



Deep Learning–Based Estimation of Radiographic Position to Automatically Set Up the X-Ray Prime Factors

C. F. Del Cerro^{1,2} · R. C. Giménez¹ · J. García-Blas³ · K. Sosenko⁴ · J. M. Ortega⁴ · M. Desco^{1,2,5,6} · M. Abella^{1,2,5}

Received: 11 June 2024 / Revised: 12 August 2024 / Accepted: 30 August 2024 / Published online: 14 October 2024
© The Author(s) 2024

Abstract

Radiation dose and image quality in radiology are influenced by the X-ray prime factors: KVp, mAs, and source-detector distance. These parameters are set by the X-ray technician prior to the acquisition considering the radiographic position. A wrong setting of these parameters may result in exposure errors, forcing the test to be repeated with the increase of the radiation dose delivered to the patient. This work presents a novel approach based on deep learning that automatically estimates the radiographic position from a photograph captured prior to X-ray exposure, which can then be used to select the optimal prime factors. We created a database using 66 radiographic positions commonly used in clinical settings, prospectively obtained during 2022 from 75 volunteers in two different X-ray facilities. The architecture for radiographic position classification was a lightweight version of *ConvNeXt* trained with fine-tuning, discriminative learning rates, and a one-cycle policy scheduler. Our resulting model achieved an accuracy of 93.17% for radiographic position classification and increased to 95.58% when considering the correct selection of prime factors, since half of the errors involved positions with the same KVp and mAs values. Most errors occurred for radiographic positions with similar patient pose in the photograph. Results suggest the feasibility of the method to facilitate the acquisition workflow reducing the occurrence of exposure errors while preventing unnecessary radiation dose delivered to patients.

Keywords Radiography · Radiographic position · Prime factors · Deep learning · Classification

Introduction

The quality of a radiograph, i.e., contrast resolution and signal-to-noise ratio, depends on the optimal tuning of the so-called prime factors: source voltage (KVp), source current

and exposure time (mAs), and distance from the source to the detector (SDD). These factors that depend on the radiographic position are typically adjusted by a technician prior to acquisition using pre-calculated values based on standards [1] and guidelines from the system manufacturer. However, a wrong

✉ M. Desco
desco@hggm.es

✉ M. Abella
mabella@ing.uc3m.es

C. F. Del Cerro
carlosfe@pa.uc3m.es

R. C. Giménez
100383011@alumnos.uc3m.es

J. García-Blas
fjblas@inf.uc3m.es

K. Sosenko
konstantin.sosenko@sedecal.com

J. M. Ortega
josemaria.ortega@sedecal.com

¹ Dept. Bioingeniería, Universidad Carlos III de Madrid, Leganés, Madrid, Spain

² Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain

³ Dept. de Informática, Universidad Carlos III de Madrid, Madrid, Spain

⁴ Sociedad Española de Electromedicina y Calidad S.A. (SEDECAL), Madrid, Spain

⁵ Centro Nacional de Investigaciones Cardiovasculares Carlos III (CNIC), Madrid, Spain

⁶ Centro de Investigación en Red en Salud Mental (CIBERSAM), Madrid, Spain

selection of these factors can result in non-diagnostic images, requiring repetition and exposing the patient to additional ionizing radiation. In fact, studies have shown that the repeat/rejection ratio in clinical practice is between 13% and 14% [2]. While most modern systems have an *Automatic Exposure Control* (AEC) that interrupts the radiation when the exposure is sufficient at the detector [3], it is not effective when prime factors are smaller than necessary. Therefore, an automatic mechanism to adjust prime factors to optimal values depending on the patient pose would be highly beneficial.

To the best of our knowledge, no other work has been published that addresses this problem. Several studies have explored the use of *convolutional neural networks* (CNN) to automatically detect radiographic positions in the acquired radiography for quality control [4, 5]. Other studies have used pretrained networks such as *VGG* [6] or *ResNet* [7] to classify chest radiographs into anteroposterior or posteroanterior views [8, 9], or to detect the radiographic position of hands [10]. However, these methods rely on acquired X-ray images for position detection, which does not prevent the patient from receiving an extra radiation dose in case of errors.

On the other hand, the identification of human body parts in photographs has been done with *You Only Look Once (YOLO)* [11], but this model is not able to distinguish between different human poses [12]. Other works described the human pose by a complete representation using key points [13], in a variety of clinical applications: estimating the position of clinicians for surgical workflow analysis [14, 15] or in-bed patients for monitoring [16], and analyzing infant motion for early detection of cerebral palsy [17]. Nevertheless, an accurate estimation of key points involves manual annotation, which is time-consuming and not necessary for the classification problem of radiographic position estimation. In [18], the authors optimized a transfer learning methodology for a CNN to classify human activities, but the application of this approach for radiographic

position estimation in a clinical setting is yet to be evaluated. In this work, we intend to fill this gap by developing an automatic method for the detection of the radiographic position of the patient to set up the X-ray prime factors. Our main contributions include the creation of an extensive database containing photographs of patients in a clinical setting for the most common radiographic positions used in clinical practice. We combined transfer learning with a cutting-edge deep-learning architecture to develop a model intended for real-time applications that uses a photograph of the patient positioned in the system to estimate the radiographic position.

Material and Methods

The proposed method estimates the radiographic position from a photograph of a patient. This information is used to determine the parameters of the X-ray source (kVp and mAs) by looking up a predefined table to ensure image quality. The following sections describe the database, the resulting model, the evaluation methodology, and the creation of an executable program to be used in a real system.

Database

We selected 66 radiographic positions commonly used in clinical settings (see Table 1), from 75 volunteers. These positions were chosen in collaboration with radiology technicians from Gregorio Marañón University Hospital (Madrid, Spain). Data acquisition was conducted in two locations, one for 50 volunteers and a different one for the remaining 25 volunteers. Although no radiation was involved in the acquisition, all participants signed a legal consent that recognizes the protection of their data as established in national and international regulations (General Data Protection Regulation, EU Regulation 2016/679).

Table 1 List of the 66 radiographic positions used in the study, categorized by body region

Body region	Positions
Head	Skull PA, Skull LAT, Sinus, Nasal Bones LAT
Spine	Cervical Spine AP, Cervical Spine LAT, Thoracic Spine AP, Thoracic Spine LAT, Lumbar Spine AP, Lumbar Spine LAT
Shoulder	Shoulder Left AP, Shoulder Left Axial, Shoulder Right AP, Shoulder Right Axial
Upper limb	Elbow Left AP, Elbow Left LAT, Elbow Right AP, Elbow Right LAT, Wrist Left AP, Wrist Left LAT, Wrist Right AP, Wrist Right LAT, Hand Left AP, Hand Left OBL, Hand Right AP, Hand Right OBL, Fingers Left 1st AP, Fingers Left 1st LAT, Fingers Left 3rd AP, Fingers Left 3rd LAT, Fingers Right 1st AP, Fingers Right 1st LAT, Fingers Right 3rd AP, Fingers Right 3rd LAT
Torso	Thorax PA, Thorax LAT, Abdomen AP, Ribs Left AP, Ribs Right AP, Pelvis AP, Hip Left AP, Hip Left Axial, Hip Right AP, Hip Right Axial, Sacro-coccyx AP, Sacro-coccyx LAT
Lower limb	Femur Left AP, Femur Left LAT, Femur Right AP, Femur Right LAT, Knee Left AP, Knee Left LAT, Knee Right AP, Knee Right LAT, Tibia Left AP, Tibia Left LAT, Tibia Right AP, Tibia Right LAT, Ankle Left AP, Ankle Left LAT, Ankle Right AP, Ankle Right LAT, Foot Left AP, Foot Left OBL, Foot Right AP, Foot Right OBL

The data acquisition setup in both facilities consisted of an X-ray table, a wall stand, and a ceiling-mounted X-ray source equipped with a YI 4 K Action camera attached to the collimator case using a monopod. To minimize the influence of background elements, we cropped the photographs to a square region with a width of 90 cm, corresponding to the width of the X-ray table (red square in Fig. 1), considering the 19 cm offset on the vertical axis of the camera with respect to the collimator light.

The pixel size to calculate the cropped area was determined using Eq. 1, as illustrated in Fig. 2.

$$\begin{aligned}
 PixSize &= FOV_{mm}/FOV_{px} = (h_s \times SOD_{cm}/LSD_{cm})/FOV_{px} \\
 &= m \times SOD_{cm} = m \times (SDD_{cm} - PDD_{cm})
 \end{aligned}
 \tag{1}$$

where h_s is the height of the camera sensor, LSD_{cm} is the distance between the lens and camera sensor, FOV_{px} is the field of view in pixels, m is a camera-dependent slope, SDD_{cm} is the distance between the source and detector, and PDD_{cm} is the distance between the patient and detector (10 cm in our case). The calculated pixel size was 0.0513 cm, 0.0798 cm, and 0.0969 cm for SDD of 100, 150, and 180, respectively.

Proposed Model

We adapted the lightweight *atto* version of the *ConvNeXt* architecture (CN) [19] from *Pytorch Image Models* (timm) [20], to accommodate the 66 radiographic positions considered in our study and enable real-time inference (Fig. 3). The loss function was cross-entropy, and we used the Ranger optimizer [21].

We conducted three experiments, resulting in a total of seven models. In the first experiment, we compared the performance of using color and grayscale images to determine the optimal color mode. In the second experiment, we used the selected color mode and evaluated the impact of three types of data augmentation and their combinations on the training set. This doubled the size of the database for one augmentation and increased it fourfold when combining all

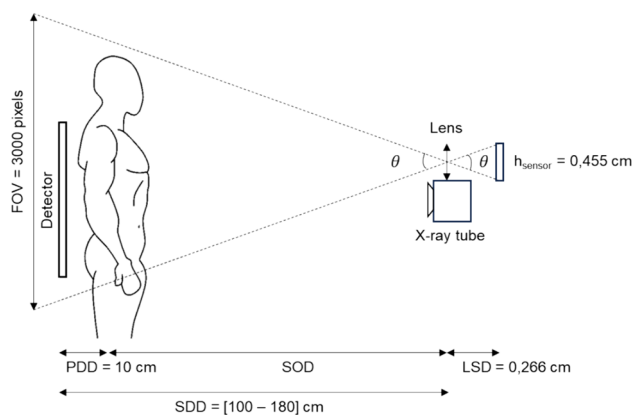


Fig. 2 Ray-tracing diagram of the FOV with respect to the camera lens and sensor size

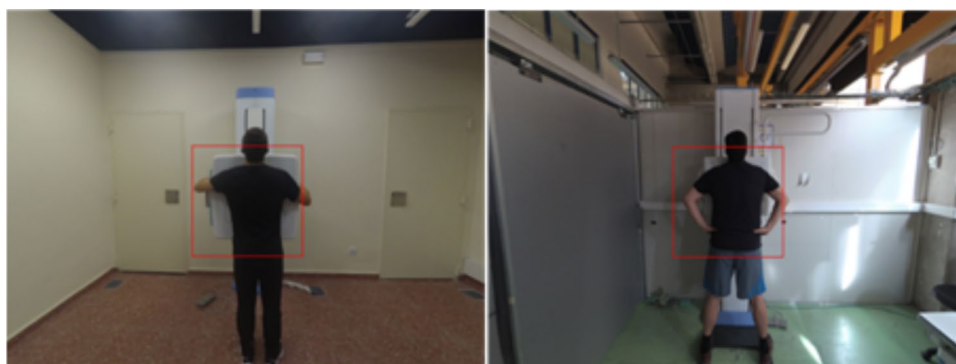
the augmentations. Data augmentations were horizontal flip, random zoom with a scale factor between 1.1 and 1.3, and random rotation of an angle θ drawn from a Gaussian distribution with a mean of 0 and a standard deviation of 10 degrees. We trained end-to-end these six pretrained models pretrained with *ImageNet-1K* [22] and obtained an optimal learning rate of 1×10^{-3} with the Leslie N. Smith test [23]. In the third experiment, we added fine-tuning (FT), discriminative learning rates (DLR) [24], and a one-cycle policy (OCP) scheduler [23]. The Leslie N. Smith indicated an optimal learning rate of 1×10^{-3} when all layers were frozen except for the last one. When training the model end-to-end, we used discriminative learning rates of 1×10^{-6} , 1×10^{-5} , and 1×10^{-4} for the first, second, and last third of our model, respectively, as determined by the same test.

Input images were rescaled to 224×224 pixels and normalized based on the mean and standard deviation of *ImageNet-1K*.

Implementation

The proposed model was implemented in *Python* [20], using *PyTorch* [25] and *Fastai* [26] libraries. We then developed

Fig. 1 Example of thorax AP position at both facilities



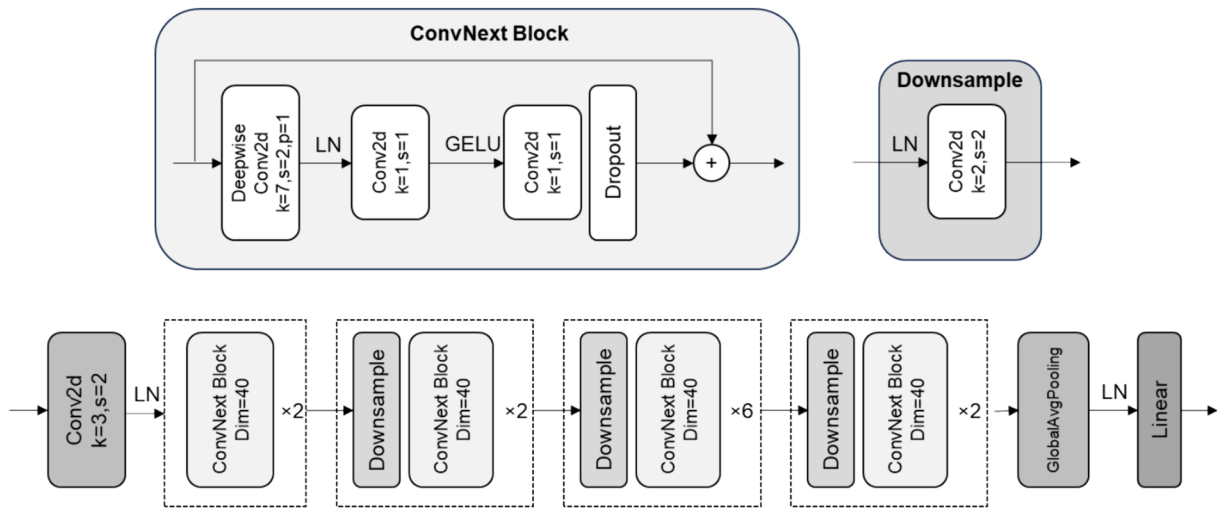
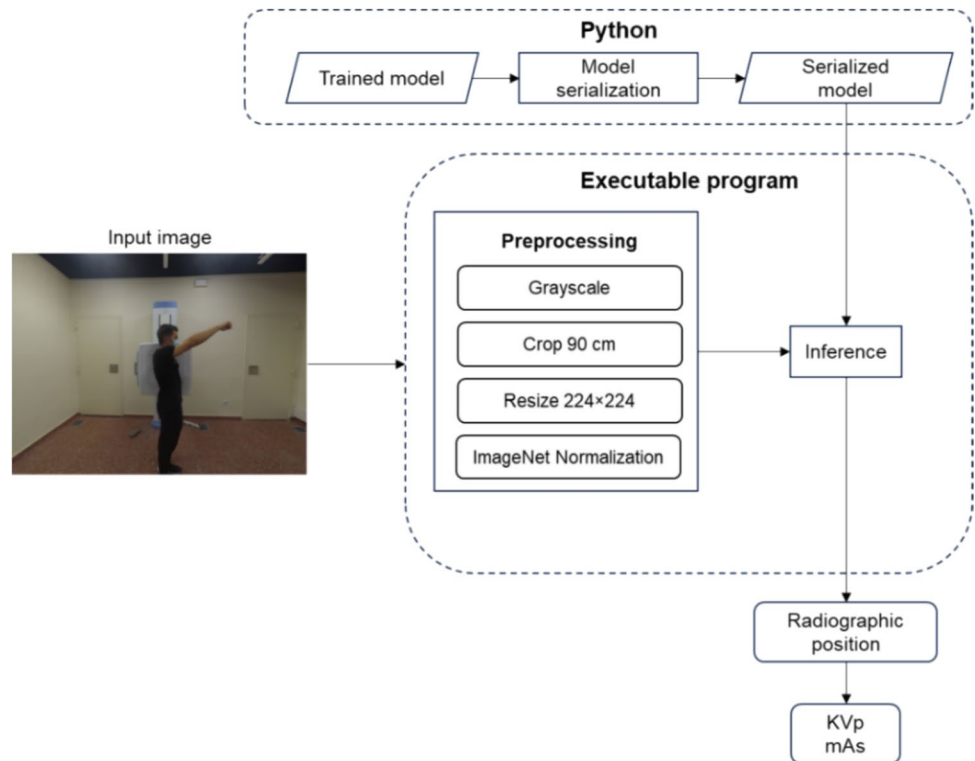


Fig. 3 Proposed architecture. Conv2d, convolution layers; LN, normalization layer; k, kernel size; s, stride; and p, padding

an executable program using *LibTorch*, a C++ library that provides an API for *PyTorch*, and *OpenCV* [27] for image preprocessing. To ensure compatibility between *PyTorch* and *LibTorch*, we used *Torch Script*. The program was compiled for Windows® machines with $\times 86-64$ processor architecture and can perform image inference using only the CPU, without requiring *Python* installation. The workflow involves

converting the photograph to grayscale, cropping it to a 90-cm square region, resizing it to 224×224 pixels, and normalizing it based on the mean and standard deviation of *ImageNet-1 K* (Fig. 4). This preprocessed image is then passed through our model to infer the radiographic position, which is used to automatically determine the appropriate KVp and mAs values based on a predefined table.

Fig. 4 Workflow of the conversion from trained models to the final executable



Evaluation Methodology

To verify the robustness of the models given the limited size of the database, we performed 5-fold cross-validation with five randomly chosen validation sets for all experiments, splitting the dataset into 60 volunteers for training and 15 for validation.

Since some radiographic positions share the same prime factors, we evaluated the accuracy of the proposed models in terms of both radiographic position identification and prime factor values. To assess classification errors, we used confusion matrices generated from the validation sets.

We compared our architecture with two lightweight state-of-the-art models with a similar number of parameters (3.7M), EfficientFormerV2 (EF) [28] and MobileNetV4 (MN) [29], both from the *timm* library. The chi-square test was used to assess differences in the accuracy rates for each model.

Results

Table 2 presents the mean and standard deviation of the 5-fold cross-validation results for each architecture, considering both grayscale (GS) and color (RGB) models. Our model, CN, achieves superior performance ($p < 0.05$) compared to EF and MN in both radiographic position identification and prime factors accuracy. While the performance

using GS and RGB models was very similar, with no significant differences between them, GS showed a shorter inference time (using only one channel).

Table 3 presents the mean and standard deviation of the 5-fold cross-validation results for different data augmentations and training policies for the proposed architecture, CN. The use of individual data augmentations or the combination of all of them did not significantly improve the accuracy of the grayscale model. However, the addition of fine-tuning, discriminative learning rates, and OCP reduced cross-entropy and improved both radiographic position and prime factors accuracy by 2.21% and 1.57% respectively ($p < 0.05$).

The left panel of Fig. 5 displays the confusion matrix, with most predictions falling along the diagonal, indicating correct classification. The right panel of Fig. 5 shows that errors in prime factors correspond to similar radiographic positions, such as Skull PA and Sinus (five errors) or Sacro-coccyx AP and Pelvis AP (four errors), which are marked with a red square.

Errors in radiographic position classification occurred in 6% of validation set cases. These errors can be categorized into four classes: similar patient pose with the same prime factors (35%), similar patient pose with different prime factors (46%), different patient poses with the same prime factors (4%), and different patient poses with different prime factors (15%). Fig. 6 provides an example of each class.

Table 2 Results obtained for each color mode with three architectures. The best result is highlighted in bold, and the second-best result is underlined. * $p < 0.05$ with respect to CN (GS)

Model	Time/epoch (s) ↓	Best epoch ↓	Cross-entropy ↓	Position acc (%) ↑	Prime factors acc. (%) ↑
CN (GS)	12	<u>7 ± 1</u>	0.27 ± 0.03	91.15 ± 1.06	<u>94.10 ± 0.93</u>
CN (RGB)	14	<u>7 ± 1</u>	<u>0.27 ± 0.04</u>	<u>91.15 ± 1.28</u>	94.57 ± 1.34
EF (GS)	12	6 ± 0	0.41 ± 0.03	86.48 ± 1.24*	92.01 ± 1.24*
EF (RGB)	14	15 ± 18	0.37 ± 0.03	88.77 ± 1.47*	93.48 ± 0.97*
MN (GS)	8	26 ± 9	0.46 ± 0.03	87.90 ± 1.20*	91.51 ± 1.15*
MN (RGB)	<u>10</u>	22 ± 12	0.43 ± 0.02	87.82 ± 0.95*	91.62 ± 0.76*

Table 3 Results obtained for the grayscale models with the proposed architecture CN(GS), adding different data augmentations and training policies. The best result is highlighted in bold, and the second-best result is underlined. * $p < 0.05$ with respect to CN (GS)

Model	Time/epoch (s) ↓	Best epoch ↓	Cross-entropy ↓	Position acc (%) ↑	Prime factors acc. (%) ↑
CN (GS + flip)	24	5 ± 1	0.27 ± 0.02	91.00 ± 0.41	93.85 ± 0.88
CN (GS + zoom)	24	4 ± 1	0.27 ± 0.03	90.98 ± 0.74	94.22 ± 0.98
CN (GS + rotation)	24	5 ± 1	0.27 ± 0.02	91.66 ± 1.00	94.49 ± 1.05
CN (GS + flip + zoom + rotation)	<u>46</u>	3 ± 2	<u>0.27 ± 0.01</u>	<u>91.77 ± 0.88</u>	<u>94.54 ± 0.39</u>
CN (GS + flip + zoom + rotation (FT + DLR + OCP))	<u>46</u>	5 ± 0	0.21 ± 0.02	93.17 ± 0.80*	95.58 ± 0.82*

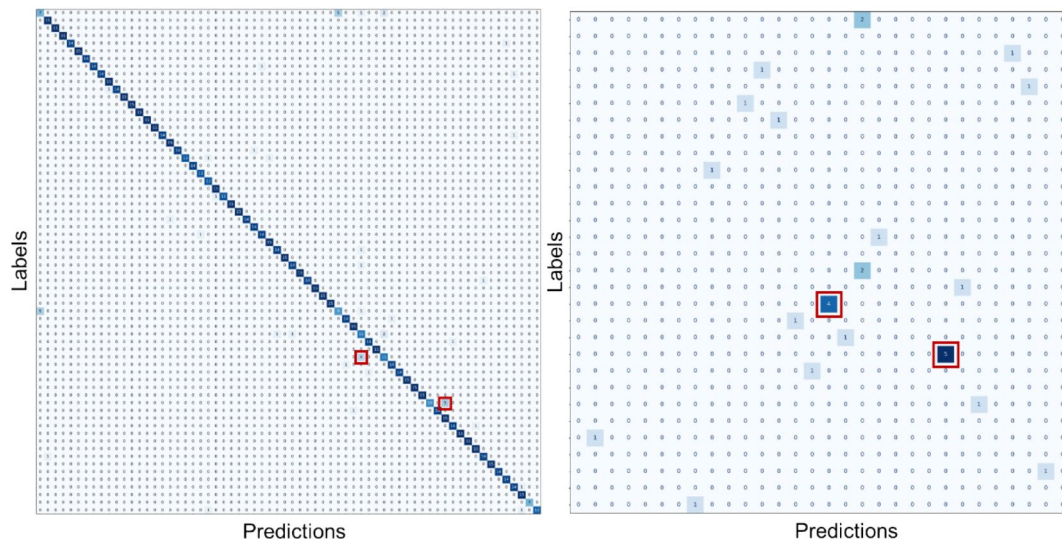


Fig. 5 Confusion matrix (left) and errors with different prime factors (right) for one of the validation sets

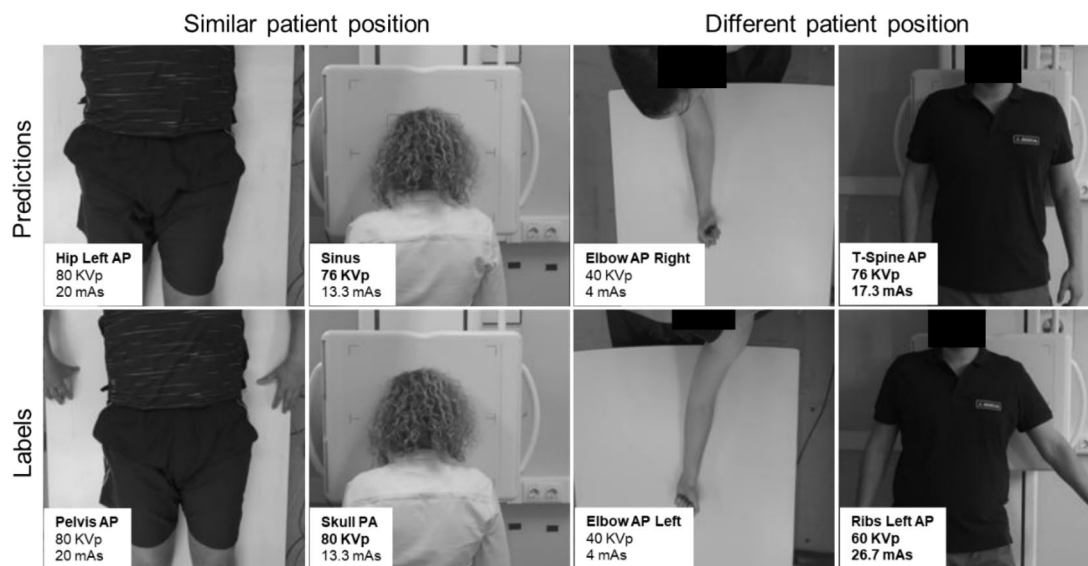


Fig. 6 Scenarios with similar (left) and different (right) patient poses. Errors are highlighted in bold

Discussion and Conclusion

We have presented a novel method that estimates the radiographic position from photographs to automatically select the prime factors in real time for X-ray acquisition. To achieve this, we created a database containing photographs of patients in a clinical setting, covering the most common radiographic positions. These photographs were taken with a camera attached to the collimator case, preprocessed, and fed into *ConvNeXt*. This architecture was chosen for

its superior balance between efficiency and accuracy, making it well suited for real-time applications.

The comparison of the proposed architecture with previously proposed architectures EfficientFormerV2 and MobileNetV4 showed statistically significant better results. The combination of data augmentation techniques and the application of training policies resulted in statistically significant improvements in performance, achieving an accuracy of 93.17% for radiographic position classification, despite the limited size of our database. Most

errors occurred for positions with a similar patient pose in the photograph. The accuracy increased to 95.88% when evaluating the correct selection of prime factors, as half of the errors involved positions with the same KVP and mAs values. Only 1% of the total images corresponded to errors with different patient poses and different prime factors. These errors may be attributed to inconsistencies in our database, as some radiographic positions were acquired at different distances in the two facilities. This problem could be minimized by increasing the database with more examples of these radiographic positions. Nevertheless, the remaining errors would not represent a major issue in a real scenario, as the acquisition would still be supervised by an X-ray technician.

Since the prime factors vary slightly depending on the anatomy of the patient, future research will evaluate the use of 3D cameras to complement the pose information with patient thickness data, estimated from the distance between the patient and the camera sensor. This will involve evaluating changes in the prime factors according to the individual morphology of the patient and creating the corresponding database.

The variation in X-ray room layouts could be a potential challenge that was partly addressed by implementing the region-of-interest cropping strategy. This could be further enhanced by increasing the database with acquisitions from more facilities or by generating synthetic data with different backgrounds [30].

The runtime and performance of our prototype demonstrate the feasibility of our method for real-time use, enabling the reduction of exposure errors and unnecessary radiation dose delivered to patients due to retakes. The implementation of our method in a real system only requires a camera, an executable installed on the control PC, and a new GUI workflow which automatically selects the positions to be confirmed by the X-ray technician. Since radiography is the most common medical imaging modality used in the clinics [31], the incorporation of this method in radiological systems can significantly impact the healthcare system in terms of patient safety and radiology department workflow optimization.

Author Contribution Conceptualization, MA, KS, JMO, and MD; methodology, MA, MD, and CFC; software, CFC, RCG, and JGB; validation, CFC, KS, JMO, and MA; formal analysis, MA and MD; investigation, MA, CFC, and RCG; resources, MA, JGB, KS, JMO, and MD; data curation, CFC, KS, and JMO; writing—original draft, CFC, RCG, JGB, and MA; writing—review and editing, KS, JMO, MA, and MD; supervision, MA and MD; project administration, MA; funding acquisition, MA, JGB, and MD. All authors have read and agreed to the published version of the manuscript.

Funding This work was supported by Ministerio de Ciencia e Innovación, Agencia Estatal de Investigación, and co-funded by

the European Regional Development Fund, “A way of making Europe”: PDC2021-121656-I00 (MULTIRAD), funded by MCIN/AEI/10.13039/501100011033 and by the European Union “NextGenerationEU”/PRTR. This is also funded by Instituto de Salud Carlos III through the projects PMPTA22/00121 and PMPTA22/00118, co-funded by the European Union “NextGenerationEU”/PRTR, and by the ASPIDE Project funded by the European Union’s Horizon 2020 Research and Innovation Programme under grant 801091. The CNIC is supported by Instituto de Salud Carlos III, Ministerio de Ciencia e Innovación, and the Pro CNIC Foundation.

Declarations

Ethics Approval This is an observational study. Approval was granted by the Ethics Committee of the University Carlos III de Madrid.

Consent to Participate All participants signed a legal consent which recognizes the protection of their data as established in the “Ley Orgánica de Protección Jurídica del Menor” (Ley O. 1/96, of January 15th), “Ley General de Sanidad” (Article 10.3, Ley 14/1986, of April 25th), “Ley Básica de Autonomía del Paciente y de Información y Documentación Clínica” (Chapters I and III Ley 41/2002, of November 14th), “Ley Orgánica 3/20 18” (of December 5th regarding the Protection of Personal Data and the Guarantee of Digital Rights), and the General Data Protection Regulation (EU Regulation 2016/679).

Consent for Publication The authors affirm that human research participants provided informed consent for the publication of the images in Figure(s) 1, 4, and 6.

Competing Interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Panzer, W., P. Shrimpton, Jessen K.: European Guidelines on Quality Criteria for Computed Tomography. Office for Official Publications of the European Communities, 2020
2. Foos, D. H., Sehnert, W. J., Reiner, B., Siegel, E. L., Segal, A., Waldman, D. L.: Digital radiography reject analysis: data collection methodology, results, and recommendations from an in-depth investigation at two hospitals. *Journal of digital imaging*: 22, 89–98, 2009
3. Bushberg, J. T., Boone J. M.: *The essential physics of medical imaging*. Lippincott Williams & Wilkins, 2011
4. Fang, X., Harris, L., Zhou, W., and Huo, D., Generalized radiographic view identification with deep learning. *Journal of Digital Imaging*: 34(1), 66–74, 2021
5. Mairhöfer, D., Laufer, M., Simon, P. M., Sieren, M., Bischof, A., Käster, T., Barth E., Barkhausen J., Martinetz, T.: An AI-based framework for diagnostic quality assessment of ankle radiographs.

- International Conference on Medical Imaging with Deep Learning, 2021
6. Simonyan, K., Zisserman A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014
 7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*: 770–778, 2016
 8. Kim, T. K., Yi, P. H., Wei, J., Shin, J. W., Hager, G., Hui, F. K., Sair, H. I., Lin, C. T.: Deep learning method for automated classification of anteroposterior and posteroanterior chest radiographs. *Journal of digital imaging*: 32, 925–930, 2019
 9. Hosch, R., Kroll, L., Nensa, F., Koitka, S.: Differentiation between anteroposterior and posteroanterior chest X-ray view position with convolutional neural networks. In *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren*: 193(2), 168–176, 2021
 10. Saun, T.J.: Automated classification of radiographic positioning of hand X-rays using a deep neural network. *Plastic Surgery*: 29(2), 75–80, 2021
 11. Wang, C. Y., Yeh, I. H., Liao, H. Y. M.: Yolov9: Learning what you want to learn using programmable gradient information. arXiv preprint arXiv:2402.13616, 2024
 12. Medaramatla, S. C., Samhitha, C. V., Pande, S. D., Vinta, S. R.: Detection of Hand Bone Fractures in X-ray Images using Hybrid YOLO NAS. *IEEE* Accessed 2024
 13. Zheng, C., Wu, W., Chen, C., Yang, T., Zhu, S., Shen, J., Kehtarnavaz N., Shah, M.: Deep learning-based human pose estimation: A survey. *ACM Computing Surveys*: 56(1), 1–37, 2023
 14. Kadkhodamohammadi, A., Gangi, A., de Mathelin, M., Padoy, N.: Articulated clinician detection using 3D pictorial structures on RGB-D data. *Medical image analysis*: 35, 215–224, 2017
 15. Srivastav V., Gangi A., and Padoy N.: Unsupervised domain adaptation for clinician pose estimation and instance segmentation in the operating room. *Medical Image Analysis*: 80, 102525, 2022
 16. Bigalke, A., Hansen, L., Diesel, J., Hennigs, C., Rostalski, P., Heinrich, M. P.: Anatomy-guided domain adaptation for 3D in-bed human pose estimation. *Medical Image Analysis*: 89, 102887, 2023
 17. Ni, H., Xue, Y., Ma, L., Zhang, Q., Li, X., Huang, S. X.: Semi-supervised body parsing and pose estimation for enhancing infant general movement assessment. *Medical Image Analysis*: 83, 102654, 2023
 18. Ogundokun, R. O., Maskeliūnas, R., Damaševičius, R.: Human posture detection using image augmentation and hyperparameter-optimized transfer learning algorithms. *Applied Sciences*: 12(19), 10156, 2022
 19. Liu, Z., Mao, H., Wu, C. Y., Feichtenhofer, C., Darrell, T., Xie, S.: A convnet for the 2020s. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*: 11976–11986, 2022
 20. Wightman, R.: PyTorch Image Models. Available at GitHub <https://github.com/huggingface/pytorch-image-models>. Accessed June 2024
 21. Wright, L.: New deep learning optimizer, ranger: Synergistic combination of radam+ lookahead for the best of both. Available at Github <https://github.com/lessw2020/Ranger-Deep-Learning-Optimizer>. Accessed Aug 2023
 22. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*: 248–255, 2009
 23. Smith, L. N.: Cyclical learning rates for training neural networks. In *2017 IEEE winter conference on applications of computer vision (WACV)*: 464–472, 2017
 24. Howard, J., Ruder, S.: Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146, 2018
 25. Paszke, A., Gross, S., Chintala S., Chanan G., Yang E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch. In *proceedings of the Conference on Neural Information Processing Systems (NIPS)*, 2017
 26. Howard, J., Guggen S.: Fastai: A layered API for deep learning. *Information*: 11(2), 108, 2020
 27. Bradski, G.: The openCV library. *Dr. Dobb's Journal: Software Tools for the Professional Programmer*: 25(11), 120–123, 2000
 28. Li, Y., Hu, J., Wen, Y., Evangelidis, G., Salah, K., Wang, Y., Tulyakov, S., Ren, J.: Rethinking vision transformers for mobilenet size and speed. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*: 16889–16900, 2023
 29. Qin, D., Leichner, C., Delakis, M., Forni, M., Luo, S., Yang, F., Wang, W., Banbury, C., Ye, C., Akin, B., Aggarwal, V., Zhu, T., Moro, D., Howard, A.: MobileNetV4-Universal Models for the Mobile Ecosystem. arXiv preprint arXiv:2404.10518, 2024
 30. Bouwmans, T., Javed, S., Sultana, M., Jung, S. K.: Deep neural network concepts for background subtraction: A systematic review and comparative evaluation. *Neural Networks*: 117, 8–66, 2019
 31. England, N., Improvement N.: Diagnostic imaging dataset statistical release, in London. Department of Health, 2023

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.