difference were used as the output. In the test set, the accuracy of the classification model was 95.79%. The mean absolute error (MAE) were 0.82, 0.88, 2.11, 2.52, and the root mean squared error (RMSE) were 1.38, 1.57, 3.33, 3.72 for 3%/3mm, 3%/2 mm, 2%/3 mm, 2%/2 mm, respectively. The trend and position of the predicted dose difference were consistent with the measured dose difference. In conclusion, the Virtual QA based on UNet++ can be used to classify the field passed or not, predict gamma pass rate for different gamma criteria, and predict dose difference. The results show that UNet++ based Virtual QA is promising in quality assurance for radiotherapy.

Keywords: deep learning, radiotherapy, quality assurance, prediction model, dose difference

INTRODUCTION

Intensity-modulated radiation therapy (IMRT), a widely used treatment modality for cancer patients, provides highly conformal dose distribution to the target while sparing surrounding healthy tissues (1). Quality assurance is performed to confirm the accuracy of dose calculation, data transmission, linear accelerator performance, radiotherapy positioning, and dosimeter response accuracy (2–6). It is essential to ensure the reliability of treatment delivery and improve patient safety. The process commonly involves comparing the calculated dose distribution or fluence with the measured dose distribution or fluence (7). However, the implementation of patient-specific QA measurement is time-consuming and places a heavy workload on medical physicists (8). Additionally, the measurement work takes a lot of clinical treatment time, which is unrealistic for busy centers. Furthermore, the QA process needs to be accomplished prior to treatment. For those QA results which do not meet the predefined “pass” criteria, re-planning and verification will cause an inevitable treatment delay. Therefore, a more streamlined, less-resourced, and automated patient-specific QA method for dose verification is necessary for radiotherapy centers.

With the development of machine learning and deep learning and its application in QA results prediction, the efficiency of patient-specific QA is expected to be improved (9–16). Valdes, Lam, Li (9–12) etc. to established prediction models for the gamma passing rate (GPR) based on the complexity parameters of the TPS plan. Other researchers (13, 14) investigated the deep learning algorithms to establish GPR prediction models based on planar dose distribution. Granville (15) used support vector machines (SVMs) to classify cold, hot, and normal plans based on the plan complexity parameters and accelerator performance parameters. Li (12) discussed the prediction model of whether the plan could pass the threshold using machine learning method. The above researchers have developed an accurate prediction model of QA results, confirming the feasibility of using machine learning or deep learning for patient-specific QA.

Previous prediction models based on machine learning or deep learning were only the results of dose verification but could not provide detailed information of dose difference (9–18). Predicting the trend and position of dose difference is an important work in automatic patient-specific QA in the near future.

In this study, we proposed a novel QA prediction model based on UNet++ using the planar dose distribution as input. A model can (a) provide the classification results whether the field QA passes; (b) predict the GPRs of different gamma criteria; (c) predict the trend and position of dose difference. The prediction model allows physicists to pre-mark potentially failed fields in a proactive way, analyze dose difference simultaneously and reduce patient delays associated with unqualified measurements. Additionally, it could reduce patient-specific QA measurements to verify data transmission and delivery accuracy combination with other tools. The model is expected to be a practical clinical tool to perform patient-specific QA accurately and provide new ideas for the development of virtual QA and process optimization.

MATERIALS AND METHODS

Data Collection

109 IMRT plans (including 568 fields) from December 2019 to May 2020 were selected. All plans were generated in the Eclipse version11 (Varian Medical Systems, Palo Alto, CA). Dose distributions were calculated using the Acuros External Beam (AXB, ver.11.0.31, Varian Medical Systems, Palo Alto, CA) with a dose calculation grid of 2.5 mm. Each plan was delivered using a linear accelerator equipped with a Varian Millennium 120 MLC. Patient-specific dose verification was performed prior to treatment using the actual angle by Portal Dosimetry. Daily dose calibration was performed during the period of data collection.

PD Mining, an in-house software developed by C #, was used to register, resample, and compare the calculated and measured dose distribution. This software was developed based on the interface of Varian ESAPI portal dosimetry and verified with the manual processing results. All patient dose data was searched through patient ID. The gamma analysis results and dose difference were obtained and exported to the local file in the form of text file automatically.

As 2%/2mm was the most sensitive criterion to detect clinically relevant errors (19, 20), it was used for establishing the classification model. If the GPR was higher than 85%, the field was considered to pass the QA, vice versa. The GPRs in the criteria of 3%/3mm, 3%/2mm, 2%/3mm were calculated at the same time. Absolute dose mode and 10% dose threshold were used for the above GPR analysis.

Data Preprocessing

Pylinac library was used to extract the original image of planar dose distribution in the resolution of 1190×1190 . Then the redundant information such as the frame and coordinate axis was cut off to get the image with the resolution 968×968 . Flipping (horizontal random lip probability: 0.5, vertical random flip probability: 0.5) and random clipping were used to prevent overfitting. The images for the training set were randomly cropped from 968×968 to 960×960 . The images in the test set were cut from the center, sizing from 968×968 to 960×960 .

UNet++ Architecture

Based on the traditional medical image processing network UNet, UNet++ enables the network to learn important features of different depths through a series of nested and dense jumping connections. By adopting deep supervision, UNet++ allows model complexity tuning to balance speed and performance optimization (21).

Figure 1 shows the architecture of the UNet++ used in this study. The downsample encoding channel of $X^{0,0} \rightarrow X^{1,0} \rightarrow X^{2,0} \rightarrow X^{3,0} \rightarrow X^{4,0}$ adopted ResNet-101 architecture as the backbone network, and then the image size was restored by the corresponding upsample decoding nodes with skip connections. The planar dose distribution was input from $X^{0,0}$, and the predicted dose difference image output of the same size was obtained from $X^{0,4}$ after passing through the UNet++ network. At the same time, $X^{0,4}$ was downsampled through three max-pooling layers and two bottleneck layers, and the

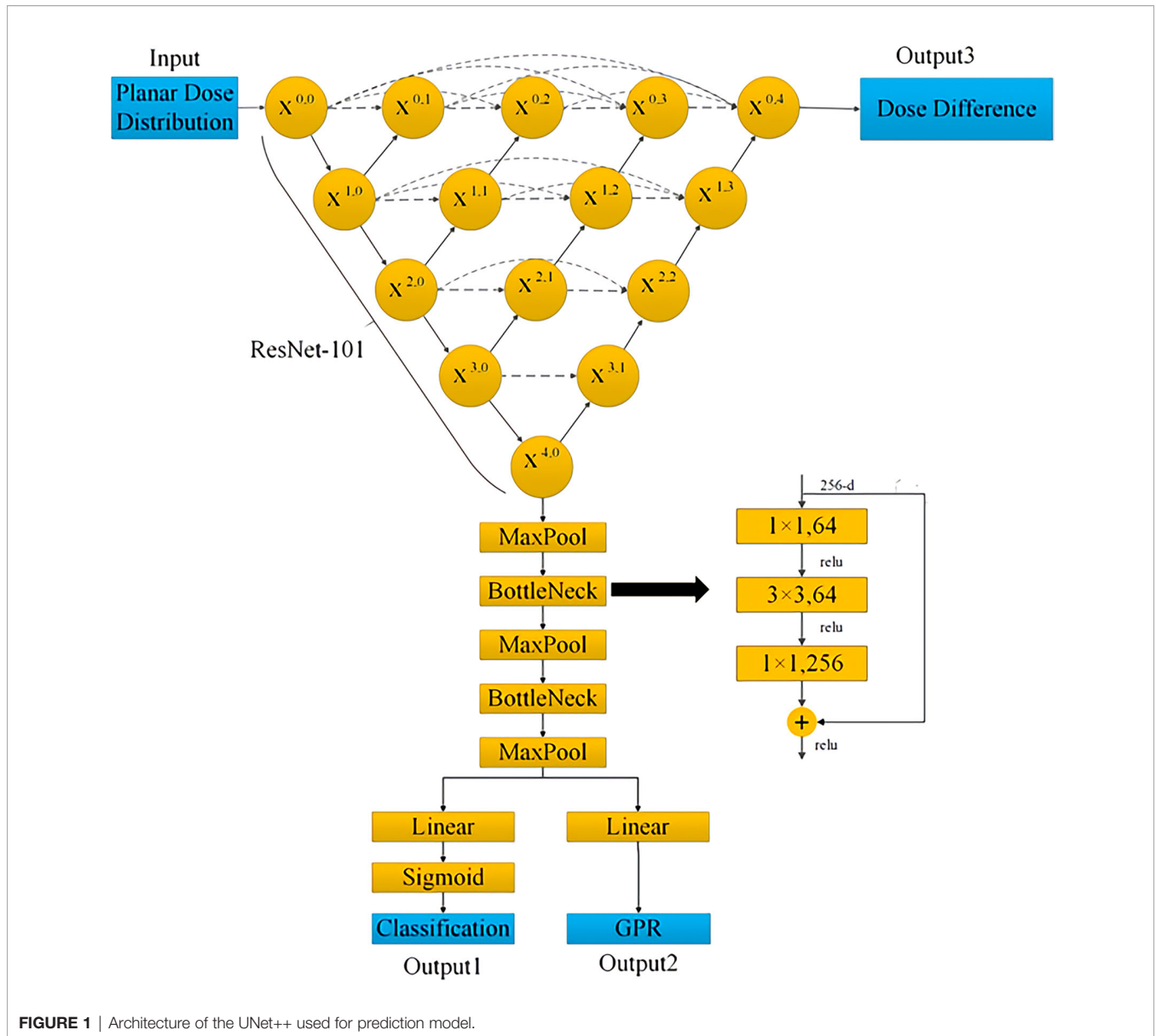


FIGURE 1 | Architecture of the UNet++ used for prediction model.

linear layer was connected in series to predict the four gamma criteria GPR and the classification results.

Model Training and Evaluation

To reduce dose difference prediction error caused by image scaling, the original resolution was adopted for the input and output. This processing would increase network memory and training time. Therefore, small batches rather than N-fold validation were selected. We randomly selected 95 fields (about 1/6) from the collected 568 samples as the test set. Four of the remaining samples (378 fields) were used as the training set and one as the validation set (95 fields).

The mean square error (MSE) loss function was used to evaluate the regression error of dose difference and GPR, and the binary cross-entropy was used to evaluate the classification error. The total loss was obtained by the weighted (1/3, 1/3, 1/3) sum of

the three errors. The commonly used Adam optimizer (22) was adopted to learn the back-propagation error. The initial learning rate was set to 0.001, and it decreased exponentially as training going on, with the dropping rate setting to 0.9. A Mini batch method was used to train the model, the batch size was set to 2, and the epoch was set to 120. The prediction model was built by the open-source pytorch library. The entire training cost about 32 hours on the NVIDIA GTX-3080 GPU.

RESULTS

Learning Curve for the Prediction Model

It is expected that more epochs would give rise to higher prediction accuracy. Thus, there typically exists a minimum number of epochs beyond which the increase in prediction

accuracy would saturate. **Figure 2** shows how the number of epochs will affect the accuracy for the prediction model. With more epochs, the loss on the training data, validation data and testing data decrease. With 60 epochs, the testing, validation and training loss converge at a stable level, indicating that increasing the epochs of training sets may not yield further improvement in the accuracy of the prediction model.

Performance of Classification Model

The proportion of GPR less than 85% was 7.37% (7/95) and 8.42% (8/95) in the validation set and test set, respectively. As shown in **Table 1**, the sensitivity of the validation set is 57.14%, the specificity is 100%, and the accuracy is 96.84%. For the test set, the sensitivity is 62.50%, the specificity is 98.85%, and the accuracy of the classification model is 95.79%. From the analysis for the failed fields in measurement, while the prediction results are pass, the GPRs for these fields are near 85%, and the predicted GPRs are higher than 85%, so the classification results are pass fields.

GPR Prediction for Different Gamma Criteria

As shown in **Table 2**, the MAE and RMSE of the validation set and test set increase with stricter gamma criteria. In the test set, the smallest MAE and RMSE are 0.82 and 1.38 under 3%/3mm.

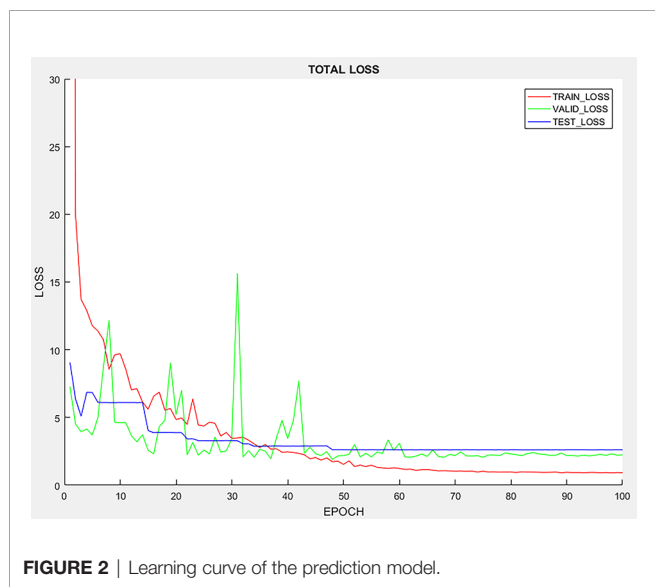


TABLE 1 | Results of classification model.

	Predicted-Fail	Predicted-Pass	
Validation set (95)			
Measured Fail	4	3	57.14%
Measured Pass	0	88	100.00%
Test set (95)			
Measured Fail	5	3	62.50%
Measured Pass	1	86	98.85%

The results are classified into four categories: failed measurement results and failed prediction results, TP; passed measurement results and passed prediction results (TN); passed measurement results and failed prediction results (FP); failed measurement results and passed prediction result passed (FN).

The 2%/2mm gamma criteria has the largest MAE and RMSE in the test set, which are 2.52 and 3.72, respectively.

The distribution of errors between the predicted GPR and the measured GPR under different gamma criteria are shown in **Figure 3**. The prediction errors among different gamma groups (90%-100%, 80%-90%, and < 80%) are compared. The accuracy of the prediction model is affected by the measured value itself. The higher the measured GPR, the smaller prediction errors between the measured and predicted GPR are observed.

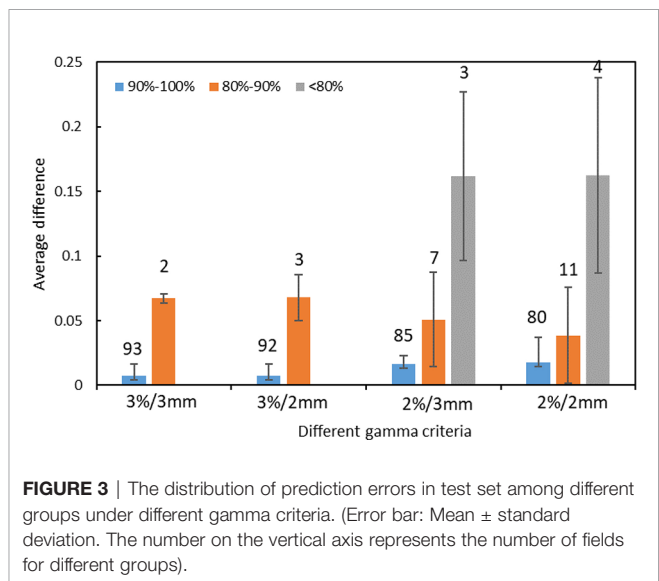
Dose Difference Prediction

In the prediction of dose difference, the dose difference and the histogram of distribution relative to TPS for the pass field and fail field are shown in **Figures 4, 5**, respectively. The position and trend of predicted dose difference are consistent with the measured dose difference.

Figure 4 shows the dose difference for a passing field of a patient. **Figure 4A** shows the dose difference between the measured planar dose and TPS calculated, and **Figure 4B** shows the dose difference between the predicted and TPS calculated planar dose. Even for a passing field, there are still some pixels with large dose difference, and their positions can be obtained. **Figure 4C** is the histogram of the dose difference between measured and TPS calculated planar dose for one

TABLE 2 | MAE and RMSE for different gamma criteria.

	MAE		RMSE	
	Validation set	Test set	Validation set	Test set
3%/3mm	0.79	0.82	1.28	1.38
3%/2mm	0.93	0.88	1.50	1.57
2%/3mm	2.01	2.11	2.31	3.33
2%/2mm	2.17	2.52	3.00	3.72



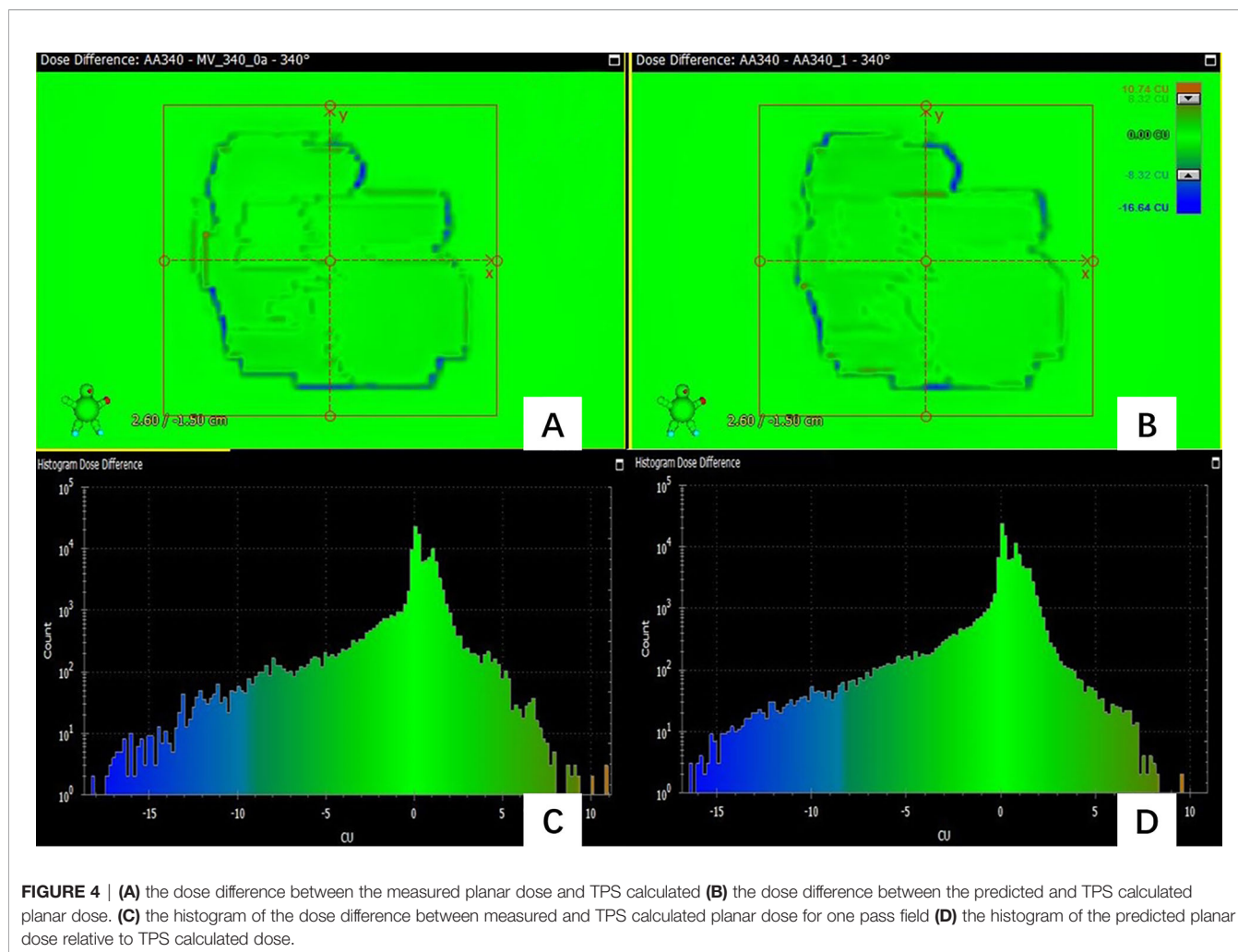


FIGURE 4 | (A) the dose difference between the measured planar dose and TPS calculated (B) the dose difference between the predicted and TPS calculated planar dose. (C) the histogram of the dose difference between measured and TPS calculated planar dose for one pass field (D) the histogram of the predicted planar dose relative to TPS calculated dose.

failed field, and **Figure 4D** is the histogram of the predicted planar dose relative to TPS calculated dose. The histogram of predicted dose difference is consistent with the measured dose difference, the maximum and the minimum dose difference are almost the same.

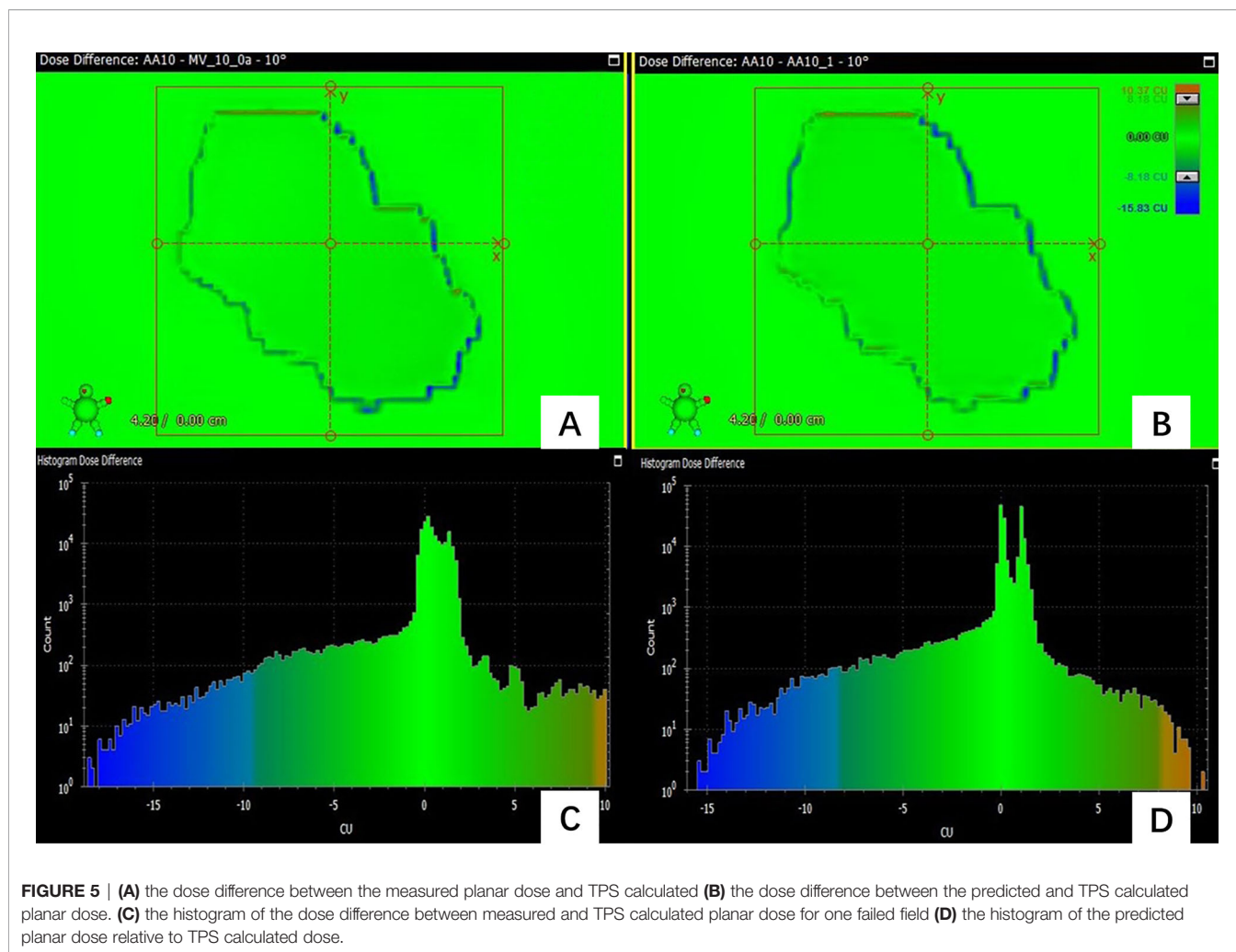
As **Figure 5** shows, the predicted dose difference and measured dose difference for one of the failed fields and the histogram of dose difference. Large dose difference mainly located in the edge, and the measured dose is lower than the TPS calculated dose. The trend and position of the predicted dose difference is consistent with the measured dose difference, indicating that the virtual QA results could be used as guidance for the analysis of dose difference and plan redesign. By comparing and analyzing the dose difference for pass field and failed field, we found that the number of pixels having the large dose difference in the failed field was larger than in the pass field.

DISCUSSIONS

IMRT is a complex technology in radiotherapy, so special QA is required to ensure the accuracy of dose delivery. In this study,

the UNet++ was used to classify QA results, GPR prediction of different gamma criteria and accurate dose difference prediction based on planar dose distribution. The accuracy of the classification model was 95.79%; there were small RMSE (1.38-3.72) and MAE (0.82-2.52) between the measured and the predicted GPR in the test set, and the trend and position of predicted dose difference were consistent with the measured dose difference. The results showed that the prospect of realizing virtual patient-specific QA with UNet++ and provides a new idea for the optimization of the individual QA process.

QA is required for every patient before treatment to ensure the accuracy of dose delivery (2–6). Dose verification depends on the equipment highly. The resolution and energy response of the detector will affect the result of dose verification. For the fields that fail the threshold, it is necessary to adjust the radiotherapy plan repeatedly, which will bring treatment delay. The process of dose verification is labor-consuming and time-consuming, so it brings more workload to the busy center. Automatic dose verification pre-measurement can mark the plans that fail the verification in advance and predict the dose difference, which is expected to be an effective method to solve the above problems.



As suggested in the TG 218 (6), the ability of the prediction model to accurately classify plans into “pass” or “fail” based on gamma criteria used is one of the most important indicators to evaluate the clinical feasibility of the model. The prediction of GPRs under different gamma criteria could provide more comprehensive information for physicists to judge whether the plan is acceptable clinically. The dose difference prediction model predicts the trend and location of dose difference, which provides direction for physicists to modify the plan. The paradigm shift of pre-measurement QA will improve the efficiency of dose verification greatly.

Since 2%/2mm is the most sensitive to clinically relevant errors (19, 20), it is selected as the basis of classification model in this study. The accuracy of failed fields was lower compared to pass fields, as the measured GPR were near 85% for some failed fields, and the predicted GPR were higher than 85%, so the predicted results of these fields were passed. Therefore, the selection of appropriate threshold plays an important role in the accuracy of the classification model and the proportion of failed fields needed to improve the accuracy of classification. The classification model gives the physicist a more intuitive result whether the field pass the QA pre-measurement.

The prediction model in this study can give the classification results according to the standard of the classification model, and also give the GPRs of different gamma criteria (3%/3mm, 3%/2mm, 2%/3mm, 2%/2mm). As an example, for one of the fail fields under the classification result, the GPR for 3%/3mm was 98.51% (Error: 0.01%), and GPR was 97.42% (Error: 0.22%) for 3%/2mm. After discussion by physicists and the doctors, the plan can be delivered. Therefore, the prediction of the GPRs of different gamma criteria could make a comprehensive judgment for the clinical enforceability of the plan.

In the prediction of GPR under different gamma criteria, the MAE and RMSE for the measured and predicted GPR of the model increase with the stricter gamma criteria, which is caused by the increase of the uncertainty of the prediction with the stricter gamma criteria. The accuracy of the model is affected by the measured GPR itself. The higher measured GPR is, the higher accuracy. This can be explained that the data with high measured GPR accounts for a large part of our model, so the accuracy of higher GPR prediction is high.

Previous studies only predicted the results of QA using machine learning or deep learning (9–18, 23, 24), but it is impossible to predict the trend and position of dose difference.

The trend and position of dose difference plays a very important role in the analysis of the error source. However, there are no relevant reports about dose difference at present. In this study, prediction of dose difference based on the planar dose distribution is fulfilled. The result of this study that the predicted dose difference is consistent with the position and trend of the measured dose difference, can provide guidance for reason analysis.

It is noting that the pre-treatment prediction model established in this study is an auxiliary tool that needs a reparatory guarantee of the accuracy of the energy calibration and delivery process of the accelerator (2, 4). The purpose of this prediction model is not intended to replace the traditional QA, but to help physicists reduce the measurement burden of patient-specific QA when verifying the dose distribution and optimize the process of QA combined with other methods (25–28). There are still some limitations in this study: 1) As most of the clinical fields are pass fields, which leads to the unbalance of data distribution. Adequate amounts of low GPR plans for model training are needed to improve the accuracy of the model in future. 2) The model is limited to our data that all the data come from the same accelerator, the same energy, the same verification equipment. To expand the universality of the model, the research on models that include a variety of energy, different types of accelerators and dose verification equipment will be done in the future.

CONCLUSION

In this study, we developed a UNet++ based prediction model for patient-specific QA. The prediction model could classify whether the field passes or fail QA, predict GPR of different gamma criteria, predict the trend and location of dose difference and mark the position. The virtual QA tool developed in this study provides a new

idea for the optimization of patient-specific QA process, and promote the development of automated patient-specific QA.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material. Further inquiries can be directed to the corresponding author.

ETHICS STATEMENT

Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

AUTHOR CONTRIBUTIONS

YH and YP: study concept and design and drafting of the manuscript. YH, KM, and RC: acquisition of data. YH and KM: analysis and interpretation of data. XM, SF, and JW: develop training model. HC, HW, HG, YS, YD, and AF: critical revision of the manuscript for important intellectual content. YH and SF: statistical analysis. WZ and ZX: study supervision. All authors contributed to the article and approved the submitted version.

FUNDING

This study was funded by the Nurture projects for basic research of Shanghai Chest Hospital.

REFERENCES

- Ma C, Wang R, Zhou S, Wang M, Yue H, Zhang Y, et al. The Structural Similarity Index for IMRT Quality Assurance: Radiomics-Based Error Classification. *Med Phys* (2021) 48:80–93. doi: 10.1002/mp.14559
- Ezzell GA, Burmeister JW, Dogan N, LoSasso TJ, Mechalakos JG, Mihailidis D, et al. IMRT Commissioning: Multiple Institution Planning and Dosimetry Comparisons, a Report From AAPM Task Group 119. *Med Phys* (2009) 36:5359–73. doi: 10.1118/1.3238104
- Van Esch A, Huyskens DP, Behrens CF, Samsoe E, Sjolín M, Bjelkengren U, et al. Implementing Rapidarc Into Clinical Routine: A Comprehensive Program From Machine QA to TPS Validation and Patient QA. *Med Phys* (2011) 38:5146–66. doi: 10.1118/1.3622672
- Klein EE, Hanley J, Bayouth J, Yin FF, Simon W, Dresser S, et al. Task Group 142 Report: Quality Assurance of Medical Accelerators. *Med Phys* (2009) 36:4197–212. doi: 10.1118/1.3190392
- Smilowitz JB, Das IJ, Feygelman V, Fraass BA, Kry SF, Marshall IR, et al. Commissioning and QA of Treatment Planning Dose Calculations - Megavoltage Photon and Electron Beams. *J Appl Clin Med Phys* (2015) 16:14–34. doi: 10.1120/jacmp.v16i5.5768
- 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) 45:e53–83. doi: 10.1002/mp.12810
- Nyflot MJ, Thammasorn P, Wootton LS, Ford EC, Chaovalitwongse WA. Deep Learning for Patient-Specific Quality Assurance: Identifying Errors in Radiotherapy Delivery by Radiomic Analysis of Gamma Images With Convolutional Neural Networks. *Med Phys* (2019) 46:456–64. doi: 10.1002/mp.13338
- Van Esch A, Bohsung J, Sorvari P, Tenhunen M, Paiusco M, Iori M, et al. Acceptance Tests and Quality Control (QC) Procedures for the Clinical Implementation of Intensity Modulated Radiotherapy (IMRT) Using Inverse Planning and the Sliding Window Technique: Experience From Five Radiotherapy Departments. *Radiother Oncol* (2002) 65:53–70. doi: 10.1016/s0167-8140(02)00174-3
- Valdes G, Scheuermann R, Hung CY, Olszanski A, Bellerive M, Solberg TD. A Mathematical Framework for Virtual IMRT QA Using Machine Learning. *Med Phys* (2016) 43:4323–34. doi: 10.1118/1.4953835
- Valdes G, Chan MF, Lim SB, Scheuermann R, Deasy JO, Solberg TD. Imrt Qa Using Machine Learning: A Multi-Institutional Validation. *J Appl Clin Med Phys* (2017) 18:279–84. doi: 10.1002/acm.2.12161
- Dao L, Zhang XZ, Li H, Deshan Y, Schott B, Zhao T, et al. Predicting Gamma Passing Rates for Portal Dosimetry-Based IMRT QA Using Machine Learning. *Med Phys* (2019) 46:4666–75. doi: 10.1002/mp.13752
- Li JQ, Wang L, Zhang X, Liu L, Li J, Chan MF, et al. Machine Learning for Patient-Specific Quality Assurance of VMAT: Prediction and Classification Accuracy. *Int J Radiat Oncol Biol Phys* (2019) 105:893–902. doi: 10.1016/j.ijrobp.2019.07.049

13. Interian Y, Rideout V, Kearney VP, Gennatas E, Morin O, Cheung J, et al. Deep Nets vs Expert Designed Features in Medical Physics: An IMRT QA Case Study. *Med Phys* (2018) 45:2672–80. doi: 10.1002/mp.12890
14. Tomori S, Kadoya N, Takayama Y, Kajikawa T, Shima K, Narazaki K, et al. A Deep Learning-Based Prediction Model for Gamma Evaluation in Patient-Specific Quality Assurance. *Med Phys* (2018) 45:4055–65. doi: 10.1002/mp.13112
15. Granville DA, Sutherland JG, Belec JG, Russa DJL. Predicting VMAT Patient-Specific QA Results Using a Support Vector Classifier Trained on Treatment Plan Characteristics and Linac QC Metrics. *Phys Med Biol* (2019) 64:09501. doi: 10.1088/1361-6560/ab142e
16. Chan MF, Witztum A, Valdes G. Integration of AI and Machine Learning in Radiotherapy Qa. *Front Artif Intell* (2020) 3:577620. doi: 10.3389/frai.2020.577620
17. Ono T, Hirashima H, Iramina H, Mukumoto N, Miyabe Y, Nakamura M, et al. Prediction of Dosimetric Accuracy for VMAT Plans Using Plan Complexity Parameters Via Machine Learning. *Med Phys* (2019) 46:3823–32. doi: 10.1002/mp.13669
18. Hirashima H, Ono T, Nakamura M, Miyabe Y, Mukumoto N, Iramina H, et al. Improvement of Prediction and Classification Performance for Gamma Passing Rate by Using Plan Complexity and Dosiomics Features. *Radiother Oncol* (2020) 153:250–7. doi: 10.1016/j.radonc.2020.07.031
19. Nelms BE, Chan MF, Jarry G, Lemire M, Lowden J, Hampton C, 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. *Med Phys* (2013) 40:111722. doi: 10.1118/1.4826166
20. Heilemann G, Poppe B, Laub W. On the Sensitivity of Common Gamma-Index Evaluation Methods to MLC Misalignments in Rapidarc Quality Assurance. *Med Phys* (2013) 40:031702. doi: 10.1118/1.4789580
21. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation. *IEEE Trans Med Imaging* (2020) 39:1856–67. doi: 10.1109/TMI.2019.2959609
22. He K, Zhang X, Ren S, Sun J. *Deep Residual Learning for Image Recognition [C]//Ieee Conference on Computer Vision & Pattern Recognition*. IEEE Computer Society (2016).
23. Wall P, Fontenot JD. Application and Comparison of Machine Learning Models for Predicting Quality Assurance Outcomes in Radiation Therapy Treatment Planning. *Inf Med Unlocked* (2020) 18:100292. doi: 10.1016/j.imu.2020.100292
24. Wang L, Li J, Zhang S, Zhang X, Zhang Q, Chan MF, et al. Multi-Task Autoencoder Based Classification-Regression Model for Patient-Specific VMAT Qa. *Phys Med Biol* (2020) 65:235023. doi: 10.1088/1361-6560/abb31c
25. Kerns JR, Childress N, Kry SF. A Multi-Institution Evaluation of MLC Log Files and Performance in IMRT Delivery. *Radiat Oncol* (2014) 11:176. doi: 10.1186/1748-717X-9-176
26. Ford EC, Terezakis S, Souranis A, Harris K, Gay H, Mutic S. Quality Control Quantification (QCQ): A Tool to Measure the Value of Quality Control Checks in Radiation Oncology. *Int J Radiat Oncol Biol Phys* (2012) 84:e263–9. doi: 10.1016/j.ijrobp.2012.04.036
27. Passarge M, Fix MK, Manser P, Stampanoni MF, Siebers JV. A Swiss Cheese Error Detection Method for Real-Time EPID-Based Quality Assurance and Error Prevention. *Med Phys* (2017) 44:1212–23. doi: 10.1002/mp.12142
28. Fan J, Li J, Chen L, Stathakis S, Luo W, Du Plessis F, et al. A Practical Monte Carlo MU Verification Tool for IMRT Quality Assurance. *Phys Med Biol* (2006) 51:2503–15. doi: 10.1088/0031-9155/51/10/010

Conflict of Interest: KM was employed by Varian Medical Systems. JW was employed by Ping An Healthcare Technology Co Ltd.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Huang, Pi, Ma, Miao, Fu, Chen, Wang, Gu, Shao, Duan, Feng, Wang, Cai, Zhuo and Xu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.