Multidisciplinary Journal of Research in Engineering and Technology

ISSN: 2348-6953

Volume 13 Issue 2s (2026) | Pages: 211-217

SkullFractureNet: A VGG16-Based Transfer Learning Pipeline with Two-Phase Fine-Tuning and Test-Time Augmentation for Automated Skull Fracture Detection in Head CT

Vedant Chaudhari[1], Vivek Bavaskar[2], Tawheed Ahmed Ansari[3], Jayashree Khanapuri[4], Prajakti Akerkar[5]

[1,2,3,4,5]KJ Somaiya Institute of Technology, Mumbai, India

[1]*vedant.chaudhari@somaiya.edu,* [2]*vivek.bavaskar@somaiya.edu,* [3]*tawheedahmed.a@somaiya.edu,*[4] *jayashreek@somaiya.edu,* [5]*prajakti.d@somaiya.edu*

Peer Review Information

Type: *Article*
Received: *27 March 2026*
Revised: *12 April 2026*
Accepted: *26 May 2026*
Published: *16 June 2026*

Abstract

This work seeks to tackle the problem of significant misdiagnosis rates in CT imaging for identifying skull fractures – as high as 14.8%, even amongst expert radiologists – thus highlighting the need for reliable automatic skull fracture detection approaches. In this study, a new SkullFractureNet approach is proposed where VGG16 is used as the convolutional neural network (CNN) backbone, pretrained on the ImageNet database, instead of the previous efficient net based approach (EfficientNetB0). The proposed approach utilizes specific architectural modifications aimed at addressing overfitting issues arising from the small number of data samples in a real-world setting: GlobalMaxPooling2D is used in place of the GlobalAveragePooling2D layer to better locate local fracture areas; the classification head size is reduced from 256 to 64; dropout rate is increased to 0.6 and 0.4, and L2 regularization is applied throughout the dense layers. In terms of the training process, the same two-phase process remains – the first phase trains the classifier head while freezing all of the VGG16 weights (learning rate = 5×10^{-5}; 40 epochs); the second phase unfreezes only the final convolutional block (block5_conv3) at a learning rate of 5×10^{-6}. Aggressive slice-level augmentation strategies (flip, rotate, zoom, brightness, contrast, noise and cutout) in combination with a test-time augmentation (TTA) approach where eight forward passes are averaged together are used.The experiments were performed on a balanced dataset of 42 samples (21 fractures vs. 21 normals) using a patient-based split approach. The results of testing on the held-out portion of the test set (6 samples) yield the following performance metrics (maximum probability aggregation): accuracy 0.667, specificity 1.000, precision 1.000, AUC-ROC 0.667, and AUC-PRC 0.756, and no false positives. It can be noted that the training process shows that the measures adopted to prevent overfitting were effective.

Keywords: Skull Fracture Detection; VGG16; Transfer Learning; Test-Time Augmentation; GlobalMaxPooling; CT Imaging; CQ500; Two-Phase Fine-Tuning; Overfitting Regularization

How to Cite This Article

Chaudhari, V., Bavaskar, V., Ansari, T. A., Khanapuri, J., & Akerkar, P. (2026). SkullFractureNet: A VGG16-based transfer learning pipeline with two-phase fine-tuning and test-time augmentation for automated skull fracture detection in head CT. *Multidisciplinary Journal of Research in Engineering and Technology*, 13(2s), 211–217.

Introduction

TBI occurs in approximately 3.17 million people each year, with skull fractures found in about 23.4% of head trauma presentations [1]. The clinical importance of skull fractures includes increased susceptibility to the development of an epidural or subdural hematoma by up to 10-30%, as well as serving as pathways for CSF leakage [2]. In a systematic review involving 26,774 participants, the prevalence of TBI was 51.04% among those with coexistent maxillofacial fractures [23], signifying their vulnerability. Although advances in medical imaging equipment continue to be made, there are still significant issues in diagnosing skull fractures radiologically. There is a persistent misdiagnosis rate of 14.8% even amongst experienced clinicians [3], and the inter-reader agreement in marking the fractures is low, with intraclass correlation coefficients (ICCs) under 0.4 [3]. A postmortem CT study analysis involving 1,538 cases yielded sensitivities of 0.89 for vault fractures and 0.87 for basilar fractures, compared to 0.96 for specificity [3], proving that there is still uncertainty in radiological diagnoses of skull fractures despite standardization in image acquisition procedures.

Automated fracture detectors have become increasingly effective in bridging this gap. Classic histogram thresholding methods were capable of detection rates of 99% [5], morphological black-hat filters were able to resolve hairline fractures as thin as 0.35mm [10], and recent deep neural networks have surpassed their predecessors in terms of discrimination capacity. Hybrid models of DenseNet161–UNet have been shown to detect fractures with an accuracy of 98.37% (AUC 0.976) [6], Feature Pyramid Networks with criss-cross attention yielded an 82% classification accuracy and 91.5% recall [7], while weakly supervised VGG-19 with Grad-CAM localization achieved 91.36% accuracy [11].

The proposed method in this paper is an updated version of the SkullFractureNet model, which is based on VGG16 transfer learning [13]. Key differences from the previous architecture, which uses EfficientNetB0 as the base network, include: (1) the substitution of GlobalMaxPooling2D for enhanced localization of spatially compact fractures; (2) the use of an aggressive augmentation process at the DICOM slice level; (3) a two-stage fine-tuning regimen where adjustments can only be done to the deepest convolutional layer during the second phase; and (4) the use of test-time augmentation to minimize prediction variance.

Related Work

Conventional Image Processing Methods

Previous computation-based techniques utilized the unique Hounsfield Unit (HU) values for cortical skull bone (232–255 HU) through multi-level Otsu's thresholding to attain an identification accuracy of roughly 99% within processing durations below two seconds [5]. In terms of comparative edge detection, Sobel filters were found superior (≈99%) compared to Canny (≈98%), Roberts (≈93%), and Prewitt (≈92%) filters in defining the borders of fracture sites [9], [16]. The black-hat morphological transformation was able to detect fractures of even 0.35 mm size [10]; however, a high false-positive error rate was attributed to vascular grooves and sutures. An integration of Fuzzy C-Means clustering with two-level Otsu thresholding resulted in intracranial hemorrhage localization accuracy of 98% and true positive ratio of 96.3% in 213 CT images with abnormal findings [24].

Deep Learning Models

An algorithm developed based on a DenseNet161–UNet combination delivered an impressive 98.37% of accuracy (AUC 0.976, precision 0.87, recall 0.88, F1 0.875), superior to the other architectures such as ResNet50–XGBoost and DeepLab frameworks [6]. The employment of a feature pyramid network architecture with criss-cross attention and augmentation increased accuracy to 82.0% while delivering 91.5% recall rate on 19,222 training images [7]. Evaluation of YOLOv8 architecture on 880 FracAtlas triplets delivered mAP50 = 0.49 and displayed poorer performance on the rare fracture classes [8], thus highlighting that deep neural networks are highly sensitive to the distribution of the training set. Weakly supervised learning with Grad-CAM produced 91.36% of accuracy for full annotation and 89.37% under 10% annotation on 19,079 CT scans [11].

Transfer Learning and Explainability

It is common practice to use transfer learning with ImageNet pre-trained network models for medical domains with relatively low numbers of annotated samples. EfficientNet models offer an efficient balance between accuracy and efficiency through a compound scaling scheme concerning depth, width, and resolution [12]. Though the use of VGG16 network models requires a larger number of parameters (~138 million), their pretrained weights have been extensively tested and deployed in various applications in the field of medical image analysis. Fine-tuning in phases starting with classification head learning and subsequently unfreezing upper backbone layers prevents catastrophic forgetting [13]. Class-discriminative heat maps produced by Grad-CAM help to localize suspect areas and support interpretation [14].

Dataset And Preprocessing

Dataset

The CQ500 dataset is a publicly available, multicenter dataset containing head computed tomography (CT) scans, which includes 491 CT scans annotated by radiologists' consensus [4]. For this study, a balanced subset was selected by selecting a total of 42 subjects: 21 subjects with fractures and 21 subjects without fractures. This balanced selection process was based on an equal number of subjects from the whole sample size of 62 subjects (including 21 subjects with fractures and 41 subjects without fractures). Only the directories labeled as 'thin' series were selected since they were acquired at the highest resolution in the axial direction.

Patient-Level Data Partitioning

Data partitioning was performed at the patient-level to avoid data leakage, which is emphasized by the low intra-observer agreement level (intraclass correlation coefficient ICC < 0.4) shown in literature regarding skull fracture classification [3], [4]. In addition, all slices from each subject were assigned to only one split during data partitioning. Random stratified sampling was applied before splitting, using a fixed seed SEED = 42 to eliminate potential biases. The details of the data partitioning process are shown in Table I below.

Table 1. *Patient-Level Dataset Split (Balanced Cohort)*

Split	Fracture Patients	Normal Patients	Total Patients
Train	13	13	26
Validation	5	5	10
Test	3	3	6

DICOM Preprocessing

DICOM files were converted to Hounsfield units (HU) using an affine transformation formula: HU = pixel_value * RescaleSlope + RescaleIntercept, where slope and intercept values came from image file metadata. A bone window was used for visualization using the parameters center = 400 HU and width = 1800 HU (range: -500 to +1300 HU). This was done based on existing CT bone window protocols for cortical bone imaging [17]. These windowed images were then linearly scaled to the range [0, 1] and resized to 224 × 224 pixels by means of bilinear interpolation [5]. Finally, each grayscale slice was replicated into three channels resulting in images with RGB color format suitable for ImageNet pre-trained models. Corrupted DICOM files were silently ignored.

Slice-Level Data Augmentation

A data augmentation pipeline was implemented in NumPy and applied to each individual slice of the DICOM files during model training: random horizontal flipping (with p=0.5 probability), rotation by an angle between [-15°, +15°], reflected padding, zoom factor between [0.85, 1.15] (center crop/zero padding), random jitter with $\alpha \in [0.85, 1.15]$ and brightness shift with $\beta \in [-0.08, +0.08]$, Gaussian noise with mean = 0 and $\sigma = 0.02$ (p=0.4), and cutout augmentation with a 32 × 32 pixels region being cut out with p = 0.4. This augmentation was much more aggressive compared to the previous EfficientNetB0 baseline and was intended to solve a severe overfitting problem when using VGG16 on the 26-patient dataset [7].

Proposed Methodology

As backbone network, VGG16 was chosen, pretrained on ImageNet dataset. One can say that the main motivation behind using VGG16 lies in well-researched intermediate features of the architecture showing great transfer capabilities in cases involving texture-heavy medical images[14]. The classification head was adjusted to minimize the quantity of parameters that could be subject to overfitting due to the small sample of train data available. GlobalMaxPooling2D replaced GlobalAveragePooling2D since one can argue that the localization of a skull fracture is rather restricted to the CT image plane. The first layer of this type of pooling keeps activated features without caring about their exact position on the grid while the second type of pooling usually diffuses activations over the whole map. The overall head design looks like the following: GlobalMaxPooling2D → BatchNormalization → Dense(256, ReLU, L2) → Dropout(0.6) → Dense(64, ReLU, L2) → Dropout(0.4) → Dense(1, Sigmoid). It should be noted that the architecture was changed from a 512→128 one to a more modest 256→64 setup, decreasing the number of parameters at risk of overfitting.

Table 2. *Model Architecture and Hyperparameter Configuration*

Hyperparameter	Value
Backbone	VGG16 (ImageNet)
Input Size	224 × 224 × 3
Pooling	GlobalMaxPooling2D
Head Architecture	Dense(256) → Dropout(0.6) → Dense(64) → Dropout(0.4)
L2 Regularization (λ)	2×10^{-4}

Batch Size	16
Phase 1 LR	5×10^{-5}
Phase 2 LR	5×10^{-6}
Phase 1 Epochs (max)	40
Phase 2 Epochs (max)	40 (early stopped)
Fine-Tuned Layer	block5_conv3 only
LossFunction	(label smoothing = 0.05)
Optimizer	Adam(clipnorm=1.0)
HU Window Center / Width	400 HU / 1800 HU
TTA Passes	8

Two-Phase Training Strategy

The overall training scheme is implemented in two sequential steps. Phase 1 (feature extraction) freezes the whole VGG16 architecture except the classifier, composed of 148,865 trainable weights, that is trained with Adam optimizer with a 5×10^{-5} learning rate and gradient clipping (clipnorm = 1.0). The reduced learning rate, compared to the 10^{-4} used in the baseline EfficientNetB0 model, was used for the sake of safety due to increased susceptibility of the smaller training sample to drastic updates at the beginning of training. Label smoothing parameter set at 0.05 was used to promote generalization by preventing overconfidence in binary cross-entropy loss function. Up to 40 epochs were run in Phase 1, whereas early stopping with a patience of 15 was used based on changes in validation loss rather than validation AUC score, which showed too much noise in preliminary trials.

Phase 2 (micro fine-tuning) involved an unfreezing of only one of five convolutional layers (block5_conv3), while all other layers remained frozen, which was done in order to avoid catastrophic forgetting with minimal risk. Specifically, Phase 2 allowed for gradient updating of higher level feature detectors, i.e., those responsible for detecting domain-specific textures and edges. The learning rate was also further reduced to 5×10^{-6} in order to incrementally adjust such features without altering the representations of the previous levels learned on ImageNet. The exact same early stopping strategy and ReduceLROnPlateau scheme (factor = 0.5 and patience = 7) were used in Phase 2.

Patient-Level Inference with Test-Time Augmentation

At inference time, the entire set of DICOM slices associated with a certain patient is considered independently. For each slice, TTA is applied, where there are eight forward passes performed for each slice, using each of them a random draw from the training augmentation process, and averaging their results. This helps reduce the variance of the probability estimate for each slice due to the randomness associated with the dropout process, and differences in anatomy along the axial direction. Three methods for patient level aggregation are considered, namely maximum pooling (maximum value among all the slice probability estimates), mean pooling, and top-5 mean pooling (the mean value of the highest probability slices).

Experimental Results

Experimental Setup

The experiments were performed using Kaggle cloud resources equipped with dual NVIDIA Tesla T4 GPUs (each with 13,757 MB VRAM, compute capability 7.5). The code was developed with TensorFlow 2.19.0 using Keras functional API, GPU-accelerated XLA compiler, and CUDA library cuDNN 9.1. DICOM files were read and preprocessed via PyDICOM, extended with support for JPEG-encoded images through PyLibJPEG package. Training histories and best-epochs checkpoints were stored to disk; weights of epoch with the lowest validation loss were loaded prior to testing the model

Training Dynamics

In phase 1 of the training, the model exhibited almost random behavior at the first epoch (accuracy 49.9%, AUC 0.49), but quickly converged to a minimum validation loss of 0.710 at epoch 4, showing that even with fully frozen VGG16 weights, the classifier can learn the discriminative features associated with fractures in a small number of steps. Further training led to a flattening and eventual upward trend in the validation loss starting at epoch 4, which reached the value of 0.776 at epoch 19 resulting in the termination of training and reloading the epoch 4 weights. Train AUC increased further to 0.847 at epoch 11, while the validation AUC did not show any improvement from epoch 4 and oscillated in the range of 0.67 - 0.70, which demonstrates overfitting in the classifier head even under the influence of regularization techniques [13].

Phase 2 started from the best phase 1 checkpoint. The first epoch of phase 2 produced a slight increase of validation loss from phase 1's best of 0.710, which showed that the increased capacity brought by unfreezing one convolutional layer had little positive impact on the performance. In all following epochs, the validation loss steadily increased until it reached the value of 0.813 at epoch 16, causing the early stopping condition to kick in. Thus, the last model is based on epoch 1 of phase 2 and is practically identical to the best phase 1 model. Figures below show the loss, AUC-ROC, and accuracy plots throughout the training; they confirm that the proposed methods prevented catastrophic divergence seen during previous experiments' iterations

Validation Set Performance

Among the 10 patients constituting the validation set (5 fracture patients and 5 normals), all the fracture patients were detected with extremely high probability (values: 0.863, 0.883, 0.822, 0.940, 0.924). All the normal subjects were misdiagnosed as having fractures (values: 0.624, 0.664, 0.688, 0.743, 0.842), making the sensitivity score perfectly accurate while specificity is zero when the threshold of 0.5 is used. The Youden J optimal threshold based on the ROC curve of the validation data is 0.863, which separates fracture patients (all above values ≥ 0.822) from normal patients, although the decision threshold is extremely high in this case.

Test Set Evaluation

The patient level assessment was performed for the test data set of 6 patients (3 fracture patients and 3 normals) using three methods of score aggregation using a threshold value of 0.863 obtained during validation.

Table 3. *Patient-Level Test Performance by Aggregation Strategy*

Aggregation	Accuracy	Sensitivity	Specificity	Precision	F1	AUC-ROC
Maximum	0.667	0.333	1.000	1.000	0.500	0.667
Mean	0.500	0.000	1.000	0.000	0.000	0.444
Top-5 Mean	0.667	0.333	1.000	1.000	0.500	0.667

Using both maximum and top-5 mean aggregation rules, all three normal patients were correctly classified without any false positives (specificity = 1.000) and only one out of three fracture patients (patient CQ500CT81 with 0.915 probability). Patient CQ500CT16 and patient CQ500CT66 who were not detected as having fractures produced maximum probabilities of 0.746 and 0.766, respectively, which do not pass the very high threshold of Youden J index (0.863) but surpass the default of 0.5. With an AUC-PRC score of 0.756, the maximum aggregation strategy shows considerable discrimination power within the limited sample of the prototype test set. In contrast, the mean aggregation rule performed worse than maximum and top-5 mean aggregation rules because averaging probabilities weakens their contribution since there is a large portion of non-fractured slices that appear unimportant in both healthy and fracture-positive patients.

Comparison with Related Work

Table 4. *Comparison of Deep Learning Methods for Skull Fracture Detection*

Method	Backbone	Dataset Size	Accuracy	AUC
Abubacker et al.	Otsu Thresholding	10 patients	99%	—
Liu et al. FPN+CCA	FPN + Attention	19,222 images	82.0%	0.972
Yang et al.	VGG-19 + Grad-CAM	19,079 scans	91.36%	—
Kodavati & Kumar	DenseNet161+U-Net	Not disclosed	98.37%	0.976
SkullFractureNet (Ours)	VGG16 + TTA	42 patients (CQ500)	0.667 (test)	0.667 test / 0.756 PRC

Discussion

Interpretation of Results

The test sensitivity of 0.333 at the optimal threshold should be considered in terms of the threshold.selection mechanism: the Youden J statistic in the validation set preferred a high cutoff of 0.863 since validation fracture favored a high cutoff of 0.863 patients had probabilities of 0.82 or more and all validation normals were between 0.62 and 0.84. This produced a threshold which optimized the trade-off between validation and could not be extrapolated to the test fracture patients (CQ500CT16 and human threshold) which maximized the trade-off between validation and could not be applied to the test fracture patients (CQ500CT16 andCQ500CT66), which had the highest probability of 0.746 and 0.766 indicating lower confidence than the validation fracture group. CQ500CT81 would be rightfully detected at a default threshold of 0.5 and no false positives of the two would be detected ordinary individuals with 0.569 and 0.724 probabilities--but CQ500CT83 would be mistakenly considered fracture at<|human|>ordinary patients with 0.569 and 0.724 probabilities--yet CQ500CT83 would be incorrectly judged as a fracture at 0.860 probability.The optimal specificity (1.000) with the maximum aggregation strategy and non-trivial AUC-PRC (0.756) indicates that the model has acquired characteristics that are consistently not present in normal class. The false negatives seem to represent real inter-patient differences in fracture severity and imaging features as opposed to systematic model failure.It is noteworthy that TTA enhanced individual-patient probability estimates by decreasing single-pass variation: without averaging patient-level scores would be noisier with the low-confidence fracture slices.

Comparison with Ef icientNetB0 Baseline

The VGG16 pipeline is a conscious compromise in the parameters efficiency VGG16 has 138M parameters, a regression parameters than EfficientNetB0 of 4M- but this was mitigated by the anti-overfitting.The earlier there was an issue of training instability on the 26-patient cohort and left tables of the implementation of EfficientNetB0 unpopulated when it was submitted. The existing pipeline generates conclusive numerical findings in three aggregation strategies and exhibits stable dynamics of training, which is a more practically useful starting point in the further expansion of a dataset.

Limitations

The main constraint is the size of the datasets. The size of the training cohort of 26 patients is many folds smaller than the datasets supporting the most effective literature techniques (19,000+ pictures) [6], [7], [11]. The test set of 6 patients restricts the statistical power. of all known measures: one reclassification of patients changes the accuracy by 16.7 percentage points. The high Youden J threshold obtained by validation is an optimization of threshold on 10 patients, which is not enough to obtain reliable threshold generalization. The cohort structure with a 1:1 balance between the classes (21 fracture, 21 normal out of 41 available normals) eliminates over 21 normal ones one half of the available normal data, which was required to give equal consideration, but diminishes the variety of normal anatomy observed.during training [4]

Future Directions

The most influential enhancement is expansion of the datasets. Introducing more CQ500 subsets or federating. The major limitation would be directly resolved by accessing other publicly available repositories of head CT. On the loss functional side,Focal Loss (2, = 0.75) will be used instead of Binary Cross-Entropy [15].To enhance performance on challenging cases fracture detection Networks Feature Pyramid Architecturally Feature Pyramid Networks provide both hairline and depressed subtypes of multi-scale fracture detection [7].

Grad-CAM visualization will be put in place in the pipeline and will be incorporated in the validation reporting in the following iteration.[14] to determine space correspondence between model attention and annotator-marked fracture sites. The semi-supervised learning with the massive amount of unlabeled CQ500 studies through pseudo-labeling.

Some training data can be effectively expanded with no extra annotation by [4] cost. The standard axial slices could be complemented with azimuthal equidistant projection and disk harmonic map representations[4] [19] by enhancing salience of basilar and frontal fractures on 2D projections.

Conclusion

In this paper, a recent version of SkullFractureNet pipeline, an automated skull fracture detector, is introduced in head CT DICOM data, replacing VGG16 with the previous EfficientNetB0 backbone and proposing a set of anti-overfitting changes [12] GlobalMaxPooling2D, trimmed classification head, higher dropout (0.6/0.4), L2 regularization, deep slice-level augmentation, constrained Phase 2 fine-tuning to block5_conv3, and test time augmentation at inference (eight passes).A balanced CQ500 subset (42 patients) experiment shows stable training dynamics, false positive zero on test set maximum aggregation, and an AUC-PRC of 0.756. These outcomes are a reproducible, numerically complete baseline which explicitly aims at precision and specificity given a constraint of an extremely small training cohort.The pipeline is developed to be improved repeatedly, and the most urgent is the expansion of the dataset and the integration of focal losses calculated improvements to clinical sensitivity goal of 0.89 as set by post-mortem CT meta-analysis [3].

Acknowledgment

The authors thank the creators of the CQ500 dataset and Kaggle for providing cloud GPU infrastructure. This work was conducted as part of the undergraduate major project at Somaiya Institute of Technology, Mumbai.

References

1. Maas, A. I., Menon, D. K., Adelson, P. D., et al. “Traumatic brain injury: Integrated approaches to improve prevention, clinical

care, and research," Lancet Neurology, vol. 16, no. 12, pp. 987–1048, Dec. 2017.
2. Aarabi, T. M., Gerstein, B., & Harrop, C. "Skull base fractures and their complications," Journal of Trauma, vol. 71, no. 6, pp. 1619–1625, Dec. 2011.
3. Henningsen, M. J., Larsen, S. T., Jacobsen, C., & Villa, C. "Sensitivity and specificity of post-mortem computed tomography in skull fracture detection—a systematic review and meta-analysis," International Journal of Legal Medicine, vol. 136, pp. 1363–1377, Mar. 2022.
4. Hadjittoouli, N., Nikolaou, C., & Pitris, C. "Azimuthal equidistant mapping and projection of head CT scans for fracture visualization," IEEE Access, vol. 13, pp. 78214–78220, May 2025.
5. Abubacker, S. N., Sulaiman, N., & Salleh, S. H. "An approach for automatic fracture detection of skull DICOM images based on neighboring pixels," Proc. IEEE International Conference on Control System, Computing Engineering, pp. 558–563, Nov. 2015.
6. Kodavati, S. K., & Kumar, Y. C. A. P. "Detection and segmentation of skull fractures via CNN and U-Net hybrid model using computed tomography images," SN Computer Science, vol. 3, no. 397, pp. 1–12, Sep. 2022.
7. Liu, G., Wen, Y., Bai, X., Zhu, W., & Li, H. "Skull fracture detection method based on improved feature pyramid network," Proc. IEEE 10th International Conference on CYBER Technology in Automation, Control, and Intelligent Systems, pp. 756–762, Oct. 2021.
8. Kumar, A., & Agarwal, D. "Automated bone fracture detection using YOLOv8," International Journal of Research in Applied Science and Engineering Technology, vol. 11, no. 5, pp. 2998–3005, May 2023.
9. Zaki, W. M. H., Husain, M., Farok, M. L. A. M. R. M., & Sulaiman, N. A. "Automated method of fracture detection in CT brain images," Proc. IEEE EMBS Conference on Biomedical Engineering and Sciences, pp. 285–289, Dec. 2016.
10. Yamada, A., Teramoto, A., Otsuka, T., Kudo, K., Anno, H., & Fujita, H. "Preliminary study on the automated skull fracture detection in CT images using black-hat transform," Proc. 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 6437–6440, Aug. 2016.
11. Yang, C.-Y., Lo, C.-H., Wang, H.-C., Chou, J.-H., & Wang, Y.-C. F. "Weakly-supervised learning for attention-guided skull fracture classification in computed tomography imaging," Proc. IEEE International Conference on Image Processing, pp. 1337–1341, Sep. 2019.
12. Tan, M., & Le, Q. V. "EfficientNet: Rethinking model scaling for convolutional neural networks," Proc. International Conference on Machine Learning (ICML), vol. 97, pp. 6105–6114, Jun. 2019.
13. Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. "How transferable are features in deep neural networks?," Advances in Neural Information Processing Systems (NeurIPS), vol. 27, pp. 3320–3328, 2014.
14. Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. "Grad-CAM: Visual explanations from deep networks via gradient-based localization," Proc. IEEE International Conference on Computer Vision (ICCV), pp. 618–626, Oct. 2017.
15. Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. "Focal loss for dense object detection," Proc. IEEE International Conference on Computer Vision (ICCV), pp. 2980–2988, Oct. 2017.
16. Canny, J. "A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.
17. Gonzalez, R. C., & Woods, R. E. Digital Image Processing, 3rd ed. Upper Saddle River, NJ: Prentice-Hall, 2008.
18. Ringl, C., Schernthaner, H., Schaefer-Prokop, R., et al. "Curved planar reformations of the skull: A new visualization technique for skull fractures," Radiology, vol. 258, no. 2, pp. 604–610, Feb. 2011.
19. Hadjittoouli, N., & Pitris, C. "Skull fracture visualization using disk harmonic maps," Proc. IEEE 22nd International Symposium on Biomedical Imaging, pp. 1–4, May 2025.
20. Daly, B., Orave, P. J., & Cibelli, L. "Occipital condyle fractures in the era of multidetector CT," Emergency Radiology, vol. 19, no. 5, pp. 407–413, Oct. 2012.
21. Le Blanc-Louvry, I., Thureau, A. C., Duval, E., Papin-Lefebvre, B., & Thiebot, G. "Postmortem computed tomography compared to forensic autopsy findings: A French experience," European Radiology, vol. 23, no. 7, pp. 1829–1835, Jul. 2013.
22. Khanuja, T., & Unni, H. N. "Computational modeling of traumatic brain injury due to impact on different sides of human head," Proc. IEEE 18th International Conference on Bioinformatics and Bioengineering, pp. 364–370, Oct. 2018.
23. Othman, A., et al. "Traumatic brain injuries and maxillofacial fractures: A systematic review and meta-analysis," Oral and Maxillofacial Surgery, vol. 27, pp. 373–385, Jun. 2023.
24. Zaki, Z., Rahiman, M. H. F., Hitam, M. S., & Sulaiman, N. "Multi-level segmentation method for serial computed tomography brain images," Journal of Telecommunications, Electronic and Computer Engineering, vol. 9, no. 1–2, pp. 49–56, 2017.