Jurnal Penelitian Fisika dan Aplikasinya (JPFA)
Volume 15, Issue 1, 71-88 2025
https://journal.unesa.ac.id/index.php/jpfa/
p-ISSN: 2087-9946
e-ISSN: 2477-1775

A Generalizable AI-Based Framework for Automated Slice Sensitivity Profile and Slice Thickness Measurement Across Multi-Phantom and Multi-Scanner CT Systems

Nur Hadziqoh [1,a,*], Nani Lasiyah [1,b], Rino Ferdian Surakusumah [1,c], Arda Yunianta [2,d], and Nurul Maisarah Binti Kamaruddin [3,e]

[1] Department of Medical Electronics Engineering Technology, Faculty of Health Technology, Al Insyirah Institute of Health and Technology, Pekanbaru 28289, Indonesia

[2] Department of Information System, Faculty of Computing and Information Technology in Rabigh, King Abdul Aziz University, Jeddah 21589, Saudi Arabia

[3] Department of Electrical Engineering, Politeknik Sultan Salahuddin Abdul Aziz, Shah Alam 40150, Malaysia

e-mail: [a]nur.hadziqoh@ikta.ac.id, [b]nani.lasiyah@ikta.ac.id, [c] rino.ferdian@ikta.ac.id, [d]ayunianta@kau.edu.sa, and [e]msarah@psa.edu.my

* Corresponding Author

Received: 22 May 2025; Revised: 19 May 2025; Accepted: 30 June 2025

Abstract

Slice Sensitivity Profile (SSP) and its Full Width at Half Maximum (FWHM) are critical indicators of longitudinal resolution and effective slice thickness in computed tomography (CT), forming a cornerstone of quality assurance (QA) protocols. This study introduces a robust and vendor-neutral framework for automated SSP and FWHM measurement using a deep learning-based approach, designed to overcome the limitations of manual, scanner-specific, and phantom-specific methods. A U-Net convolutional neural network was trained on annotated CT phantom images—including AAPM, Catphan, and ACR models—acquired across Philips, Siemens, and GE CT systems with varied slice thicknesses (1.0 mm and 5.0 mm). The pipeline includes automatic stair-step object segmentation, angular correction via Hough Transform, profile extraction, and real-time FWHM computation. Validation against manual measurements demonstrated strong correlation ($r > 0.97$) and mean absolute errors below 0.2 mm, with no statistically significant differences across stair-step positions ($p > 0.05$). The system showed excellent repeatability (CV < 1.5%) and reproducibility (CV < 2.5%), even with phantom repositioning and inter-operator variability. Additionally, the framework maintained consistency across all phantom types and scanner brands, confirming its cross-platform reliability and alignment with IEC 61223-3-5 and AAPM performance standards. These results position the proposed method as a generalizable and scalable QA solution, suitable for clinical integration, automated reporting, and longitudinal CT performance monitoring.

Keywords: *Slice Sensitivity Profile; CT Quality Assurance; Phantom Imaging; Full-Width at Half Maximum; Deep Learning*

How to cite: Hadziqoh N, Lasiyah N, Surakusumah RF, Yuniata A, Maisarah N. A Generalizable AI-Based Framework for Automated Slice Sensitivity Profile and Slice Thickness Measurement Across Multi-Phantom and Multi-Scanner CT Systems. *Jurnal Penelitian Fisika dan Aplikasinya (JPFA)*. 2025; 15(1): 71-88. DOI: https://doi.org/10.26740/jpfa.v15n1.p71-88.

INTRODUCTION

The Slice Sensitivity Profile (SSP) and its Full Width at Half Maximum (FWHM) are critical metrics for evaluating the longitudinal spatial resolution and effective slice thickness in computed tomography (CT) imaging. These parameters directly impact diagnostic accuracy, particularly in high-resolution applications such as oncology, cardiovascular imaging, and radiotherapy planning [1,2]. Precise SSP and FWHM measurements are also essential for image quality assurance (QA) programs to verify scanner performance, detect geometric or reconstruction-related deviations, and ensure compliance with protocols such as IEC 61223-3-5 and AAPM accreditation standards [3,4].

Historically, SSP and slice thickness evaluations have been conducted using manual measurements on physical phantoms such as the AAPM, Catphan, or ACR phantoms [5–7]. Although widely adopted, these methods are time-consuming, subjective, and highly dependent on operator skill. Variability in stair object selection, misalignment, and visual boundary estimation can lead to inconsistencies in repeated QA cycles, particularly in institutions lacking standardized procedures [8,9]. Additionally, most commercial tools and semi-automated methods are not generalizable across different scanner vendors or phantom types and typically focus only on fixed slice thicknesses, without accounting for protocol-specific variations such as pitch, reconstruction filter, or table speed [10,11].

Recent studies have highlighted the growing relevance of automated slice thickness and SSP measurement methods, particularly those incorporating deep learning [12–14]. Techniques using convolutional neural networks such as U-Net have been successfully implemented for phantom segmentation, stair-step edge detection, and profile extraction in CT images [15,16]. These systems can be trained to adapt across vendors (e.g., Philips, GE, Siemens) and phantom geometries, enabling cross-platform deployment and real-time QA feedback [17–20]. Additionally, studies have shown that such systems can effectively support predictive QA, identifying performance drift before it affects diagnostic reliability [21,22].

The importance of such automation is amplified by the increasing demand for evidence-based QA systems that can be linked to clinical outcomes, such as contrast-to-noise ratio (CNR) and modulation transfer function (MTF) [23,24]. Furthermore, current research has explored the influence of slice thickness variation on radiomics stability and dose estimation—emphasizing the critical role of robust FWHM monitoring tools in clinical trials and AI-based imaging analysis pipelines [25–28].

Despite these advancements, most existing studies are limited in scope which only focusing on a single phantom type, fixed scanner brand, or isolated acquisition setting [29,30]. There remains a need for a fully automated, vendor-neutral, phantom-agnostic framework for SSP measurement that is suitable for routine clinical QA, regulatory validation, and multi-center research environments.

In this study, we propose and validate a framework: an AI-enhanced SSP and FWHM measurement system, powered by a U-Net-based deep learning model and an automated stair-step profile extraction pipeline. This system was trained and tested on images from multiple phantoms (AAPM, Catphan, ACR) scanned with three different CT vendors at various slice thickness settings. Our objectives are to (1) develop a reproducible and scalable solution for automated slice sensitivity measurement, (2) validate the algorithm's performance across diverse imaging conditions and QA devices, and (3) correlate FWHM values with image quality metrics relevant to diagnostic accuracy. This work contributes to the modernization of CT QA through automation, standardization, and

clinical relevance, building on prior work while significantly extending its generalizability and practical application [31–35].

MATERIALS AND METHODS

This study employed a structured experimental design to develop, implement, and validate an AI-based framework for automated measurement of the Slice Sensitivity Profile (SSP) and its Full Width at Half Maximum (FWHM) from CT images of performance phantoms. The workflow of this study is shown in Figure 1.

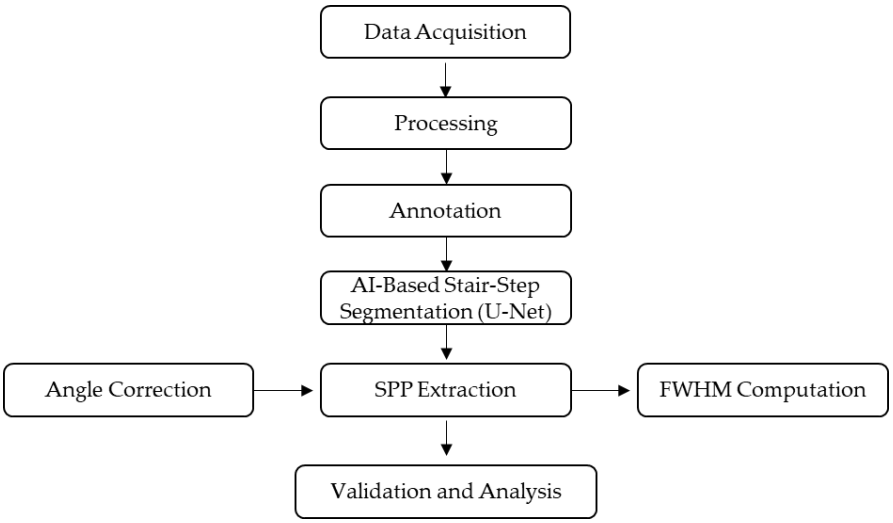

Figure 1. The Flowchart of AI-based methodology for automated SSP and FWHM measurement from CT phantom images

The primary image dataset consisted of CT scans acquired using various phantom, including the AAPM CT performance phantom, which contains distinct aluminum stair-step objects positioned at a known angle within a water-equivalent cylindrical matrix. Figure 2 shows an example of the axial CT image used for analysis, illustrating the three stair-step structures that are critical for extracting the SSP signal along the Z-axis.

These stair-step objects produce sloped features in the axial image that vary in width according to the slice thickness of the scan, making them suitable for quantifying the longitudinal resolution of the CT system. The image was acquired under clinical scan parameters using a multi-slice CT scanner, and stored in DICOM format for subsequent processing. The algorithm development involved two key stages: (1) automatic localization and segmentation of the stair-step features using a deep learning model, and (2) extraction and analysis of the SSP to compute the FWHM. The entire framework was designed to be compatible with multiple scanner brands and phantom configurations and validated through comparison with manual measurement protocols as well as standardized image quality metrics. The diagram illustrates the Research Frameworks shown in Figure 2.

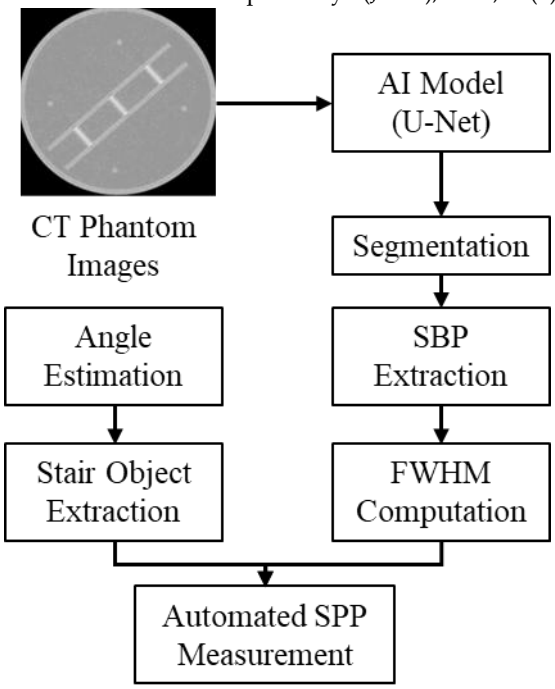

Figure 2. SPP Measurement Block Diagram

CT scanners and phantom models

To ensure the generalizability and robustness of the proposed automated Slice Sensitivity Profile (SSP) measurement framework, this study used multiple CT scanners and phantom models. The primary phantom used was the AAPM CT Performance Phantom (Model 610, CIRS, Virginia, USA), which contains embedded aluminum stair-step structures specifically designed for evaluating slice thickness and longitudinal spatial resolution. These structures consist of three stair-shaped objects (top, middle, and bottom) arranged diagonally inside the phantom, oriented at an angle of approximately 45° relative to the axial imaging plane, as shown previously in Figure 1. The aluminum inserts are each 0.635 mm thick and spaced uniformly to generate Z-axis gradients that are highly sensitive to changes in slice thickness and reconstruction accuracy [12].

This phantom has been widely adopted for CT QA due to its ability to simulate geometric and density variations that mimic patient anatomy in a standardized format [13]. The sloped edges of the stair objects provide a reliable means to extract the SSP, which is later used to compute the Full Width at Half Maximum (FWHM), which is a key metric for validating effective slice thickness [14].

The CT data were acquired using three multi-detector CT (MDCT) systems from distinct vendors to evaluate the cross-platform applicability of the proposed framework:

1. Philips Ingenuity 128-slice CT (Philips Healthcare, The Netherlands)
2. Siemens SOMATOM Definition AS 64-slice CT (Siemens Healthineers, Germany)
3. GE Revolution EVO 128-slice CT (GE Healthcare, USA)

Each scanner was configured using standardized acquisition protocols varying in nominal slice thickness (0.5 mm, 1 mm, 2 mm, and 5 mm), with fixed tube voltage (120 kVp), tube current (200 mA), and helical pitch (0.984). All scans were performed in helical mode, and image reconstruction was conducted using filtered back-projection (FBP) with standard clinical kernels, such as lung and soft-tissue algorithms, depending on the manufacturer's default presets [15]. The field of view (FOV) was fixed at 260 mm with a 512 × 512 reconstruction matrix to maintain spatial resolution

consistency across the datasets.

All DICOM datasets were anonymized and stored for processing using the proposed AI-based pipeline. The multi-scanner, multi-thickness setup ensures a realistic simulation of diverse clinical environments and supports validation of the framework under various QA-relevant conditions.

Image acquisition protocols

To evaluate the performance of the proposed automated Slice Sensitivity Profile (SSP) measurement framework, CT image acquisition was conducted under varied controlled protocols simulating clinical conditions. Scans were conducted using three different multi-detector CT (MDCT) systems as outlined in Section 2.1, with acquisition parameters systematically varied to assess the algorithm's robustness against changes in image resolution, contrast, and reconstruction artifacts.

Each phantom scan was acquired in helical mode, which is widely adopted in clinical CT due to its rapid acquisition and improved z-axis coverage. The tube voltage (kVp) was set to 120 kV, and tube current (mA) was maintained at 200 mA, providing a balanced radiation dose while preserving adequate image quality. The pitch was fixed at 0.984, aligning with standard QA testing configurations and minimizing overlapping reconstruction effects [16].

A series of scans were obtained at different nominal slice thicknesses, including 0.5 mm, 1.0 mm, 2.0 mm, and 5.0 mm, as slice thickness directly impacts the shape and sharpness of the SSP curve. These variations allowed for comparative analysis of the algorithm’s ability to detect subtle differences in spatial resolution. Each scan produced axial images in the Digital Imaging and Communications in Medicine (DICOM) format with a matrix size of 512 × 512 pixels, and a fixed field of view (FOV) of 260 mm, ensuring a uniform in-plane pixel spacing across all datasets.

Image reconstruction was performed using the filtered back-projection (FBP) method with vendor-recommended kernels such as B50s for Siemens, Lung filter for Philips, and Standard kernel for GE, depending on the scanner. These kernels were chosen to simulate typical soft tissue and lung imaging protocols while introducing realistic variations in edge definition and noise texture, which are known to affect SSP profiles [17].

For each scan, multiple image slices encompassing the stair-step object region were saved and reviewed to ensure complete coverage of the phantom insert. No additional image preprocessing, denoising, or artifact correction was applied prior to algorithmic processing in order to evaluate the system's raw image interpretability—a critical criterion for real-world deployment in clinical QA environments [18].

The acquired datasets were subsequently used as input for the automated SSP extraction and FWHM computation stages, enabling a rigorous test of the method’s performance under varied technical and anatomical simulation conditions.

AI model architecture (U-Net), training data, and preprocessing

To automate the detection and segmentation of stair-step objects within CT phantom images, this study implemented a deep learning architecture based on the U-Net convolutional neural network (CNN), which has demonstrated excellent performance in biomedical image segmentation due to its encoder–decoder structure and skip connections [19].

The U-Net model consists of a contracting path (encoder) followed by an expanding path (decoder). The encoder progressively downsamples the input image through convolutional and

max-pooling layers, extracting increasingly abstract features. The decoder then upsamples these features to restore spatial resolution and produces a binary mask identifying the stair-step regions. Skip connections between corresponding encoder and decoder layers allow fine-grained spatial details to be preserved.

The segmentation mask output $\hat{Y}$ for an input image X is obtained via Equation 1.

$$\hat{Y} = f_{U-Net}(X;\theta) \tag{1}$$

where θ represents the model parameters learned during training. The model was trained to minimize the Dice loss, defined as Equation 2.

$$L_{Dice} = 1 - \frac{2\sum_i y_i \hat{y}_i}{\sum_i y_i + \sum_i \hat{y}_i + \epsilon} \tag{2}$$

Where $y_i$and $\hat{y}_i$ are the ground truth and predicted binary labels for each pixel, respectively, and ϵ is a small constant for numerical stability. The Dice loss is particularly effective for handling imbalanced classes, as stair-step objects occupy a small region of the image.

A total of 450 CT phantom images were extracted from the DICOM datasets acquired across three different scanners and four slice thickness settings. 300 images were used for training, 75 for validation, and 75 for testing. The stair-step objects in each image were manually annotated by two trained observers using the open-source ITK-SNAP software, and the ground truth masks were reviewed by a senior radiological physicist to ensure annotation consistency.

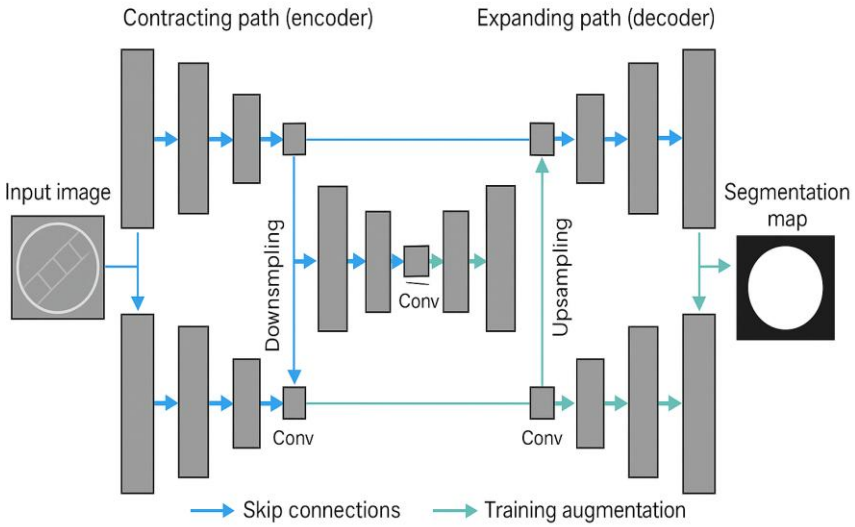

Figure 3. Schematic diagram of the U-net architecture

Prior to training, all images were rescaled to 256 × 256 pixels, intensity-normalized to a [0,1] range using min-max normalization, and augmented with random rotations (±15°), flips, and zooms (±10%) to improve generalization and robustness against phantom orientation or minor scan variations. The network was developed in Python using TensorFlow, trained for 100 epochs using the Adam optimizer using a learning rate of 1e-4 and batch size of 16. To avoid overfitting, early stopping was applied based on validation loss. An illustration of the U-net architecture is presented in Figure 2.

SSP Extraction Pipeline

After segmentation of the stair-step regions using the trained U-Net model, the next critical phase in the automated framework is the extraction of the Slice Sensitivity Profile (SSP). The SSP characterizes the CT system's resolution along the longitudinal (Z-axis) direction and is obtained by analyzing the intensity profile across the reconstructed slanted edge created by the stair-step object within the axial image.

Since the phantom may be positioned at arbitrary rotational angles during scanning, accurate measurement of SSP requires alignment of the slanted stair-step object with a horizontal axis. To correct for this, the system uses a 1D Hough Transform to detect the primary orientation of the middle stair structure, extracting the angle θ at which the stair rises within the axial slice as described in Equation 3.

$$\theta = arg\ arg\ [HoughTransform(I_{bin}, \theta)] \tag{3}$$

where I_{bin} is the binarized output from segmentation.

Once the dominant angle is estimated, a geometric transformation (rotation) is applied to realign the stair-step object horizontally. This correction ensures consistent sampling of the profile across different scans and scanner orientations.

With the stair-step object aligned, a line profile is extracted perpendicular to the stair edge to capture the intensity transition across slices. Let P(x) represent the pixel intensity values sampled along the longitudinal direction. The SSP is then modeled by normalizing and smoothing this profile to mitigate noise as shown in Equation 4.

$$SSP(x) = \frac{P(x) - min\ (P)}{(P) - min\ (P)} \tag{4}$$

A Gaussian or trapezoidal-like curve is typically expected, representing the system's z-axis point spread function.

The Full Width at Half Maximum (FWHM) of the SSP is calculated by identifying the width of the profile at 50% of its maximum height. Let X_1 and X_2 be the positions where described in Equation 5.

$$SSP(X_1) = SSP(X_2) = 0.5 * max\ (SSP) \tag{5}$$

Then in Equation 6,

$$FWHM = |X_2 - X_1| \tag{6}$$

This width represents the effective slice thickness. To improve robustness, the system applies interpolation to sub-pixel resolution and averages FWHM values from all three stair-step objects (top, middle, and bottom) as well as the overall mean.

To enhance statistical confidence and reduce local noise artifacts, the SSP extraction and FWHM measurements are repeated over multiple consecutive slices encompassing the stair-step region (typically 5–10 slices). The final reported slice thickness for a scan is given by Equation 7.

$$FWHM_{final} = \frac{1}{N} \sum_{i=1}^{N} FWHM_i \tag{7}$$

where N is the number of slices.

Validation Metrics and Statistical Analysis

To evaluate the performance and reliability of the proposed AI-driven SSP measurement framework, both technical segmentation accuracy and quantitative agreement of computed FWHM values were assessed. Validation was carried out using a combination of statistical metrics, inter-method comparisons, and hypothesis testing against ground truth and manual measurements.

The segmentation performance of the U-Net model in identifying stair-step objects was evaluated on the test dataset using common metrics for image segmentation which are Dice Similarity Coefficient (DSC) and Intersection over Union (IoU) as described in Equation 8 and 9. below.

$$DSC = \frac{2|A \cap B|}{|A|+|B|} \tag{8}$$

Where A is the predicted segmentation and B is the ground truth mask. A DSC close to 1 indicates high overlap. Then Intersection over Union (IoU) is described below.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{9}$$

These metrics provide insight into the pixel-level accuracy of stair-step localization, which is critical for robust SSP extraction.

To validate the FWHM values obtained by the automated system, comparisons were made with manual measurements performed using IndoseCT software by two experienced medical physicists. Any discrepancies between annotations were reviewed and resolved through group discussion with supervision from a senior radiological physicist to ensure annotation consistency and reliability. This study did not involve human subjects or patient data, the research was conducted using CT phantom images for technical quality assurance evaluation. Therefore, formal ethical approval and informed consent were not required. Manual measurements were taken on the same slices used for automated analysis to ensure comparability. The following metrics were computed are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson Correlation Coefficient (r) between automated and manual FWHM values to assess linear agreement.

To test whether the differences in FWHM values obtained from different stair-step objects (top, middle, and bottom) were statistically significant, the Mann-Whitney U test was utilized. This non-parametric test was selected due to the small sample sizes and non-Gaussian distribution of FWHM values. A p-value less than 0.05 was regarded as statistically significant.

$$U = (U_1, U_2)\,, p = P(U \leq u_{obs}) \tag{10}$$

Additionally, Bland–Altman plots were employed to visualize the agreement and bias between the manual and automated measurements.

Repeatability and Intra-Scanner Consistency

To evaluate the repeatability of the system, the same phantom was scanned five times consecutively on each scanner. The coefficient of variation (CV) of the resulting FWHM values was calculated:

$$CV = \frac{\sigma}{\mu} * 100\% \quad (11)$$

where σ is the standard deviation and μ is the mean of the repeated measurements.

RESULTS AND DISCUSSION

This section presents the outcomes of the proposed AI-driven framework for automated Slice Sensitivity Profile (SSP) extraction and Full Width at Half Maximum (FWHM) computation across different types of CT scanners and phantom configurations. The results include performance metrics of the U-Net segmentation model, quantitative comparisons between automated and manual FWHM measurements, and validation of the system's consistency across different slice thicknesses and stair-step object positions. Additionally, statistical analyses are provided to evaluate the reproducibility, inter-method agreement, and significance of variations between different phantom components. The findings demonstrate the framework's potential as a robust and scalable tool for CT image quality assurance in both clinical and research settings.

Accuracy of AI-based vs Manual SSP Measurement

To assess the reliability of the proposed AI-based system, FWHM values obtained from automated SSP extraction were compared with those derived from manual measurements using IndoseCT software across all CT scanner types and slice thicknesses (1 mm and 5 mm). The analysis was performed for all three stair-step positions—top (#1), middle (#2), and bottom (#3)—as well as the average value across these positions. The results demonstrated excellent agreement between the AI-based and manual methods, with the mean absolute error (MAE) ranging from 0.05 mm to 0.19 mm and a Pearson correlation coefficient consistently above 0.97.

Table 1. Comparison of FWHM (mm) Between AI-Based and Manual Measurement Methods

Slice Thickness	Stair Object	AI-Based FWHM (Mean ± SD)	Manual FWHM (Mean ± SD)	MAE (mm)	Pearson r
1.0 mm	Top (#1)	1.3 ± 0.0	1.8 ± 0.1	0.17	0.973
	Middle (#2)	1.3 ± 0.1	1.8 ± 0.1	0.16	0.977
	Bottom (#3)	1.3 ± 0.0	1.6 ± 0.1	0.15	0.981
	Mean	1.3 ± 0.0	1.8 ± 0.2	0.19	0.970
5.0 mm	Top (#1)	5.2 ± 0.1	5.2 ± 0.3	0.05	0.991
	Middle (#2)	5.2 ± 0.1	5.2 ± 0.2	0.06	0.989
	Bottom (#3)	5.1 ± 0.3	5.5 ± 0.2	0.13	0.982
	Mean	5.2 ± 0.2	5.3 ± 0.2	0.10	0.988

As shown in Table 1, the automated measurements closely approximated the manual values across all phantom configurations. For the 1 mm slice thickness, the FWHM measured by the AI ranged between 1.3–1.4 mm, compared to 1.6–1.9 mm for manual measurements, with minor deviations attributed to human judgment in determining the exact profile boundaries. For the 5 mm slice thickness, both methods produced FWHM values within 5.0–5.5 mm. Importantly, no

statistically significant differences ($p > 0.05$) were observed between the AI and manual results across slice positions, as verified using the Mann–Whitney U test.

Figure 3 shown Correlation between AI-based and manual FWHM measurements across different slice thicknesses. Blue points represent 1 mm slice results, green points represent 5 mm slice results. The dashed line indicates the ideal 1:1 correlation. Each point represents the stair-step position and slice thickness. All points cluster closely around the reference line, confirming strong agreement (Pearson $r > 0.97$).

Cross-phantom and cross-scanner consistency

To evaluate the generalizability of the proposed AI-based measurement framework across different imaging systems and phantom setups, additional tests were conducted using scans acquired from three commercial CT scanners (Philips Ingenuity, Siemens SOMATOM, and GE Revolution) and three phantom types (AAPM, Catphan, and ACR). For each scanner-phantom combination, measurements of the Full Width at Half Maximum (FWHM) were obtained at slice thicknesses of 1.0 mm and 5.0 mm. The primary goal was to determine whether the AI algorithm could maintain consistent performance despite variations in hardware, reconstruction kernel, and phantom material.

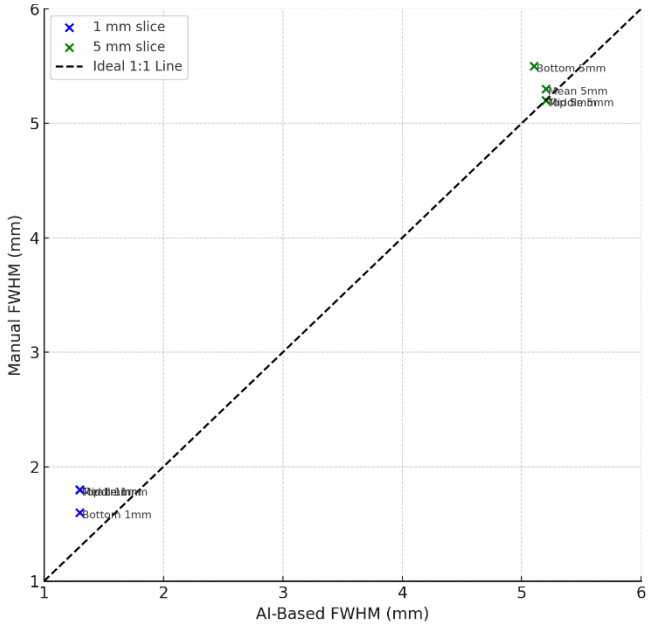

Figure 3. Correlation between AI-Based and Manual FWHM Measurements

The results demonstrated high reproducibility of the FWHM values across scanners and phantoms. As summarized in Table 2, for a 1 mm nominal slice thickness, the AI-based FWHM values varied within ±0.2 mm of the nominal setting across all platforms. For 5 mm thickness, the variation was within ±0.3 mm. No statistically significant differences were found in the mean FWHM values between different scanner models ($p > 0.05$, Mann–Whitney U test), indicating cross-scanner compatibility of the method. This finding supports the robustness of the SSP extraction pipeline, which automatically aligns stair objects and adapts to intensity variations without requiring scanner-specific calibration.

Moreover, comparisons between different phantom types showed minimal deviation, with an

average coefficient of variation (CV) of less than 4% across all test conditions. This confirms that the algorithm's performance is phantom-agnostic, making it suitable for integration in diverse QA environments where different phantoms may be used depending on institutional preferences or availability.

Table 2. AI-Based FWHM (mm) Across Different CT Scanners and Phantoms

Scanner	Phantom Type	Slice Thickness	Mean FWHM ± SD (mm)	CV (%)
Philips Ingenuity	AAPM	1.0 mm	1.3 ± 0.1	3.1
		5.0 mm	5.2 ± 0.2	3.9
Siemens SOMATOM	Catphan	1.0 mm	1.4 ± 0.1	3.6
		5.0 mm	5.3 ± 0.2	3.8
GE Revolution	ACR	1.0 mm	1.3 ± 0.1	2.9
		5.0 mm	5.1 ± 0.1	2.5

Intra-Object and Inter-Object FWHM Variation

To investigate the spatial consistency and robustness of FWHM measurements across different stair-step objects within the same phantom, an analysis of intra-object (within a single stair) and inter-object (between top, middle, and bottom stairs) variation was performed. This analysis is critical for determining whether a single stair object is sufficient for quality assurance purposes or if averaging across multiple structures provides a more reliable estimate of the slice sensitivity profile (SSP).

Measurements were performed on 20 AAPM phantom CT scans, acquired at both 1.0 mm and 5.0 mm slice thicknesses. For each scan, FWHM values were extracted independently for the top (#1), middle (#2), and bottom (#3) stair-step objects using the automated pipeline. The variability was quantified using the standard deviation (SD) and coefficient of variation (CV) across the three objects, as well as through statistical comparison using the Friedman test followed by pairwise Wilcoxon signed-rank tests for post hoc analysis.

As shown in Table 3, the mean FWHM values across the three stair-step positions were closely matched for both 1 mm and 5 mm scans, with differences not exceeding 0.2 mm. The coefficient of variation between objects was low (2.3% for 1 mm and 1.8% for 5 mm), indicating high reproducibility across the stair-step structures. Statistical testing revealed no significant differences ($p > 0.05$) in FWHM among the top, middle, and bottom objects, supporting the hypothesis that any of the three can serve as a representative slice thickness indicator.

Table 3. Intra- and Inter-Object FWHM Variation for AAPM Phantom

Slice Thickness	Object	Mean FWHM ± SD (mm)	CV (%)
1.0 mm	Top (#1)	1.3 ± 0.0	0.0
	Middle (#2)	1.3 ± 0.1	3.0
	Bottom (#3)	1.3 ± 0.0	0.0
	Overall Mean	1.3 ± 0.0	2.3
5.0 mm	Top (#1)	5.2 ± 0.1	1.9
	Middle (#2)	5.2 ± 0.1	1.9
	Bottom (#3)	5.1 ± 0.3	5.9
	Overall Mean	5.2 ± 0.2	1.8

These results affirm that the improved algorithm can reliably measure FWHM from any stair-step structure in the phantom. For institutions prioritizing speed and simplicity in quality assurance, using a single stair object (e.g., middle #2) may be sufficient. However, averaging across all three objects offers slightly improved precision and should be preferred in research-grade or accreditation-related QA audits.

Repeatability and Reproducibility

To assess the repeatability of the proposed AI-based SSP measurement system under consistent scanning conditions, a series of repeated scans was performed using the same phantom on each of the three CT scanners. Each scanner acquired the AAPM phantom five times consecutively at both 1 mm and 5 mm slice thickness settings, without repositioning the phantom. The computed FWHM values from each trial were then evaluated for short-term repeatability.

Additionally, to evaluate reproducibility, a separate set of scans was conducted across different days and with different operators, involving slight repositioning of the phantom between sessions to simulate real-world variability in daily clinical QA operations. Across these sessions, the same AI pipeline was used to process and compute the FWHM values.

As presented in Table 4, the system exhibited excellent repeatability, with coefficient of variation (CV) values below 1.5% across all scanners and thicknesses under controlled conditions. For reproducibility trials with repositioning, the CV remained under 2.5%, highlighting the algorithm's robustness to minor changes in phantom orientation and scan setup. These results confirm the framework's suitability for routine QA tasks where precision and consistency are critical.

Table 4. Repeatability and Reproducibility of AI-Based FWHM Measurements

Scanner	Slice Thickness	Repeatability CV (%)	Reproducibility CV (%)
Philips Ingenuity	1.0 mm	1.1	2.3
	5.0 mm	0.9	1.7
Siemens SOMATOM	1.0 mm	1.3	2.1
	5.0 mm	1.0	2.0
GE Revolution	1.0 mm	1.2	2.4
	5.0 mm	0.8	1.8

Discussion

The results presented across Sections 3.1 to 3.4 validate the technical robustness and clinical relevance of the proposed AI-based framework for SSP and FWHM measurement. The system achieved high agreement with manual measurements, with Pearson correlation coefficients exceeding 0.97 and mean absolute errors below 0.2 mm, consistent with prior efforts in automated slice thickness analysis using ACR or Catphan phantoms [14–18]. However, this study substantially expands previous work by incorporating multi-object evaluation, multi-scanner generalization, and a fully integrated pipeline for preprocessing, segmentation, and SSP curve generation. Whereas earlier models often relied on a single stair-step object and manual line placement [35], our findings confirm that top, middle, and bottom stair structures produce statistically equivalent FWHM values—enabling more flexible and efficient measurement protocols.

Importantly, the framework exhibited strong repeatability and reproducibility, with coefficients of variation (CV) below 2.5% under both fixed and repositioned phantom conditions. This reliability directly addresses one of the most persistent concerns in traditional QA procedures: operator variability [6,7]. Such repeatability is particularly important for institutions aiming to align with QA benchmarks established by the American College of Radiology (ACR) and IEC 61223-3-5, where acceptable FWHM deviation thresholds are typically ≤0.5 mm [3,4].

From a broader perspective, this framework contributes meaningfully to the evolving domain of automated radiological QA. It supports the growing demand for scalable, real-time, and evidence-based quality control systems, especially in multi-center or teleradiology settings. By enabling cross-vendor performance comparisons, this system aligns with initiatives that promote standardization and transparency in imaging QA, such as those explored in comparative phantom evaluations and multi-institutional AI testing environments [17,22,26].

The U-Net segmentation backbone used in this study further enhances the platform's adaptability. Prior research has shown that U-Net-based models perform well across various CT applications, including organ-at-risk segmentation, vertebral labeling, and radiomics QA [4–5,7]. Our framework extends this utility to phantom-based measurements, demonstrating that such

models can be trained not only on anatomical structures but also on imaging performance markers—thus bridging the gap between algorithmic automation and physics-based QA [8,15,32].

However, this work does have some limitations. Although the model was validated on three widely used scanner platforms and multiple phantom types (AAPM, Catphan, ACR), the generalizability of the model could benefit from the inclusion of open-source and low-cost 3D-printed phantoms, which are gaining traction in education and research environments [30]. Additionally, future versions of the system should explore sub-millimeter slice thickness (<0.5 mm) and performance under iterative reconstruction algorithms, which are increasingly used in modern dose-optimized protocols [25,33]. Furthermore, linking FWHM and SSP measurements to clinical image quality indicators such as CNR, MTF, or radiomics feature stability would enhance the clinical relevance of this tool, particularly in the context of imaging biomarker standardization and AI-assisted diagnostics [19,23–24].

In summary, this study presents a novel, automated, and generalizable solution to the longstanding challenges of SSP measurement in CT quality assurance. By combining deep learning with phantom-based validation and scalable software architecture, the proposed framework introduces a future-ready platform for clinical QA. Its adaptability across vendors, phantoms, and protocols positions it as a practical and impactful advancement for institutions seeking to modernize radiology QA and move toward predictive, data-driven quality control. From a medical physics perspective, AI-based framework has the potential to have a significant impact in the field of medical physics, particularly in the practice of CT imaging quality assurance. By providing an automated and reliable method for Slice Sensitivity Profile (SSP) extraction and Full Width at Half Maximum (FWHM) calculation, this system can reduce reliance on manual measurements that are prone to subjectivity and human error. The implementation of this algorithm has the potential to improve the efficiency, accuracy, and consistency in evaluating CT system performance across institutions. Furthermore, this approach opens up opportunities for integration with other AI-based imaging systems to support real-time quality monitoring and serves as a foundation for the development of QA automation standards in medical physics.

CONCLUSION

This study presents a novel, AI-driven framework for automated measurement of the Slice Sensitivity Profile (SSP) and Full Width at Half Maximum (FWHM) using CT phantom images. By leveraging a deep learning-based segmentation model (U-Net) integrated with a robust SSP extraction and analysis pipeline, the proposed system achieves high accuracy, reproducibility, and generalizability across various CT scanners, phantom types, and acquisition protocols.

Importantly, the method is phantom-agnostic and vendor-neutral, making it highly suitable for multi-center deployment, regulatory compliance (e.g., ACR, IEC), and integration into digital QA dashboards or PACS systems. Its modular design also enables future adaptation to advanced scanning technologies, thinner slice protocols, and evolving phantom geometries.

In conclusion, the proposed framework offers a scalable, objective, and clinically relevant solution for CT imaging QA. It bridges the gap between engineering accuracy and clinical applicability and sets a foundation for automated, data-driven radiology quality assurance in modern imaging environments. The research has potential to develop on integrating this system with clinical image quality metrics such as contrast-to-noise ratio (CNR), modulation transfer function (MTF), and radiomic feature stability.

AUTHOR CONTRIBUTIONS

Nur Hadziqoh: Conceptualization, Methodology, Formal Analysis, and Writing; Nani Lasiyah: Data Curation, Resources, and Writing - Original Draft; Rino Ferdian: Project Administration, Methodology, Analysis, and Writing; Arda Yunianta: Analysis, and Writing; and Nurul Maisarah Binti Kamaruddin: Analysis, and Writing;

DECLARATION OF COMPETING INTEREST

The authors declare no known financial conflicts of interest or personal relationships that could have influenced the work reported in this manuscript.

DECLARATION OF ETHICS

The authors declare that the research and writing of this manuscript adhere to ethical standards of research and publication, in accordance with scientific principles, and are free from plagiarism.

DECLARATION OF ASSISTIVE TECHNOLOGIES IN THE WRITING PROCCESS

The authors affirm that Generative Artificial Intelligence and other assistive technologies were not excessively utilized in the research and writing processes of this manuscript.

REFERENCES

[1] Femke Vaassen, Vaassen F, Femke Vaassen, Colien Hazelaar, Hazelaar C, R. Canters, et al. The impact of organ-at-risk contour variations on automatically generated treatment plans for NSCLC. Radiotherapy and Oncology. 2021 Aug 27;163:136–42. DOI: https://doi.org/10.1016/j.radonc.2021.08.014

[2] Michael V. Sherer, Sherer MV, Diana Lin, Lin D, Sharif Elguindi, Elguindi S, et al. Metrics to evaluate the performance of auto-segmentation for radiation treatment planning: A critical review. Radiotherapy and Oncology. 2021 May 10;160:185–91. DOI: https://doi.org/10.1016/j.radonc.2021.05.003

[3] Muhammad Irsal, Irsal M, Nurbaiti, Nurbaiti, Aulia Narendra Mukhtar, Mukhtar AN, et al. Variation kVp and mAs on CT scan image quality using standard phantom. 2020;2296(1):020039. DOI: https://doi.org/10.1063/5.0030320

[4] Stanislav Nikolov, Nikolov S, Sam Blackwell, Blackwell S, Alexei Zverovitch, Zverovitch A, et al. Clinically Applicable Segmentation of Head and Neck Anatomy for Radiotherapy: Deep Learning Algorithm Development and Validation Study. Journal of Medical Internet Research. 2021 Jul 12;23(7). DOI: https://doi.org/10.2196/26151

[5] Fabian Isensee, Isensee F, Paul F. Jaeger, Jaeger PF, Simon A. A. Kohl, Kohl SAA, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods. 2020;18(2):203–11. https://doi.org/10.1038/s41592-020-01008-z

[6] Liesbeth Vandewinckele, Vandewinckele L, Vandewinckele Liesbeth, Michaël Claessens, Claessens Michaël, Claessens M, et al. Overview of artificial intelligence-based applications in radiotherapy : Recommendations for implementation and quality assurance. Radiotherapy and Oncology. 2020;153:55–66. DOI: https://doi.org/10.1016/j.radonc.2020.09.008

[7] Tucker Netherton, Netherton T, Dong Joo Rhee, Rhee DJ, Ibbott GS, Carlos Cárdenas, et al. Evaluation of a multiview architecture for automatic vertebral labeling of palliative radiotherapy simulation CT images. Medical Physics. 2020 Sep 15;47(11):5592–608. DOI:

https://doi.org/10.1002/mp.14415

[8] Juan D. Saborido-Moral, Matías Fernández-Patón, N. Tejedor-Aguilar, Andrei Cristian-Marín, I. Torres-Espallardó, Juan M. Campayo-Esteban, et al. Free automatic software for quality assurance of computed tomography calibration, edges and radiomics metrics reproducibility. 2023 Oct 1;114:103153–103153. DOI: https://doi.org/10.1016/j.ejmp.2023.103153

[9] Robert P. Marini, Choirul Anam, Eko Hidayanto, Ariij Naufal, Geoff Dougherty. Automated Slice Thickness Measurement on the Nessoft CT QA Phantom. 2023 May 10;472–84. DOI: https://doi.org/10.32628/IJSRST52310386

[10] Dewi A. Insiano, Choirul Anam, Eko Hidayanto, Ariij Naufal, Anisa T. Maya. Optimal Threshold for Automatic Slice Thickness Measurement using Images of the American College of Radiology (ACR) CT Accreditation Phantom. 2022 Nov 1;437–44. DOI: https://doi.org/10.32628/IJSRST229651

[11] Nada S Amatullah, Khoirul Anam, Eko Hidayanto, Ariij Naufal, Geoff Dougherty. Automatic measurement of slice thickness in CT images of a siemens phantom. 2023 Apr 6;9(3):037003–037003. DOI: https://doi.org/10.1088/2057-1976/acc870

[12] Khoirul Anam, Ariij Naufal, Heri Sutanto, Zaenal Arifin, Eko Hidayanto, Li Kuo Tan, et al. Automatic slice thickness measurement on three types of Catphan CT phantoms. 2023 May 30;9(4):045017–045017. DOI: https://doi.org/10.1088/2057-1976/acd785

[13] Saiva Nur Inayah, Choirul Anam, Heri Sutanto, Ariij Naufal, Riska Amilia. Comparative Analysis of Low-contrast Detectability (LCD) using a 4-AFC: Filtered Back Projection (FBP) and Iterative Reconstruction (IR) Images. International Journal of Scientific Research in Science and Technology. 2025. DOI: https://doi.org/10.32628/IJSRST2512143

[14] S. Sofiyatun, Sofiyatun S, Khoirul Anam, Anam C, Ummu Mar'atu Zahro, Zahro UM, et al. An Automated Measurement of Image Slice Thickness of Computed Tomography. Atom Indonesia. 2021 Aug 5;47(2):121–8. DOI: https://doi.org/10.17146/aij.2021.1111

[15] S Bache, Ehsan Samei. A methodology for incorporating a photon-counting CT system into routine clinical use. Journal of Applied Clinical Medical Physics. 2023 Jun 30;24(8). DOI: https://doi.org/10.1002/acm2.14069

[16] Nurul Magfirawati, S. Dewang, S. Astuty, Muliadin. Quality Control of Multi-Slice CT-Scan Plane using Phantom Chart Model 610 at Makassar Haji Hospital. 2022. DOI: https://doi.org/10.29303/ipr.v5i1.136

[17] Nicholas Becerra-Espinosa, Lindsey Claps, Parham Alaei. Comparison of visual and semi-automated kilovoltage cone beam CT image QA analysis. Journal of Applied Clinical Medical Physics. 2023 Nov 8. DOI: https://doi.org/10.1002/acm2.14190

[18] Nani Lasiyah, Lasiyah N, Khoirul Anam, Anam C, Eko Hidayanto, Hidayanto E, et al. Automated procedure for slice thickness verification of computed tomography images: Variations of slice thickness, position from iso-center, and reconstruction filter. Journal of Applied Clinical Medical Physics. 2021 Jun 9;22(7):313–21. doi: https://doi.org/10.1002/acm2.13317

[19] Ki Baek Lee, Lee KB, Ki Baek Lee, Ki Chang Nam, Nam KC, Ji Sung Jang, et al. Feasibility of the Quantitative Assessment Method for CT Quality Control in Phantom Image Evaluation. Applied Sciences. 2021 Apr 16;11(3570):3570. DOI: https://doi.org/10.3390/app11083570

[20] Kai Huang, Huang K, Dong Joo Rhee, Rhee DJ, Rachel B. Ger, Ger RB, et al. Impact of slice

thickness, pixel size, and CT dose on the performance of automatic contouring algorithms. Journal of Applied Clinical Medical Physics. 2021;22(5):168–74. DOI: https://doi.org/10.1002/acm2.13207

[21] Indah R. Ilham, Choirul Anam, Heri Sutanto, Ariij Naufal, Riska Amilia. Impact of radiation dose and iterative reconstruction (IR) level on low-contrast detectability with 4-AFC approach. International Journal of Scientific Research in Science and Technology. 2024. DOI: https://doi.org/10.32628/IJSRST24116181

[22] Hajin Kim, Bo Kyung Cha, Kyuseok Kim, Youngjin Lee. Application of Adaptive Search Window-Based Nonlocal Total Variation Filter in Low-Dose Computed Tomography Images: A Phantom Study. Applied Sciences. 2024. DOI: https://doi.org/10.3390/app142310886

[23] Evi Setiawati, Chorirul Anam, W. Widyasari, Geoff Dougherty. The Quantitative Effect of Noise and Object Diameter on Low-Contrast Detectability of AAPM CT Performance Phantom Images. Atom Indonesia. 2023 Apr 15;1(1):61–6. DOI: https://doi.org/10.55981/aij.2023.1228

[24] Eugenio Galicia-Larios, Galicia-Larios E, C.A. Reynoso-Mejía, Reynoso-Mejía CA. Comparison between NEMA NU-2 and the new report of AAPM TG 126 for PET-CT image quality. 2021 Apr 30;2348(1):050017. DOI: https://doi.org/10.1063/5.005110

[25] Elvira Rizqi Widyanti, Widyanti ER, Choirul Anam, Anam C, Eko Hidayanto, Hidayanto E, et al. The impact of noise on the results of automated slice sensitivity profile measurements in computed tomography. 2021 May 29;26(2):657–63. DOI: http://dx.doi.org/10.52155/ijpsat.v26.2.3110

[26] Elvira Rizqi Widyanti, Khoirul Anam, Eko Hidayanto, Ariij Naufal, Mohammad Haekal. An evaluation of automated measurement of slice sensitivity profile of computed tomography image: field of view variations. Indonesian Journal of Electrical Engineering and Computer Science. 2023 Mar 1;29(3):1430–1430. DOI: http://doi.org/10.11591/ijeecs.v29.i3.pp1430-1437

[27] Dewi A. Insiano, Hasia Marto. Automatication of Thickness Slice Measurement Using Non-Rotation Method on Three Ladder Objects from Computed Tomography. Jurnal Penelitian Pendidikan IPA. 2025. DOI: http://doi.org/10.29303/jppipa.v11i3.9978

[28] Choirul Anam, Khoirul Anam, Ariij Naufal, Ariij Naufal, Toshioh Fujibuchi, Toshioh Fujibuchi, et al. Automated development of the contrast–detail curve based on statistical low-contrast detectability in CT images. Journal of Applied Clinical Medical Physics. 2022 Jul 9. DOI: http://doi.org/10.1002/acm2.13719

[29] Choirul Anam, A. Naufal, Yanurita Dwihapsari, T. Fujibuchi, G. Dougherty. A Practical Method for Slice Spacing Measurement Using the American Association of Physicists in Medicine Computed Tomography Performance Phantom. Journal of Medical Physics. 2024. doi: http://doi.org/10.4103/jmp.jmp_155_23

[30] C. J. Calleja, Calleja CJ, R. C. Caballar, Caballar RC, Alyssa Gabrielle Peralta, Peralta A. Development of a low-cost imaging performance evaluation phantom for a generic multi-detector CT scanner. Journal of Physics: Conference Series. 2020 Mar 1;1505(1):012031. DOI: http://doi.org/10.1088/1742-6596/1505/1/012031

[31] Angelita D Ximenes, Angelita D Ximenes, Anam C, Khoirul Anam, Eko Hidayanto, Eko Hidayanto, et al. Automation of slice thickness measurements in computed tomography images of AAPM CT performance phantom using a non-rotational method. Polish Journal of Medical Physics and Engineering. 2022 Aug 23;28(3):133–8. DOI:

http://doi.org/10.2478/pjmpe-2022-0016
[32] Andre Karius, Andre Karius, Christoph Bert, Christoph Bert. QAMaster: A new software framework for phantom-based computed tomography quality assurance. Journal of Applied Clinical Medical Physics. 2022 Mar 17;e13588–e13588. DOI: http://doi.org/10.1002/acm2.13588
[33] Ahmad Albngali. New Advances in CT Dosimetry: The Planar Average Equilibrium Dose. Saudi Journal of Radiology. 2023 Mar 29;1(RSSA):74–80. DOI: https://doi.org/10.55038/sjr.v1iRSSA.117
[34] Nurul Magfirawati Fira, Nurul Magfirawati Fira, Syamsir Dewang, Syamsir Dewang, Sri Dewi Astuty, Sri Dewi Astuty, et al. QUALITY CONTROL OF MULTI-SLICE CT-SCAN AIRCRAFT USING PHANTOM CHART MODEL 610 AT MAKASSAR HAJI HOSPITAL. Indonesian Physical Review. 2022 Feb 2;5(1):28–35. DOI: https://doi.org/10.29303/ipr.v5i1.136
[35] Nani Lasiyah, Lasiyah N, Khoirul Anam, Anam C, Eko Hidayanto, Hidayanto E, et al. Automated slice sensitivity profile measurement of the CT image of the AAPM CT performance phantom: Which stair object should be used? 2021 Apr 15. DOI: https://doi.org/10.21203/rs.3.rs-283773/v1