applied sciences

MDPI

Article

Deep Learning-Based Detection of Intracranial Hemorrhages in Postmortem Computed Tomography: Comparative Study of 15 Transfer-Learned Models

Rentaro Matsumoto [1], Hidetoshi Matsuo [2,*], Marie Sugimoto [3], Takaaki Matsunaga [2], Mizuho Nishio [2], Atsushi K. Kono [4], Gentaro Yamasaki [3], Motonori Takahashi [3], Takeshi Kondo [3], Yasuhiro Ueno [3], Ryuichi Katada [3] and Takamichi Murakami [2]

[1] Shinko Hospital, Kobe 651-0072, Japan
[2] Department of Radiology, Graduate School of Medicine, Kobe University, Kobe 650-0017, Japan
[3] Department of Legal Medicine, Graduate School of Medicine, Kobe University, Kobe 650-0017, Japan
[4] Department of Radiology, Faculty of Medicine, Kindai University, Osaka 589-8511, Japan
* Correspondence: hidetoshi.matsuo@bear.kobe-u.ac.jp

Featured Application

This study demonstrates that a transfer-learned intracranial hemorrhage detector trained on non-postmortem data can accurately detect hemorrhages using postmortem head computed tomography. In forensic workflows, the model can be used to triage cases by prioritizing probable hemorrhages and reducing unnecessary cranial openings, potentially lowering the workload and turnaround time of autopsy teams.

Abstract

With the increasing use of postmortem imaging, deep learning (DL)-based automated analysis may assist in the detection of intracranial hemorrhages. However, limited postmortem data complicate model training. This study aims to assess the accuracy of DL models in detecting intracranial hemorrhages in postmortem head computed tomography (CT) scans using transfer learning. A total of 75,000 labeled head CT images from the Radiological Society of North America Intracranial Hemorrhage Detection Challenge serve as the training data for the 15 DL models. Each model is fine-tuned via transfer learning. A total of 134 postmortem cases with hemorrhage status confirmed by autopsy serve as the external test set. Model performance is evaluated using the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, training time, inference time, and number of parameters. Spearman's rank correlation coefficients are calculated for these metrics. DenseNet201 achieves the highest AUC (0.907), with the AUCs of the 15 models ranging from 0.862 to 0.907. A longer inference time moderately correlates with higher AUC (Spearman's $\rho = 0.586$, $p = 0.022$), whereas the number of parameters is not positively correlated with performance ($\rho = -0.472$, $p = 0.076$). The sensitivity and specificity are 0.828 and 0.871, respectively. Transfer learning using a large non-postmortem dataset enables accurate intracranial hemorrhage detection using postmortem CT, potentially reducing the autopsy workload. The results demonstrate that models with fewer parameters often perform comparably to more complex models, emphasizing the need to balance accuracy with computational efficiency.

Keywords: deep learning; postmortem imaging; intracranial hemorrhage; transfer learning; forensic radiology

check for updates

Academic Editors: Mehmed M. Kantardzic, Mehmet Akif Gulum and Hanqing Hu

Received: 5 September 2025
Revised: 24 September 2025
Accepted: 27 September 2025
Published: 28 September 2025

Citation: Matsumoto, R.; Matsuo, H.; Sugimoto, M.; Matsunaga, T.; Nishio, M.; Kono, A.K.; Yamasaki, G.; Takahashi, M.; Kondo, T.; Ueno, Y.; et al. Deep Learning-Based Detection of Intracranial Hemorrhages in Postmortem Computed Tomography: Comparative Study of 15 Transfer-Learned Models. *Appl. Sci.* **2025**, *15*, 10513. https://doi.org/10.3390/app151910513

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

In recent years, postmortem imaging techniques, such as computed tomography (CT) and magnetic resonance imaging (MRI), have been increasingly used to investigate causes of death [1]. These non-invasive methods play an essential role in cases involving suspicious deaths, suspected abuse, or potential medical malpractice. The primary advantages of postmortem imaging include the ability to examine the body without dissection and the relative ease of obtaining consent from the family of the deceased compared to traditional autopsy. Additionally, using imaging as a preliminary screening tool can help reduce unnecessary autopsies [2]. However, limitations remain, such as lower diagnostic accuracy than autopsy and challenges in diagnosing certain conditions, such as pulmonary embolism [2]. Early identification of an intracranial hemorrhage directly informs the decision to perform autopsy and the prioritization of autopsy planning; therefore, an objective and reproducible automated assessment of postmortem CT is required.

Artificial intelligence (AI) technologies, particularly deep learning (DL), have been widely adopted in various fields, from object recognition to conversational AI, and are becoming indispensable in daily life. Research on the application of image processing and DL to the interpretation of medical images, including CT and MRI, has become increasingly active, with progress being made in the field of medical imaging [3–5]. DL-based support for image diagnosis has the potential to improve the efficiency of interpretation tasks of radiologists, enhance diagnostic accuracy, and reduce the risk of oversight [6–8]. In particular, postmortem imaging is often conducted by non-specialists in radiology; thus, applying DL to postmortem imaging could improve diagnostic performance and reduce both labor and costs.

To build a reliable DL model, it is essential to collect a substantial amount of high-quality labeled data [9]. However, the number of postmortem imaging cases is significantly lower than that of standard CT and MRI scans, leading to insufficient data for training DL models. Transfer learning, which leverages knowledge from models pretrained on large datasets, is a promising solution to the data-scarcity problem [10–12]. Although it has been highly successful in clinical image interpretation, transfer learning has not been systematically explored for use in postmortem imaging. One reason is that postmortem changes alter image appearance; therefore, postmortem CT does not necessarily share the same domain as standard clinical imaging. Any limitations from domain mismatch must therefore be weighed against substantial gains in the available training data. To address this issue, this study aims to overcome the data scarcity problem by using publicly available datasets of non-postmortem medical images for training.

Different DL models exhibit different structures and parameters. For practical implementation, it is crucial to optimize predictive accuracy and reduce costs. Therefore, this study evaluates the efficiency of different models by comparing their accuracy, training time, inference time, and number of parameters.

This study aims to construct a model that can accurately detect intracranial hemorrhages in postmortem imaging, where data are limited, by employing transfer learning. By comparing diagnostic accuracy and computational costs, this study aims to provide insights into the effectiveness of DL in postmortem imaging.

Postmortem CT has long been recognized as a useful adjunct or potential alternative to conventional autopsy. Alongside advances in medical imaging AI, automated intracranial hemorrhage detection on clinical CT has been extensively investigated [13–15]. However, owing to the postmortem CT-specific patterns of findings and acquisition conditions, the generalizability of models trained solely on in vivo data has not been adequately verified. Under this practical constraint of in vivo→postmortem CT transfer, this study provides complementary evidence through a systematic external evaluation against the forensic autopsy reference standard.

2. Materials and Methods

This study conformed to the principles of the Declaration of Helsinki and the Ethical Guidelines for Medical and Health Research Involving Human Subjects in Japan. This study was approved by the institutional review board and conducted in accordance with the guidelines of the committee (approval number: B200131). The requirement for informed consent was waived due to the retrospective design of the study.

2.1. Postmortem Imaging Dataset

From cases autopsied at a university of legal medicine between 1 January 2011, and 31 December 2020, 224 cases were selected, in which detailed autopsy records were available and radiographic images provided by investigative or medical institutions were accessible. Among them, 134 patients with postmortem axial head CT images acquired after the estimated time of death were included in the test dataset. Of the 134 patients, 107 (80%) were male and 27 (20%) were female, aged 0–97 years with an average age of 58 years. The age distribution was as follows (years): 0–9, n = 6; 10–19, n = 6; 20–29, n = 11; 30–39, n = 9; 40–49, n = 13; 50–59, n = 15; 60–69, n = 22; 70–79, n = 27; 80–89, n = 19; and 90–97, n = 4. Head hemorrhage was observed during conventional autopsy in 64 patients (48%).

2.2. Non-Postmortem Imaging Dataset

To train the DL model for intracranial hemorrhage detection, the Intracranial Hemorrhage Detection Challenge Dataset (RSNA2019 Dataset) [16], published by the Radiological Society of North America (RSNA), was used as the training dataset. This dataset includes over 75,000 head CT axial slice images, each labeled with six types of hemorrhages: five specific types (epidural, intraparenchymal, intraventricular, subarachnoid, and subdural) and a label indicating the presence of any hemorrhage (any). The dataset was divided randomly into training (80%) and validation (20%) groups. Splitting was performed on a patient basis to avoid subject leakage across the subsets. A validation dataset was used for early discontinuation. Multiple DL models available in Pytorch image models (timm) [17] were fine-tuned. Slice-level targets comprised six binary labels (epidural, intraparenchymal, intraventricular, subarachnoid, subdural, and "any"). The models were trained in a multi-label setting to predict all six labels simultaneously.

2.3. Software and Hardware Environment

All experiments were conducted using Python 3.10. Model implementation, training, and inference used PyTorch 2.2.0 and the timm v0.5.4 library. Image preprocessing and DICOM handling were performed using numpy 1.26.4 and pydicom 2.4.4. The experiments were conducted on a workstation with an Intel Core i7-9800X (3.80 GHz) CPU (Intel Corporation, Santa Clara, CA, USA), 64 GB of RAM, and an NVIDIA TITAN RTX GPU (NVIDIA Corporation, Santa Clara, CA, USA).

2.4. DL Models

Fifteen DL models were selected from the timm library. Table 1 lists the models, their release dates, and number of parameters. A classification head was added to each model to output six per-label logits (epidural, intraparenchymal, intraventricular, subarachnoid, subdural, and "any"), enabling probability estimates for the presence of intracranial hemorrhage. Each backbone was adapted with a six-unit classification head (output dimension = 6) to produce per-label probabilities for the five subtypes and "any". During training, logits were optimized using binary cross-entropy with logit loss; at inference, a sigmoid was applied to obtain probabilities. Figure 1 illustrates a representative DL model.

Table 1. Comparison of Parameter Counts and Release Dates.

Model	Parameter	Date
densenet201 [18]	1.81×10^7	25 August 2016
dm_nfnet_f0 [19]	6.84×10^7	11 February 2021
efficientnet_b2 [20]	7.71×10^6	28 May 2019
efficientnetv2_rw_m [21]	5.11×10^7	01 April 2021
efficientnetv2_rw_s [21]	2.22×10^7	01 April 2021
mobilenetv3_rw [22]	4.21×10^6	06 May 2019
resnetv2_101×1_bitm [23]	4.25×10^7	16 March 2016
resnetv2_101×1_bitm_in21k [23]	4.25×10^7	16 March 2016
resnext101_32×8d [24]	8.68×10^7	16 November 2016
resnext101_64×4d [24]	8.14×10^7	16 November 2016
swin_large_patch4_window7_224 [25]	6.23×10^6	25 May 2021
vgg16 [26]	1.34×10^8	04 September 2014
vgg16_bn [26]	1.34×10^8	04 September 2014
vit_base_patch32_224 [27]	8.74×10^7	22 October 2020
vit_large_patch16_224 [27]	3.03×10^8	20 October 2020

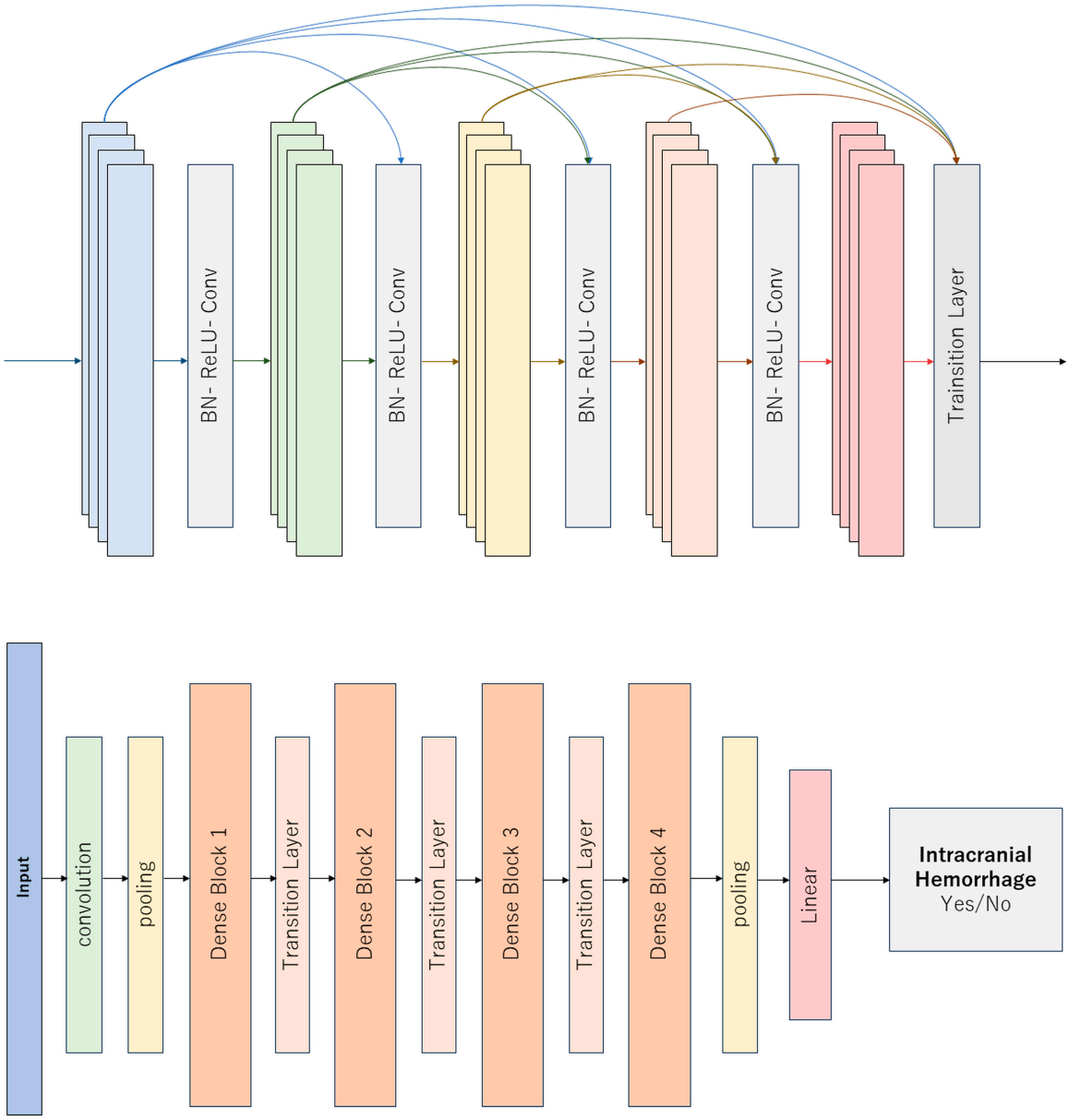


Figure 1. Overview of DenseNet 201.

2.5. Image Processing Steps for Each Case

Head CT images from the Digital Imaging and Communications in Medicine files were first loaded for each case. The original images were then adjusted to three standard head-CT window settings: brain (WL/WW = 40/80 HU), subdural (80/200 HU), and bone (600/2800 HU) [16], resized to 224 × 224; and converted into arrays of size 3 × 224 × 224. All models were trained end-to-end with a binary cross-entropy with logit loss (BCEWithLogitsLoss) averaged across the six outputs, using the Adam optimizer (learning rate 2×10^{-5}), batch size 32, and 5 epochs early stopping based on validation loss; identical hyperparameters were used for all backbones. During training, random augmentation (N = 2, magnitude = 9) was applied to the 3-channel PIL images before tensor conversion [28], and no augmentation was used for validation or testing.

2.6. Inference on Postmortem Images

A test dataset consisting of postmortem head CT cases was used to evaluate the fine-tuned DL models from the timm library. The reference standard was the presence or absence of forensic intracranial hemorrhage, as determined by autopsy. For each axial slice, the models produced six sigmoid outputs (five hemorrhage subtypes: epidural, intraparenchymal, intraventricular, subarachnoid, and subdural; and an "any" output). For patient-level evaluation, the "any-hemorrhage" output aggregated to a case-level hemorrhage score was used by taking the maximum probability across all slices in the case (max pooling). This case-level score was then compared with the autopsy-based labels for evaluation.

2.7. Evaluation Methods

Following training, the validation loss was calculated and the total training time was recorded. The inference time of the test set was also measured. Subsequently, based on the hemorrhage score generated by the DL model and the presence or absence of hemorrhage as determined by autopsy, a receiver operating characteristic (ROC) curve was created for patient-based evaluation, and the area under the ROC curve (ROC AUC) was computed. Furthermore, the Youden index was used to determine the threshold value, and the sensitivity and specificity at that threshold were calculated. ROC curves, the operating threshold, and sensitivity/specificity were computed using the case-level "any-hemorrhage" score.

2.8. Statistical Analysis

Spearman's rank correlation coefficients were calculated to investigate the relationships between the four metrics (ROC AUC, training time, inference time, and number of parameters). Scatter plots were created for each pairwise combination of the metrics.

2.9. Reader Study by a Radiology Resident

To provide a human benchmark, a fourth-year radiology resident independently reviewed the same postmortem head CT test set used for the model evaluation. The cases were anonymized and presented in random order. The reader was blinded to the autopsy findings, clinical information, and model outputs. For each case, the reader assigned a five-point likelihood score for the presence of any intracranial hemorrhage (1 = definitely absent, 2 = probably absent, 3 = indeterminate, 4 = probably present, and 5 = definitely present). Using these ordinal scores and the autopsy-based reference standard, a patient-level ROC curve was constructed, and the ROC AUC was computed. For comparison with the DL models, an optimal cut-off was determined using the Youden index, and the sensitivity and specificity at that threshold were calculated.

3. Results

Table 2 lists the training and inference times, ROC AUC, sensitivity, and specificity determined using the Youden index for each model. The training time ranged from approximately 4×10^4 to 1.75×10^5 s, while the inference time ranged from approximately 100 to 200 s. Overall, the ROC AUC values were in the high 80–90% range, with the highest value being 90.7% for DenseNet201. The confusion matrix for the DenseNet201-based model is presented in Table 3, and its ROC curve is shown in Figure 2.

Table 2. Performance Metrics Comparison.

Model	ROC AUC	ROC AUC 95% CI	Training Time (seconds)	Inference Time (seconds)	Parameter	Sensitivity	Specificity	Accuracy	F1 Score
densenet201	0.907	0.854–0.960	5.43×10^4	193	1.81×10^7	0.828	0.871	0.850	0.841
dm_nfnet_f0	0.883	0.824–0.942	4.90×10^4	139	6.84×10^7	0.828	0.814	0.821	0.815
efficientnet_b2	0.881	0.821–0.941	4.13×10^4	120	7.71×10^6	0.859	0.814	0.835	0.832
efficientnetv2_rw_m	0.883	0.824–0.942	6.80×10^4	194	5.11×10^7	0.703	0.957	0.836	0.803
efficientnetv2_rw_s	0.873	0.811–0.935	4.48×10^4	151	2.22×10^7	0.672	0.986	0.836	0.797
mobilenetv3_rw	0.870	0.807–0.933	5.64×10^4	95	4.21×10^6	0.656	0.957	0.813	0.770
resnetv2_101×1_bitm	0.895	0.839–0.951	4.83×10^4	153	4.25×10^7	0.734	0.957	0.850	0.824
resnetv2_101×1_bitm_in21k	0.884	0.825–0.943	4.82×10^4	154	4.25×10^7	0.781	0.886	0.836	0.820
resnext101_32×8d	0.876	0.815–0.937	8.89×10^4	176	8.68×10^7	0.750	0.886	0.821	0.800
resnext101_64×4d	0.873	0.811–0.935	1.61×10^5	197	8.14×10^7	0.734	0.843	0.791	0.770
swin_large_patch4_window7_224	0.892	0.835–0.949	1.41×10^5	170	6.23×10^6	0.906	0.700	0.798	0.811
vgg16	0.870	0.807–0.933	5.10×10^4	103	1.34×10^8	0.750	0.900	0.828	0.807
vgg16_bn	0.862	0.798–0.926	4.72×10^4	104	1.34×10^8	0.750	0.871	0.813	0.793
vit_base_patch32_224	0.872	0.810–0.934	4.67×10^4	105	8.74×10^7	0.750	0.871	0.813	0.793
vit_base_patch16_224	0.879	0.819–0.939	1.75×10^5	167	3.03×10^8	0.672	0.929	0.806	0.768
Radiology Resident	0.810	0.736–0.884				0.600	1.000	0.809	0.750

Table 3. Confusion matrix of the densenet201 model.

Confusion Matrix		Predicted	
		Positive	**Negative**
Actual	**Positive**	53	11
	Negative	9	61

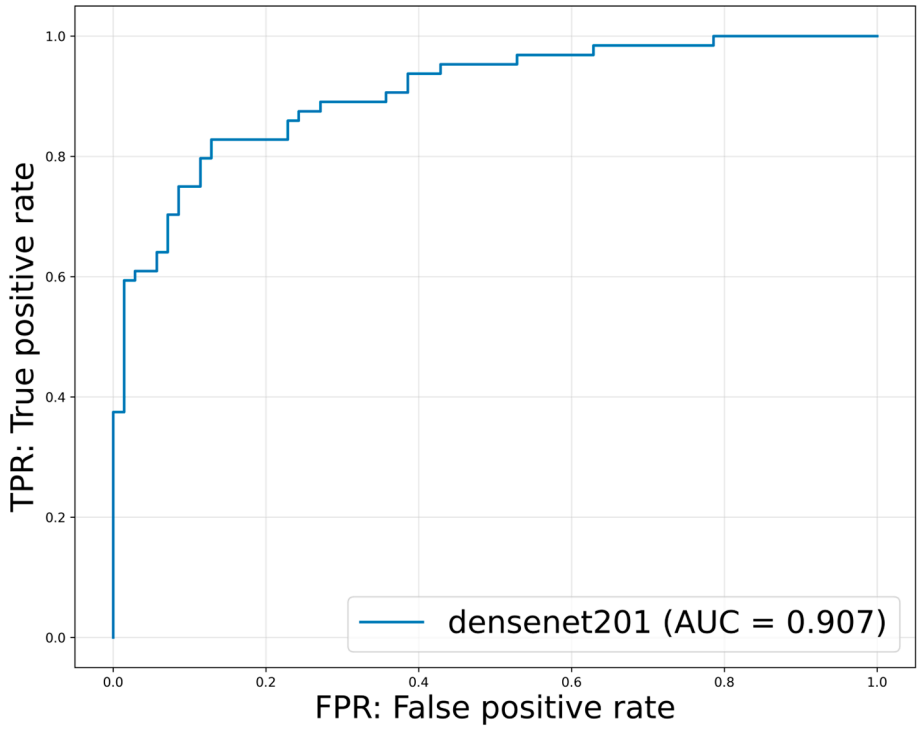


Figure 2. ROC curve for DenseNet201.

Figure 3 shows the relationship between the number of parameters and the ROC AUC. Spearman's rank correlation coefficient between the ROC AUC and the number of parameters was −0.472. Due to the limited number of models (n = 15), p = 0.076, there was no statistically significant correlation. Although models with more parameters are generally considered more complex and perform better, a negative correlation was observed in this study. These results do not account for architectural factors that affect the parameter counts, among other variables, which may partly explain the observed relationship.

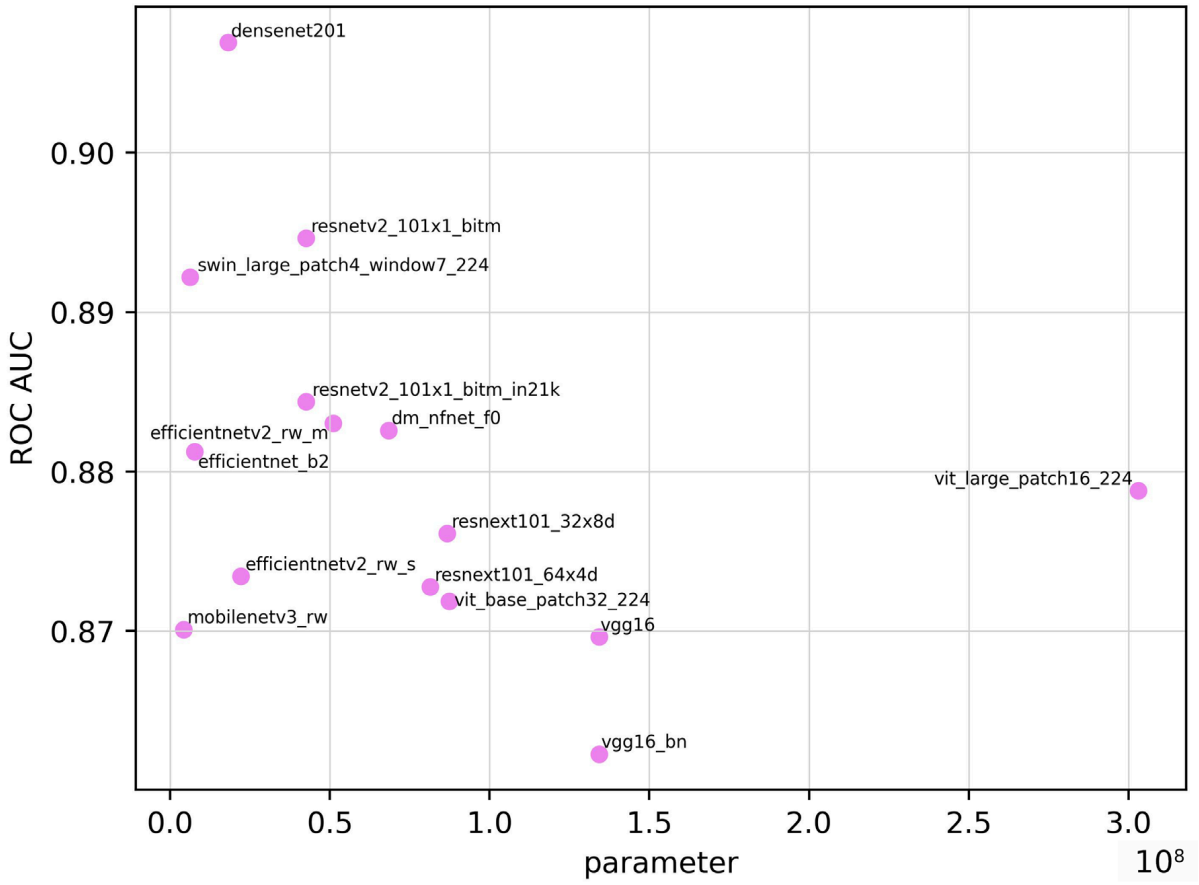


Figure 3. Relationship between ROC AUC and number of parameters.

Figure 4 shows the relationship between the inference time and the ROC AUC. Spearman's rank correlation coefficient was 0.586 (p = 0.022). These results indicate a moderate positive correlation between the ROC AUC and inference time.

Figure 5 shows the relationship between training time and inference time. Spearman's rank correlation coefficient between the training time and inference time was 0.579 (p = 0.024), suggesting a moderate positive correlation. Therefore, the models requiring longer training times tended to exhibit longer inference times. The correlation coefficients for the other pairs (ROC AUC vs. training time, training time vs. number of parameters, and inference time vs. number of parameters) were 0.129, 0.144, and −0.038, respectively. However, none of these relationships were statistically significant. On the same test set, the radiology resident achieved an ROC AUC of 0.810 (95% CI 0.746–0.876), which was lower than the range observed for the DL models (0.862–0.907).

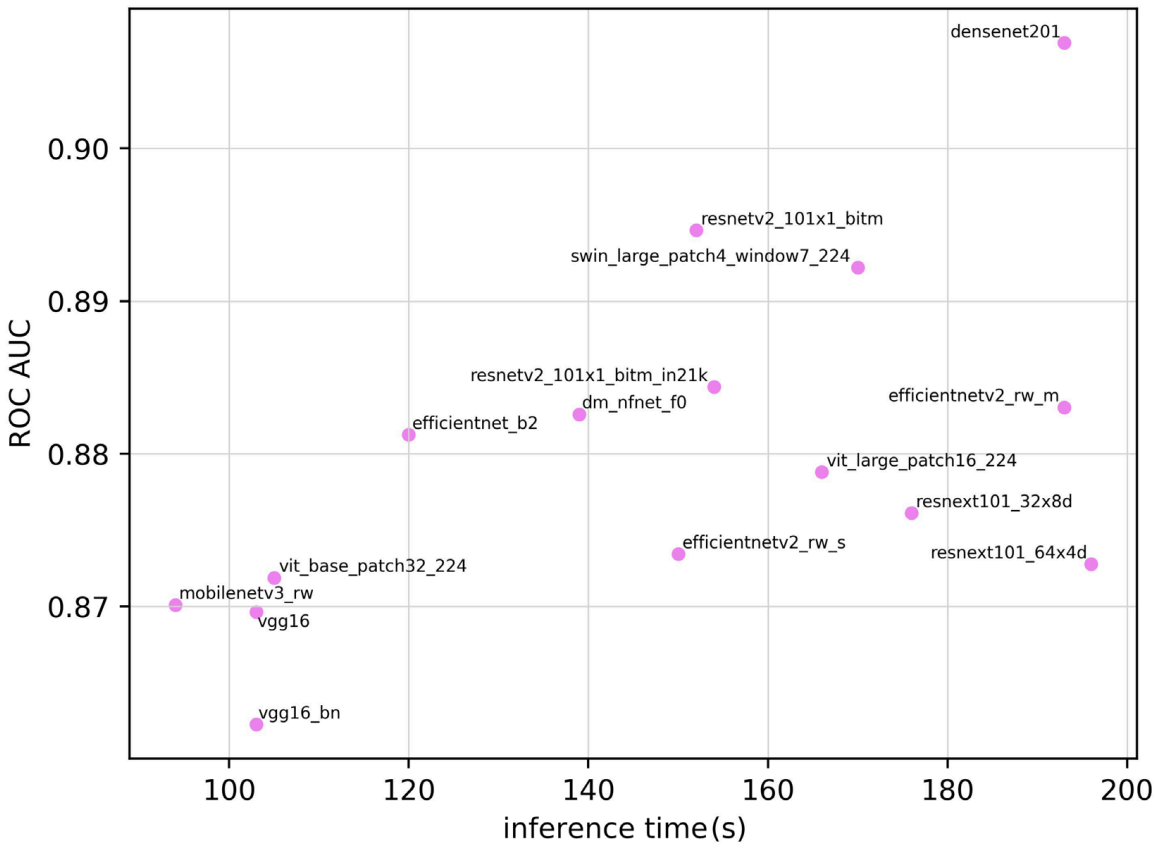


Figure 4. Relationship between the ROC AUC and inference time.

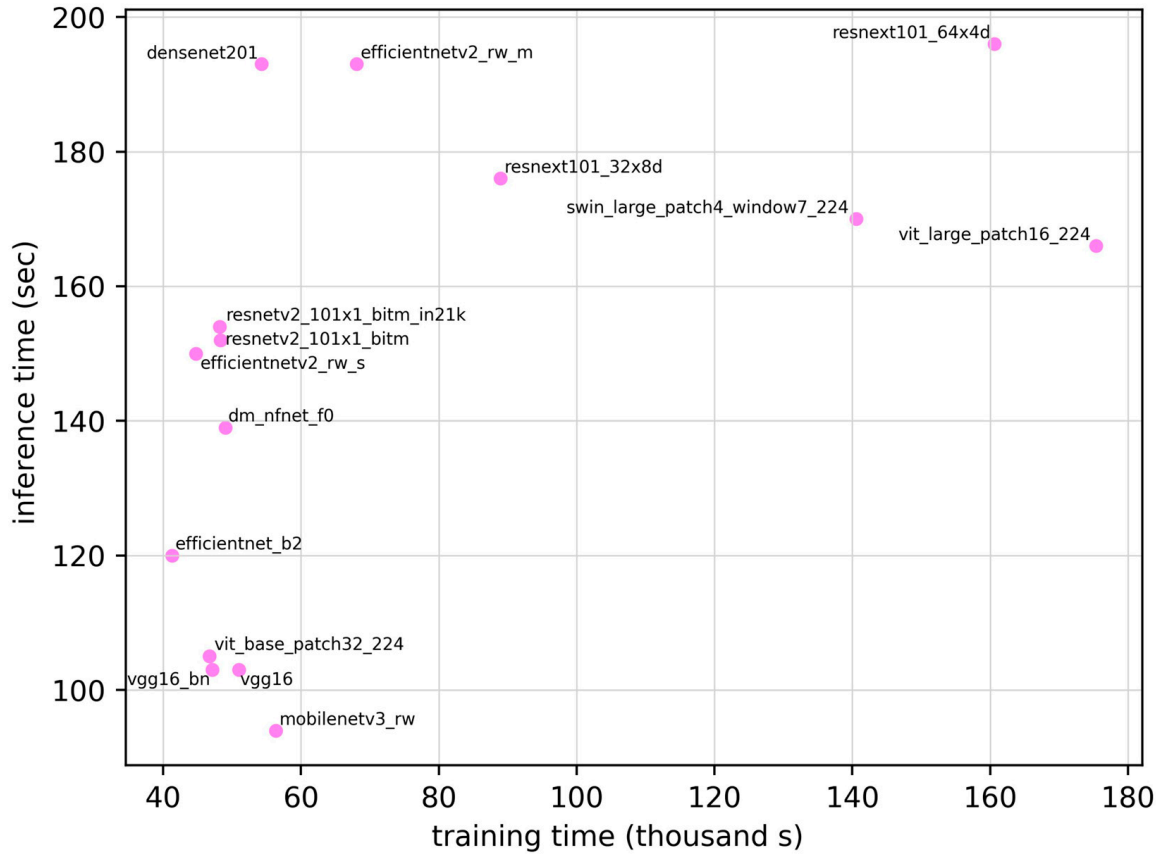


Figure 5. Relationship between training time and inference time.

4. Discussion

In this study, transfer learning was performed on 15 DL models using head CT images from non-postmortem CT examinations as training data. The resulting trained models were evaluated using postmortem head CT images as an external validation dataset. When conducting binary classification (presence or absence of hemorrhage) on a patient basis, the ROC AUCs of the 15 DL models ranged from 0.862 to 0.907. Among the models, DenseNet201 performed the best, attaining an ROC AUC of 90.7%. A moderate negative correlation was observed between the ROC AUC and the number of parameters, whereas a moderate positive correlation was observed between the ROC AUC and the inference time, as well as between the training time and inference time. Notably, when benchmarked against a radiology resident who reviewed the same cases, all DL models achieved higher patient-level ROC AUCs (0.862–0.907) than that of the resident (0.810; 95% CI, 0.736–0.884), indicating trainee-level or better performance on this dataset.

One plausible reason that DenseNet201 achieved the best performance (AUC 0.907) is that its dense connections promote feature reuse and stabilize gradient flow, which may have facilitated capturing high-attenuation regions and edge contrast represented by the three input windows (brain/subdural/bone). Note that identical training settings were applied to all backbones; therefore, differences in the degree of per-model optimization may remain.

Although numerous studies have compared the performance of multiple models using general image datasets, such as ImageNet, most medical imaging studies have only compared a limited number of models. The strength of this study lies in the comparison of 15 models under identical conditions. In addition to the model performance, the inference time, which is often a critical factor in practical applications, was also considered. The results demonstrated a negative correlation between the ROC AUC and the number of parameters and positive correlations between the ROC AUC and inference time and between the training time and inference time. However, practical deployment must also consider additional factors, such as the memory footprint, power consumption, and constraints of the intended application.

In addition, a uniform batch size and learning rate were maintained across all models, which is a major contribution of this study. The correlations observed between the ROC AUC and the number of parameters, as well as between the ROC AUC and inference time, may be useful for model selection. However, it is important to note that the hyperparameters employed here may not be optimal for each model, and the results may change with more fine-tuned adjustments.

The new-generation Transformer-based models used in this study, specifically ViT and Swin, are relatively new. Although they tend to have fewer parameters, their structures are more complex, which often results in higher computational requirements. This complexity may partially explain the negative correlation between the number of parameters and the ROC AUC.

In this study, the postmortem CT test set comprised consecutive cases submitted to a single department of legal medicine without deliberate selection, thus reflecting a typical use case. Furthermore, the best DL model achieved an ROC AUC of 90.7%. Because cranial processing for autopsy is highly labor-intensive, the findings of this study suggest that combining postmortem imaging with DL models could potentially reduce the workload of forensic pathologists. This human benchmark comparison suggests that even under a domain shift to postmortem imaging, the models perform favorably relative to a trainee reader, supporting their potential as decision-support tools in forensic workflows.

The training dataset used in this study consisted of head CT images from living patients; therefore, the model did not consider the characteristic changes in normal findings or hemorrhages commonly observed in postmortem CT images. To further improve accuracy, it is necessary to train the model using actual postmortem imaging data. High-priority next steps include multi-institutional, systematic postmortem CT data curation with prospective external validation, together with additional training that incorporates domain-adaptation strategies (e.g., self-/semi-supervised learning and test-time adaptation). Coupling these with site-specific calibration and threshold optimization may improve both generalization and operational reliability.

This study had several limitations. First, the number of training cases was small and the resulting accuracy may be insufficient for clinical applications. Furthermore, only a single-center external validation was conducted, and further validation at additional institutions is required. Beyond the limitations of a single institution and reader, potential biases include variability in autopsy interpretation, differences in the interval between death and imaging, heterogeneity in scanners and reconstruction protocols, and artifacts related to postmortem changes. Future studies should be conducted to record and stratify confounders. In this study, the training focused exclusively on intracranial hemorrhage, and thus did not include conditions, such as cerebral infarction [29], other intracranial pathologies, or pathologies in other organs [7]. CT scans submitted from multiple hospitals to the Department of Legal Medicine were used and detailed scanner information could not be obtained. The convolutional neural network and Transformer architectures evaluated were selected by the authors as representative examples; therefore, the selection was somewhat subjective and not exhaustive. Finally, the reader comparison involved a single radiology resident; performance relative to board-certified radiologists and inter-reader variability were not assessed.

5. Conclusions

In conclusion, this study developed a model that exhibited a relatively high accuracy in discriminating intracranial hemorrhages on postmortem head CT scans using transfer learning. This study compared 15 DL models on an autopsy-verified postmortem CT test set, and the results can assist in guiding model selection for the application of AI in healthcare. To improve the ROC AUC further, more extensive training data are required, including the integration of datasets from multiple medical institutions and an increased number of postmortem CT cases. In addition, a more comprehensive model is required for practical applications in postmortem imaging. These findings address the domain-mismatch issue raised in the Introduction issue: models trained solely on in vivo CT can be generalized to postmortem CT when adapted by transfer learning and judged against autopsies. This autopsy-anchored systematic external evaluation complements the clinical intracranial hemorrhage literature and supports its use as a triage aid in forensic workflows.

Author Contributions: R.M.: Data curation, Investigation, Software, Visualization, Writing—original draft, and review & editing. H.M.: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing—original draft, and review & editing. M.S.: Conceptualization, Investigation, Resources, and Writing—review & editing. T.M. (Takaaki Matsunaga): Conceptualization, Methodology, and Writing—review & editing. M.N.: Conceptualization, Funding acquisition, Methodology, Project administration, Supervision, and Writing—review & editing. A.K.K.: Conceptualization, Data curation, and Writing—review & editing. G.Y.: Data curation, Resources, Validation, and Writing—review & editing. M.T.: Data curation, Investigation, and Writing—review & editing. T.K.: Data curation, Investigation, and Writing—review & editing. Y.U.: Data curation, Investigation, and Writing—review & editing. R.K.: Data curation, Investigation, and Writing—review & editing. T.M. (Takamichi Murakami):

Conceptualization, Supervision, Methodology, and Writing—review & editing. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by JSPS KAKENHI (Grant Numbers: 23K17229 and 23KK0148).

Institutional Review Board Statement: This study conformed to the Declaration of Helsinki and the Ethical Guidelines for Medical and Health Research Involving Human Subjects in Japan (https://www.mhlw.go.jp/file/06-Seisakujouhou-10600000-Daijinkanboukouseikagakuka/0000080278.pdf accessed on 1 September 2025). This study was approved by the institutional review board and conducted in accordance with the guidelines of the committee (approval number: B200131).

Informed Consent Statement: The requirement for informed consent was waived due to the retrospective design of the study.

Data Availability Statement: Due to privacy concerns, the data cannot be made publicly available. Individuals interested in accessing the data are requested to contact the authors directly. The datasets used in this study have different availability restrictions due to their nature and source: **Non-Postmortem Dataset:** The RSNA Intracranial Hemorrhage Detection Challenge Dataset (RSNA2019) used for model training is publicly available through the Radiological Society of North America at https://www.rsna.org/rsnai/ai-image-challenge/rsna-intracranial-hemorrhage-detection-challenge-2019 (accessed on 1 September 2025). **Postmortem Dataset:** The data supporting the findings of this study are available upon request from the corresponding author, H.M. These data are not publicly available because of privacy concerns.

Acknowledgments: During the preparation of this manuscript, the authors used ChatGPT (o1, OpenAI accessed on 20 April 2024) to translate drafts written in Japanese into English. Following the use of this tool, the authors carefully reviewed and revised the translated content as necessary. The authors are responsible for the manuscript content.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial intelligence
AUC	Area under the curve
CT	Computed tomography
DL	Deep learning
MRI	Magnetic resonance imaging
ROC	Receiver operating characteristic
ROC AUC	Area under the ROC curve
RSNA	Radiological Society of North America

References

1. Bolliger, S.A.; Thali, M.J. Imaging and virtual autopsy: Looking back and forward. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **2015**, *370*, 20140253. [CrossRef]
2. Roberts, I.S.D.; Benamore, R.E.; Benbow, E.W.; Lee, S.H.; Harris, J.N.; Jackson, A.; Mallett, S.; Patankar, T.; Peebles, C.; Roobottom, C.; et al. Post-mortem imaging as an alternative to autopsy in the diagnosis of adult deaths: A validation study. *Lancet* **2012**, *379*, 136–142. [CrossRef]
3. Zhou, L.Q.; Wang, J.Y.; Yu, S.Y.; Wu, G.G.; Wei, Q.; Deng, Y.B.; Wu, X.L.; Cui, X.W.; Dietrich, C.F. Artificial intelligence in medical imaging of the liver. *World J. Gastroenterol.* **2019**, *25*, 672–682. [CrossRef] [PubMed]
4. Dey, D.; Slomka, P.J.; Leeson, P.; Comaniciu, D.; Shrestha, S.; Sengupta, P.P.; Marwick, T.H. Artificial intelligence in cardiovascular imaging: JACC state-of-the-art review. *J. Am. Coll. Cardiol.* **2019**, *73*, 1317–1335. [CrossRef] [PubMed]
5. Cui, Y.; Zhu, J.; Duan, Z.; Liao, Z.; Wang, S.; Liu, W. Artificial intelligence in spinal imaging: Current status and future directions. *Int. J. Environ. Res. Public Health* **2022**, *19*, 11708. [CrossRef]

6. Matsuo, H.; Kitajima, K.; Kono, A.K.; Kuribayashi, K.; Kijima, T.; Hashimoto, M.; Hasegawa, S.; Yamakado, K.; Murakami, T. Prognosis prediction of patients with malignant pleural mesothelioma using conditional variational autoencoder on 3D PET images and clinical data. *Med. Phys.* **2023**, *50*, 7548–7557. [CrossRef]
7. Matsuo, H.; Nishio, M.; Kanda, T.; Kojita, Y.; Kono, A.K.; Hori, M.; Teshima, M.; Otsuki, N.; Nibu, K.I.; Murakami, T. Diagnostic accuracy of deep-learning with anomaly detection for a small amount of imbalanced data: Discriminating malignant parotid tumors in MRI. *Sci. Rep.* **2020**, *10*, 19388. [CrossRef]
8. Nishio, M.; Noguchi, S.; Matsuo, H.; Murakami, T. Automatic classification between COVID-19 pneumonia, non-COVID-19 pneumonia, and the healthy on chest X-ray image: Combination of data augmentation methods. *Sci. Rep.* **2020**, *10*, 17532. [CrossRef]
9. Kumari, R.; Nikki, S.; Beg, R.; Ranjan, S.; Gope, S.K.; Mallick, R.R.; Dutta, A. A review of image detection, recognition and classification with the help of machine learning and artificial intelligence. *SSRN J.* **2020**. [CrossRef]
10. Kim, H.E.; Cosa-Linan, A.; Santhanam, N.; Jannesari, M.; Maros, M.E.; Ganslandt, T. Transfer learning for medical image classification: A literature review. *BMC Med. Imaging* **2022**, *22*, 69. [CrossRef]
11. Alzubaidi, L.; Al-Amidie, M.; Al-Asadi, A.; Humaidi, A.J.; Al-Shamma, O.; Fadhel, M.A.; Zhang, J.; Santamaría, J.; Duan, Y. Novel Transfer Learning Approach for Medical Imaging with Limited Labeled Data. *Cancers* **2021**, *13*, 1590. [CrossRef]
12. Li, M.; Jiang, Y.; Zhang, Y.; Zhu, H. Medical image analysis using deep learning algorithms. *Front. Public Health* **2023**, *11*, 1273253. [CrossRef]
13. Thali, M.J.; Yen, K.; Schweitzer, W.; Vock, P.; Boesch, C.; Ozdoba, C.; Schroth, G.; Ith, M.; Sonnenschein, M.; Doernhoefer, T.; et al. Virtopsy, a new imaging horizon in forensic pathology: Virtual autopsy by postmortem multislice computed tomography (MSCT) and magnetic resonance imaging (MRI)—A feasibility study. *J. Forensic Sci.* **2003**, *48*, 386–403. [CrossRef]
14. Arbabshirani, M.R.; Fornwalt, B.K.; Mongelluzzo, G.J.; Suever, J.D.; Geise, B.D.; Patel, A.A.; Moore, G.J. Advanced machine learning in action: Identification of intracranial hemorrhage on computed tomography scans of the head with clinical workflow integration. *npj Digit. Med.* **2018**, *1*, 9. [CrossRef] [PubMed]
15. Wang, X.; Shen, T.; Yang, S.; Lan, J.; Xu, Y.; Wang, M.; Zhang, J.; Han, X. A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans. *NeuroImage Clin.* **2021**, *32*, 102785. [CrossRef] [PubMed]
16. Flanders, A.E.; Prevedello, L.M.; Shih, G.; Halabi, S.S.; Kalpathy-Cramer, J.; Ball, R.; Mongan, J.T.; Stein, A.; Kitamura, F.C.; Lungren, M.P.; et al. RSNA-ASNR 2019 Brain Hemorrhage CT Annotators, Construction of a machine learning dataset through collaboration: The RSNA 2019 brain CT hemorrhage challenge. *Radiol. Artif. Intell.* **2020**, *2*, e190211. [CrossRef]
17. PyTorch-Image-Models: The Largest Collection of PyTorch Image Encoders/Backbones. Including Train, Eval, Inference, Export Scripts, and Pretrained Weights—ResNet, ResNeXT, EfficientNet, NFNet, Vision Transformer (ViT), MobileNetV4, MobileNet-V3 & V2, RegNet, DPN, CSPNet, Swin Transformer, MaxViT, CoAtNet, ConvNeXt, and More. GitHub. Available online: https://github.com/huggingface/pytorch-image-models (accessed on 26 January 2025).
18. Huang, G.; Liu, Z.; van der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. *arXiv* **2016**, arXiv:1608.06993. Available online: http://arxiv.org/abs/1608.06993 (accessed on 26 January 2025).
19. Brock, A.; De, S.; Smith, S.L.; Simonyan, K. High-performance large-scale image recognition without normalization. *arXiv* **2021**, arXiv:2102.06171. Available online: http://arxiv.org/abs/2102.06171 (accessed on 26 January 2025).
20. Tan, M.; Le, Q.V. EfficientNet: Rethinking model scaling for convolutional Neural Networks. *arXiv* **2019**, arXiv:1905.11946. Available online: http://arxiv.org/abs/1905.11946 (accessed on 26 January 2025).
21. Tan, M.; Le, Q.V. EfficientNetV2: Smaller models and faster training. *arXiv* **2021**, arXiv:2104.00298. Available online: http://arxiv.org/abs/2104.00298 (accessed on 26 January 2025).
22. Howard, A.; Sandler, M.; Chu, G.; Chen, L.C.; Chen, B.; Tan, M.; Wang, W.; Zhu, Y.; Pang, R.; Vasudevan, V.; et al. Searching for MobileNetV3. *arXiv* **2019**, arXiv:1905.02244. Available online: http://arxiv.org/abs/1905.02244 (accessed on 26 January 2025).
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Identity mappings in deep residual networks. *arXiv* **2016**, arXiv:1603.05027. Available online: http://arxiv.org/abs/1603.05027 (accessed on 26 January 2025).
24. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. *arXiv* **2016**, arXiv:1611.05431. Available online: http://arxiv.org/abs/1611.05431 (accessed on 26 January 2025).
25. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin Transformer: Hierarchical vision Transformer using shifted windows. *arXiv* **2021**, arXiv:2103.14030. Available online: http://arxiv.org/abs/2103.14030 (accessed on 26 January 2025).
26. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556. Available online: http://arxiv.org/abs/1409.1556 (accessed on 26 January 2025).
27. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterhiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. *arXiv* **2020**, arXiv:2010.11929. Available online: http://arxiv.org/abs/2010.11929 (accessed on 26 January 2025).

28. Cubuk, E.D.; Zoph, B.; Shlens, J.; Le, Q.V. RandAugment: Practical automated data augmentation with a reduced search space. *arXiv* **2019**, arXiv:1909.13719. Available online: http://arxiv.org/abs/1909.13719 (accessed on 26 January 2025). [CrossRef]
29. Nishio, M.; Koyasu, S.; Noguchi, S.; Kiguchi, T.; Nakatsu, K.; Akasaka, T.; Yamada, H.; Itoh, K. Automatic detection of acute ischemic stroke using non-contrast computed tomography and two-stage deep learning model. *Comput. Methods Programs Biomed.* **2020**, *196*, 105711. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.