Evaluation of a Deep Learning Model for Skull Fracture Detection in Brain CT Scans: A Performance Assessment on an Independent Dataset

Department of Statistics, Faculty of Science, Chiang Mai University

Author: Ratima Khamfu **Advisor:** Assoc. Prof. Dr.Patrinee Traisathit and Dr.Pimwarat Srikummoon

ABSTRACT

This study evaluates the effectiveness of a deep learning model in detecting skull fractures from brain CT images. The model was developed using a dataset from Maharaj Nakorn Chiang Mai Hospital and tested on an independent dataset comprising 147 randomly selected patients with traumatic brain injuries admitted to Lampang Hospital between 2022 and 2024. The evaluation aimed to assess the model's performance on previously unseen data.

The model's diagnostic performance was measured using accuracy, precision, sensitivity, and specificity. Results showed a high sensitivity of 0.952, indicating its effectiveness in detecting skull fractures, while the specificity was 0.744. With an accuracy of 0.850, the model demonstrated strong predictive capability for both positive and negative cases. However, its precision of 0.759 suggests some misclassification in detecting depressed skull fractures.

Given these findings, the deep learning model shows promise as a screening tool for identifying skull fractures that require urgent surgical intervention. However, for optimal diagnosis and treatment planning, its use should be complemented by expert radiological assessment.

INTRODUCTION

Head injuries result from external force impacting the head or body, causing trauma to the scalp, skull, brain, or nerves. Road accidents are a major cause of head injuries, with the skull being one of the most frequently affected structures.

Patients with brain injuries are diagnosed using brain computed tomography (CT) images, which are interpreted by radiologists. Referral for consultation and treatment by a surgeon is considered if any of the following CT abnormalities are detected: midline shift, hematomas, or depressed skull fractures. Artificial intelligence (AI) is increasingly being used to detect abnormalities in CT images of patients with traumatic brain injuries. However, studies evaluating AI models using test datasets from different hospitals remain scarce due to variations in CT imaging processes among medical institutions.

In this study, an independent dataset from Lampang Hospital was used to evaluate the performance of a skull fracture detection model originally developed with data from Maharaj Nakorn Chiang Mai Hospital. Assessing the model's performance on this dataset will help determine its effectiveness and reliability for practical applications in detecting skull fractures on brain CT images. This study is part of the research project "Development and Transfer of Technology: Artificial Intelligence Techniques in the Diagnosis of Patients with Brain Injury."

OBJECTIVE

To evaluate the effectiveness of a deep learning model in detecting skull fractures from brain CT images using a test dataset from Lampang Hospital.

METHODS

- **Study design:** Prospective Study
- **Study population:** Patients with traumatic brain injuries admitted to Lampang Hospital
- **Data collection:** Data was obtained from Lampang Hospital under the research project "Development and transfer of technology, artificial intelligence techniques in the diagnosis of patients with brain injury".The artificial intelligence model was developed using a dataset from Maharaj Nakorn Chiang Mai Hospital.
- **Data Cleansing :** Checking for duplicates and ensuring data completeness
- **Statistical analysis:**

Table 1. confusion matrix.

Actual values (Diagnosis result by radiologist)			
Depressed skull fracture	skull fracture		
TP (True Positive)	FP (False Positive)	Depressed skull fracture	Predicted values (Predicted results from AI models)
FN (False Negative)	TN (True Negative)	skull fracture	

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\text{Precission} = \frac{TP}{TP + FP}$$

$$\text{Specificity} = \frac{TN}{TN + FP}$$

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

RESULTS

Among the 147 patients, the majority were male, with a mean age of 42.82 years. Based on radiologists' classifications of skull fractures from brain CT scans, most patients had depressed skull fractures (42.85%), and the majority of these cases required surgical intervention (42.86%).

Table 2. Diagnostic Performance of the Deep Learning Model

Diagnostic Performance Value	Test Dataset from Maharaj Nakorn Chiang Mai Hospital	Test Dataset from Lampang Hospital
Accuracy	0.957	0.850
Precision	0.977	0.759
Sensitivity	0.933	0.952
Specificity	0.980	0.744

The results in Table 2 indicate that the model achieved an accuracy of 0.850, a precision of 0.759, a sensitivity of 0.952, and a specificity of 0.744 when tested with 147 cases from Lampang Hospital.

Table 3. Confusion Matrix for Skull Fracture Detection (Lampang Hospital Dataset)

Actual values			
Depressed skull fracture	skull fracture	n = 147	
65	19	Depressed skull fracture	Predicted values
3	60	skull fracture	

The results in Table 3 show the model's performance values using the Lampang Hospital dataset, calculated by comparing radiologists' diagnoses with the AI model's predictions.

CONCLUSION

The deep learning model exhibited high sensitivity (0.952), indicating its effectiveness in detecting skull fractures. However, its specificity was 0.744, and its precision of 0.759 suggests some misclassification, particularly in detecting depressed skull fractures. When comparing performance on test datasets from Maharaj Nakorn Chiang Mai Hospital and Lampang Hospital, the model's effectiveness was notably lower on the Lampang dataset, with only sensitivity remaining similar. This difference is likely due to the higher image sharpness of the bone window CT scans from Lampang Hospital compared to the bone window images adapted from brain window scans in the Maharaj Nakorn Chiang Mai dataset. Consequently, this affected the model's performance, leading to detections of other skull regions instead of focusing solely on fractures. Despite these variations, the deep learning model demonstrates potential as a screening tool for identifying skull fractures requiring urgent surgical intervention. However, expert radiological assessment remains essential for optimal diagnosis and treatment planning.

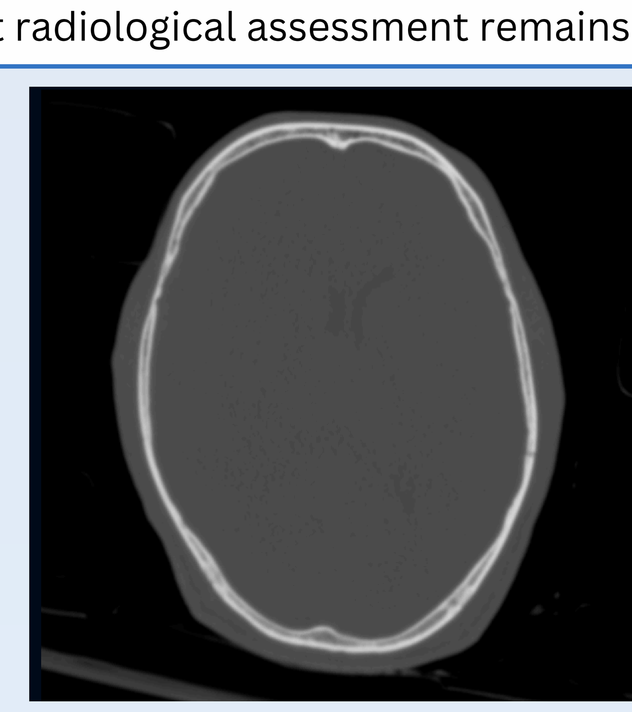

Fig 1. The bone window CT images adapted from brain window scans in the Maharaj Nakorn Chiang Mai dataset

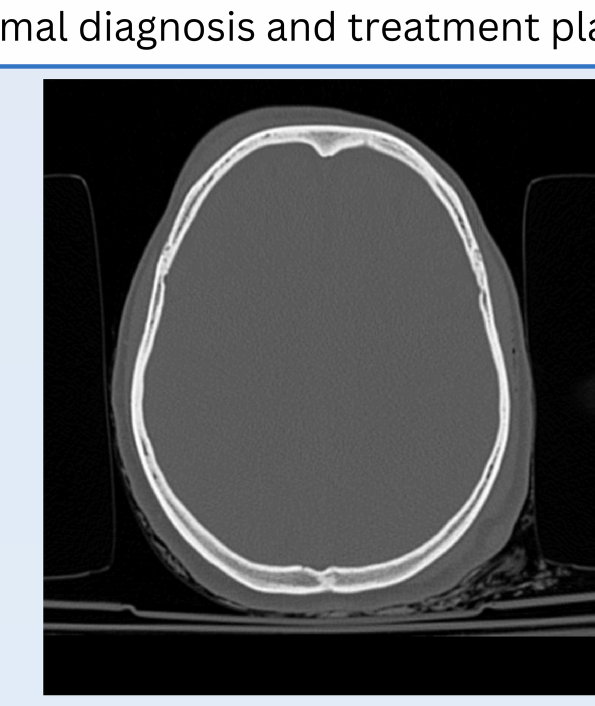

Fig 2. The bone window CT scans from Lampang Hospital

REFERENCES :

[1] Nirase Fathima Abubacker et al. (2013). **An approach for an automatic fracture detection of skull dicom images based on neighboring pixels.** 2013 13th International Conference on Intellient Systems Design and Applications, 177-181. IEEE.

[2] Hicks et al. (2022). **On evaluation metrics for medical applications of artificial intelligence.** Scientific Reports, 12(1), 5979. doi:10.1038/s41598-022-09954-8

[3] Prakash et al. (2018). **Depressed fractures of skull: An institutional series of 453 patients and brief review of literature.** Asian Journal of Neurosurgery, 13(2), 222–226. doi:10.4103/ajns.AJNS_168_16

[4] Wei Shan et al. (2021). **Automated identification of skull fractures with deep learning: A comparison between object detection and segmentation approach.** Frontiers in Neurology, 12, 687931. doi:10.3389/fneur.2021.687931

[5] National Institute of Neurological Disorders and Stroke. (n.d.). **Traumatic brain injury (TBI).** from https://www.ninds.nih.gov/health-information/disorders/traumatic-brain-injury-tbi?search-term=Traumatic%20Brain%20Injury

ACKNOWLEDGMENTS

The authors would like to thank Assoc. Prof. Salita Angkurawaranon, M.D., Ms. Phattanan Thap-asa, and the other professors who contributed to this study. This research was supported by the project "Development and Transfer of Technology: Artificial Intelligence Techniques in the Diagnosis of Patients with Brain Injury" which was funded by "The Program Management Unit for Area-based Development (PMU-A), The Office of National Higher Education Science Research and Innovation Policy Council (NXPO)" and the Department of Radiology, Faculty of Medicine, Chiang Mai University.