



Diagnosing Drowning in Postmortem CT Images Using Artificial Intelligence

Terumasa Ogawara,¹ Akihito Usui,² Noriyasu Homma³ and Masato Funayama¹

¹Department of Forensic Medicine, Tohoku University Graduate School of Medicine, Sendai, Miyagi, Japan

²Department of Diagnostic Image Analysis, Tohoku University Graduate School of Medicine, Sendai, Miyagi, Japan

³Department of Radiological Imaging and Informatics, Tohoku University Graduate School of Medicine, Sendai, Miyagi, Japan

Imaging features of the lung in postmortem computed tomography (CT) scans have been reported in drowning cases. However, it is difficult for forensic pathologists with limited experience to distinguish subtle differences in CT images. In this study, artificial intelligence (AI) with deep learning capability was used to diagnose drowning in postmortem CT images, and its performance was evaluated. The samples consisted of high-resolution CT images of the chest of 153 drowned and 160 non-drowned bodies captured by an 8- or 64-row multislice CT system. The images were captured with an image slice thickness of 1.0 mm and spacing of 30 mm, and 28 images were typically captured. A modified AlexNet was used as the AI architecture. The output result was the drowning probability for each component image. To evaluate the performance of the proposed model, the area under the receiver operating characteristic curve (AUC) was analyzed, and the AUC value of 0.95 was obtained. This indicates that the proposed AI architecture is a useful and powerful complementary testing approach for diagnosing drowning in postmortem CT images. Notably, the accuracy was 81% (62/77) for cases in which resuscitation was performed, and 92% (216/236) for cases in which resuscitation was not attempted. Therefore, the proposed AI method should not be used to diagnose the cause of death when aggressive cardiopulmonary resuscitation was performed. Additionally, because honeycomb lungs are likely to exhibit different morphologies, emphysema cases should also be treated with caution when the proposed AI method is used to diagnose drowning.

Keywords: artificial intelligence; autopsy; deep learning; drowning; postmortem computed tomography

Tohoku J. Exp. Med., 2023 January, 259 (1), 65-75.

doi: 10.1620/tjem.2022.J097

Introduction

The diagnosis of drowning is one of the most difficult tasks for forensic pathologists (DiMaio and DiMaio 2001; Ludes and Fornes 2003; Lunetta and Modell 2005; Spitz 2006; Levy 2011). Although various supportive findings, such as foam in the airways, pulmonary edema, sand in the airways, and diatoms in the organs, may be present in drowning, these autopsy findings are nonspecific to drowning (DiMaio and DiMaio 2001; Ludes and Fornes 2003; Lunetta and Modell 2005; Spitz 2006; Levy 2011). For example, foam in the airways or pulmonary edema is also observed in deaths caused by cardiovascular diseases, drug overdose, or strangulation (DiMaio and DiMaio 2001;

Ludes and Fornes 2003; Lunetta and Modell 2005; Spitz 2006; Levy 2011; Stephenson et al. 2019). Sand in the airways suggests the aspiration of water, but this is not always indicative of drowning (Spitz 2006). The reliability of the diatom test remains controversial because of false-positive and false-negative results (DiMaio and DiMaio 2001; Ludes and Fornes 2003; Lunetta and Modell 2005; Spitz 2006). Therefore, the diagnosis of drowning requires not only autopsy examination but also scene investigation and the exclusion of other causes of death (DiMaio and DiMaio 2001; Ludes and Fornes 2003; Lunetta and Modell 2005; Spitz 2006; Levy 2011; Stephenson et al. 2019).

In forensic medicine, postmortem computed tomography (CT) has recently become popular as an adjunctive

Received August 23, 2022; revised and accepted November 3, 2022; J-STAGE Advance online publication November 17, 2022

Correspondence: Terumasa Ogawara, Department of Forensic Medicine, Tohoku University Graduate School of Medicine, 2-1 Seiryomachi, Aoba-ku, Sendai, Miyagi 980-8575, Japan.

e-mail: terumasa.ogawara.t3@dc.tohoku.ac.jp

©2023 Tohoku University Medical Press. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC-BY-NC-ND 4.0). Anyone may download, reuse, copy, reprint, or distribute the article without modifications or adaptations for non-profit purposes if they cite the original authors and source properly.
<https://creativecommons.org/licenses/by-nc-nd/4.0/>

means of diagnosing the cause of death. It is typically not possible to perform imaging using contrast media for dead bodies, unlike living bodies. However, simple CT images alone are an effective means of confirming hard tissue damage and identifying hemorrhagic injuries and diseases. Additionally, inflammation and emphysema are relatively easy to detect in the lungs. In drowning cases, pulmonary CT imaging features have also been reported. For example, Levy et al. (2007) evaluated the postmortem CT images of 28 drowning cases, and observed ground-glass opacity with interstitial thickening in 25 out of 28 cases. Christie et al. (2008) reported pulmonary edema caused by fluid exudation from the interstitium in half of 10 drowning death cases. More recently, Usui et al. (2014) analyzed 100 drowning death cases and reported ground-glass opacity with a thickened interstitium in 34 cases and a centrilobular distribution of ill-defined nodules along the airways in 38 cases, with 10 cases exhibiting a mixture of the two types and the remaining 18 cases exhibiting other patterns. However, it is difficult for a forensic pathologist with little or no experience to distinguish subtle differences in CT images. Additionally, in cadavers laid on their backs, congestion and edema caused by postmortem blood hypostasis strongly appear on the dorsal surface of the lungs with the passage of time after death. Because these changes and decomposition never appear in living persons, radiologists who are not experienced in reading the CT images of cadavers may be unable to interpret the images.

Artificial intelligence (AI) can be a powerful aid to forensic pathologists when diagnosing drowning through the examination of postmortem CT images. With the development of computer techniques such as deep learning, AI has improved dramatically over the last decade. Since Krizhevsky's deep learning model called AlexNet achieved the best results in the ImageNet Large-Scale Visual Recognition Challenge 2012 competition, deep learning, which is a subtype of machine learning, has attracted substantial attention because of its great capability in recognizing images (Krizhevsky et al. 2012; Litjens et al. 2017). Deep learning is characterized by automatic feature extraction, which solves two major problems of conventional AI: the time-consuming and often very difficult processing of data input features and the selection bias that occurs when selecting the input features (van Ginneken 2017). These deep learning characteristics have led to extensive research on the application of AI to image recognition. In clinical radiology, computer-aided diagnosis using deep learning has become popular and many researchers have reported the high performance of deep learning on the classification of medical images. For example, AI has been reported to be useful for the classification of tuberculosis in chest radiography (Lakhani and Sundaram 2017), and for the detection and quantification of pneumothorax in chest CT images (Rohrich et al. 2020). If these technologies can be applied to the diagnosis of drowning in the field of forensics, they are expected to replace experienced diagnostic radiologists

in the examination of CT images.

Currently, there are few studies on the application of AI to postmortem CT images, and only studies on the detection of hemorrhagic pericardial fluid (Ebert et al. 2017) and the identification of fatal head injuries (Garland et al. 2020) have been conducted to date. Therefore, in this study, AI was used to conduct drowning diagnosis and the feasibility of using this approach in practice was investigated.

Materials and Methods

Materials

From autopsy cases in which CT images were taken at the Autopsy Imaging Center, Tohoku University Graduate School of Medicine, from June 2012 to January 2021, 153 cases (95 males and 58 females) diagnosed as drowning cases were selected. Two unidentified adult males were excluded, and the mean age of the remaining males and females was 61.9 (range 18-93) and 69.1 (range 37-96) years, respectively. Nineteen of these deaths occurred in bathtubs. As controls, 160 cases (101 males and 59 females) were selected as the non-drowning group. One unidentified adult male was excluded, and the mean age of the remaining males and females was 54.5 (range 19-91) and 57.7 (range 8-91) years, respectively. In the non-drowning group, the causes of death were cardiovascular disease (n = 49), asphyxia other than drowning (n = 19), infection (n = 16), poisoning (n = 15), head injury (n = 15), blood loss (n = 11), alcoholic and diabetic ketoacidosis (n = 11), and other causes (n = 24) (Table 1). The exclusion criteria for both groups were advanced decomposition, infants, severe carbonization, hypothermia, and severe chest trauma. Among these cases, cardiopulmonary resuscitation (CPR) was performed in 29 drowning and 48 non-drowning cases.

CT imaging equipment and imaging conditions

The CT system used in this study was an 8- or 64-row multislice CT system (Aquilion; Toshiba Medical Systems, Tokyo, Japan). The type of multislice CT scanner used in

Table 1. Details of causes of death classified as other* in non-drowning cases.

Cause of death	N
Carbon monoxide poisoning	5
Cervical spine injury, cervical cord injury	4
Fatty liver damage	4
Pulmonary thromboembolism	3
Tension pneumothorax	3
Heatstroke	2
Death by fire	1
Electrocution	1
Malnutrition, dehydration and ketosis	1
Total	24

*See the text for other non-drowning cases.

this study was made by the same manufacturer and had the same image reconstruction functions. The $1\text{ mm} \times 4\text{-row}$ slice configuration mode was used for CT imaging for both multislice CT systems; thus, there was no difference in image quality. The image size was 512×512 and high-resolution CT images of the chest were used. Twenty-eight images with an image slice thickness of 1.0 mm and intervals of 30 mm were captured at seven levels (i.e., pulmonary apex, aortic arch, tracheal bifurcation, hilar region, left ventricle, lung base, and the lowermost part of the left lung) (Fig. 1). However, depending on the case, the levels were changed from 5 to 9. Four images were collected at each level, but in some cases, only two or three images were collected at the lowest level because of missing data.

Deep convolutional neural network and its analysis

The proposed model is a modification of AlexNet (Krizhesky et al. 2012), which consists of five convolutional layers, three pooling layers, and three fully connected layers. The convolutional layers extract features from the input image, the pooling layers reduce the features by extracting the necessary features, and the fully connected layers classify the input data (Krizhesky et al. 2012; Ali et al. 2018; Hosny et al. 2020). In this study, the model was pretrained as described below and then the last fully coupled layer was replaced by two new artificial neurons because the classification considered only two classes: drowning and non-drowning. The output result was the drowning probability for each component image. If the drowning probability of the image was close to 1, the image was more likely to be a drowning image; if the drowning probability of the image was close to 0, the image was more

likely to be a non-drowning image. Furthermore, the arithmetic mean of all images was used to determine whether the overall case was a drowning or non-drowning case. A case with a value greater than a threshold t ($0 \leq t \leq 1$) was considered as a drowning case, and a case with a value less than the threshold was considered as a non-drowning case. In the present study, the additive mean of the values of the drowning probability calculated from each image was used to determine whether the value was greater than or equal to 0.50 for drowning and less than 0.50 for non-drowning. This is because a drowning probability of 0.50 for an image means that the drowning and non-drowning properties of the input image are equal.

Transfer learning and ten-fold cross-validation were used to compensate for the small size of the dataset used in this study. Although large datasets are generally required for AI training, collecting a sufficient amount of data is difficult in medical fields (Dawud et al. 2019). Transfer learning, a technique that uses a pretrained model, makes it possible to train an AI model with limited data in a short time (Choy et al. 2018; Dawud et al. 2019). The proposed model was pretrained on ImageNet, which is a large-scale database of approximately 1.3 million non-medical images (Deng et al. 2009). After transfer learning, ten-fold cross-validation was conducted to reduce random noise and evaluate overfitting. In this method, a dataset is divided into ten non-overlapping groups of approximately the same size; nine groups are used for training and the remaining group is used for testing. This process is repeated ten times, with each group used only once for testing (Molinario et al. 2005). In this study, each group included 14 to 16 drowning cases and 16 non-drowning cases. A workstation with a

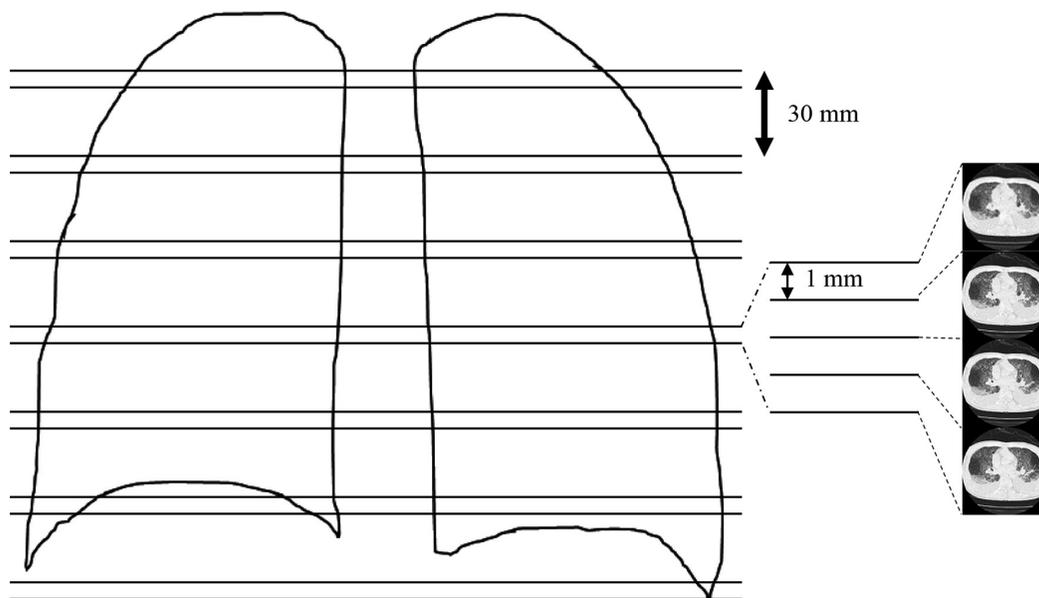


Fig. 1. Schematic representation of high-resolution chest CT scans.

Scanning was performed from the apex pulmonis to the basis pulmonis, and four images with a slice thickness of 1 mm were captured every 30 mm at seven levels (i.e., pulmonary apex, aortic arch, tracheal bifurcation, hilar region, left ventricle, lung base, and lowermost part of the left lung), and sometimes at five to nine levels.

graphics processing unit (NVIDIA GeForce RTX 2080 Ti, Santa Clara, CA, USA) was used for analysis.

To evaluate the performance of the model, the area under the receiver operating characteristic (ROC) curve was analyzed, and the sensitivity, accuracy, and specificity were obtained. The ROC curve is a plot of sensitivity versus 1 – specificity when the threshold t changes between 0 and 1, and the area under the ROC curve (AUC) value represents the diagnostic performance of the test (Obuchowski and Bullen 2018). In this study, the AUC was calculated for each of the ten results obtained using ten-fold cross-validation.

Statistical analysis was conducted using a binomial test. The statistical significance was assessed using the cut-off p value of 0.05. Table 2 shows the items to be considered in this study as a 2×2 contingency table.

Ethics committee approval

The ethics committee of Tohoku University Graduate School of Medicine approved the analysis of the CT images used in this study (2020-1-466).

Results

Fig. 2 shows the drowning probability distribution of drowning and non-drowning deaths. This figure indicates that a threshold of 0.5 is not a significant problem. The proposed AI model had a sensitivity of 86.3% (132/153), specificity of 91.3% (146/160), and accuracy of 88.8% (278/313). From the ROC analysis, all 10 AUC values were greater than 0.9, and their mean and median values were 0.95 and 0.96, respectively (Fig. 3). To investigate the effect of resuscitation on the AI system’s diagnostic performance, only cases in which resuscitation was attempted were analyzed, and the sensitivity, specificity, and accuracy were obtained as 72% (21/29), 85% (41/48), and 81%

Table 2. The items to be considered are presented as a 2×2 contingency table.

Cases diagnosed as drowning at autopsy	
True-positive cases (AI diagnoses drowning with a drowning probability of 0.5 or greater)	False-negative cases (AI diagnosed as non-drowning with probability of drowning less than 0.5)
Cases diagnosed as non-drowning at autopsy	
True-negative cases (AI diagnosed as non-drowning with probability of drowning less than 0.5)	False-positive cases (AI diagnoses drowning with a drowning probability of 0.5 or greater)

The exclusion criteria for both groups were advanced decomposition, infants, severe carbonization, hypothermia, and severe chest trauma. See the text for the drowning probability. AI, artificial intelligence.

Number of cases

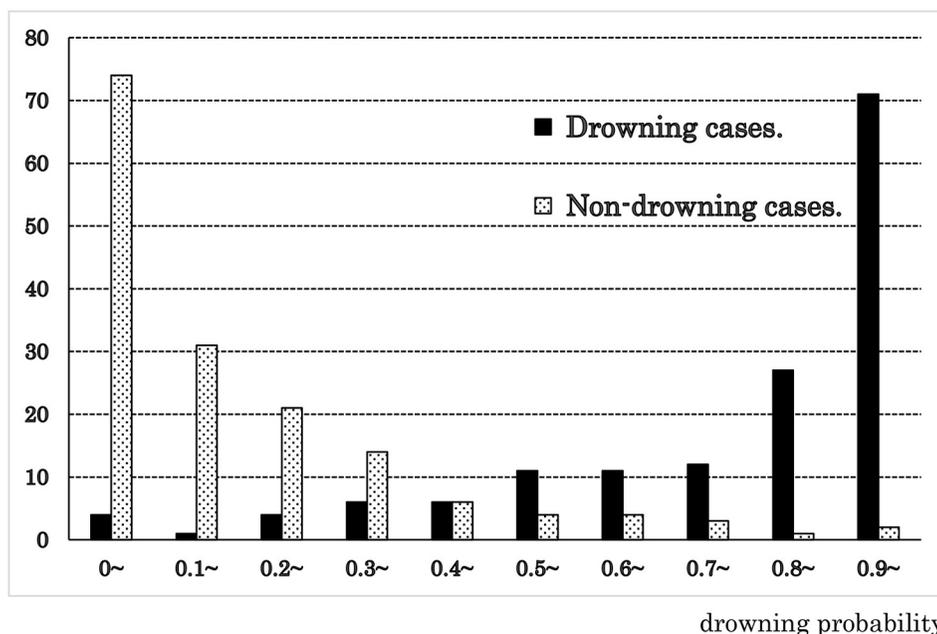


Fig. 2. Probability distribution of drowning and non-drowning cases.

(62/77), respectively, which are lower than those of the all-cases group.

In the following, one false-positive and one false-negative case in which CPR was attempted are discussed. In the false-positive example, Case 2 (Table 3), the patient underwent CPR for one hour. The drowning probability of this case was 0.78, which means that the individual was diagnosed to have died by drowning, even though this was a non-drowning case (cause of death was peritonitis caused by perforation of the gastrointestinal tract). The drowning

probability for each slice image was high (Fig. 4) and ranged from 0.77 to 1.00 for the remaining 20 images, except for the four images at the lowest level (0.10 to 0.20), where the liver and stomach constituted the majority of the images. Therefore, the AI system identified this case as a drowning case. In the false-negative case, Case 21 (Table 4), the individual was found submerged in a domestic waterway. Approximately 1.5 hours had passed since her death, and although she was transported to the emergency room, she was pronounced dead one hour later. The

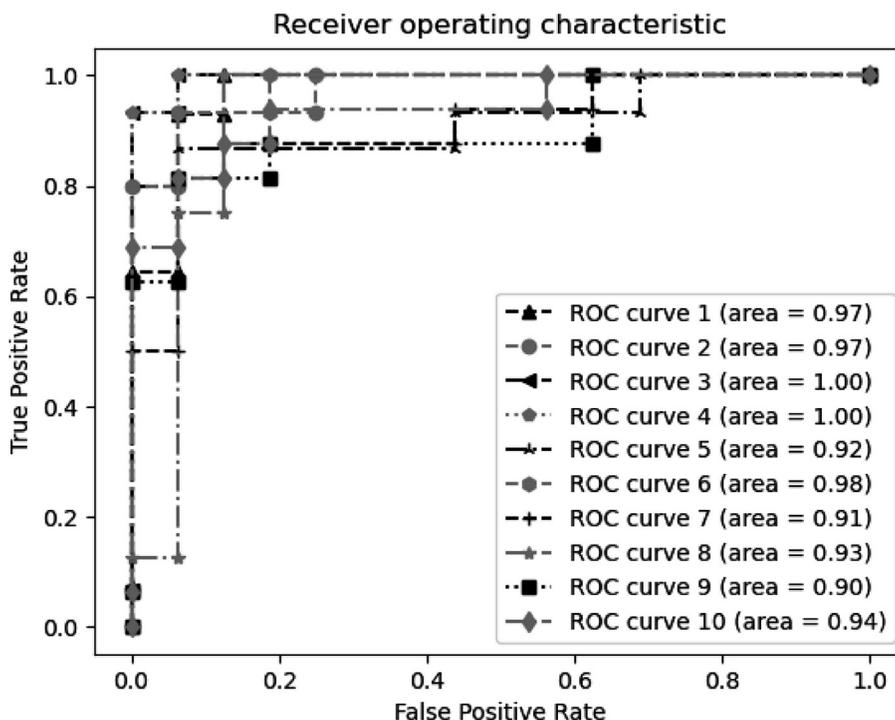


Fig. 3. Receiver operating characteristic (ROC) curves are shown for each of the ten-fold cross-validations. The area under the curve (AUC) calculated for each curve is also shown.

Table 3. False-positive cases.

Case No.	Age	Sex	Cause of death	Resuscitation	Drowning probability
1	80s	M	Pneumonia with pulmonary fibrosis and bronchiectasis	+	0.87
2	40s	M	Perforation peritonitis after blunt abdominal trauma	+	0.78
3	40s	M	Asphyxia due to aspiration of gastric contents after a drug overdose	+	0.72
4	80s	F	Asphyxia due to foreign bodies in the air way	+	0.67
5	70s	F	Blood loss due to sharp force injuries to the neck	+	0.66
6	60s	M	Bronchopneumonia	+	0.57
7	20s	M	Carbon monoxide poisoning	+	0.54
8	20s	M	Acute bronchitis after a drug overdose	-	0.99
9	60s	M	Heart failure due to heart hypertrophy	-	0.96
10	90s	F	Aortic stenosis	-	0.74
11	50s	M	Drug overdose	-	0.63
12	90s	F	Pulmonary thromboembolism	-	0.60
13	50s	M	Peritonitis due to gastric perforation	-	0.57
14	60s	F	Blood loss due to gastric hemorrhage with ketoacidosis	-	0.51

M, male; F, female; +, attempted resuscitation; -, resuscitation not attempted.

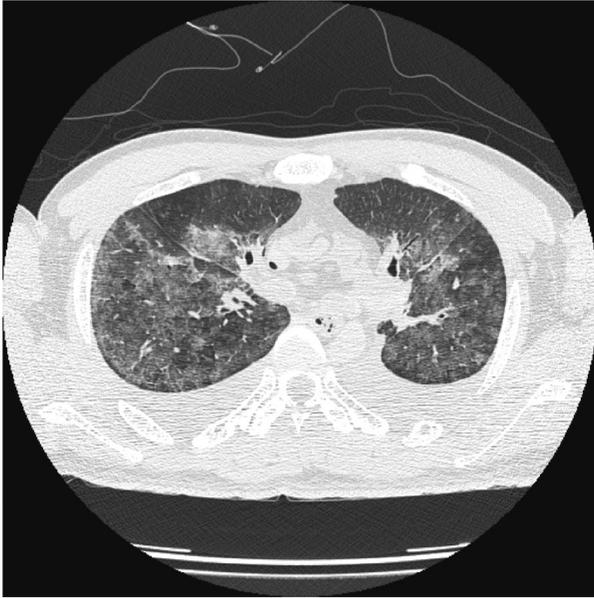


Fig. 4. False-positive case (Case 2) in which emergency resuscitation was performed. This image is the first slice of the third level, and its drowning probability was calculated as 0.994. The cause of death was peritonitis. The overall drowning probability of this case was 0.78.

autopsy revealed that the left lung weighed 620 g, whereas the right lung weighed 790 g and had severe edema. The AI system calculated the drowning probability as 0.28 and diagnosed the death as non-drowning. The drowning probability of each image was low, ranging from 0.04 to 0.33 for 11 out of 12 images in the first to third levels, and from 0.00 to 0.36 for 10 out of 12 images in the fifth to seventh levels. However, at the fourth level, all four images had high values ranging from 0.75 to 1.00. A radiologist with experience in postmortem image interpretation commented that there was no clear difference between the four images at the fourth level and the 21 images that the AI system determined to be non-drowning cases (Fig. 5).

Of the 160 cases in the non-drowning group, 14 were false-positives (Table 3). Of these 14 cases, resuscitation was attempted in seven cases. For the remaining seven false-positives, the drowning probability ranged from 0.51 to 0.99. As presented in Table 3, the causes of death in the misclassified cases included heart diseases and poisoning, which are known to be accompanied by pulmonary edema, as nonspecific findings. Therefore, to investigate the effect of these entities on the AI system's diagnosis, the false-positive rate was calculated again. After excluding cases in which CPR was attempted, the false-positive rates for death caused by heart disease, poisoning, and asphyxia (strangulation in four cases, airway obstruction in three cases, chest compression in one case, and positional asphyxia in one case) were obtained as 5.2% (2/38), 13% (2/15), and 0% (0/9), respectively.

Two of the seven false-positive cases are illustrated in the following, with drowning probabilities in the 0.7 range

Table 4. False-negative cases.

Case No.	Age	Sex	Resuscitation	Drowning probability
15	30s	M	+	0.49
16	70s	M	+	0.47
17	50s	M	+	0.45
18	50s	M	+	0.37
19	70s	M	+	0.34
20	30s	M	+	0.32
21	70s	F	+	0.28
22	80s	F	+	0.12
23	90s	F	-	0.45
24	80s	M	-	0.42
25	80s	M	-	0.40
26	40s	F	-	0.37
27	90s	F	-	0.34
28	60s	M	-	0.32
29	50s	F	-	0.27
30	30s	M	-	0.26
31	70s	M	-	0.25
32	80s	M	-	0.07
33	50s	M	-	0.02
34	40s	M	-	0.02
35	30s	M	-	0.01

M, male; F, female; +, attempted resuscitation; -, resuscitation not attempted.

and nearly 1. Case 10 (Table 3) was classified as a drowning case by the AI system because the drowning probability was calculated as 0.74. In this case, a woman in her nineties was found lying in the prone position on a slope half a day after the time she was last known to be alive. She was placed in the supine position and autopsied 16 hours later. The cause of death was diagnosed as aortic valve stenosis. The lung weight was 460 g for both the right and left sides, and moderate to severe edema was observed along with lung aeration. In the pulmonary CT imaging, out of a total of 28 images at seven levels, one image from the second level, four images from the third level, and one image from the sixth level were in the range of 0.00-0.49, whereas the other 22 images were in the range of 0.54-1.00. The radiologist commented that there existed edema-like shadows, which were observed on the ventral and dorsal sides of the lungs and thought to have been caused by postmortem positional changes. However, the shadows did not differ significantly from image to image, which makes it difficult to attribute the difference in the drowning probability to the difference in the gross findings (Fig. 6). Case 8 (Table 3) has a drowning probability of 0.99 (0.90-1.00), and is a confirmed drowning case according to the AI. This case is a male in his twenties, found in the supine position 10 hours after the time he was last known to be alive, and autopsied 33 hours later. The immediate cause of death was diagnosed as circulatory failure, and was assumed to be related

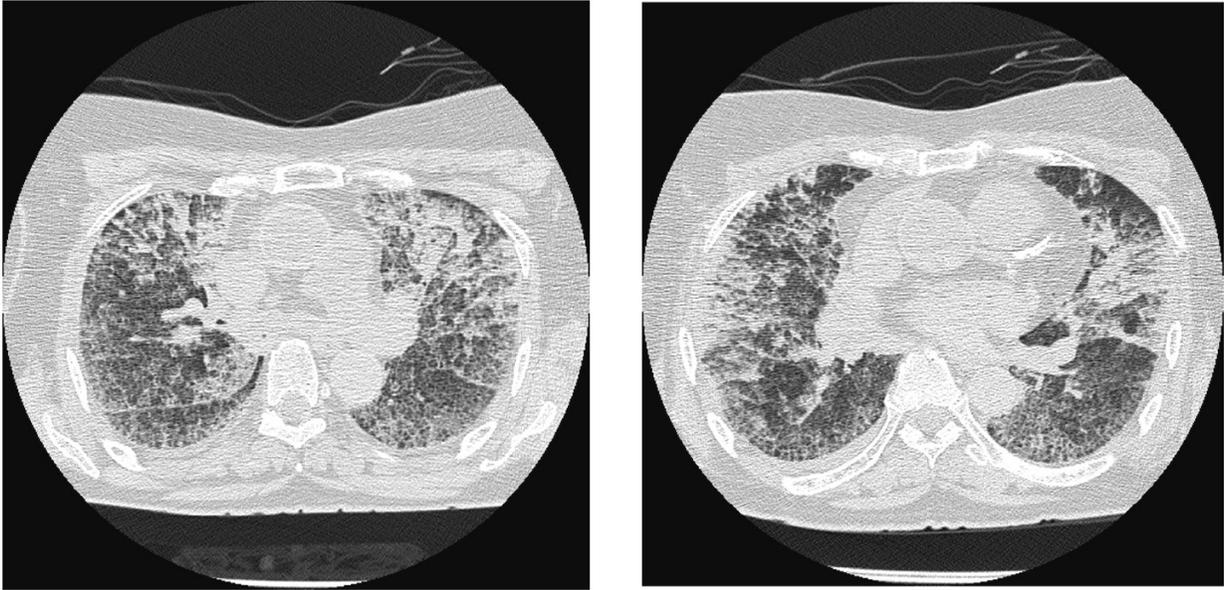


Fig. 5. False-negative case (Case 21) in which emergency resuscitation was performed.

The overall drowning probability of this case was 0.28. First slice image of the third level (left, drowning probability of 0.329) and that of the fourth level (right, drowning probability of 0.748). Even a radiologist experienced in examining postmortem CT images was unable to identify a qualitative difference between the left and right images.

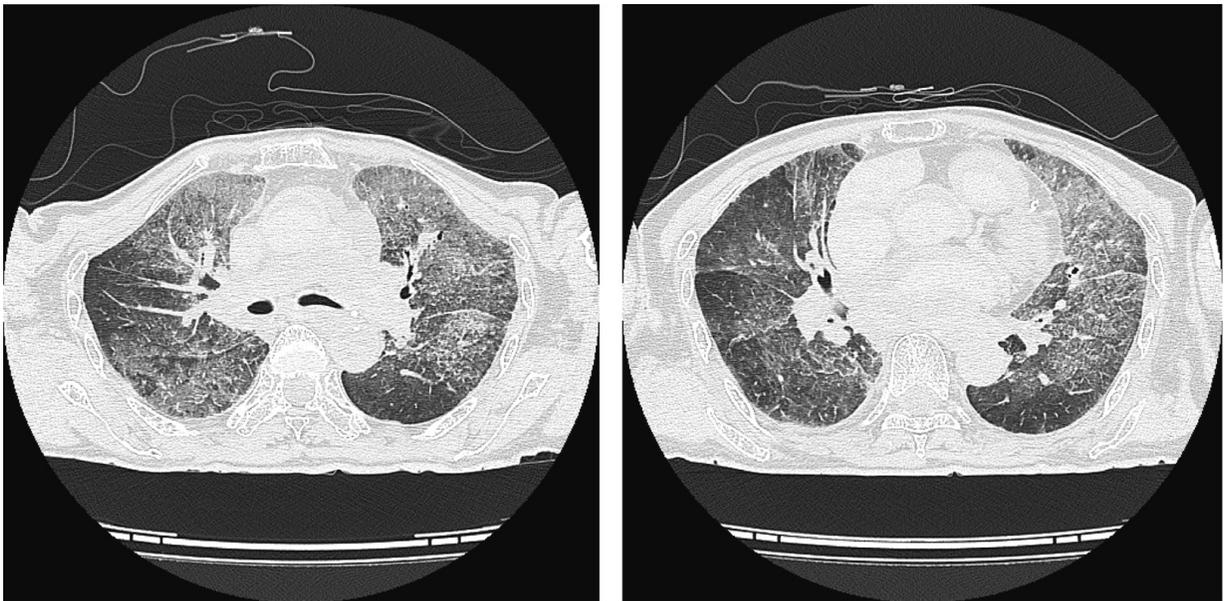


Fig. 6. False-positive case (Case 10) without emergency resuscitation.

The cause of death was aortic stenosis. The overall drowning probability of this case was 0.74. First slice image of the third level (left, drowning probability of 0.001) and that of the fourth level (right, drowning probability of 0.898). A radiologist identified edema-like shadows on the ventral and dorsal sides of the lungs, which were attributed to postmortem repositioning. The body was found in the prone position on a slope and was then placed on its back. However, it was difficult to identify qualitative differences between the two images.

to drug ingestion in the moderate intoxication range and moderate bronchitis. The lung weight was large, that is, 700 g on the left side and 690 g on the right side, and a small amount of foam existed in the airway along with some food residue. Edema-like shadows were observed in the pulmonary images (Fig. 7).

Of the 153 drowning cases, 21 were false-negatives (Table 4). In eight of these cases, CPR was performed. The drowning probability in the 13 cases in which CPR was not performed ranged from 0.01 to 0.45. Of these 13 cases, honeycomb lung was found in three cases, with drowning probabilities of 0.34, 0.40, and 0.42. Furthermore, the

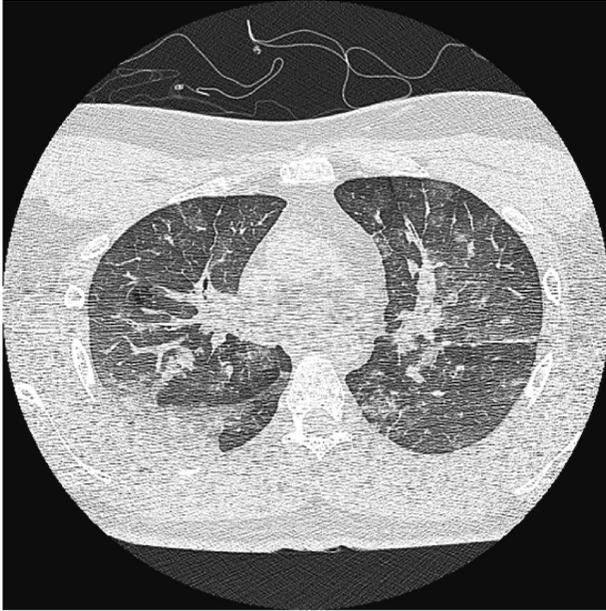


Fig. 7. False-positive case without emergency resuscitation (Case 8).

This image is the first slice of the third level, and its drowning probability was calculated as 0.998. The overall drowning probability of this case was 0.99. The pulmonary images exhibited edema-like shadows.

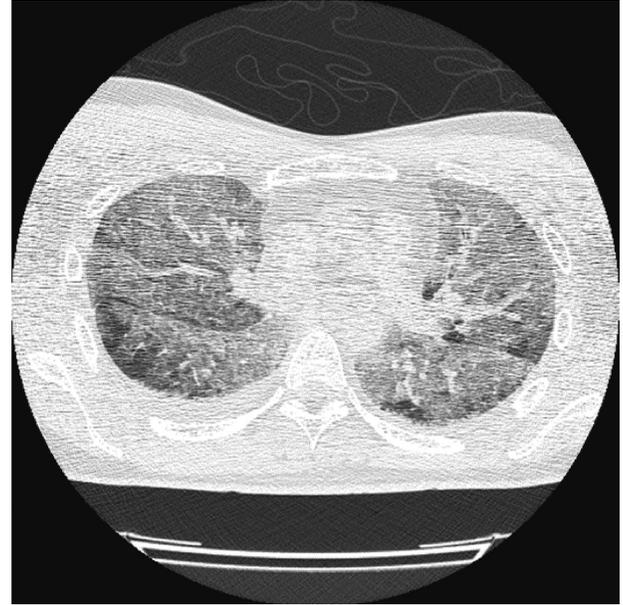


Fig. 8. False-negative case without emergency resuscitation (Case 35).

The body was found face down in a river. This image is the first slice of the third layer, and its drowning probability was calculated as 0.002. The overall drowning probability of this case was 0.01. The lung weights were extremely large (left 1,000 g; right 1,200 g), and the lung was highly edematous. Although drowning aspiration cannot be ruled out, this is not a typical drowning CT image.

drowning probability was less than 0.1 in four cases. These included one case of decomposition (included in this study because the decomposition appeared to be minor), one case of suspected preceding myocardial ischemia, and one case in which drowning was diagnosed as the immediate cause of death, even though the patient had a fatal upper cervical cord injury. In the remaining case, the lung weights were extremely large (left 1,000 g; right 1,200 g), and a gross diagnosis of drowning was made. However, the radiologist's examination of the lung images indicated that there existed relatively uniform edema-like shadows over the entire lung, which is not a typical drowning finding. In this case, Case 35 (Table 4), a male in his thirties, was found in the prone position in a river at a water temperature of 4°C in winter. Although he had been dead for approximately five days, there were few postmortem changes. The lung weights were extremely large (left 1,000 g; right 1,200 g) with severe edema. The pleural effusion was 70 ml on the left and 100 ml on the right. Although the gross findings suggest that the edema in the lungs was caused by the massive aspiration of drowning fluid, the drowning probability for all seven levels of 28 images was extremely low at 0.01. By gross examination, this appeared to be a drowned lung. However, the radiologist commented that, although drowning aspiration was undeniable, this was not a typical CT image of drowning (Fig. 8).

Discussion

In this study, an average AUC of 0.95 was achieved by analyzing postmortem high-resolution CT chest images

using an AI approach with deep learning. Generally, AUC values, which are a criterion of discrimination accuracy, are classified as poor for AUC values of 0.6-0.7, fair for AUC values of 0.7-0.8, good for AUC values of 0.8-0.9, and excellent for AUC values of 0.9-1.0 (Lindberg et al. 2015). Hence, the AUC value of 0.95 indicates that the proposed model is an excellent classifier. Van Hoyweghen et al. (2015) used postmortem CT imaging to differentiate between drowning and other asphyxiation cases. They compared several findings, including the ground-glass opacity of the lungs, and did not identify significant differences in any of the findings, except for the height of the right diaphragm. However, in their study, the sample size was small (14 drowning cases and 11 non-drowning cases), and the presence or absence of CPR or decomposition was unknown. Additionally, the image findings of ground-glass opacity were scored only based on presence or absence. However, it is difficult to numerically evaluate the density and distribution of ground-glass opacity findings based on naked-eye observations.

An advantage of AI is that findings can be evaluated numerically. In this study, this numerical evaluation is called the drowning probability and has a threshold of 0.5; that is, a value of 0.5 or greater indicates drowning, whereas a value less than 0.5 indicates non-drowning. Considering the probability distribution of all cases, there seems to be no problem in using 0.5 as the threshold. Clearly, it is

debatable whether this approach is appropriate. However, if the value is at least close to 1, the image is close to a drowning chest CT image; if the value is close to 0, the image can be considered as a non-drowning image.

Notably, although a highly accurate diagnosis based on image recognition was achieved in this study, false-positive or false-negative results were obtained in some cases. In forensic cases, errors in determining the cause of death can severely impair subsequent criminal investigations. In deep learning, image features are extracted using AI, and it is difficult for researchers to understand the features and extraction process (Ebert et al. 2017; Hosny et al. 2020). However, it may be possible to verify whether a certain trend exists in drowning diagnosis using deep learning, and if so, the trend type can be analyzed by comparing the drowning probability, that is, the AI system's assessment value for each CT image, to actual CT images, autopsy findings, and the circumstances at the time of discovery.

It should be mentioned that, before investigating the causes of false-positive and false-negative results, it is necessary to discuss the presence or absence of CPR in the cases in question. When CPR was performed, the AI system's diagnostic performance decreased for both drowning and non-drowning cases. This suggests that CPR may be a reason for false-positive and false-negative results because it can result in the appearance of intracardiac macrovascular and intrahepatic air, dilated gastrointestinal tract, rib and sternal fractures, and pulmonary edema caused by infusion. In the authors' experience, the longer the duration of CPR, the more air is lost from the lungs, the higher the infusion volume, and the stronger the findings of pulmonary edema. However, as shown in Case 2 (Fig. 4) and Case 21 (Fig. 5), it is not currently known how the AI system makes the drowning/non-drowning diagnosis, at least when active resuscitative actions, including a large amount of infusion and cardiac massage, are performed. Therefore, the authors believe that the results obtained by the AI system should not be used as a diagnostic aid.

In this study, 14 out of 160 cases in the non-drowning group were determined to be false-positives, and in seven of these cases, CPR had not been performed. From these, cases of heart disease (Case 10; Fig. 6) and poisoning (Case 8; Fig. 7) were extracted and the CT images were compared. Sudden deaths caused by cardiovascular diseases, as in Case 10, or deaths involving poisoning, as in Case 8, may exhibit severe congestion and edema in the lungs, with large amounts of fine bubbles in the airways. In comparisons of the lungs in cases in which the cause of death is drowning with those in which the cause of death is cardiovascular disease or poisoning, forensic pathologists have often experienced that the former generally contained air and were relatively weakly congested (the fluid leaking from the lung cut surface was bloody with few red cells), whereas the latter contained little air and were often highly congested. In this study, excluding cases in which CPR was performed, there were only two false-positive cases out

of 38 cases in which the cause of death was circulatory failure, such as sudden death from ischemic heart disease and other cardiovascular diseases, including Case 10 mentioned above. Furthermore, of the 15 deaths caused by poisoning, such as carbon monoxide poisoning, there were two false-positives, including Case 8 mentioned above. All nine asphyxial deaths, excluding drowning cases, were determined as true negatives. Therefore, in most cases of sudden cardiac death, acute poisoning, and asphyxia other than drowning, the AI system distinguished between the chest images associated with these cases and those associated with drowning.

Of the 21 cases with false-negative AI results despite the drowning diagnosis by autopsy, CPR was not performed in 13 cases. In three of these 13 cases, honeycomb lung was observed. In honeycomb lung, cystic dilation that consists of the collapse and fusion of the fibrotic alveoli, and the dilation of the alveolar ducts is observed. In such cases, diagnosis using AI may be difficult because alveolar contraction and dilation are suppressed. A woman in her nineties, who was found in a river in the supine position (Case 27; Table 4), was considered as an example case of honeycomb lung in this study. The drowning probability for this case was 0.34. However, the drowning probability per image at each level varied widely, with a drowning probability above 0.5 for three images at the third level and all four images at the sixth level, and a drowning probability in the range of 0.03-0.47 for the remaining 17 images. The lung weights were large (left 610 g; right 620 g), the pulmonary edema was severe, and a large amount of fine frothy fluid existed at the lung cut surfaces. There was also fine frothy fluid in the airway. Therefore, the drowning diagnosis may have been based on the gross findings. Notably, Usui et al. (2014), who grossly classified postmortem CT lung images of drowned lungs, classified honeycomb lung as an atypical group in their study. Furthermore, in this study, two false-positive cases with drowning probabilities of 0.61 and 0.57 also included honeycomb lung. Therefore, honeycomb lungs are likely to exhibit different morphologies at each level of aspiration imaging, depending on the stage of the disease, and honeycomb lung cases must be treated with extreme care in drowning diagnosis using the proposed AI system. Out of 10 cases, excluding the three honeycomb lung cases, four had a drowning probability of less than 0.1. One of these cases (Case 35; Fig. 8), illustrated in the previous section, had a high degree of edema on gross findings and may have been diagnosed rather aggressively as a drowned lung. It is unclear why the AI system would diagnose these cases as negative. Future analysis using software such as Gradient-weighted Class Activation Mapping (Grad-CAM), which highlights feature areas as heat maps, will be performed.

As discussed previously, false-positives and false-negatives exist in the AI system's diagnosis performed by the analysis of postmortem chest CT images, as in clinical imaging (Ali et al. 2018). In particular, false-positives are

directly related to the risk of not identifying a crime. For example, if a person is killed by neck compression and then dumped into water, and the AI system determines before an autopsy that death was caused by drowning, police investigators and police physicians with little experience may easily misidentify the death as accidental or suicide without conducting a sufficient background investigation or physical examination. In this study, the majority of poisoning and asphyxiation deaths were determined to be non-drowning deaths, with the exception of cases in which CPR was performed, which indicates the usefulness of the AI system. However, further analytical studies on false-positive cases should be conducted in future work. By contrast, the number of false-negative cases was somewhat high in this study. Despite this, when a body is found underwater and there is no evidence of drowning or the evidence of drowning in the images is weak, a criminal investigation will be conducted vigorously. Therefore, the number of false-negatives is not considered to be a major problem in practice, including in criminal investigations.

Finally, the most common case of drowning in Japan is the death of an elderly person in a bathtub. However, with exception of cases in which postmortem changes have progressed, very few of these cases are autopsied. Nineteen cases were analyzed in this study, and excluding emergency resuscitation cases, the number was only 11. Of these cases, two were false negatives and nine were true positives, with probability distributions for the latter in the 0.6 range in two cases, in the 0.7 range in two cases, in the 0.8 range in one case, and in the 0.9 range in four cases, showing a trend toward lower probability values when compared with the probability distributions for all drowning cases in Fig. 2. By contrast, one false-negative case was Case 34 (Table 4): Case 34 was a male in his forties, who was found lying face down in a bathtub with hot water one hour after the last time he was known to be alive. The temperature of the water was 60°C when he was found dead. Some areas of the body surface exhibited skin slippage. The lung weight was 490 g on the left side and 540 g on the right side. The lungs were congested, foam fluid leaked from the cracked surfaces of the lungs, and the airways were filled with fine foam fluid. The airway and lung findings were considered to be indicative of drowning aspiration. However, the drowning probability for all 24 images at all six levels was extremely low at 0.02. After examining the lung CT images, the radiologist commented that there were minor edema-like shadows on the ventral and dorsal surfaces, but also commented that the shadows were heterogeneous, which suggests drowning aspiration. In this case, there was relatively strong stenosis in all three coronary arteries, and contractile zone necrosis was observed only in some myocardial areas, which suggests that it is possible that an ischemic heart attack had already occurred prior to bathing. Reportedly, some diseases are responsible for the death of elderly persons in bathtubs (DiMaio and DiMaio 2001; Ludes and Fornes 2003; Spitz 2006). In many cases,

ischemic heart disease is diagnosed without an autopsy simply because there are no bubbles in the nasal or oral cavities, but an autopsy should be performed to confirm the presence or absence of the disease involved. With regard to deaths in the bathtub, more cases of bathtub deaths need to be collected to analyze the differences in the imaging patterns of the lungs, in addition to the presence or absence of drowning suction, compared with deaths outdoors.

In conclusion, the results obtained by the AI system's analysis of chest CT images of drowned individuals had an AUC value of 0.95, which indicates extremely high diagnostic performance. Therefore, AI analysis may be useful as a supplementary test for drowning diagnosis for deceased individuals found in water. Because several false-positive and false-negative results were obtained for cases in which CPR was performed, it is considered necessary to exclude cases in which CPR was vigorously performed. Additionally, honeycomb lungs are likely to exhibit different morphologies at each level of aspiration imaging, depending on the severity, and emphysema cases should also be treated with caution when AI is used in drowning diagnosis.

Acknowledgments

The authors thank Dr. Yusuke Kawasumi, M.D., for his comments on the findings in the postmortem CT images.

Conflict of Interest

The authors declare no conflict of interest.

References

- Ali, I., Hart, G.R., Gunabushanam, G., Liang, Y., Muhammad, W., Nartowt, B., Kane, M., Ma, X. & Deng, J. (2018) Lung nodule detection via deep reinforcement learning. *Front. Oncol.*, **8**, 108.
- Choy, G., Khalilzadeh, O., Michalski, M., Do, S., Samir, A.E., Pianykh, O.S., Geis, J.R., Pandharipande, P.V., Brink, J.A. & Dreyer, K.J. (2018) Current applications and future impact of machine learning in radiology. *Radiology*, **288**, 318-328.
- Christe, A., Aghayev, E., Jackowski, C., Thali, M.J. & Vock, P. (2008) Drowning--post-mortem imaging findings by computed tomography. *Eur. Radiol.*, **18**, 283-290.
- Dawud, A.M., Yurtkan, K. & Oztoprak, H. (2019) Application of deep learning in neuroradiology: brain haemorrhage classification using transfer learning. *Comput. Intell. Neurosci.*, **2019**, 4629859.
- Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. & Fei-Fei, L. (2009) ImageNet: A Large-Scale Hierarchical Image Database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, pp. 248-255.
- DiMaio, V. & DiMaio, D. (2001) Death by drowning. In *Forensic Pathology*, 2nd ed., CRC Press LLC, Boca Raton, FL, pp. 399-407.
- Ebert, L.C., Heimer, J., Schweitzer, W., Sieberth, T., Leipner, A., Thali, M. & Ampanozi, G. (2017) Automatic detection of hemorrhagic pericardial effusion on PMCT using deep learning - a feasibility study. *Forensic Sci. Med. Pathol.*, **13**, 426-431.
- Garland, J., Ondruschka, B., Stables, S., Morrow, P., Keshu, K., Glenn, C. & Tse, R. (2020) Identifying fatal head injuries on postmortem computed tomography using convolutional neural

- network/deep learning: a feasibility study. *J. Forensic Sci.*, **65**, 2019-2022.
- Hosny, K.M., Kassem, M.A. & Fouad, M.M. (2020) Classification of skin lesions into seven classes using transfer learning with AlexNet. *J. Digit. Imaging*, **33**, 1325-1334.
- Krizhchysky, A., Sutskever, I. & Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems, vol.1*, pp. 1097-1105.
- Lakhani, P. & Sundaram, B. (2017) Deep learning at chest radiography: automated classification of pulmonary tuberculosis by using convolutional neural networks. *Radiology*, **284**, 574-582.
- Levy, A.D. (2011) Death by Drowning and Bodies Found in Water. In *Essentials of Forensic Imaging*, 1st ed., editd by Levy, A.D. & Harcke H.T. Jr., CRC Press LCC, Boca Raton, FL., pp. 183-199.
- Levy, A.D., Harcke, H.T., Getz, J.M., Mallak, C.T., Caruso, J.L., Pearse, L., Frazier, A.A. & Galvin, J.R. (2007) Virtual autopsy: two- and three-dimensional multidetector CT findings in drowning with autopsy comparison. *Radiology*, **243**, 862-868.
- Lindberg, L., Grubb, D., Dencker, D., Finnhult, M. & Olsson, S.G. (2015) Detection of mouth alcohol during breath alcohol analysis. *Forensic Sci. Int.*, **249**, 66-72.
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J., van Ginneken, B. & Sanchez, C.I. (2017) A survey on deep learning in medical image analysis. *Med. Image Anal.*, **42**, 60-88.
- Ludes, B. & Fornes, P. (2003) Drowning. In *Forensic Medicine: Clinical and Pathological Aspects*, 1st ed., edited by Payne-James, J., Busutti, A. & Smock, W., Greenwich Medical Media Ltd., London, pp. 247-257.
- Lunetta, P. & Modell, J.H. (2005) Macroscopical, Microscopical, and Laboratory Findings in Drowning Victims. In *Forensic Pathology Reviews, vol. 3*, 1st ed, edited by Tsokos, M., Human Press Inc., Totowa, NJ, pp. 3-77.
- Molinaro, A.M., Simon, R. & Pfeiffer, R.M. (2005) Prediction error estimation: a comparison of resampling methods. *Bioinformatics*, **21**, 3301-3307.
- Obuchowski, N.A. & Bullen, J.A. (2018) Receiver operating characteristic (ROC) curves: review of methods with applications in diagnostic medicine. *Phys. Med. Biol.*, **63**, 07TR01.
- Rohrich, S., Schlegl, T., Bardach, C., Prosch, H. & Langs, G. (2020) Deep learning detection and quantification of pneumothorax in heterogeneous routine chest computed tomography. *Eur. Radiol. Exp.*, **4**, 26.
- Spitz, D.J. (2006) Investigation of Bodies in Water. In *Spitz and Fisher's Medicolegal Investigation of Death: Guidelines for the Application of Pathology to Crime Investigation*, 4th ed., edited by Spitz, W.U. & Spitz, D.J., Charles C. Thomas, Springfield, IL, pp. 846-881.
- Stephenson, L., Van den Heuvel, C. & Byard, R.W. (2019) The persistent problem of drowning - a difficult diagnosis with inconclusive tests. *J. Forensic Leg. Med.*, **66**, 79-85.
- Usui, A., Kawasumi, Y., Funayama, M. & Saito, H. (2014) Post-mortem lung features in drowning cases on computed tomography. *Jpn. J. Radiol.*, **32**, 414-420.
- van Ginneken, B. (2017) Fifty years of computer analysis in chest imaging: rule-based, machine learning, deep learning. *Radiol. Phys. Technol.*, **10**, 23-32.
- Van Hoyweghen, A.J., Jacobs, W., Op de Beeck, B. & Parizel, P.M. (2015) Can post-mortem CT reliably distinguish between drowning and non-drowning asphyxiation? *Int. J. Legal Med.*, **129**, 159-164.
-