

Legal Medicine

Development of a deep learning algorithm for age estimation on CT images of the vertebral column --Manuscript Draft--

Manuscript Number:	LEGMED-D-23-00322R1
Article Type:	Research Paper
Keywords:	Cadaver; deep learning; CT; Spine
Corresponding Author:	Wataru Fukumoto Hiroshima University JAPAN
First Author:	Ikuo Kawashita, PhD
Order of Authors:	Ikuo Kawashita, PhD Wataru Fukumoto, MD, PhD Hidenori Mitani, MD, PhD Keigo Narita, MD, PhD Keigo Chosa, MD, PhD Yuko Nakamura, MD, PhD Masataka Nagao, MD, PhD Kazuo Awai, MD, PhD
Abstract:	<p>Purpose;The accurate age estimation of cadavers is essential for their identification. However, conventional methods fail to yield adequate age estimation especially in elderly cadavers. We developed a deep learning algorithm for age estimation on CT images of the vertebral column and checked its accuracy.Method;For the development of our deep learning algorithm, we included 1,120 CT data of the vertebral column of 140 patients for each of 8 age decades. The deep learning model of regression analysis based on Visual Geometry Group-16 (VGG16) was improved in its estimation accuracy by bagging. To verify its accuracy, we applied our deep learning algorithm to estimate the age of 219 cadavers who had undergone postmortem CT (PMCT). The mean difference and the mean absolute error (MAE), the standard error of the estimate (SEE) between the known- and the estimated age, were calculated. Correlation analysis using the intraclass correlation coefficient (ICC) and Bland-Altman analysis were performed to assess differences between the known- and the estimated age.Results;For the 219 cadavers, the mean difference between the known- and the estimated age was 0.30 years; it was 4.36 years for the MAE, and 5.48 years for the SEE. The ICC (2,1) was 0.96 (95% confidence interval: 0.95 - 0.97, $p < 0.001$). Bland-Altman analysis showed that there were no proportional or fixed errors ($p = 0.08$ and 0.41).Conclusions;Our deep learning algorithm for estimating the age of 219 cadavers on CT images of the vertebral column was more accurate than conventional methods and highly useful.</p>
Suggested Reviewers:	Yohsuke Makino, M.D., Ph.D. Associate Professor, The University of Tokyo ymakino@m.u-tokyo.ac.jp Hideki Hyodoh, M.D., Ph.D. Professor, University of Fukui hyodoh@u-fukui.ac.jp
Response to Reviewers:	

Dear Editor :

Enclosed is a manuscript entitled “Development of a deep-learning algorithm for age estimation on CT images of the vertebral column” for consideration for publication in the Legal Medicine.

There is a need for a reliable age estimation method for cadaver identification. We developed a deep-learning algorithm for age estimation of cadavers on postmortem CT (PMCT) images of the vertebral column and checked its accuracy. Our deep-learning algorithm for estimating the age of 219 cadavers using PMCT images of the vertebral column was accurate and highly useful. To the best of our knowledge, our deep-learning algorithm was the most superior method for estimating the age of cadavers in terms of accuracy and the adaptive age range.

Thank you for your consideration of our manuscript. We look forward to hearing from you.

Sincerely yours,

Wataru Fukumoto, MD, PhD

Corresponding author:

Wataru Fukumoto, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health Science, Hiroshima University

Center for Cause of Death Investigation Research, Graduate School of Biomedical and Health Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: wfukumoto@hiroshima-u.ac.jp

Statement to the editor

- This paper is submitted only to the Legal Medicine.
- This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
- This retrospective study of anonymized preexisting image data was approved by our institutional review board; prior informed consent was waived.

First author

Ikuo Kawashita, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health
Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: ikawa190@hiroshima-u.ac.jp

Corresponding author:

Wataru Fukumoto, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health
Science, Hiroshima University

Center for Cause of Death Investigation Research, Graduate School of Biomedical and
Health Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: wfukumoto@hiroshima-u.ac.jp

Co-authors:

Hidegori Mitani, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health
Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: hmitani@hiroshima-u.ac.jp

Keigo Narita, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health
Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: yuttfagc@hiroshima-u.ac.jp

Keigo Chosa, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health
Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: chosa@hiroshima-u.ac.jp

Yuko Nakamura, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health

Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: yukon@hiroshima-u.ac.jp

Masataka Nagao, MD, PhD

Center for Cause of Death Investigation Research, Graduate School of Biomedical and Health Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: nagao@hiroshima-u.ac.jp

Kazuo Awai, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health Science, Hiroshima University

Center for Cause of Death Investigation Research, Graduate School of Biomedical and Health Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: awai@hiroshima-u.ac.jp

Reviewer #1 comments

1. Please clarify the hardware specifications for the deep learning model, such as CPU, memory, and GPU.

Response:

Thank you very much for your valuable comments.

These procedures were conducted using a desktop computer equipped with an Intel Core i9 processor, 64GB of RAM, and an NVIDIA GeForce RTX 3090 graphics card, i.e.:

Material and Method

Development of the algorithm

“These procedures were conducted using a desktop computer equipped with an Intel Core i9 processor, 64GB of RAM, and an NVIDIA GeForce RTX 3090 graphics card.”
(Page 11, lines 7-9)

2. According to your explanation of the deep learning structure, you employed the VGG16 model and used the Sony NNC commercial software. You mentioned that Sony NNC was used to optimize the model to suppress overfitting, and you added a batch normalization layer and a dropout layer. However, I believe that Sony NNC is not code-based software; it's more like a tool for stacking or removing layers from the model. Did you create the VGG16 model using Python-based tools such as Google Colab? Or was the entire model produced using Sony NNC software? Please clarify how you constructed the model.

Response:

In this study, for convenience, the entire VGG16 model was developed using Sony NNC software. Sony NNC software uses a GUI to build the model, but internally it runs an open-source Python library called nnabla, whose code can be checked. We confirmed that similar results can be obtained in a Python environment. We replaced the Sony NNC software with the Python nnabla code – it can be found in the attached file “Age_Estimation_VGG16_code”.

3. One of the advantages of using the VGG16 model is the possibility of transfer learning. Is there any particular reason why you did not utilize transfer learning? It may have the potential to improve the model's performance.

Response:

Although transfer learning can improve the estimation accuracy, we were able to obtain sufficiently accurate age estimations because of ensemble learning. Currently there are no public databases similar to the vertebral images used in this study. Therefore, we added no further transfer learning. In our next study we will attempt to improve the accuracy of age estimation methods using transfer learning.

4. You used the SHAP model as an explainable AI model, but there is a lack of explanation about this model in your paper. There are several explainable AI models, such as Grad-CAM, and I'm not sure why you chose the SHAP model. My understanding is that the SHAP model is used to represent the contribution of each variable with red/blue colors when the model estimates results. However, in your paper, you mentioned that older areas are shown in red and younger areas in blue. Does this mean that older areas (in red) are indicative of better age estimation? When I examine Figure 6c and 7c, I notice differences in the distribution of red and blue dots in the images. In Figure 6c, the red and blue dots seem to be concentrated in the L-spine, especially in osteophytes, while in Figure 7c, the dots are spread over all areas, and there are differences in the number of dots. Can you explain the results of the SHAP model in more detail? Please provide additional information about the principles of the SHAP model in the Materials and Methods section and discuss the valuable results in the Discussion section.

Response:

The SHAP model was used as an explanatory AI model to visualize what parts of the vertebral image contributed to older or younger age estimates. When the SHAP model is applied to regression analysis, regions that contribute to an increase in the objective variable are indicated in red dots and regions that contribute to a decrease in the objective variable are indicated in blue dots. Thus, regions that contributed to older age estimates are shown in red dots and regions that contributed to younger age estimates are shown in blue dots. In addition, the higher their contribution, the more intensely the red and blue densities are displayed. – i.e.:

Materials and Methods

Evaluation of our age estimation results

We visualized what parts of the vertebral image contributed to older or younger age

estimates using SHapley Additive exPlanations (SHAP), one of the explainable AI methods [28]. When the SHAP model is applied to regression analysis, regions that contribute to an increase in the objective variable are indicated in red dots and regions that contribute to a decrease in the objective variable are indicated in blue dots. Thus, regions that contributed to older age estimates are shown in red dots and regions that contributed to younger age estimates are shown in blue dots. In addition, the higher the contribution, the more intensely the red and blue densities are displayed. (Page 12, lines 1-8)

Discussion

The inability to explain the results is a problem when deep learning is applied. We used SHAP to confirm what parts of the vertebral image contributed to the age estimation. On the SHAP image in Figure 6, the red dots are clustered on the osteophytes of the lumbar vertebrae, indicating that osteophyte formation contributes to the older age estimates. The blue dots, on the other hand, are clustered on the intervertebral discs, indicating that the retention of the intervertebral space contributes to the younger age estimates. On the SHAP image in Figure 7, numerous red dots are seen on the osteophytes of the vertebrae. This contributed to the older age estimates. The observations on the SHAP images were consistent with the biological processes of vision and cognition observed in aging. (Page 17-18, lines 15-6)

5. In your paper, you only presented well-estimated cases. Please include cases where the model's estimates were incorrect and provide explanations for these inaccuracies.

Response:

We added two cases with incorrect age estimates and explained the causes based on the SHAP images in the Discussion section – i.e.:

Results

Cases where the age estimation was not correct are presented in Figures 8 and 9. (Page 14, lines 14-15)

Figure Legends

Figure 8

Images of a middle-aged male who died suddenly on his bed [a: axial CT image, b:

maximum intensity projection (MIP) image of the vertebral column, c: SHapley Additive exPlanations (SHAP) image].

Despite his age, few morphological changes and osteophytes were observed on the vertebrae. The SHAP image revealed that a normal vertebral morphology and the retention of the intervertebral space contribute to a decrease in the estimated age. The known- and the estimated age (52.44 and 40.61 years, respectively) were more than 10 years apart.

Figure 9

Images of a female who was strangled to death [a: axial CT image, b: maximum intensity projection (MIP) image of the vertebral column, c: SHapley Additive exPlanations (SHAP) image].

Severe morphological changes and osteophytes are observed on the lumbar vertebrae relative to her age. The SHAP image revealed that severe morphological changes contributed strongly to an increase in the estimated age. Therefore, the known- and the estimated age (72.26 and 87.57 years, respectively) were more than 15 years apart.

Discussion

On the other hand, there were several cases in which the age estimation was incorrect. In Figure 8, there were few osteophytes on the vertebrae despite the middle age of the subject; this may have contributed to the underestimation of the estimated age. In Figure 9, the estimated age was overestimated due to severe vertebral deformation and osteophyte formation relative to her actual age. Thus, when the vertebral morphology and the age are discrepant, the accuracy of age estimation is reduced. This is a limitation of our age estimation method. (Page 18, lines 6-12)

Reviewer #2 comments

Major points

1. I have some concerns about the image pre-processing process. The authors emphasize that this method is objective and fast, but doesn't subjectivity arise when selecting the spine during the image pre-processing process? Also, doesn't the pre-processing process take time? Please consider and make any necessary modifications to resolve these concerns.

Response:

Thank you very much for your valuable comments.

For image pre-processing, we used MIP images because they can be easily reconstructed with available software. In addition, to simplify the manual technique and to minimize subjective selection of the vertebral region, we set the region of interest to be 10 cm wide and 10 cm deep around the posterior margin of the vertebral body. With some training, the time required for these procedures is 10–20 minutes per case. Therefore, we suggest that our method is simpler and easier than previous methods using artificial intelligence. We modified the Discussion as follows:

Discussion

However, their image pre-processing is complicated and their method may not be practical for non-specialists in medical imaging. For image pre-processing, we used MIP images because they can be easily reconstructed with available software. In addition, to simplify the manual technique and minimize the subjective selection of the vertebral region, we set the region of interest to be 10 cm wide and 10 cm deep around the posterior margin of the vertebral body. With some training, the time required for these procedures is 10–20 minutes per case. Therefore, we suggest that our method is simpler and easier than previous methods using artificial intelligence. (Page 17, lines 2-9)

2. Looking at the MIP images in Figures 6 and 7, it appears that the mandible, hyoid bone, and thyroid cartilage are also included in the MIP images in this study, aren't they? If so, for example, would the results be the same in a cadaver with only the vertebrae remaining, without the mandible, hyoid bone, or thyroid cartilage? I would like this point to be included in the discussion.

Response:

As mentioned above, to simplify the manual technique and to minimize subjective selection of the vertebral region, we set the region of interest to be 10 cm wide and 10 cm deep around the posterior margin of the vertebral body. Consequently, there were some cases in which the MIP images included parts other than the vertebrae, e.g. the mandible, hyoid bone, and thyroid cartilage. We re-analyzed the age estimations of 176 cases using only the vertebrae. The mean difference between the original- and the re-analyzed cases was 0.64 years. There was no significant difference between the results obtained in our 219 cases and the 176 re-analyzed cases ($p = 0.54$, Welch t-test). On the SHAP images, red or blue dots were observed mainly on the vertebrae. Therefore, we concluded that the influence of the mandible, hyoid bone, and thyroid

cartilage on the age estimation was minimal. However, as the possibility that other parts contributed slightly to the age estimation cannot be ruled out, we added a statement to the limitation section, i.e.:

Limitations

There were some cases in which, to simplify image pre-processing and minimize subjectivity, the MIP images included parts other than the vertebrae, e.g. the mandible, hyoid bone, and thyroid cartilage. Although the possibility that those parts contributed slightly to the age estimation cannot be ruled out, on the SHAP images, red or blue dots were primarily located on the vertebrae. (Page 19-20, lines 15-2)

3. One of the problems with AI is that it cannot be used without the same AI. This method also cannot be used unless the method is open. This point is inferior to the conventional method. Regarding this point, I would like an explanation or excuse to be added in the limitations etc.

Response:

To comply, we added a statement to the study limitations section, i.e.:

Limitations

We evaluated our age estimation method based on data from a single center and our method has not yet been presented to the public. We will evaluate its generalizability by using data from various centers and CT scans. (Page 19, lines 10-12)

4. The authors have introduced two cases of explainable AI, but what about other cases? The black box problem is also a disadvantage of AI method compared with conventional methods, so please discuss this point and also consider what conclusions can be reached with XAI in this study.

Response:

The inability to explain the results is a problem when deep learning is applied. We added to the Discussion:

Discussion

The inability to explain the results is a problem when deep learning is applied. We used SHAP to confirm what parts of the vertebral image contributed to the age estimation.

On the SHAP image in Figure 6, the red dots are clustered on the osteophytes of the lumbar vertebrae, indicating that osteophyte formation contributes to the older age estimates. The blue dots, on the other hand, are clustered on the intervertebral discs, indicating that the retention of the intervertebral space contributes to the younger age estimates. On the SHAP image in Figure 7, numerous red dots are seen on the osteophytes of the vertebrae. This contributed to the older age estimates. The observations on the SHAP images were consistent with the biological processes of vision and cognition observed in aging. On the other hand, there were several cases in which the age estimation was incorrect. In Figure 8, there were few osteophytes on the vertebrae despite the middle age of the subject; this may have contributed to the underestimation of the estimated age. In Figure 9, the estimated age was overestimated due to severe vertebral deformation and osteophyte formation relative to her actual age. Thus, when the vertebral morphology and the age are discrepant, the accuracy of age estimation is reduced. This is a limitation of our age estimation method. (Page 17-18, lines 15-12)

5. As I point out many grammatical things as follows, professional English proofreading is recommended.

Response:

Our professional editor of scientific manuscripts reviewed our responses and used the internationally accepted book 'The Elements of Style' by Strunk and White for guidance.

Minor points

Highlights

1. I think "deep-learning" in the second line of the highlight is wrong as this hyphen is used for an adjective.

Response:

We complied.

Abstract

2. In the Abstract, the authors mention "1120 CT images", but I think this is not appropriate. Do you only use one CT image in each case??

Response:

We replaced “images” with “data”.

3. SSE should be SEE in the results.

Response:

We corrected.

4. In conclusion, it is stated that this method is "accurate", but this is misleading.

Please use a more accurate term, e.g., "more accurate than conventional methods".

Response:

We complied.

Key words

5. I think the Keyword "Tomography, X-Ray Computed" is a mistake.

Response:

We corrected.

Abbreviations

6. I think "field-of-view" should be "field of view".

Response:

We revised.

Introduction

7. In the 3rd paragraph, the phrase "deep-learning" is used about three times, but I think deep learning is more appropriate for all of them.

Response:

We revised.

Materials and Methods

Training data

8. In the 1st line, the phrase "deep-learning" is used about three times, but I think deep

learning is more appropriate.

Response:

We revised.

9. When indicating a range of numbers, such as "20.10 - 99.75", dashes should be used rather than hyphens.

Response:

We complied.

10. As for CT conditions, if SD is set for AEC, please write its value as well. The term "field-of-view" should be "field of view". The term "soft kernel" should be "soft tissue kernel".

Response:

The 64-row-detector CT scanner (SOMATOM Definition AS; SIEMENS) adapted CARE Dose4D as the AEC system. We selected as the quality reference 40 mAs and rewrote the statement. The other terms were also reviewed.

Materials and Methods

Training data

Helical scans were performed at a tube voltage of 120 kV; the tube current was regulated with the quality reference '40 mAs' of CARE Dose4D, an automatic exposure control (AEC) system. (Page 7-8, lines 17-1)

Test data

11. When indicating a range of numbers, such as "21.13 - 94.40" or "1 - 90" dashes should be used rather than hyphens.

Response:

We complied.

12. On the 9th line in the 1st paragraph, the description "The cadavers were divided into damaged, i.e. severely burned ore decomposed bodies" seems uncomfortable, as when forensic pathologists say "damaged," we can assume various other corpses,

such as traumatized bodies, dismembered, mummified, and skeletonized bodies. So, please explain that "damaged" is just a definition for this study, and that only burned and decomposed bodies were included in the definition as a result.

Response:

We revised as follows:

Materials and Methods

Test data

The cadavers were divided into damaged and undamaged. Corpses were defined damaged when the condition of the body surface rendered identification difficult in this study. Of the 219 corpses, 151 were undamaged; 41 of the 68 damaged cadavers had been severely burned and 27 were severely decomposed. (Page 8, lines 15-18)

13. As for CT conditions, is the "preset noise level" same as SD? If so, please make it clear. The term "soft kernel" should be "soft tissue kernel".

Response:

We complied.

Image pre-processing

14. Please change "1mm³" into "1 mm³".

Response:

We complied.

15. The word "Seconds" should be "Secondly" or "Second"

Response:

We revised.

16. What application is used to perform this pre-processing? If possible, please write.

Response:

We used Image J 1.53i (National Institutes of Health) and added this information in the Materials and Methods section.

Materials and Methods

Image pre-processing

These procedures were conducted using Image J 1.53i (National Institutes of Health).
(Page 10, lines 4-5)

17. The authors described "MIP was applied in the frontal and lateral direction". Does lateral direction mean from the right, from the left, or both? Please specify.

Response:

We revised it as follows.

Materials and Methods

Image pre-processing

MIP was applied in the frontal and lateral direction from left to right on the mask-processed spine perimeter area to uncomplicate the image data and to emphasize the bone structure. (Page 9-10, lines 17-2)

18. Please show us the definition of the spine that creates the MIP here. Do you mean from the cervical vertebrae to the lumbar vertebrae? Does it include the sacral vertebrae?

Response:

The cephalocaudal extent of the vertebrae was defined as reaching from the cervical vertebrae to the coccyx within the range of the CT images, i.e.:

Materials and Methods

Image pre-processing

The cephalocaudal extent of the vertebrae was defined as reaching from the cervical vertebrae to the coccyx within the range of the CT images. (Page 9, lines 16-17)

Development of the algorithm

19. I do not understand "Loss of function was changed from cross-entropy to squared error". Does this mean that the authors calculate the cross-entropy first and then calculate the squared error? Or the authors mean that only squared error is used as a loss of function? If so, why mention cross-entropy?

Response:

Although VGG was originally adapted for image classification and cross entropy was used, we adapted it for our regression analysis. In other words, we used only the squared error as a loss of function and now state as follows:

Materials and Methods

Development of the algorithm

For regression analysis, the squared error was used as the loss of function. (Page 10, lines 11-12)

20. From figure 2, there are 2 batch normalization layers and 2 drop out layers.

However, in the manuscript the authors described "a batch normalization layer and a dropout layer were added". Please explain the discrepancy.

Response:

The entire VGG16 model was developed using Sony NNC software. To suppress overlearning, we added a batch normalization layer and a dropout layer to the original VGG16 model.

Materials and Methods

Development of the algorithm

Our final model constructed with software is shown in Figure 2. (Page 10, lines 14-15)

Evaluation of our age estimation results

21. Please indicate the formula for mean difference.

Response:

We added the formula applied to obtain the mean difference.

Materials and Methods

Evaluation of our age estimation results

mean difference = $\frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n}$ (Page 11, lines 17)

Statistical Analysis

22. Do not use the term "gender".

Response:

We revised.

23. When indicating a range of numbers, such as "0 - 0.50" for ICC, dashes should be used rather than hyphens.

Response:

We complied.

Results

24. In the 1st line, SSE should be SEE.

25. In the last line, "figures 6,7" should be "figures 6 and 7".

Response:

We complied.

26. In the last line, it is stated that the age estimation is "accurate", but this is misleading. Please use a more accurate term, e.g., "age estimation in good agreement".

Response:

We complied.

Discussion

27. In the 1st paragraph, SSE should be SEE.

28. In the 2nd paragraph, I don't think the hyphen after 4.03 is necessary.

29. In the 2nd and 3rd paragraph, do not use the term "gender".

30. In the 3rd paragraph, I don't think the hyphen after sternum is necessary.

31. In the 3rd paragraph, when indicating a range of numbers, such as "12 - 14 years", dashes should be used rather than hyphens.

32. In the 4th paragraph, on the 2nd line, "deep-learning" should be "deep learning".

33. In the 4th paragraph, please add "years" after 7.92, 8.70, 7.25, and 7.16.

34. In the 5th paragraph, on 10th line, "severe" should be "severely". Also, I don't think the hyphen after burned is necessary.

35. In conclusion, it is stated that this method is "accurately", but this is misleading.

Please use a more accurate term, e.g., "more accurately than conventional methods".

Response:

We complied with all remarks.

Figure Legends

Figure 1

36. Please change "1mm³" into "1 mm³".

Figure legends

Figure 6

37. I don't think the hyphen after 44.22 is necessary.

Figure 7

38. I don't think the hyphen after 82.95 is necessary.

Response:

We complied all.

```

import nnabla as nn
import nnabla.functions as F
import nnabla.parametric_functions as PF

def network(x, y, test=False):
    # Input: x -> 1,400,250
    # Convolution -> 64,400,250
    h = PF.convolution(x, 64, (3,3), (1,1), name='Convolution')
    # ReLU
    h = F.relu(h, True)
    # Convolution_2
    h = PF.convolution(h, 64, (3,3), (1,1), name='Convolution_2')
    # ReLU_2
    h = F.relu(h, True)
    # MaxPooling -> 64,200,125
    h = F.max_pooling(h, (2,2), (2,2))
    # Convolution_3 -> 128,200,125
    h = PF.convolution(h, 128, (3,3), (1,1), name='Convolution_3')
    # ReLU_3
    h = F.relu(h, True)
    # Convolution_4
    h = PF.convolution(h, 128, (3,3), (1,1), name='Convolution_4')
    # ReLU_4
    h = F.relu(h, True)
    # MaxPooling_2 -> 128,100,62
    h = F.max_pooling(h, (2,2), (2,2))
    # Convolution_5 -> 256,100,62
    h = PF.convolution(h, 256, (3,3), (1,1), name='Convolution_5')
    # ReLU_5
    h = F.relu(h, True)
    # Convolution_6
    h = PF.convolution(h, 256, (3,3), (1,1), name='Convolution_6')
    # ReLU_6
    h = F.relu(h, True)
    # Convolution_7
    h = PF.convolution(h, 256, (3,3), (1,1), name='Convolution_7')
    # ReLU_7
    h = F.relu(h, True)
    # MaxPooling_3 -> 256,50,31
    h = F.max_pooling(h, (2,2), (2,2))
    # Convolution_8 -> 512,50,31
    h = PF.convolution(h, 512, (3,3), (1,1), name='Convolution_8')
    # BatchNormalization_2
    h = PF.batch_normalization(h, (1,), 0.9, 0.0001, not test,
name='BatchNormalization_2')
    # ReLU_8
    h = F.relu(h, True)
    # Convolution_9
    h = PF.convolution(h, 512, (3,3), (1,1), name='Convolution_9')
    # ReLU_9
    h = F.relu(h, True)
    # Convolution_10
    h = PF.convolution(h, 512, (3,3), (1,1), name='Convolution_10')
    # ReLU_10

```

```

h = F.relu(h, True)
# MaxPooling_4 -> 512,25,15
h = F.max_pooling(h, (2,2), (2,2))
# Convolution_11
h = PF.convolution(h, 512, (3,3), (1,1), name='Convolution_11')
# ReLU_11
h = F.relu(h, True)
# Convolution_12
h = PF.convolution(h, 512, (3,3), (1,1), name='Convolution_12')
# ReLU_12
h = F.relu(h, True)
# Convolution_13
h = PF.convolution(h, 512, (3,3), (1,1), name='Convolution_13')
# ReLU_13
h = F.relu(h, True)
# MaxPooling_5 -> 512,12,7
h = F.max_pooling(h, (2,2), (2,2))
# Affine -> 4096
h = PF.affine(h, (4096,), name='Affine')
# ReLU_14
h = F.relu(h, True)
# Dropout
if not test:
    h = F.dropout(h, 0.5, 0)
# Affine_2 -> 512
h = PF.affine(h, (512,), name='Affine_2')
# BatchNormalization
h = PF.batch_normalization(h, (1,), 0.9, 0.0001, not test,
name='BatchNormalization')
# ReLU_15
h = F.relu(h, True)
# Dropout_2
if not test:
    h = F.dropout(h, 0.5, 0)
# Affine_3 -> 1
h = PF.affine(h, (1,), name='Affine_3')
# SquaredError
h = F.squared_error(h, y)
return h

```

Highlights

- There is a need for a reliable age estimation method for cadaver identification.
- We developed a deep learning for age estimation on CT images of vertebral columns.
- Our deep-learning algorithm estimated the age of cadavers accurately.
- Our age estimation yielded highly accurate information even in elderly cadavers.
- The age of severely damaged cadavers was also accurately estimated.

Title: Development of a deep-learning algorithm for age estimation on CT images of the vertebral column

Authors: Ikuo Kawashita, PhD,¹ Wataru Fukumoto, MD, PhD,^{1,2} Hidenori Mitani, MD, PhD,¹ Keigo Narita, MD, PhD,¹ Keigo Chosa, MD, PhD,¹ Yuko Nakamura, MD, PhD,¹ Masataka Nagao, MD, PhD,² Kazuo Awai, MD, PhD^{1,2}

¹ Department of Diagnostic Radiology, Graduate School of Biomedical and Health Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

² Center for Cause of Death Investigation Research, Graduate School of Biomedical and Health Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

Corresponding author:

Wataru Fukumoto, MD, PhD (ORCID: 0000-0001-7645-8446)

Department of Diagnostic Radiology, Graduate School of Biomedical and Health Science, Hiroshima University

Center for Cause of Death Investigation Research, Graduate School of Biomedical and Health Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: wfukumoto@hiroshima-u.ac.jp

TEL: +81-82-257-5257

FAX: +81-82-257-5259

First author

Ikuo Kawashita, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health Science,
Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: ikawa190@hiroshima-u.ac.jp

Corresponding author:

Wataru Fukumoto, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health Science,
Hiroshima University

Center for Cause of Death Investigation Research, Graduate School of Biomedical and Health
Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: wfukumoto@hiroshima-u.ac.jp

Co-authors:

Hidegori Mitani, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health Science,
Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: hmitani@hiroshima-u.ac.jp

Keigo Narita, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health Science,
Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: yuttfage@hiroshima-u.ac.jp

Keigo Chosa, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health Science,
Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: chosa@hiroshima-u.ac.jp

Yuko Nakamura, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health Science,
Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: yukon@hiroshima-u.ac.jp

Masataka Nagao, MD, PhD

Center for Cause of Death Investigation Research, Graduate School of Biomedical and Health
Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: nagao@hiroshima-u.ac.jp

Kazuo Awai, MD, PhD

Department of Diagnostic Radiology, Graduate School of Biomedical and Health Science,
Hiroshima University

Center for Cause of Death Investigation Research, Graduate School of Biomedical and Health
Science, Hiroshima University

1-2-3 Kasumi, Minamiku, Hiroshima 734-8551, JAPAN

E-mail: awai@hiroshima-u.ac.jp

Statements and Declarations

- This paper is submitted only to the Legal Medicine.
- This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.
- This retrospective study of anonymized preexisting image data was approved by our institutional review board; prior informed consent was waived.

CRediT author statement

Ikuo Kawashita: Methodology, Software, Validation, Formal analysis

Wataru Fukumoto: Conceptualization, Methodology, Investigation, Data Curation, Writing -
Original Draft

Hidegori Mitani: Investigation, Data Curation

Keigo Narita: Investigation, Data Curation

Keigo Chosa: Methodology, Data Curation

Yuko Nakamura: Writing- Reviewing and Editing

Masataka Nagao: Resources, Methodology, Supervision

Kazuo Awai: Conceptualization, Formal analysis, Supervision

1

Title

2 Development of a deep learning algorithm for age estimation on CT images of the

3 vertebral column

4

Abstract

Purpose;

The accurate age estimation of cadavers is essential for their identification. However, conventional methods fail to yield adequate age estimation especially in elderly cadavers. We developed a deep learning algorithm for age estimation on CT images of the vertebral column and checked its accuracy.

Method;

For the development of our deep learning algorithm, we included 1,120 CT data of the vertebral column of 140 patients for each of 8 age decades. The deep learning model of regression analysis based on Visual Geometry Group-16 (VGG16) was improved in its estimation accuracy by bagging. To verify its accuracy, we applied our deep learning algorithm to estimate the age of 219 cadavers who had undergone postmortem CT (PMCT). The mean difference and the mean absolute error (MAE), the standard error of the estimate (SEE) between the known- and the estimated age, were calculated.

Correlation analysis using the intraclass correlation coefficient (ICC) and Bland-Altman analysis were performed to assess differences between the known- and the estimated age.

Results;

For the 219 cadavers, the mean difference between the known- and the estimated age was 0.30 years; it was 4.36 years for the MAE, and 5.48 years for the SEE. The ICC (2,1) was 0.96 (95% confidence interval: 0.95 - 0.97, $p < 0.001$). Bland-Altman analysis showed that there were no proportional or fixed errors ($p = 0.08$ and 0.41).

Conclusions;

Our deep learning algorithm for estimating the age of 219 cadavers on CT images of the vertebral column was more accurate than conventional methods and highly useful.

Keywords:

Cadaver; Deep Learning; CT; Spine

Abbreviations:

VGG16: Visual Geometry Group-16

PMCT: postmortem computed tomography

MAE: mean absolute error

SEE: standard error of the estimate

ICC: intraclass correlation coefficient

MIP: maximum intensity projection

- 1 PET-CT: positron emission tomography CT
- 2 AEC: automatic exposure control
- 3 FOV: field of view
- 4 CI: confidence interval

Introduction

The identification of individuals for unidentified cadavers is an important task in forensic investigations. The age estimation at the time of death is one of the most essential processes in individual identification [1-3]. Age-related morphological changes in the pubic symphysis [4], auricular surface [5], cranial sutures [6], and teeth [7] have been used to estimate the age of cadavers. However, these conventional age-estimation methods require expertise and their accuracy, especially with respect to cadavers of the elderly, is limited [8-10]. Therefore, the development of a reliable age-estimation method is needed in the forensic science area.

The age estimation using postmortem CT (PMCT) have been developed since CT scan enables rapid, non-destructive observation of the bone without requiring any preparation [8, 10-16]. As multiplanar reconstruction-, 3D-, and maximum intensity projection (MIP) CT images yield clear and objective information on the bone structure, making it possible for inexperienced examiners to perform bone examinations [17, 18]. Osteophyte formation on the vertebral column occurs and increases with age, and the development of vertebral osteophytes has been shown to be a general indicator of age [19, 20]. Chiba et al. [8] estimated the age of cadavers by evaluating osteophytes on the thoracic and lumbar vertebrae on PMCT scans, however, their method requires visual

1 scoring and is time consuming.

2 Deep learning, a branch of machine learning that mimics the functioning of the
3 human brain is increasingly used to develop algorithms for medical imaging tasks,
4 including detection, segmentation and classification [21-24]. As deep learning can
5 automatically detects features without human interference by training using large sets of
6 labeled data, it has great potential to improve diagnostic accuracy, obtain objective
7 findings and reduce the time required [23, 25]. Deep learning could be used to develop
8 better than the currently available age estimation methods.

9 In this study, we developed a deep learning algorithm for age estimation on CT
10 images of the vertebral column and checked its accuracy for cadavers.

11

Materials and Methods

This retrospective study of anonymized preexisting image data was approved by our institutional review board; prior informed consent was waived.

Training data

As training data for deep learning, we randomly prepared CT images of 140 anonymized patients (70 males, 70 females) for each of 8 decades of life (20–90 years) which were acquired during fluorodeoxyglucose positron emission tomography CT (PET-CT) examination between 2012 and 2021 in our hospital. The training dataset therefore consisted of CT images on 1,120 patients (140 patients for each of 8 age decades) (Table 1). Their mean age and age range were 59.78 years and 20.10–99.75 years, respectively. Since age estimation methods for children and teenagers are relatively well established [26], this study focused on those aged 20 and older. Patients with bone metastases or devices implanted in the vertebral column were excluded before the data extraction process.

All CT images acquired during PET-CT examinations were obtained on a 64-row-detector CT scanner (SOMATOM Definition AS; SIEMENS). Helical scans were performed at a tube voltage of 120 kV; the tube current was regulated with the quality

reference '40 mAs' of CARE Dose4D, an automatic exposure control (AEC) system.

The other scanning parameters were rotation time 0.5 sec, beam collimation 0.6 x 64 mm, helical pitch 1.5, display field of view (FOV) 50 x 50 cm², slice thickness 3 mm, and soft tissue kernel.

Test data

We enrolled 253 cadavers subjected to PMCT examinations between October 2018 and June 2022 at our institution to verify the accuracy of our age estimation algorithm. Among 253 cadavers, we excluded 34 (younger than 20 years, n = 12; severe vertebral fracture and dislocation, n = 12; age undetermined after forensic investigation, n = 9; vertebral column anomalies due to congenital muscular dystrophy, n = 1).

Demographic information on the remaining 219 cadavers (137 males, 82 females; mean age 65.14 years, range 21.13–94.40 years) and the cause of death are shown in Table 2.

When the date of death was unclear, the age at the time of death was estimated based on forensic investigation. The cadavers were divided into damaged and undamaged.

Corpses were defined damaged when the condition of the body surface rendered identification difficult in this study. Of the 219 corpses, 151 were undamaged; 41 of the 68 damaged cadavers had been severely burned and 27 were severely decomposed. The

median interval between the date of death and CT imaging was 5 days (range 1–90 days).

All cadavers were placed in body bags and subjected to whole-body scanning on a 16-row-detector CT scanner (Aquilion Lightning; Canon Medical Systems).

Helical scans were acquired at a tube voltage of 120 kV. The tube current was regulated by AEC, the preset noise level was 8 Hounsfield units with standard deviation. The other scanning parameters were rotation time 0.75 sec, beam collimation 0.5 x 16 mm, helical pitch 0.98, display FOV 50 x 50 cm², slice thickness 1 mm, and soft tissue kernel.

Image pre-processing

The image pre-processing was performed on CT images of training- and test data (Figure 1). First, isotropic voxelization (voxel size 1 mm³) was performed to unify the image sizes. Second, bicubic interpolation was applied to reduce the blur due to interpolation. After orienting the body, a 10-cm wide and 10-cm deep area along the spine was masked. The cephalocaudal extent of the vertebrae was defined as reaching from the cervical vertebrae to the coccyx within the range of the CT images. Third, MIP was applied in the frontal and lateral direction from left to right on the mask-processed

spine perimeter area to uncomplicate the image data and to emphasize the bone structure. Lastly, the MIP images were combined and rescaled to a matrix size of 250 x 400 (pixel size 2.0 x 2.0 mm), 8 bit grayscale (HU = 0 to 2000 → 0 to 255) to save memory for the deep learning model. These procedures were conducted using Image J 1.53i (National Institutes of Health).

Development of the algorithm

Rather than the more multilayered and complex network structure that requires a large amount of training data, we selected the deep learning model based on the Visual Geometry Group-16 (VGG16) because it provides sufficient recognition accuracy with a relatively simple network structure. For regression analysis, the squared error was used as the loss of function. In addition, the network structure was optimized with commercial software (Neural Network Console; SONY) to suppress overlearning, and a batch normalization layer and a dropout layer were added. Our final model constructed with software is shown in Figure 2.

The model was trained by the mini-batch gradient descent algorithm with batch size 64 and squared error as the loss function. The optimization method was Adam

(Alpha:0.001, Beta1:0.9, Beta2:0.999, Epsilon:1E-8), and the maximum number of epochs was 1000 and early stop is performed by monitoring the validation error.

Bagging, a method of ensemble learning, was employed to reduce variance in the age-estimation results (Figure 3) [27]. The 11-fold cross-validation method was applied to create 11 different datasets (1,000 for training and 120 for validation); each model was trained individually and combined with the averaging process.

These procedures were conducted using a desktop computer equipped with an Intel Core i9 processor, 64GB of RAM, and an NVIDIA GeForce RTX 3090 graphics card.

Evaluation of our age estimation results

We estimated the age of the 219 cadavers using developed deep learning algorithm. The estimated age at the time of death was expressed to the second decimal place and the known- and the estimated age at the time of death were compared. The mean difference, the mean absolute error (MAE), and the standard error of the estimate (SEE) between the known- and the estimated age at the time of death were calculated

with the formula; $mean\ difference = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n}$, $MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$, $SEE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-2}}$, where y is the known age, \hat{y} the estimated age, and n the number of cases.

We visualized what parts of the vertebral image contributed to older or younger age estimates using SHapley Additive exPlanations (SHAP), one of the explainable AI methods [28]. When the SHAP model is applied to regression analysis, regions that contribute to an increase in the objective variable are indicated in red dots and regions that contribute to a decrease in the objective variable are indicated in blue dots. Thus, regions that contributed to older age estimates are shown in red dots and regions that contributed to younger age estimates are shown in blue dots. In addition, the higher the contribution, the more intensely the red and blue densities are displayed.

Statistical analysis

We applied the Welch t-test to determine the difference between the known- and the estimated age of damaged- and undamaged cadavers, and sex of cadavers. The correlation between the known- and the estimated age at the time of death was assessed with the intraclass correlation coefficient (ICC); 0–0.50 indicated poor-, 0.51–0.75 moderate-, 0.76–0.90 good-, and 0.91 or greater excellent reliability [29]. Bland-Altman analysis was performed to identify the difference between the known- and the estimated age. We used software for statistical analysis (R version 4.1.2); $p < 0.05$ was considered to indicate statistically significant differences. Our calculations showed that, at an ICC

- 1 of 0.8, a sample of 20 subjects provided 90% power with a two-sided type I error of
- 2 0.05.

Results

The mean difference, the MAE, and the SEE between the known- and the estimated age at the time of death for 219 cadavers were 0.30, 4.36, and 5.48 years, respectively. For 151 undamaged cadavers they were 0.44, 4.50 and 5.61 years and for 68 damaged cadavers they were 1.95, 4.03 and 5.23 years ($p = 0.002, 0.32, 0.47$, respectively). For male 137 cadavers they were 0.71, 4.67 and 5.74 years and for 82 female cadavers they were -0.38, 3.83 and 5.08 years ($p = 0.14, 0.07, 0.20$, respectively).

The ICC (2,1) between the known- and estimated age at the time of death was 0.96 [95% confidence interval (CI) 0.95–0.97, $p < 0.001$]; reliability was excellent (Figure 4). By Bland-Altman analysis the mean difference was 0.30 years; Bland-Altman plots revealed no proportional or fixed errors ($p = 0.08$ and $p = 0.41$); the limit of agreement ranged from -9.14 to 9.74 years (Figure 5).

Representative cases with age estimation in good agreement with known age are shown in Figures 6 and 7. Cases where the age estimation was not correct are presented in Figures 8 and 9.

Discussion

Although the accurate age estimation of unidentified cadavers is essential, conventional methods that assess morphological bone changes are unsatisfactory, especially when the cadaver is of an elderly subject [8-10]. We developed a deep learning algorithm for age estimation and applied it to estimate the age of 219 cadavers. The MAE and SEE showed high accuracy (4.36 and 5.48 years, respectively) and the ICC was 0.96. Bland-Altman analysis revealed that the estimated age of cadavers of elderly- and young subjects was highly accurate; the limit of agreement was less than approximately 10 years. This indicates that our algorithm correctly estimated the decade of life of cadavers. To the best of our knowledge, our deep learning algorithm applied to CT images of the vertebral column was the superior method for estimating the age of cadavers in terms of accuracy and the adaptive age range.

As in undamaged cadavers, our age estimation was accurate even in 68 severely burned or severely decomposed cadavers; the MAE was 4.03 and the SEE was 5.23 years. Although, compared to undamaged cadavers, the estimated age of damaged cadavers was slightly higher than the known age, the mean difference was 1.95 years and of little relevance. When CT images were subjected to MIP, the effect of factors such as air in the vertebral column due to decomposition may have been reduced. Also,

our findings were not affected by the cadaver's sex. Thus, our age estimation method was applicable equivalently regardless of the condition- or the sex of the cadavers.

Methods using PMCT to estimate the age of cadavers have been reported [8, 10-16]. Chiba et al. [8] evaluated the age based on the subjective inspection of vertebral osteophytes on CT images; their SEE was approximately 10 years. According to Monum et al. [14], age estimation based on sternum and rib ossification observed on CT images returned an SEE of 12–14 years. Others reported that measuring the bone density of the pubic symphysis on CT images was sufficiently accurate; their absolute error ranged from 9–16 years [11]. Forensically, analysis of the HU value of the proximal femur on CT images was reported to be useful for sex identification and age estimation [13]. However, unlike ours, these earlier methods involve subjective visual scoring and manual measurements of the bone density by experts and are time-consuming and labor intensive.

A new approach for age estimation using CT image and machine learning or deep learning have been developed to improve accuracy, obtain objective findings and reduce the time required [30-32]. Imaizumi K. et al. applied machine learning for age estimation on PMCT images of several bones; based on a MAE of 7.92 years their results were good [30]. Zhang Y. et al. who applied machine learning to the radiomic

features of the pubic symphysis estimated the age of adults accurately; their MAE was 8.70 years [32]. However, their image pre-processing is complicated and their method may not be practical for non-specialists in medical imaging. For image pre-processing, we used MIP images because they can be easily reconstructed with available software. In addition, to simplify the manual technique and minimize the subjective selection of the vertebral region, we set the region of interest to be 10 cm wide and 10 cm deep around the posterior margin of the vertebral body. With some training, the time required for these procedures is 10–20 minutes per case. Therefore, we suggest that our method is simpler and easier than previous methods using artificial intelligence. Although Kondou H. et al. also developed a deep learning algorithm for age estimation on PMCT images of vertebral columns with a MAE of 7.25 years for male and 7.16 years for female, they excluded completely putrefied cases [31]. Our deep learning algorithm could estimate the age even in severely decomposed cadavers with a MAE of 4.03 and ours were more accurate and practical.

The inability to explain the results is a problem when deep learning is applied. We used SHAP to confirm what parts of the vertebral image contributed to the age estimation. On the SHAP image in Figure 6, the red dots are clustered on the osteophytes of the lumbar vertebrae, indicating that osteophyte formation contributes to

the older age estimates. The blue dots, on the other hand, are clustered on the intervertebral discs, indicating that the retention of the intervertebral space contributes to the younger age estimates. On the SHAP image in Figure 7, numerous red dots are seen on the osteophytes of the vertebrae. This contributed to the older age estimates. The observations on the SHAP images were consistent with the biological processes of vision and cognition observed in aging. On the other hand, there were several cases in which the age estimation was incorrect. In Figure 8, there were few osteophytes on the vertebrae despite the middle age of the subject; this may have contributed to the underestimation of the estimated age. In Figure 9, the estimated age was overestimated due to severe vertebral deformation and osteophyte formation relative to her actual age. Thus, when the vertebral morphology and the age are discrepant, the accuracy of age estimation is reduced. This is a limitation of our age estimation method.

The vertebral column is relatively new parts of the bone applied to an age estimation [8, 19, 20]. Although it is well-known that osteophytes of the vertebral column develop with age and represent one of the most common age-related change, the age estimation using vertebral column is not yet a well-established due to lack of sufficient evidences. However, the age estimation using vertebral column have some advantages over other parts of bone. First, the age estimation using vertebral column

showed the highest accuracy even in cadavers of the elderly from earlier study [8, 20]. This is significant since the number of unidentified elderly cadavers is increasing with the aging of society [9]. Second, vertebral column is one of the last part to remain in severely burned and decomposed body which require individual identification [20]. Therefore, the age estimation on CT images of the vertebral column yielded accurate information on cadavers of young and elderly subjects, irrespective of their condition and it was of forensic value.

Our study has some limitations. It was retrospective, all study subjects were Japanese, and cadavers with vertebral column lesions or implanted medical device were excluded. We evaluated our age estimation method based on data from a single center and our method has not yet been presented to the public. We will evaluate its generalizability by using data from various centers and CT scans. Although the CT scanner and the slice thicknesses of the training- and test CT data were different, for image pre-processing the slice thickness of the MIP images was standardized to the median value of 2 mm to reduce differences in the scanning conditions. There were some cases in which, to simplify image pre-processing and minimize subjectivity, the MIP images included parts other than the vertebrae, e.g. the mandible, hyoid bone, and thyroid cartilage. Although the possibility that those parts contributed slightly to the age

1 estimation cannot be ruled out, on the SHAP images, red or blue dots were primarily
2 located on the vertebrae. Lastly, differences in the CT images of living subjects and
3 cadavers may have affected our findings since the morphology of the vertebral column
4 of damaged cadavers may have changed after death. Nonetheless, we can report that an
5 accurate age estimation was obtained even in severely damaged cadavers.

6 In conclusion, our deep learning algorithm for age estimation on CT images of
7 vertebral column could estimate age more accurately than conventional methods.

8

References

1. Cunha E, Baccino E, Martrille L, Ramsthaler F, Prieto J, Schuliar Y, et al. The problem of aging human remains and living individuals: a review. *Forensic Sci Int* 2009;193:1-13
2. Baccino E, Ubelaker DH, Hayek LA, Zerilli A. Evaluation of seven methods of estimating age at death from mature human skeletal remains. *J Forensic Sci* 1999;44:931-936
3. Konigsberg LW, Herrmann NP, Wescott DJ, Kimmerle EH. Estimation and evidence in forensic anthropology: age-at-death. *J Forensic Sci* 2008;53:541-557
4. Savall F, Rerolle C, Herin F, Dedouit F, Rouge D, Telmon N, et al. Reliability of the Suchey-Brooks method for a French contemporary population. *Forensic Sci Int* 2016;266:586 e581-586 e585
5. Buckberry JL, Chamberlain AT. Age estimation from the auricular surface of the ilium: a revised method. *Am J Phys Anthropol* 2002;119:231-239
6. Meindl RS, Lovejoy CO. Ectocranial suture closure: a revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *Am J Phys Anthropol* 1985;68:57-66
7. Liversidge HM. Controversies in age estimation from developing teeth. *Ann Hum Biol* 2015;42:397-406
8. Chiba F, Inokuchi G, Hoshioka Y, Sakuma A, Makino Y, Torimitsu S, et al. Age estimation by evaluation of osteophytes in thoracic and lumbar vertebrae using postmortem CT images in a modern Japanese population. *Int J Legal Med* 2021
9. Nomura M, McLean S, Miyamori D, Kakiuchi Y, Ikegaya H. Isolation and unnatural death of elderly people in the aging Japanese society. *Sci Justice* 2016;56:80-83
10. Warriar V, Kanchan T, Shedge R, Krishan K, Singh S. Computed tomographic age estimation from the pubic symphysis using the Suchey-Brooks method: A Systematic Review and Meta-analysis. *Forensic Sci Int* 2021;325:110811
11. Bascou A, Dubourg O, Telmon N, Dedouit F, Saint-Martin P, Savall F. Age estimation based on computed tomography exploration: a combined method. *Int J Legal Med* 2021;135:2447-2455
12. Belghith M, Marchand E, Ben Khelil M, Rouge-Maillart C, Blum A, Martrille L. Age estimation based on the acetabulum using global illumination rendering with computed tomography. *Int J Legal Med* 2021
13. Ford JM, Kumm TR, Decker SJ. An Analysis of Hounsfield Unit Values and

- Volumetrics from Computerized Tomography of the Proximal Femur for Sex and Age Estimation. *J Forensic Sci* 2020;65:591-596
14. Monum T, Makino Y, Prasitwattanaseree S, Yajima D, Chiba F, Torimitsu S, et al. Age estimation from ossification of sternum and true ribs using 3D post-mortem CT images in a Japanese population. *Leg Med (Tokyo)* 2020;43:101663
 15. Torimitsu S, Makino Y, Saitoh H, Ishii N, Inokuchi G, Motomura A, et al. Age estimation based on maturation of the medial clavicular epiphysis in a Japanese population using multidetector computed tomography. *Leg Med (Tokyo)* 2019;37:28-32
 16. Villa C, Hansen MN, Buckberry J, Cattaneo C, Lynnerup N. Forensic age estimation based on the trabecular bone changes of the pelvic bone using post-mortem CT. *Forensic Sci Int* 2013;233:393-402
 17. Dedouit F, Telmon N, Costagliola R, Otal P, Joffre F, Rouge D. Virtual anthropology and forensic identification: report of one case. *Forensic Sci Int* 2007;173:182-187
 18. Zech WD, Hatch G, Siegenthaler L, Thali MJ, Losch S. Sex determination from os sacrum by postmortem CT. *Forensic Sci Int* 2012;221:39-43
 19. Praneatpolgrang S, Prasitwattanaseree S, Mahakkanukrauh P. Age estimation equations using vertebral osteophyte formation in a Thai population: comparison and modified osteophyte scoring method. *Anat Cell Biol* 2019;52:149-160
 20. Watanabe S, Terazawa K. Age estimation from the degree of osteophyte formation of vertebral columns in Japanese. *Leg Med (Tokyo)* 2006;8:156-160
 21. Park S, Lee SM, Kim W, Park H, Jung KH, Do KH, et al. Computer-aided Detection of Subsolid Nodules at Chest CT: Improved Performance with Deep Learning-based CT Section Thickness Reduction. *Radiology* 2021;299:211-219
 22. Perez AA, Noe-Kim V, Lubner MG, Graffy PM, Garrett JW, Elton DC, et al. Deep Learning CT-based Quantitative Visualization Tool for Liver Volume Estimation: Defining Normal and Hepatomegaly. *Radiology* 2021:210531
 23. Yang J, Xie M, Hu C, Alwalid O, Xu Y, Liu J, et al. Deep Learning for Detecting Cerebral Aneurysms with CT Angiography. *Radiology* 2021;298:155-163
 24. Yasaka K, Akai H, Kunimatsu A, Abe O, Kiryu S. Liver Fibrosis: Deep Convolutional Neural Network for Staging by Using Gadoteric Acid-enhanced Hepatobiliary Phase MR Images. *Radiology* 2018;287:146-155
 25. Ahn Y, Lee SM, Noh HN, Kim W, Choe J, Do KH, et al. Use of a Commercially Available Deep Learning Algorithm to Measure the Solid Portions of Lung Cancer Manifesting as Subsolid Lesions at CT: Comparisons with Radiologists

and Invasive Component Size at Pathologic Examination. *Radiology* 2021;299:202-210

26. Bjork MB, Kvaal SI. CT and MR imaging used in age estimation: a systematic review. *J Forensic Odontostomatol* 2018;36:14-25
27. Breiman, L. Bagging Predictors. *Machine Learning* 1996;26:123–140
28. S.M. Lundberg, S.-I. Lee. A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems* 2017; 4768–4777
29. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;15:155-163
30. Imaizumi K, Usui S, Taniguchi K, Ogawa Y, Nagata T, Kaga K, et al. Development of an age estimation method for bones based on machine learning using post-mortem computed tomography images of bones. *Forensic Imaging* 2021;26
31. Kondou H, Morohashi R, Ichioka H, Bandou R, Matsunari R, Kawamoto M, et al. Deep Neural Networks-Based Age Estimation of Cadavers Using CT Imaging of Vertebrae. *Int J Environ Res Public Health* 2023;20
32. Zhang Y, Wang Z, Liao Y, Li T, Xu X, Wu W, et al. A machine-learning approach using pubic CT based on radiomics to estimate adult ages. *Eur J Radiol* 2022;156:110516

Figure Legends

Figure 1

Image pre-processing

(a) Isotropic voxelization: Isotropic voxelization (voxel size 1 mm³) was performed to

unify the image sizes.

(b) Mask processing of the spine region: The spine region was masked (width 10 cm,

depth 10 cm)

(c) Maximum intensity projection (MIP): MIP was applied in the frontal and lateral

direction from left to right on the mask-processed spine perimeter area.

(d) Image rescaling: MIP images were combined and rescaled to a matrix size of 250 x

400 (pixel size 2.0 x 2.0 mm), 8 bit grayscale (HU = 0 to 2000 → 0 to 255).

- 1 Figure 2
- 2 Structure of the deep learning model.
- 3 The deep learning model based on VGG16 was selected because it provides sufficient
- 4 recognition accuracy with a relatively simple network structure. The loss function was
- 5 changed from cross-entropy to squared error to perform regression analysis. In addition,
- 6 the network structure was optimized to suppress overlearning, and a batch
- 7 normalization- and a dropout layer were added.
- 8

1 Figure 3

2 Bagging

3 Bagging, a method of ensemble learning, was used to reduce variance in the age-
4 estimation results. To create 11 different datasets (1000 for training and 120 for
5 validation), 11-fold cross-validation was applied. Each model was trained individually
6 and combined using the averaging process.

7

1 Figure 4

2 The correlation between the actual- and the estimated age at the time of death of 219

3 subjects. The intraclass correlation coefficient (2,1) between the known- and the

4 estimated age at the time of death was 0.96 (95% confidence interval: 0.95 - 0.97, $p <$

5 0.001). The reliability of findings obtained with our algorithm was excellent.

6

1 Figure 5

2 Bland-Altman analysis to determine the difference between the known- and the

3 estimated age at the time of death of 219 study subjects. The mean difference was 0.30

4 years and there were no proportional or fixed errors ($p = 0.08$ and 0.41). The limit of

5 agreement was from -9.14 to 9.74 years.

6

Figure 6

Images of a male who died from carbon monoxide poisoning [a: axial CT image, b: maximum intensity projection (MIP) image of the vertebral column, c: SHapley Additive exPlanations (SHAP) image].

Mild morphological changes and osteophytes are observed on the lumbar vertebrae. The SHAP image revealed that osteophyte formation contributed to an increase in the estimated age, while retention of the intervertebral space contributed to a decrease in estimated age. The known- and the estimated age (44.22 and 43.63 years, respectively) were in excellent agreement.

Figure 7

Images of a male found dead in his home [a: axial CT image, b: maximum intensity projection (MIP) image of the vertebral column, c: SHapley Additive exPlanations (SHAP) image].

Severe morphological changes and osteophytes are observed on lumbar vertebrae. The SHAP image revealed that severe morphological changes contributed to an increase in the estimated age. The known- and the estimated age (82.95 and 81.31 years, respectively) were in good agreement. Our deep learning algorithm estimated the age even in cadavers of the elderly.

Figure 8

Images of a middle-aged male who died suddenly on his bed [a: axial CT image, b: maximum intensity projection (MIP) image of the vertebral column, c: SHapley Additive exPlanations (SHAP) image].

Despite his age, few morphological changes and osteophytes were observed on the vertebrae. The SHAP image revealed that a normal vertebral morphology and the retention of the intervertebral space contribute to a decrease in the estimated age. The known- and the estimated age (52.44 and 40.61 years, respectively) were more than 10 years apart.

1 Figure 9

2 Images of a female who was strangled to death [a: axial CT image, b: maximum

3 intensity projection (MIP) image of the vertebral column, c: SHapley Additive

4 exPlanations (SHAP) image].

5 Severe morphological changes and osteophytes are observed on the lumbar vertebrae

6 relative to her age. The SHAP image revealed that severe morphological changes

7 contributed strongly to an increase in the estimated age. Therefore, the known- and the

8 estimated age (72.26 and 87.57 years, respectively) were more than 15 years apart.

9

10

11

1

1 **Title**

2 Development of a deep learning algorithm for age estimation on CT images of the

3 vertebral column

4

Abstract

Purpose;

The accurate age estimation of cadavers is essential for their identification. However, conventional methods fail to yield adequate age estimation especially in elderly cadavers. We developed a deep learning algorithm for age estimation on CT images of the vertebral column and checked its accuracy.

Method;

For the development of our deep learning algorithm, we included 1,120 CT data of the vertebral column of 140 patients for each of 8 age decades. The deep learning model of regression analysis based on Visual Geometry Group-16 (VGG16) was improved in its estimation accuracy by bagging. To verify its accuracy, we applied our deep learning algorithm to estimate the age of 219 cadavers who had undergone postmortem CT (PMCT). The mean difference and the mean absolute error (MAE), the standard error of the estimate (SEE) between the known- and the estimated age, were calculated. Correlation analysis using the intraclass correlation coefficient (ICC) and Bland-Altman analysis were performed to assess differences between the known- and the estimated age.

Results;

Commented [福本航1]: Reviewer #2 comment
Minor points 2

For the 219 cadavers, the mean difference between the known- and the estimated age was 0.30 years; it was 4.36 years for the MAE, and 5.48 years for the SEE. The ICC (2,1) was 0.96 (95% confidence interval: 0.95 - 0.97, $p < 0.001$). Bland-Altman analysis showed that there were no proportional or fixed errors ($p = 0.08$ and 0.41).

Conclusions;

Our deep learning algorithm for estimating the age of 219 cadavers on CT images of the vertebral column was more accurate than conventional methods and highly useful.

Keywords:

Cadaver; Deep Learning; CT; Spine

Abbreviations:

VGG16: Visual Geometry Group-16

PMCT: postmortem computed tomography

MAE: mean absolute error

SEE: standard error of the estimate

ICC: intraclass correlation coefficient

MIP: maximum intensity projection

Commented [福本航2]: Reviewer #2 comment
Minor points 3

Commented [福本航3]: Reviewer #2 comment
Minor points 4

Commented [福本航4]: Reviewer #2 comment
Minor points 5

- 1 PET-CT: positron emission tomography CT
- 2 AEC: automatic exposure control
- 3 FOV: field of view
- 4 CI: confidence interval

Commented [福本航5]: Reviewer #2 comment
Minor points 6

Introduction

The identification of individuals for unidentified cadavers is an important task in forensic investigations. The age estimation at the time of death is one of the most essential processes in individual identification [1-3]. Age-related morphological changes in the pubic symphysis [4], auricular surface [5], cranial sutures [6], and teeth [7] have been used to estimate the age of cadavers. However, these conventional age-estimation methods require expertise and their accuracy, especially with respect to cadavers of the elderly, is limited [8-10]. Therefore, the development of a reliable age-estimation method is needed in the forensic science area.

The age estimation using postmortem CT (PMCT) have been developed since CT scan enables rapid, non-destructive observation of the bone without requiring any preparation [8, 10-16]. As multiplanar reconstruction-, 3D-, and maximum intensity projection (MIP) CT images yield clear and objective information on the bone structure, making it possible for inexperienced examiners to perform bone examinations [17, 18]. Osteophyte formation on the vertebral column occurs and increases with age, and the development of vertebral osteophytes has been shown to be a general indicator of age [19, 20]. Chiba et al. [8] estimated the age of cadavers by evaluating osteophytes on the thoracic and lumbar vertebrae on PMCT scans, however, their method requires visual

1 scoring and is time consuming.

2 **Deep learning**, a branch of machine learning that mimics the functioning of the
3 human brain is increasingly used to develop algorithms for medical imaging tasks,
4 including detection, segmentation and classification [21-24]. As deep learning can
5 automatically detects features without human interference by training using large sets of
6 labeled data, it has great potential to improve diagnostic accuracy, obtain objective
7 findings and reduce the time required [23, 25]. Deep learning could be used to develop
8 better than the currently available age estimation methods.

9 In this study, we developed a deep learning algorithm for age estimation on CT
10 images of the vertebral column and checked its accuracy for cadavers.

11

Commented [福本航6]: Reviewer #2 comment
Minor points 7

Materials and Methods

This retrospective study of anonymized preexisting image data was approved by our institutional review board; prior informed consent was waived.

Training data

As training data for deep learning, we randomly prepared CT images of 140 anonymized patients (70 males, 70 females) for each of 8 decades of life (20–90 years) which were acquired during fluorodeoxyglucose positron emission tomography CT (PET-CT) examination between 2012 and 2021 in our hospital. The training dataset therefore consisted of CT images on 1,120 patients (140 patients for each of 8 age decades) (Table 1). Their mean age and age range were 59.78 years and 20.10–99.75 years, respectively. Since age estimation methods for children and teenagers are relatively well established [26], this study focused on those aged 20 and older. Patients with bone metastases or devices implanted in the vertebral column were excluded before the data extraction process.

All CT images acquired during PET-CT examinations were obtained on a 64-row-detector CT scanner (SOMATOM Definition AS; SIEMENS). Helical scans were performed at a tube voltage of 120 kV; the tube current was regulated with the quality

Commented [福本航7]: Reviewer #2 comment
Minor points 8

Commented [福本航8]: Reviewer #2 comment
Minor points 9

reference '40 mAs' of CARE Dose4D, an automatic exposure control (AEC) system.

Commented [福本航9]: Reviewer #2 comment
Minor points 10

The other scanning parameters were rotation time 0.5 sec, beam collimation 0.6 x 64

mm, helical pitch 1.5, display field of view (FOV) 50 x 50 cm², slice thickness 3 mm,

Commented [福本航10]: Reviewer #2 comment
Minor points 10

and soft tissue kernel.

Commented [福本航11]: Reviewer #2 comment
Minor points 10

Test data

We enrolled 253 cadavers subjected to PMCT examinations between October 2018 and June 2022 at our institution to verify the accuracy of our age estimation algorithm. Among 253 cadavers, we excluded 34 (younger than 20 years, n = 12; severe vertebral fracture and dislocation, n = 12; age undetermined after forensic investigation, n = 9; vertebral column anomalies due to congenital muscular dystrophy, n = 1).

Demographic information on the remaining 219 cadavers (137 males, 82 females; mean age 65.14 years, range 21.13–94.40 years) and the cause of death are shown in Table 2.

Commented [福本航12]: Reviewer #2 comment
Minor points 11

When the date of death was unclear, the age at the time of death was estimated based on

forensic investigation. The cadavers were divided into damaged and undamaged.

Corpses were defined damaged when the condition of the body surface rendered

identification difficult in this study. Of the 219 corpses, 151 were undamaged; 41 of the

68 damaged cadavers had been severely burned and 27 were severely decomposed. The

Commented [福本航13]: Reviewer #2 comment
Minor points 12

1 median interval between the date of death and CT imaging was 5 days (range 1–90
2 days).

Commented [福本航14]: Reviewer #2 comment
Minor points 11

3 All cadavers were placed in body bags and subjected to whole-body scanning
4 on a 16-row-detector CT scanner (Aquilion Lightning; Canon Medical Systems).

5 Helical scans were acquired at a tube voltage of 120 kV. The tube current was regulated

6 by AEC, the preset noise level was 8 Hounsfield units with standard deviation. The

Commented [福本航15]: Reviewer #2 comment
Minor points 13

7 other scanning parameters were rotation time 0.75 sec, beam collimation 0.5 x 16 mm,

8 helical pitch 0.98, display FOV 50 x 50 cm², slice thickness 1 mm, and soft tissue

9 kernel.

Commented [福本航16]: Reviewer #2 comment
Minor points 13

10

11 **Image pre-processing**

12 The image pre-processing was performed on CT images of training- and test

13 data (Figure 1). First, isotropic voxelization (voxel size 1 mm³) was performed to unify

Commented [福本航17]: Reviewer #2 comment
Minor points 14

14 the image sizes. Second, bicubic interpolation was applied to reduce the blur due to

Commented [福本航18]: Reviewer #2 comment
Minor points 15

15 interpolation. After orienting the body, a 10-cm wide and 10-cm deep area along the

16 spine was masked. The cephalocaudal extent of the vertebrae was defined as reaching

17 from the cervical vertebrae to the coccyx within the range of the CT images. Third, MIP

Commented [福本航19]: Reviewer #2 comment
Minor points 18

18 was applied in the frontal and lateral direction from left to right on the mask-processed

spine perimeter area to uncomplicate the image data and to emphasize the bone structure. Lastly, the MIP images were combined and rescaled to a matrix size of 250 x 400 (pixel size 2.0 x 2.0 mm), 8 bit grayscale (HU = 0 to 2000 → 0 to 255) to save memory for the deep learning model. These procedures were conducted using Image J 1.53i (National Institutes of Health).

Commented [福本航20]: Reviewer #2 comment
Minor points 17

Commented [福本航21]: Reviewer #2 comment
Minor points 16

Development of the algorithm

Rather than the more multilayered and complex network structure that requires a large amount of training data, we selected the deep learning model based on the Visual Geometry Group-16 (VGG16) because it provides sufficient recognition accuracy with a relatively simple network structure. For regression analysis, the squared error was used as the loss of function. In addition, the network structure was optimized with commercial software (Neural Network Console; SONY) to suppress overlearning, and a batch normalization layer and a dropout layer were added. Our final model constructed with software is shown in Figure 2.

Commented [福本航22]: Reviewer #2 comment
Minor points 19

Commented [福本航23]: Reviewer #2 comment
Minor points 20

The model was trained by the mini-batch gradient descent algorithm with batch size 64 and squared error as the loss function. The optimization method was Adam

(Alpha:0.001, Beta1:0.9, Beta2:0.999, Epsilon:1E-8), and the maximum number of epochs was 1000 and early stop is performed by monitoring the validation error.

Bagging, a method of ensemble learning, was employed to reduce variance in the age-estimation results (Figure 3) [27]. The 11-fold cross-validation method was applied to create 11 different datasets (1,000 for training and 120 for validation); each model was trained individually and combined with the averaging process.

These procedures were conducted using a desktop computer equipped with an Intel Core i9 processor, 64GB of RAM, and an NVIDIA GeForce RTX 3090 graphics card.

Commented [福本航24]: Reviewer #1 comment 1

Evaluation of our age estimation results

We estimated the age of the 219 cadavers using developed deep learning algorithm. The estimated age at the time of death was expressed to the second decimal place and the known- and the estimated age at the time of death were compared. The mean difference, the mean absolute error (MAE), and the standard error of the estimate (SEE) between the known- and the estimated age at the time of death were calculated

with the formula; $mean\ difference = \frac{\sum_{i=1}^n (\hat{y}_i - y_i)}{n}$, $MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$, $SEE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-2}}$, where y is the known age, \hat{y} the estimated age, and n the number of cases.

Commented [福本航25]: Reviewer #2 comment
Minor points 21

We visualized what parts of the vertebral image contributed to older or younger age estimates using SHapley Additive exPlanations (SHAP), one of the explainable AI methods [28]. When the SHAP model is applied to regression analysis, regions that contribute to an increase in the objective variable are indicated in red dots and regions that contribute to a decrease in the objective variable are indicated in blue dots. Thus, regions that contributed to older age estimates are shown in red dots and regions that contributed to younger age estimates are shown in blue dots. In addition, the higher the contribution, the more intensely the red and blue densities are displayed.

Commented [福本航26]: Reviewer #1 comment 4

Statistical analysis

We applied the Welch t-test to determine the difference between the known- and the estimated age of damaged- and undamaged cadavers, and sex of cadavers. The correlation between the known- and the estimated age at the time of death was assessed with the intraclass correlation coefficient (ICC); 0–0.50 indicated poor-, 0.51–0.75 moderate-, 0.76–0.90 good-, and 0.91 or greater excellent reliability [29]. Bland-Altman analysis was performed to identify the difference between the known- and the estimated age. We used software for statistical analysis (R version 4.1.2); $p < 0.05$ was considered to indicate statistically significant differences. Our calculations showed that, at an ICC

Commented [福本航27]: Reviewer #2 comment
Minor points 22

Commented [福本航28]: Reviewer #2 comment
Minor points 23

- 1 of 0.8, a sample of 20 subjects provided 90% power with a two-sided type I error of
- 2 0.05.

Results

The mean difference, the MAE, and the SEE between the known- and the estimated age at the time of death for 219 cadavers were 0.30, 4.36, and 5.48 years, respectively. For 151 undamaged cadavers they were 0.44, 4.50 and 5.61 years and for 68 damaged cadavers they were 1.95, 4.03 and 5.23 years ($p = 0.002, 0.32, 0.47$, respectively). For male 137 cadavers they were 0.71, 4.67 and 5.74 years and for 82 female cadavers they were -0.38, 3.83 and 5.08 years ($p = 0.14, 0.07, 0.20$, respectively).

The ICC (2,1) between the known- and estimated age at the time of death was 0.96 [95% confidence interval (CI) 0.95–0.97, $p < 0.001$]; reliability was excellent (Figure 4). By Bland-Altman analysis the mean difference was 0.30 years; Bland-Altman plots revealed no proportional or fixed errors ($p = 0.08$ and $p = 0.41$); the limit of agreement ranged from -9.14 to 9.74 years (Figure 5).

Representative cases with age estimation in good agreement with known age are shown in Figures 6 and 7. Cases where the age estimation was not correct are presented in Figures 8 and 9.

Commented [福本航29]: Reviewer #2 comment
Minor points 24

Commented [福本航30]: Reviewer #2 comment
Minor points 26

Commented [福本航31]: Reviewer #2 comment
Minor points 25

Commented [福本航32]: Reviewer #1 comment 5

Discussion

Although the accurate age estimation of unidentified cadavers is essential, conventional methods that assess morphological bone changes are unsatisfactory, especially when the cadaver is of an elderly subject [8-10]. We developed a deep learning algorithm for age estimation and applied it to estimate the age of 219 cadavers. The MAE and SEE showed high accuracy (4.36 and 5.48 years, respectively) and the ICC was 0.96. Bland-Altman analysis revealed that the estimated age of cadavers of elderly- and young subjects was highly accurate; the limit of agreement was less than approximately 10 years. This indicates that our algorithm correctly estimated the decade of life of cadavers. To the best of our knowledge, our deep learning algorithm applied to CT images of the vertebral column was the superior method for estimating the age of cadavers in terms of accuracy and the adaptive age range.

As in undamaged cadavers, our age estimation was accurate even in 68 severely burned or severely decomposed cadavers; the MAE was 4.03 and the SEE was 5.23 years. Although, compared to undamaged cadavers, the estimated age of damaged cadavers was slightly higher than the known age, the mean difference was 1.95 years and of little relevance. When CT images were subjected to MIP, the effect of factors such as air in the vertebral column due to decomposition may have been reduced. Also,

Commented [福本航33]: Reviewer #2 comment
Minor points 27

Commented [福本航34]: Reviewer #2 comment
Minor points 28

our findings were not affected by the cadaver's sex. Thus, our age estimation method was applicable equivalently regardless of the condition- or the sex of the cadavers.

Commented [福本航35]: Reviewer #2 comment
Minor points 29

Commented [福本航36]: Reviewer #2 comment
Minor points 29

Methods using PMCT to estimate the age of cadavers have been reported [8, 10-16]. Chiba et al. [8] evaluated the age based on the subjective inspection of vertebral osteophytes on CT images; their SEE was approximately 10 years. According to

Monum et al. [14], age estimation based on sternum and rib ossification observed on CT images returned an SEE of 12-14 years. Others reported that measuring the bone

Commented [福本航37]: Reviewer #2 comment
Minor points 30

Commented [福本航38]: Reviewer #2 comment
Minor points 31

density of the pubic symphysis on CT images was sufficiently accurate; their absolute error ranged from 9-16 years [11]. Forensically, analysis of the HU value of the

Commented [福本航39]: Reviewer #2 comment
Minor points 31

proximal femur on CT images was reported to be useful for sex identification and age estimation [13]. However, unlike ours, these earlier methods involve subjective visual

Commented [福本航40]: Reviewer #2 comment
Minor points 29

scoring and manual measurements of the bone density by experts and are time-consuming and labor intensive.

A new approach for age estimation using CT image and machine learning or deep learning have been developed to improve accuracy, obtain objective findings and reduce the time required [30-32]. Imaizumi K. et al. applied machine learning for age estimation on PMCT images of several bones; based on a MAE of 7.92 years their results were good [30]. Zhang Y. et al. who applied machine learning to the radiomic

Commented [福本航41]: Reviewer #2 comment
Minor points 32

Commented [福本航42]: Reviewer #2 comment
Minor points 33

features of the pubic symphysis estimated the age of adults accurately; their MAE was 8.70 years [32]. However, their image pre-processing is complicated and their method may not be practical for non-specialists in medical imaging. For image pre-processing, we used MIP images because they can be easily reconstructed with available software. In addition, to simplify the manual technique and minimize the subjective selection of the vertebral region, we set the region of interest to be 10 cm wide and 10 cm deep around the posterior margin of the vertebral body. With some training, the time required for these procedures is 10–20 minutes per case. Therefore, we suggest that our method is simpler and easier than previous methods using artificial intelligence. Although Kondou H. et al. also developed a deep learning algorithm for age estimation on PMCT images of vertebral columns with a MAE of 7.25 years for male and 7.16 years for female, they excluded completely putrefied cases [31]. Our deep learning algorithm could estimate the age even in severely decomposed cadavers with a MAE of 4.03 and ours were more accurate and practical.

The inability to explain the results is a problem when deep learning is applied. We used SHAP to confirm what parts of the vertebral image contributed to the age estimation. On the SHAP image in Figure 6, the red dots are clustered on the osteophytes of the lumbar vertebrae, indicating that osteophyte formation contributes to

Commented [福本航43]: Reviewer #2 comment
Minor points 33

Commented [福本航44]: Reviewer #2 comment
Major points 1

Commented [福本航45]: Reviewer #2 comment
Minor points 33

Commented [福本航46]: Reviewer #2 comment
Minor points 33

the older age estimates. The blue dots, on the other hand, are clustered on the intervertebral discs, indicating that the retention of the intervertebral space contributes to the younger age estimates. On the SHAP image in Figure 7, numerous red dots are seen on the osteophytes of the vertebrae. This contributed to the older age estimates. The observations on the SHAP images were consistent with the biological processes of vision and cognition observed in aging. On the other hand, there were several cases in which the age estimation was incorrect. In Figure 8, there were few osteophytes on the vertebrae despite the middle age of the subject; this may have contributed to the underestimation of the estimated age. In Figure 9, the estimated age was overestimated due to severe vertebral deformation and osteophyte formation relative to her actual age. Thus, when the vertebral morphology and the age are discrepant, the accuracy of age estimation is reduced. This is a limitation of our age estimation method.

The vertebral column is relatively new parts of the bone applied to an age estimation [8, 19, 20]. Although it is well-known that osteophytes of the vertebral column develop with age and represent one of the most common age-related change, the age estimation using vertebral column is not yet a well-established due to lack of sufficient evidences. However, the age estimation using vertebral column have some advantages over other parts of bone. First, the age estimation using vertebral column

Commented [福本航47]: Reviewer #1 comment 4,5
Reviewer #2 comment
Minor points 4

showed the highest accuracy even in cadavers of the elderly from earlier study [8, 20].

This is significant since the number of unidentified elderly cadavers is increasing with

the aging of society [9]. Second, vertebral column is one of the last part to remain in

severely burned and decomposed body which require individual identification [20].

Therefore, the age estimation on CT images of the vertebral column yielded accurate

information on cadavers of young and elderly subjects, irrespective of their condition

and it was of forensic value.

Our study has some limitations. It was retrospective, all study subjects were

Japanese, and cadavers with vertebral column lesions or implanted medical device were

excluded. We evaluated our age estimation method based on data from a single center

and our method has not yet been presented to the public. We will evaluate its

generalizability by using data from various centers and CT scans. Although the CT

scanner and the slice thicknesses of the training- and test CT data were different, for

image pre-processing the slice thickness of the MIP images was standardized to the

median value of 2 mm to reduce differences in the scanning conditions. There were

some cases in which, to simplify image pre-processing and minimize subjectivity, the

MIP images included parts other than the vertebrae, e.g. the mandible, hyoid bone, and

thyroid cartilage. Although the possibility that those parts contributed slightly to the age

Commented [福本航48]: Reviewer #2 comment
Minor points 34

Commented [福本航49]: Reviewer #2 comment
Major points 3

estimation cannot be ruled out, on the SHAP images, red or blue dots were primarily located on the vertebrae. Lastly, differences in the CT images of living subjects and cadavers may have affected our findings since the morphology of the vertebral column of damaged cadavers may have changed after death. Nonetheless, we can report that an accurate age estimation was obtained even in severely damaged cadavers.

In conclusion, our deep learning algorithm for age estimation on CT images of vertebral column could estimate age more accurately than conventional methods.

Commented [福本航50]: Reviewer #2 comment
Major points 2

Commented [福本航51]: Reviewer #2 comment
Minor points 35

References

1. Cunha E, Baccino E, Martrille L, Ramsthaler F, Prieto J, Schuliar Y, et al. The problem of aging human remains and living individuals: a review. *Forensic Sci Int* 2009;193:1-13
2. Baccino E, Ubelaker DH, Hayek LA, Zerilli A. Evaluation of seven methods of estimating age at death from mature human skeletal remains. *J Forensic Sci* 1999;44:931-936
3. Konigsberg LW, Herrmann NP, Wescott DJ, Kimmerle EH. Estimation and evidence in forensic anthropology: age-at-death. *J Forensic Sci* 2008;53:541-557
4. Savall F, Rerolle C, Herin F, Dedouit F, Rouge D, Telmon N, et al. Reliability of the Suchey-Brooks method for a French contemporary population. *Forensic Sci Int* 2016;266:586 e581-586 e585
5. Buckberry JL, Chamberlain AT. Age estimation from the auricular surface of the ilium: a revised method. *Am J Phys Anthropol* 2002;119:231-239
6. Meindl RS, Lovejoy CO. Ectocranial suture closure: a revised method for the determination of skeletal age at death based on the lateral-anterior sutures. *Am J Phys Anthropol* 1985;68:57-66
7. Liversidge HM. Controversies in age estimation from developing teeth. *Ann Hum Biol* 2015;42:397-406
8. Chiba F, Inokuchi G, Hoshioka Y, Sakuma A, Makino Y, Torimitsu S, et al. Age estimation by evaluation of osteophytes in thoracic and lumbar vertebrae using postmortem CT images in a modern Japanese population. *Int J Legal Med* 2021
9. Nomura M, McLean S, Miyamori D, Kakiuchi Y, Ikegaya H. Isolation and unnatural death of elderly people in the aging Japanese society. *Sci Justice* 2016;56:80-83
10. Warriar V, Kanchan T, Shedge R, Krishan K, Singh S. Computed tomographic age estimation from the pubic symphysis using the Suchey-Brooks method: A Systematic Review and Meta-analysis. *Forensic Sci Int* 2021;325:110811
11. Bascou A, Dubourg O, Telmon N, Dedouit F, Saint-Martin P, Savall F. Age estimation based on computed tomography exploration: a combined method. *Int J Legal Med* 2021;135:2447-2455
12. Belghith M, Marchand E, Ben Khelil M, Rouge-Maillart C, Blum A, Martrille L. Age estimation based on the acetabulum using global illumination rendering with computed tomography. *Int J Legal Med* 2021
13. Ford JM, Kumm TR, Decker SJ. An Analysis of Hounsfield Unit Values and

- 1 Volumetrics from Computerized Tomography of the Proximal Femur for Sex
2 and Age Estimation. *J Forensic Sci* 2020;65:591-596
- 3 14. Monum T, Makino Y, Prasitwattanaseree S, Yajima D, Chiba F, Torimitsu S, et
4 al. Age estimation from ossification of sternum and true ribs using 3D post-
5 mortem CT images in a Japanese population. *Leg Med (Tokyo)* 2020;43:101663
- 6 15. Torimitsu S, Makino Y, Saitoh H, Ishii N, Inokuchi G, Motomura A, et al. Age
7 estimation based on maturation of the medial clavicular epiphysis in a Japanese
8 population using multidetector computed tomography. *Leg Med (Tokyo)*
9 2019;37:28-32
- 10 16. Villa C, Hansen MN, Buckberry J, Cattaneo C, Lynnerup N. Forensic age
11 estimation based on the trabecular bone changes of the pelvic bone using post-
12 mortem CT. *Forensic Sci Int* 2013;233:393-402
- 13 17. Dedouit F, Telmon N, Costagliola R, Otal P, Joffre F, Rouge D. Virtual
14 anthropology and forensic identification: report of one case. *Forensic Sci Int*
15 2007;173:182-187
- 16 18. Zech WD, Hatch G, Siegenthaler L, Thali MJ, Losch S. Sex determination from
17 os sacrum by postmortem CT. *Forensic Sci Int* 2012;221:39-43
- 18 19. Praneatpolgrang S, Prasitwattanaseree S, Mahakkanukrauh P. Age estimation
19 equations using vertebral osteophyte formation in a Thai population: comparison
20 and modified osteophyte scoring method. *Anat Cell Biol* 2019;52:149-160
- 21 20. Watanabe S, Terazawa K. Age estimation from the degree of osteophyte
22 formation of vertebral columns in Japanese. *Leg Med (Tokyo)* 2006;8:156-160
- 23 21. Park S, Lee SM, Kim W, Park H, Jung KH, Do KH, et al. Computer-aided
24 Detection of Subsolid Nodules at Chest CT: Improved Performance with Deep
25 Learning-based CT Section Thickness Reduction. *Radiology* 2021;299:211-219
- 26 22. Perez AA, Noe-Kim V, Lubner MG, Graffy PM, Garrett JW, Elton DC, et al.
27 Deep Learning CT-based Quantitative Visualization Tool for Liver Volume
28 Estimation: Defining Normal and Hepatomegaly. *Radiology* 2021;210531
- 29 23. Yang J, Xie M, Hu C, Alwalid O, Xu Y, Liu J, et al. Deep Learning for Detecting
30 Cerebral Aneurysms with CT Angiography. *Radiology* 2021;298:155-163
- 31 24. Yasaka K, Akai H, Kunitatsu A, Abe O, Kiryu S. Liver Fibrosis: Deep
32 Convolutional Neural Network for Staging by Using Gadoteric Acid-enhanced
33 Hepatobiliary Phase MR Images. *Radiology* 2018;287:146-155
- 34 25. Ahn Y, Lee SM, Noh HN, Kim W, Choe J, Do KH, et al. Use of a Commercially
35 Available Deep Learning Algorithm to Measure the Solid Portions of Lung
36 Cancer Manifesting as Subsolid Lesions at CT: Comparisons with Radiologists

- and Invasive Component Size at Pathologic Examination. *Radiology* 2021;299:202-210
26. Bjork MB, Kvaal SI. CT and MR imaging used in age estimation: a systematic review. *J Forensic Odontostomatol* 2018;36:14-25
27. Breiman, L. Bagging Predictors. *Machine Learning* 1996;26:123–140
28. S.M. Lundberg, S.-I. Lee. A unified approach to interpreting model predictions. *Proceedings of the 31st International Conference on Neural Information Processing Systems* 2017; 4768–4777
29. Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;15:155-163
30. Imaizumi K, Usui S, Taniguchi K, Ogawa Y, Nagata T, Kaga K, et al. Development of an age estimation method for bones based on machine learning using post-mortem computed tomography images of bones. *Forensic Imaging* 2021;26
31. Kondou H, Morohashi R, Ichioka H, Bandou R, Matsunari R, Kawamoto M, et al. Deep Neural Networks-Based Age Estimation of Cadavers Using CT Imaging of Vertebrae. *Int J Environ Res Public Health* 2023;20
32. Zhang Y, Wang Z, Liao Y, Li T, Xu X, Wu W, et al. A machine-learning approach using pubic CT based on radiomics to estimate adult ages. *Eur J Radiol* 2022;156:110516

Figure Legends

Figure 1

Image pre-processing

(a) Isotropic voxelization: Isotropic voxelization (voxel size 1 mm³) was performed to unify the image sizes.

(b) Mask processing of the spine region: The spine region was masked (width 10 cm, depth 10 cm)

(c) Maximum intensity projection (MIP): MIP was applied in the frontal and lateral direction from left to right on the mask-processed spine perimeter area.

(d) Image rescaling: MIP images were combined and rescaled to a matrix size of 250 x 400 (pixel size 2.0 x 2.0 mm), 8 bit grayscale (HU = 0 to 2000 → 0 to 255).

Commented [福本航52]: Reviewer #2 comment
Minor points 36

1 Figure 2

2 Structure of the deep learning model.

3 The deep learning model based on VGG16 was selected because it provides sufficient

4 recognition accuracy with a relatively simple network structure. The loss function was

5 changed from cross-entropy to squared error to perform regression analysis. In addition,

6 the network structure was optimized to suppress overlearning, and a batch

7 normalization- and a dropout layer were added.

8

1 Figure 3

2 Bagging

3 Bagging, a method of ensemble learning, was used to reduce variance in the age-
4 estimation results. To create 11 different datasets (1000 for training and 120 for
5 validation), 11-fold cross-validation was applied. Each model was trained individually
6 and combined using the averaging process.

7

1 Figure 4

2 The correlation between the actual- and the estimated age at the time of death of 219
3 subjects. The intraclass correlation coefficient (2,1) between the known- and the
4 estimated age at the time of death was 0.96 (95% confidence interval: 0.95 - 0.97, $p <$
5 0.001). The reliability of findings obtained with our algorithm was excellent.

6

1 Figure 5

2 Bland-Altman analysis to determine the difference between the known- and the

3 estimated age at the time of death of 219 study subjects. The mean difference was 0.30

4 years and there were no proportional or fixed errors ($p = 0.08$ and 0.41). The limit of

5 agreement was from -9.14 to 9.74 years.

6

1 Figure 6

2 Images of a male who died from carbon monoxide poisoning [a: axial CT image, b:

3 maximum intensity projection (MIP) image of the vertebral column, c: SHapley

4 Additive exPlanations (SHAP) image].

5 Mild morphological changes and osteophytes are observed on the lumbar vertebrae. The

6 SHAP image revealed that osteophyte formation contributed to an increase in the

7 estimated age, while retention of the intervertebral space contributed to a decrease in

8 estimated age. The known- and the estimated age (44.22 and 43.63 years, respectively)

9 were in excellent agreement.

Commented [福本航53]: Reviewer #2 comment
Minor points 37

1 Figure 7

2 Images of a male found dead in his home [a: axial CT image, b: maximum intensity
3 projection (MIP) image of the vertebral column, c: SHapley Additive exPlanations
4 (SHAP) image].

5 Severe morphological changes and osteophytes are observed on lumbar vertebrae. The
6 SHAP image revealed that severe morphological changes contributed to an increase in
7 the estimated age. The known- and the estimated age (82.95 and 81.31 years,
8 respectively) were in good agreement. Our deep learning algorithm estimated the age
9 even in cadavers of the elderly.

Commented [福本航54]: Reviewer #2 comment
Minor points 38

10

Figure 8

Images of a middle-aged male who died suddenly on his bed [a: axial CT image, b: maximum intensity projection (MIP) image of the vertebral column, c: SHapley Additive exPlanations (SHAP) image].

Despite his age, few morphological changes and osteophytes were observed on the vertebrae. The SHAP image revealed that a normal vertebral morphology and the retention of the intervertebral space contribute to a decrease in the estimated age. The known- and the estimated age (52.44 and 40.61 years, respectively) were more than 10 years apart.

Commented [福本航55]: Reviewer #1 comment 5

1 Figure 9

2 Images of a female who was strangled to death [a: axial CT image, b: maximum
3 intensity projection (MIP) image of the vertebral column, c: SHapley Additive
4 exPlanations (SHAP) image].

5 Severe morphological changes and osteophytes are observed on the lumbar vertebrae
6 relative to her age. The SHAP image revealed that severe morphological changes
7 contributed strongly to an increase in the estimated age. Therefore, the known- and the
8 estimated age (72.26 and 87.57 years, respectively) were more than 15 years apart.

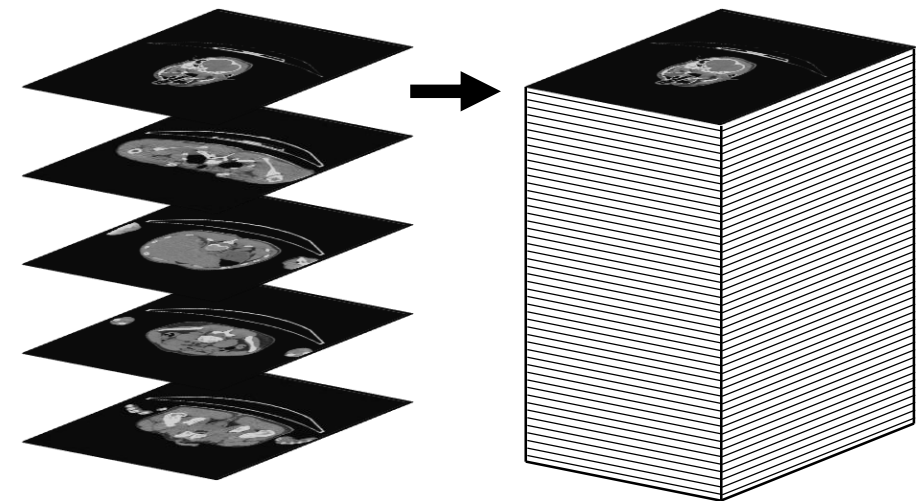
9

10

11

Figure_1

tri-cubic interpolation

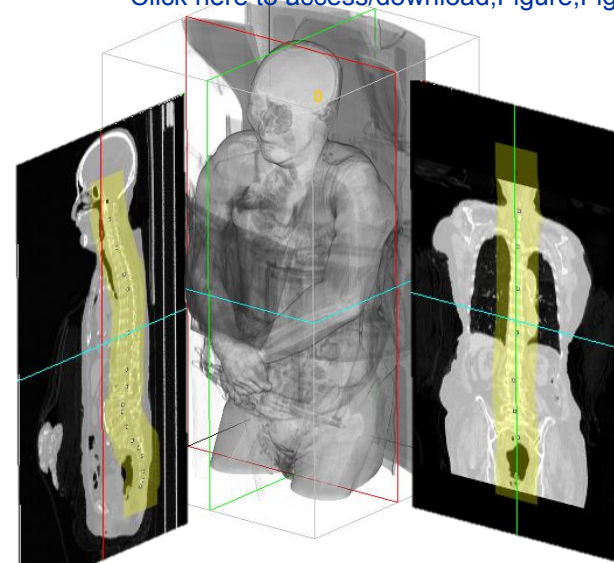


axial CT images
(thickness: 1 or 3mm)

isotropic volume data
(voxel size: 1 mm³)

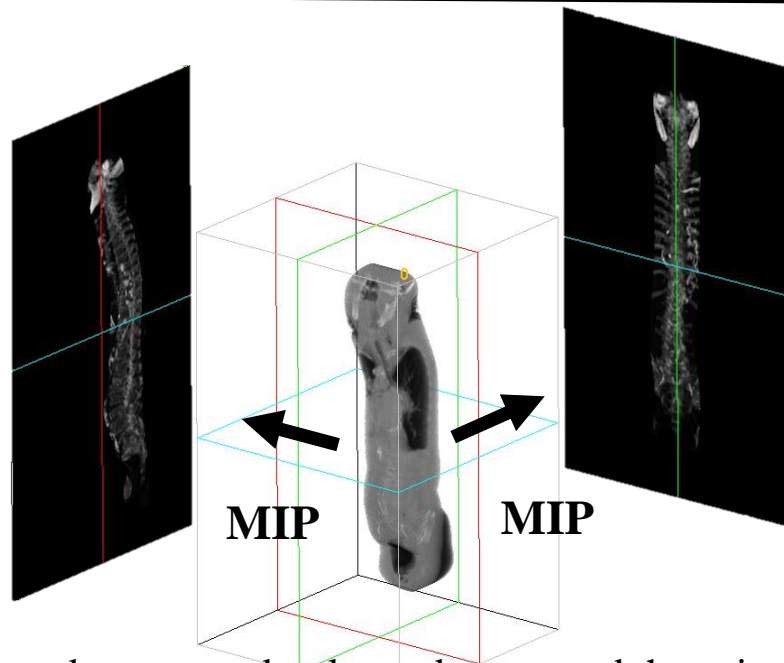
a

[Click here to access/download;Figure;Figure_1.pdf](#)



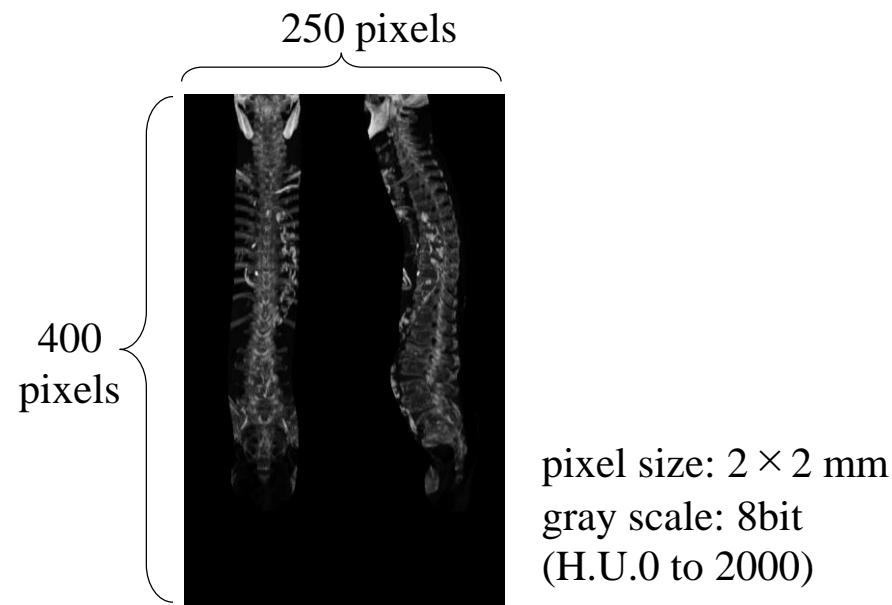
masking process along the spine
(width: 10cm, depth: 10cm)

b



mask-processed volume data around the spine

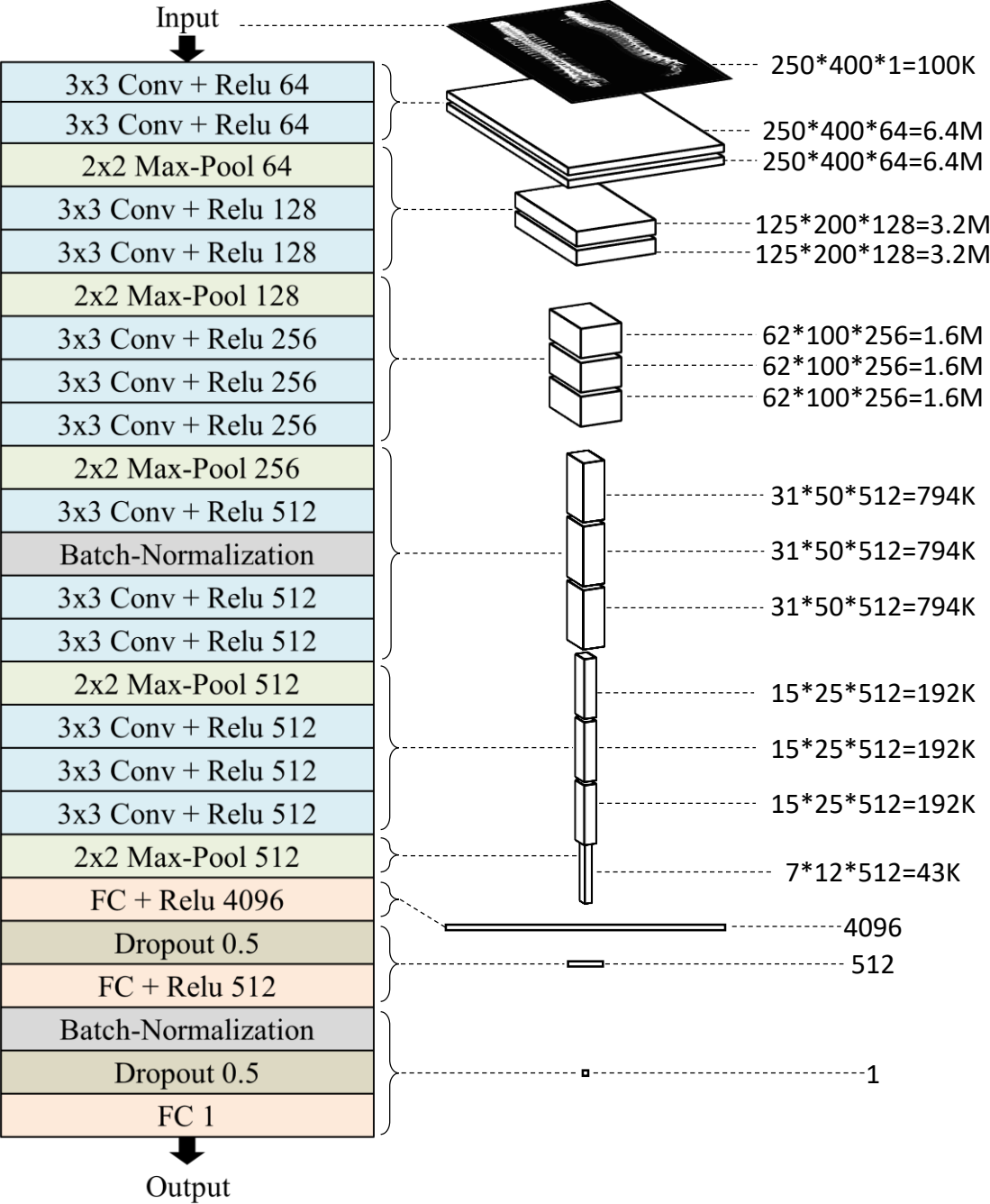
c

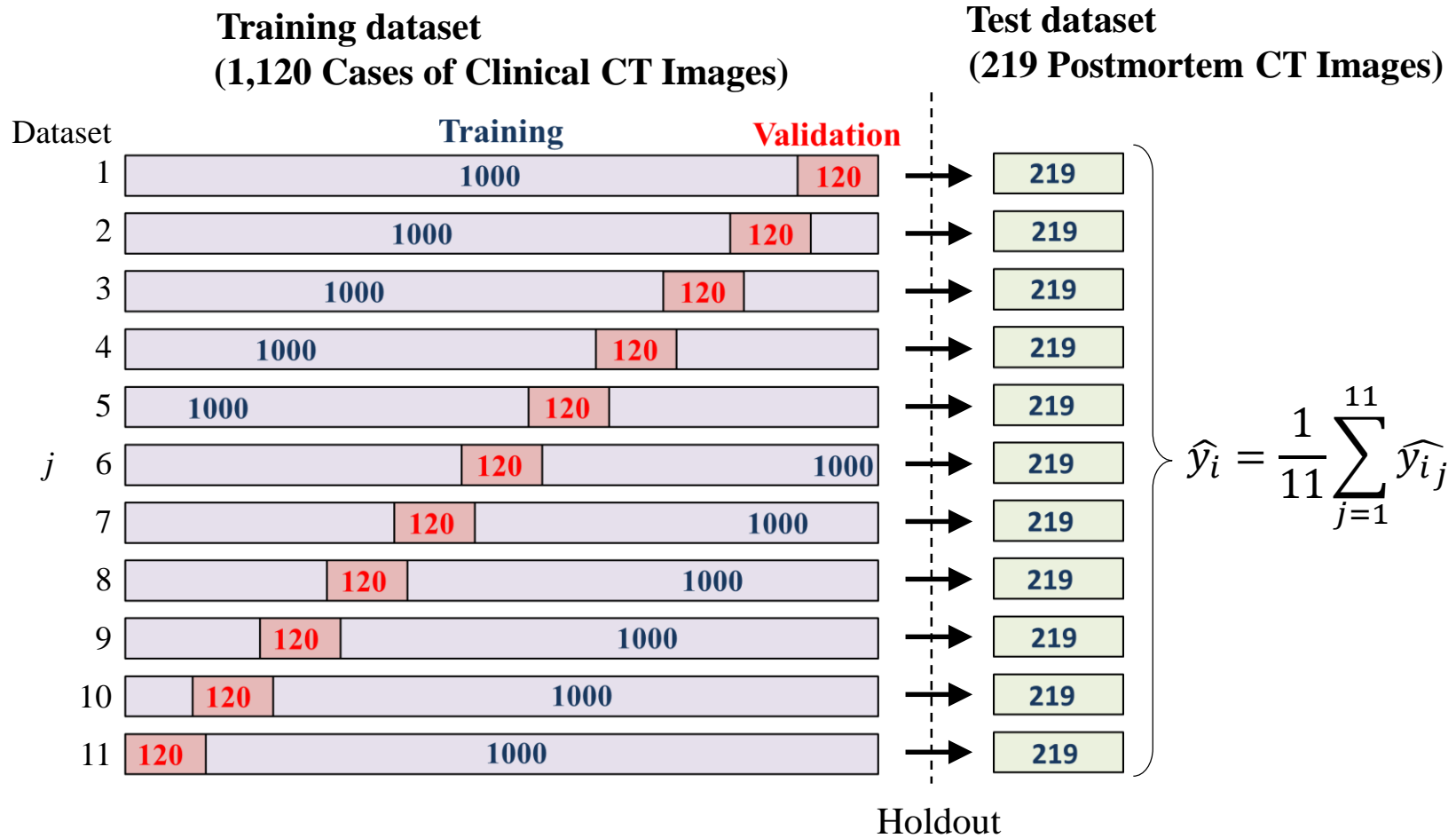


preprocessed image

d

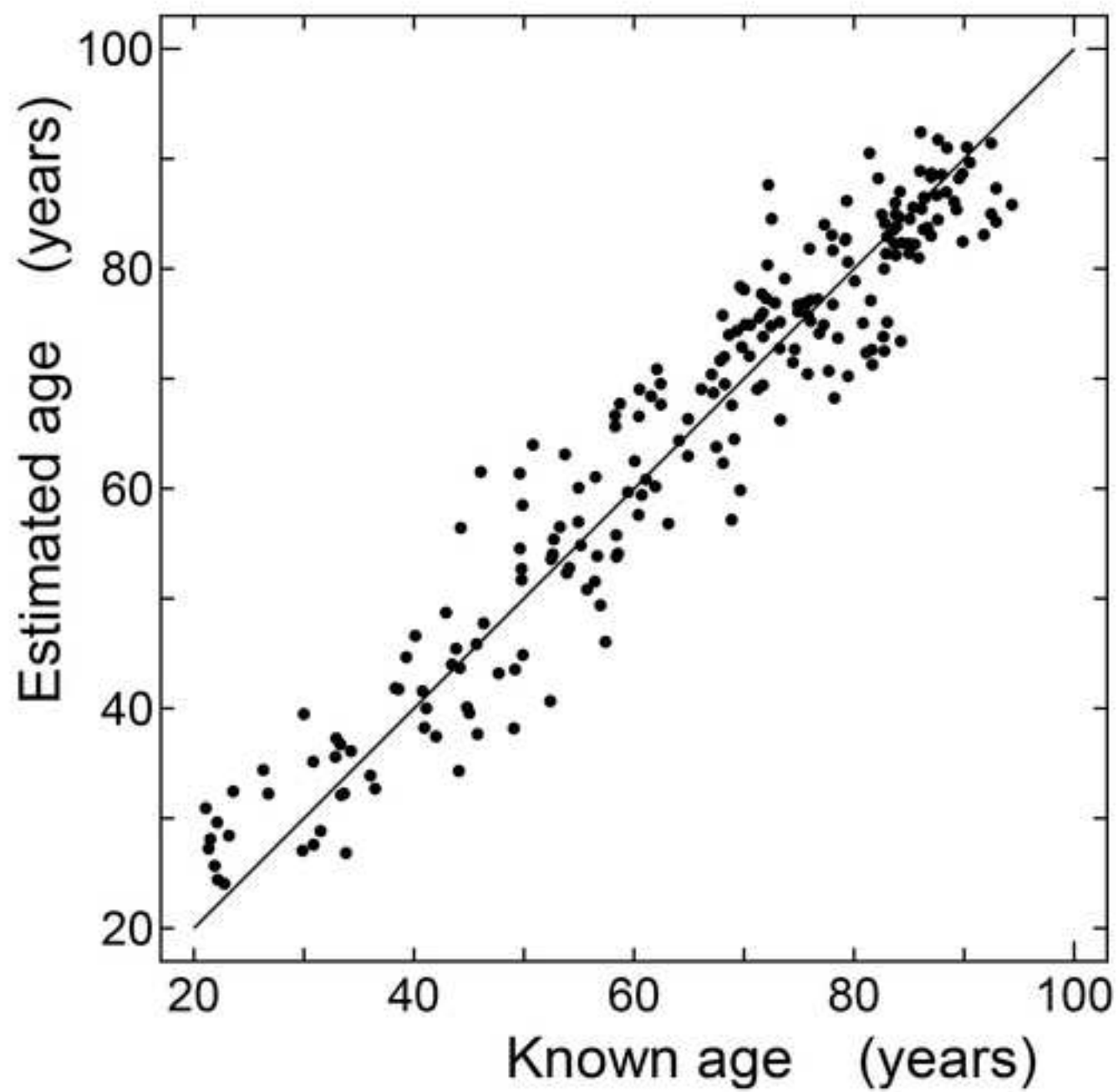
Figure_2

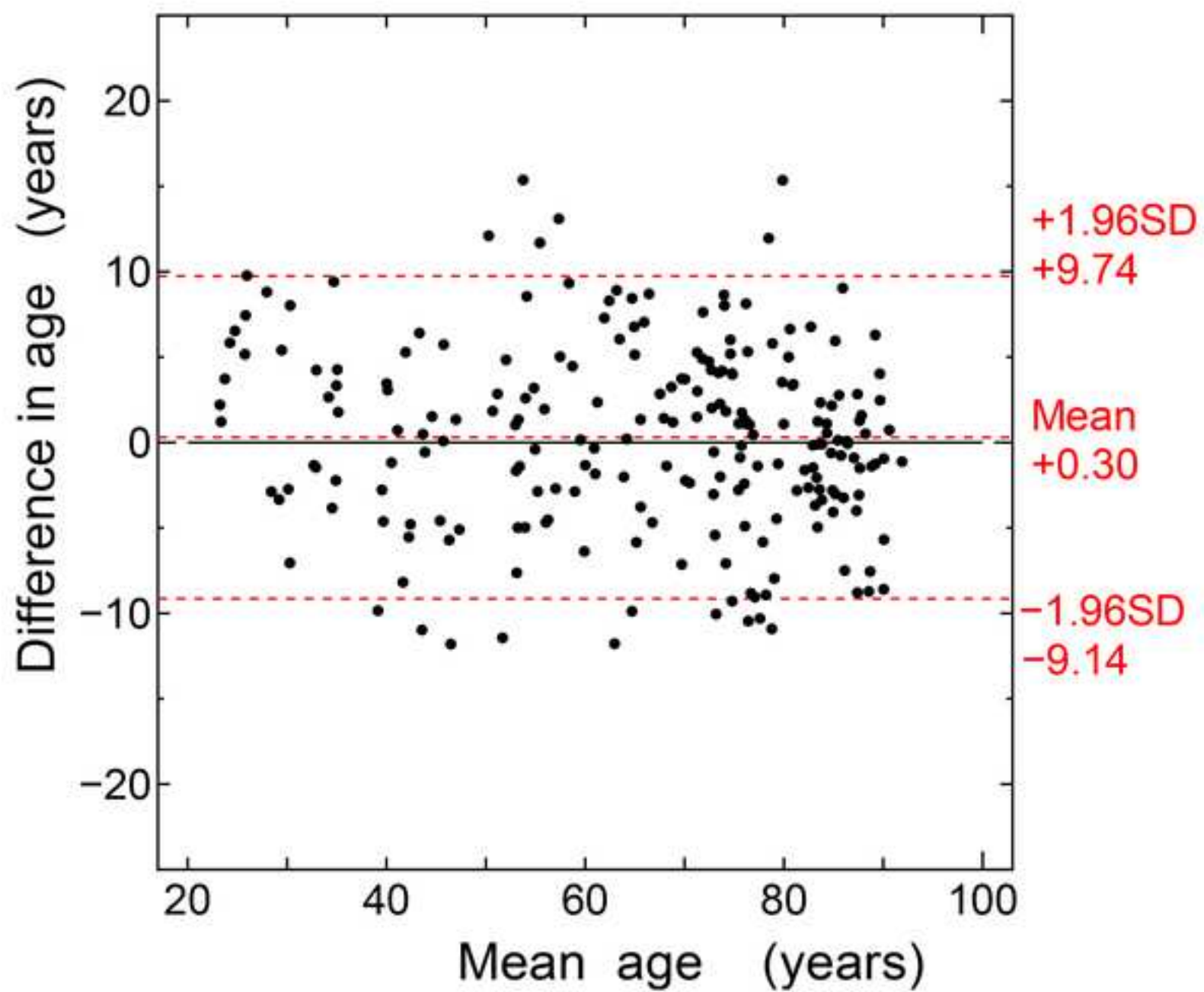




\hat{y}_i : Estimated age at death for case i

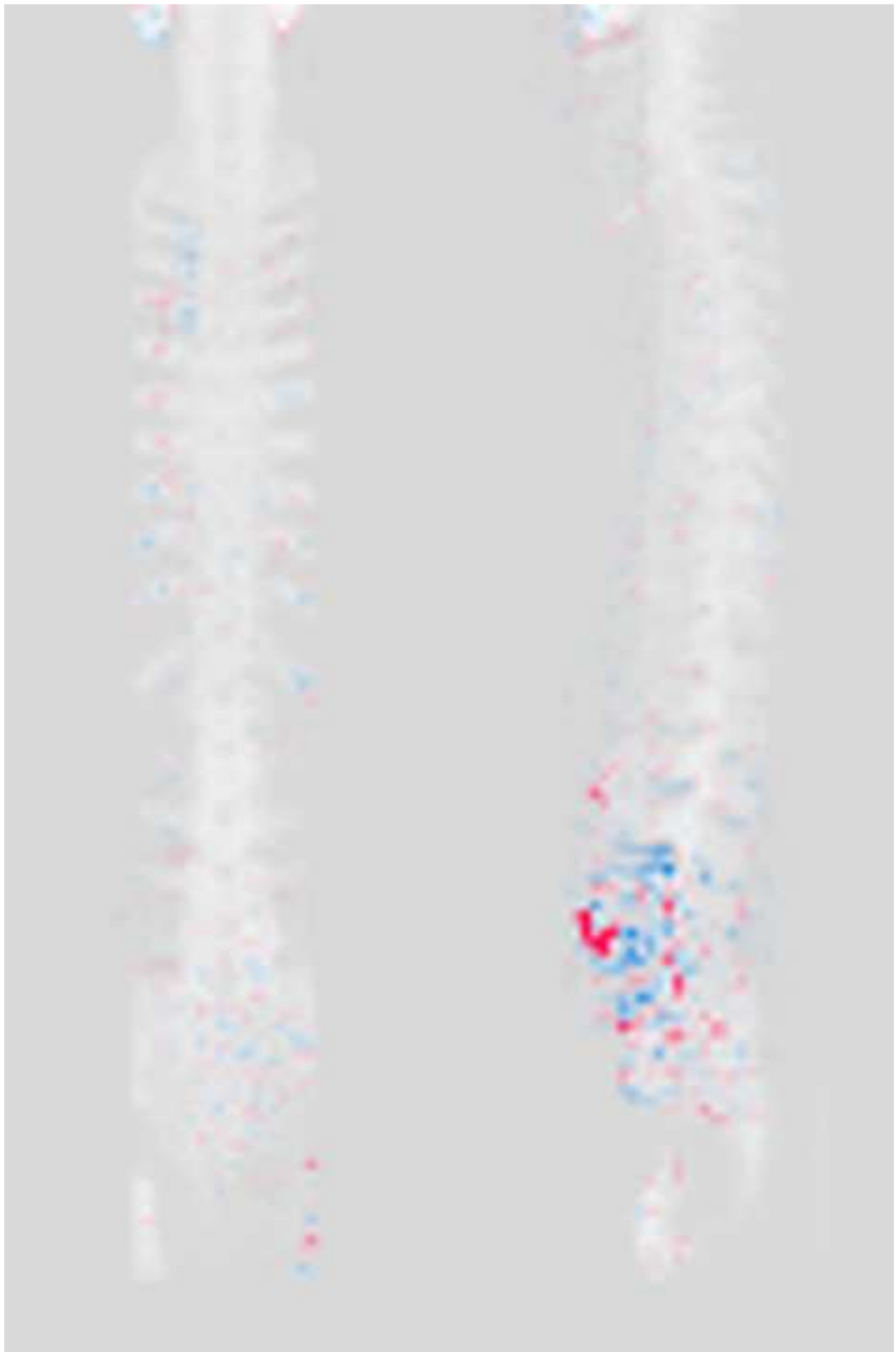
\hat{y}_{ij} : Estimated age at death for case i using the training dataset j

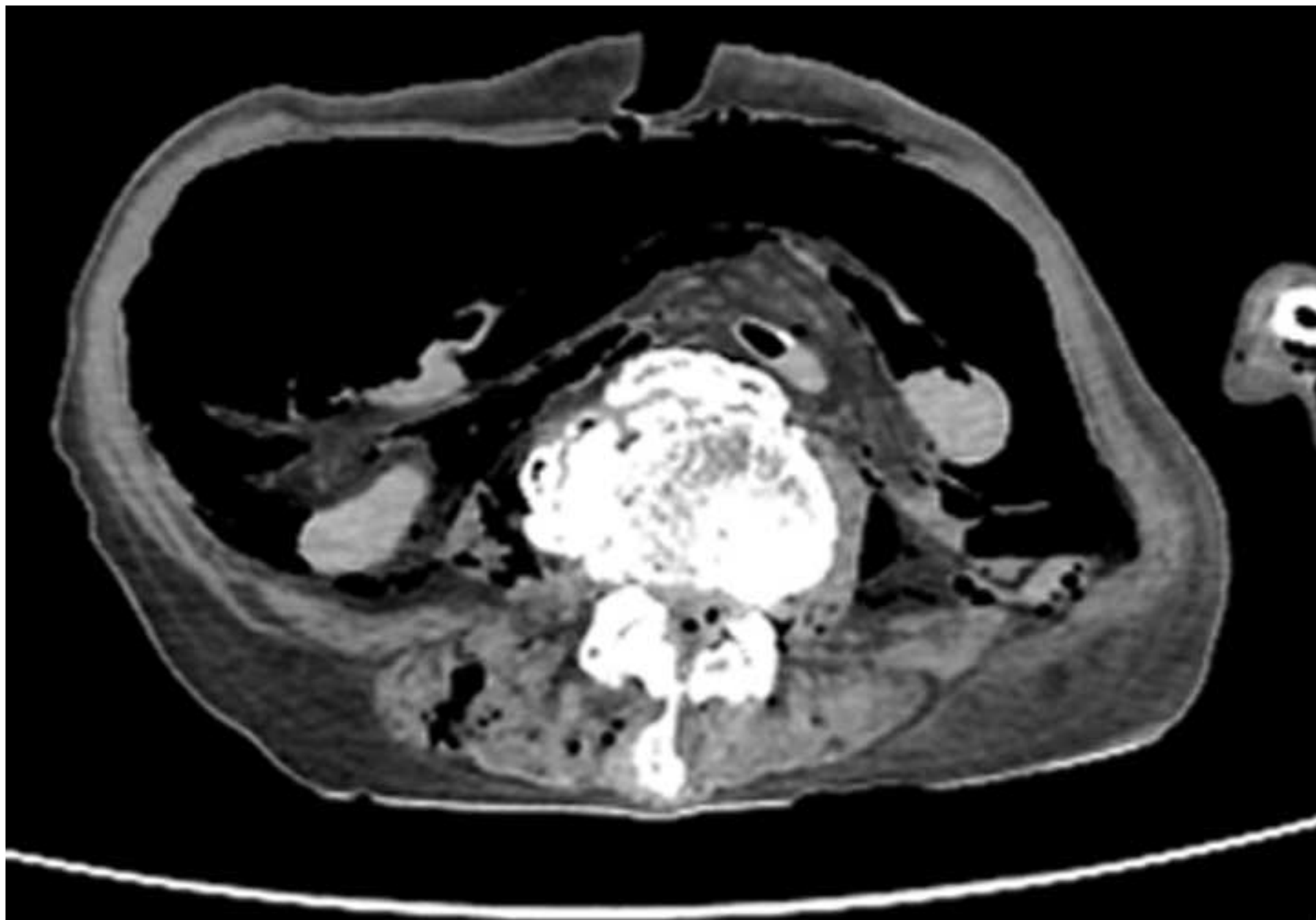




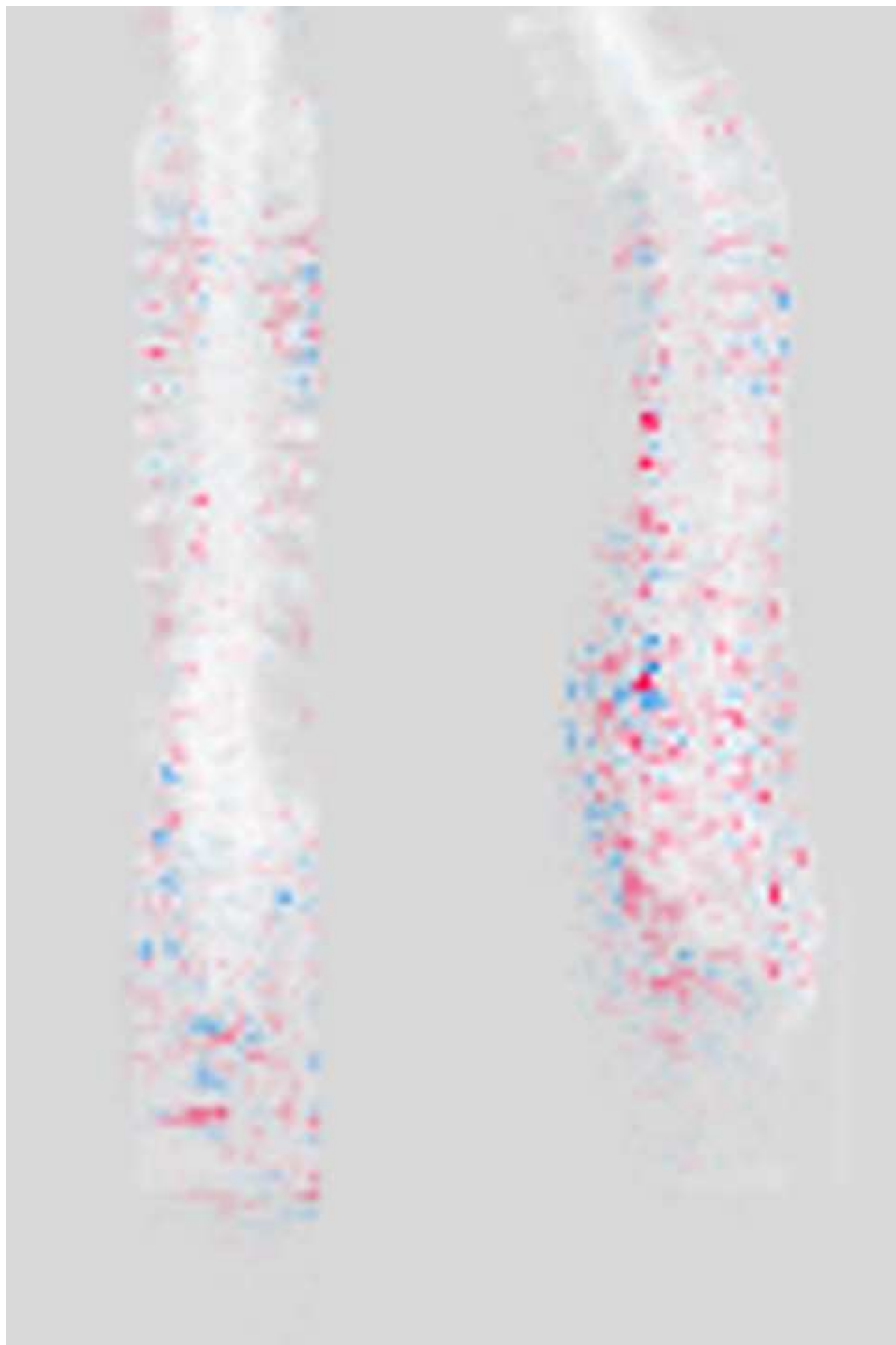






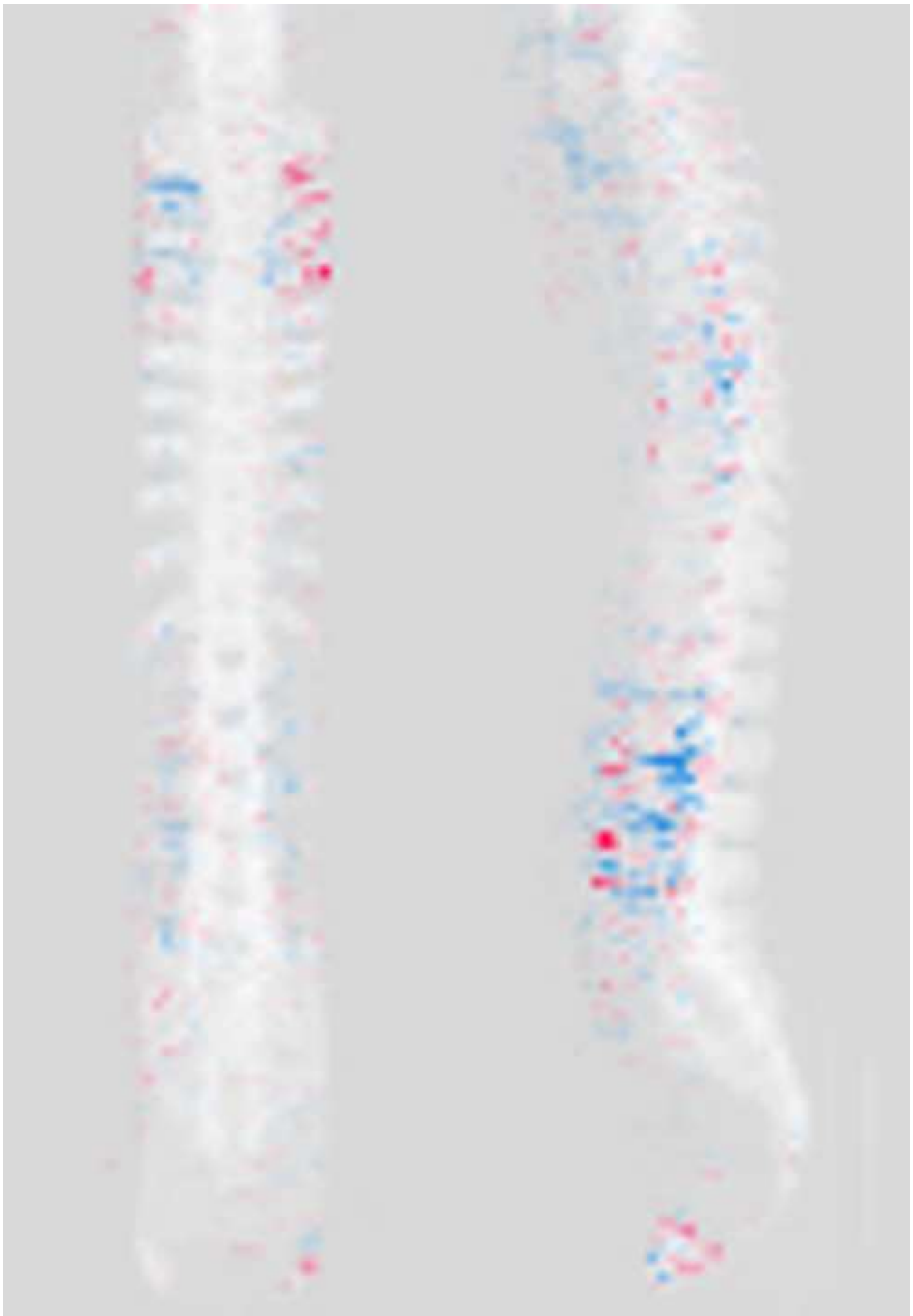


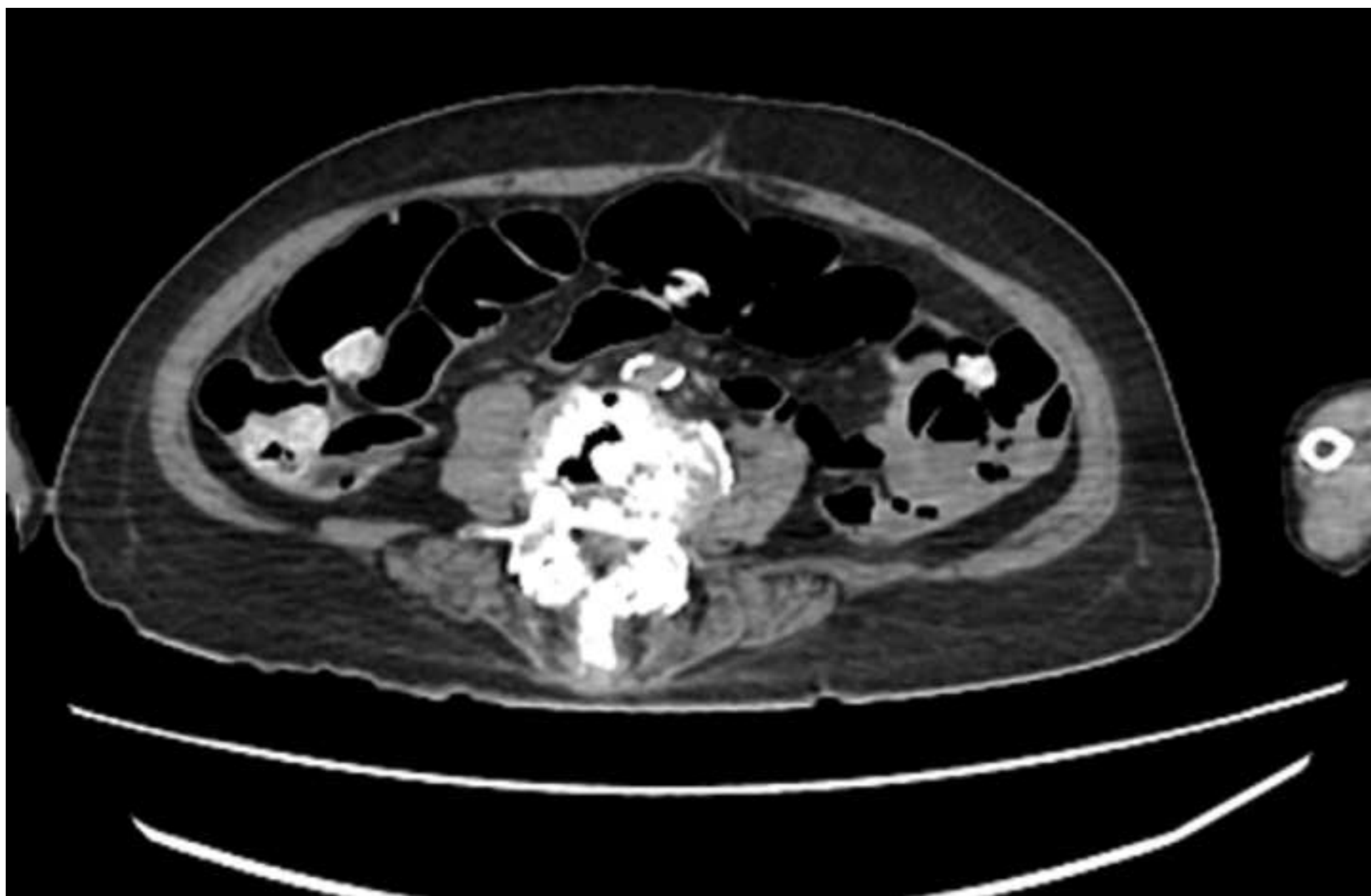




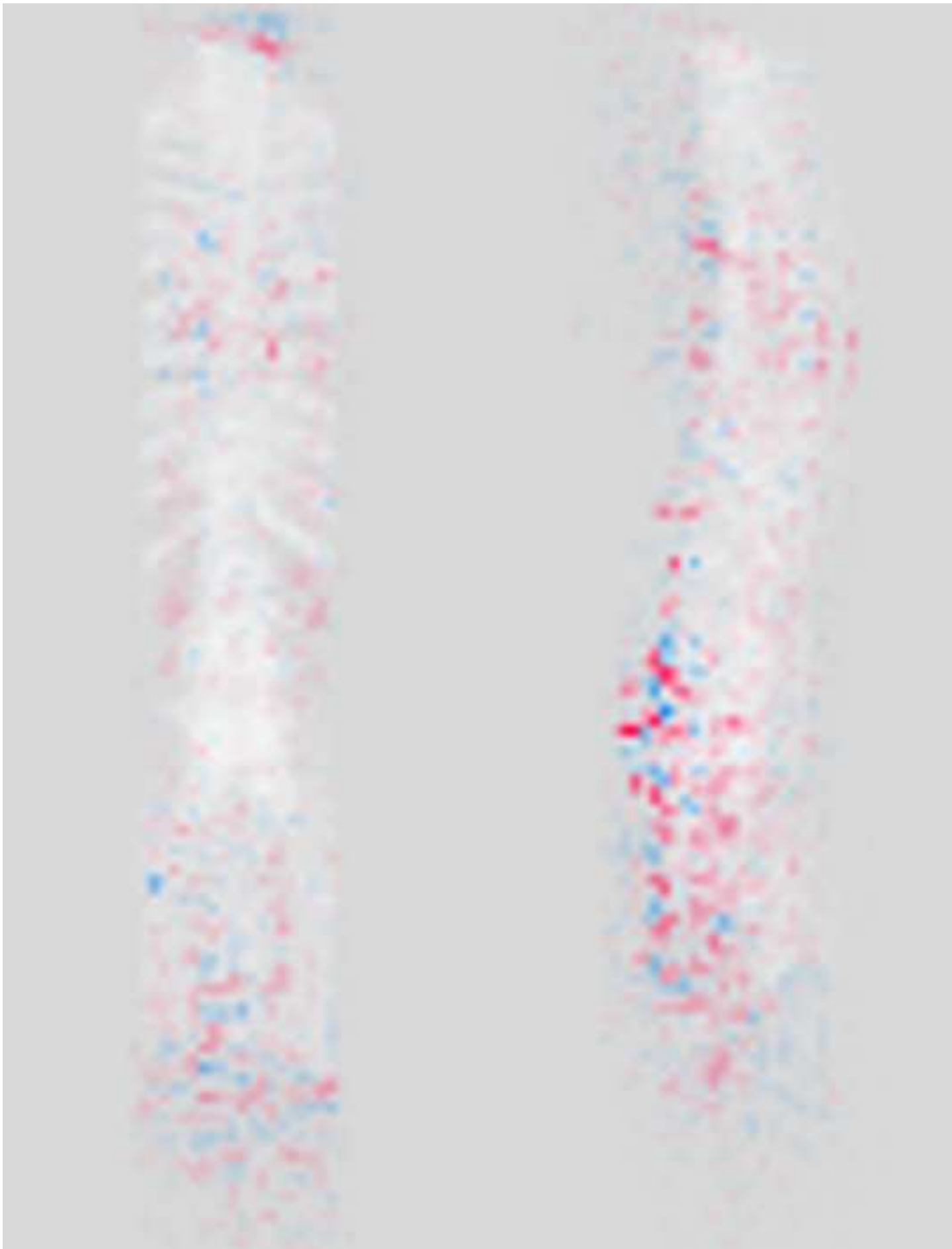












Tables

Table 1. Age distribution data of 1,120 patients

Age	Mean age in years (SD, 95% age range)		
	All (n=1,120)	Male (n=560)	Female (n=560)
20s (20-29)	25.06 (3.00, 20.50-29.84)	24.55 (3.10, 20.50-29.36)	25.58 (2.81, 20.90-29.92)
30s (30-39)	35.22 (2.97, 30.22-39.78)	35.99 (2.83, 30.67-39.78)	34.46 (2.91, 30.22-39.04)
40s (40-49)	44.90 (2.96, 40.16-49.67)	45.02 (3.00, 40.16-49.67)	44.79 (2.91, 40.41-49.54)
50s (50-59)	54.92 (3.07, 50.17-59.86)	54.82 (3.24, 50.17-59.92)	55.03 (2.88, 50.29-59.75)
60s (60-69)	65.76 (2.84, 60.34-69.87)	65.96 (2.70, 60.77-69.89)	65.56 (2.97, 60.34-69.72)
70s (70-79)	74.71 (2.74, 70.02-79.73)	74.42 (2.74, 70.30-79.73)	74.99 (2.71, 70.38-79.69)
80s (80-89)	85.17 (2.61, 80.76-89.84)	84.75 (2.42, 80.85-89.36)	85.60 (2.73, 81.00-89.84)
90s (90-99)	92.46 (2.05, 90.05-98.29)	92.27 (1.75, 90.08-95.67)	92.66 (2.29, 90.14-98.34)
Total	59.78 (22.60, 21.69-93.79)	59.72 (22.48, 21.23-93.55)	59.90 (22.69, 22.61-94.09)

SD: standard deviation

Table 2. Demographic information of 219 cadavers

Number of cadavers	n=219
Mean age (years)	65.14
Gender (Male/Female)	137/82
Damaged-/Undamaged body	68/151
Cause of death	
Burning	47
Drowning	32
Hanging	20
Trauma	21
Stabbing	14
Drug or alcohol intoxication	10
Others	9
Unknown	66