Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Deep Learning Model Based on Transcranial Sonography for Detection of Acute Intracranial Hemorrhage in Adults: A Comparative Study With Sonologists of Different Experience Levels

Jianbo Gao
Chinese People' s Liberation Army General Hospital
Xiazhen Qin
Chinese People ' s Liberation Army General Hospital
Xiaomeng Zou
Chinese People' s Liberation Army General Hospital
Yueming Li
Chinese People' s Liberation Army General Hospital
Yongqi Fang
Chinese People' s Liberation Army General Hospital
Naiqin Fu
Chinese People' s Liberation Army General Hospital
Ruilan Niu
Chinese People' s Liberation Army General Hospital
Chiyuan Ma
Chinese People' s Liberation Army General Hospital
Xingyao Cui
Chinese People' s Liberation Army General Hospital
Xu Li
Chinese People ' s Liberation Army General Hospital
Lianhua Zhu
Chinese People' s Liberation Army General Hospital
Zhili Wang
wzllg@sina.com

Chinese People' s Liberation Army General Hospital

Research Article

Keywords: acute intracranial hemorrhage, artificial intelligence, deep learning, diagnostic accuracy, transcranial sonography

Posted Date: November 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11612041/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

Background

To evaluate the diagnostic performance of a deep learning model for acute intracranial hemorrhage (ICH) on transcranial sonography (TCS) versus sonologists with different levels of experience.

Methods

We retrospectively studied 395 adults with suspected acute ICH undergoing handheld TCS, using noncontrast head CT as the reference standard (194 ICH). Participants were randomized 8:2 to training (n = 317) and test (n = 78) sets. An EfficientNet-B2 model was trained on cine-loop frames, with frame probabilities averaged per participant. Six blinded sonologists (2 each with > 10, 5–10, and < 5 years' experience) read the images. Accuracy, sensitivity, specificity, positive and negative predictive values (PPV, NPV), and area under the receiver operating characteristic curve (AUC) were calculated, with calibration and decision curve analyses.

Results

In the test set (38 ICH, 40 non-ICH), model AUC was 0.956 (95% confidence interval, 0.892–0.996), accuracy 91.03%, sensitivity 89.47%, specificity 92.50%, PPV 91.89%, and NPV 90.24%. For senior, mid-level, and junior sonologists, respectively: accuracy 91.67%/78.85%/66.03%, sensitivity 94.74%/82.89%/69.74%, specificity 88.75%/75.00%/62.50%, PPV 88.89%/75.90%/63.86%, and NPV 94.67%/82.19%/68.49%. Model accuracy matched senior sonologists ($P > 0.05$) and exceeded mid-level and junior sonologists ($P = 0.008$, $P < 0.001$); sensitivity and specificity matched senior and mid-level sonologists (all $P > 0.05$) and exceeded junior sonologists ($P = 0.007$, $P = 0.001$). Calibration was good, and decision curve analysis demonstrated a positive net benefit.

Conclusions

The TCS-based model showed good discrimination of acute ICH, matching senior and outperforming junior sonologists, and may aid screening in prehospital, bedside, and resource-limited settings.

Background

Acute intracranial hemorrhage (ICH) is a common neurologic emergency characterized by abrupt onset, rapid progression, and high rates of disability and death [1, 2]. Early recognition of ICH and assessment of the hemorrhage burden facilitate emergency triage, timely initiation of blood pressure lowering and anticoagulation reversal, and improved outcomes [3]. Noncontrast computed tomography (NCCT) of the head is highly sensitive for acute hemorrhage and remains the first-line imaging modality for diagnosing

acute ICH. In prehospital care, bedside evaluation, and resource-limited settings, however, access to CT is constrained by equipment availability and the feasibility of patient transport.

Transcranial sonography (TCS) is a noninvasive, real-time, bedside neuroimaging technique that has been shown to detect intracranial hemorrhage [4–6]. A systematic review and meta-analysis of 44 studies comprising 2,426 patients reported a pooled sensitivity of 88%, specificity of 81%, and accuracy of 94% for TCS detection of ICH [7]. The development of portable handheld ultrasound devices has expanded the potential applications of TCS. These compact, battery-powered devices connect to smartphones for real-time transmission of high-resolution images. Increasingly affordable and easy to operate, they have been adopted for trauma FAST/e-FAST assessment, pericardial effusion screening, pneumothorax diagnosis, and cardiac function evaluation, among other extracranial applications, and can be deployed in ambulances, emergency departments, and resource-limited settings. Despite published methodological recommendations [8], TCS still lacks a uniform, broadly applicable scanning protocol, and both image acquisition and interpretation remain highly operator dependent [9], which has limited its clinical use.

Artificial intelligence (AI), particularly deep learning, has been widely applied to medical image classification, detection, and segmentation. For ICH detection on NCCT, deep learning models have achieved performance comparable or superior to that of radiologists; multiple systematic reviews and meta-analyses have reported pooled sensitivity of 0.89–0.92, specificity of 0.91–0.94, and an area under the receiver operating characteristic curve (AUC) of 0.94–0.96 [10, 11]. However, AI-assisted TCS for the identification of acute ICH remains exploratory [12], and reports on model diagnostic performance and on comparisons with clinicians are limited. EfficientNet balances network depth, width, and input image resolution through a compound scaling mechanism [13], and has performed well in deep learning analyses of ultrasound video [14], making it well suited to medical image recognition in small-sample settings. A binary transfer learning model based on the EfficientNet backbone may therefore enable rapid screening for intracranial hemorrhage on adult TCS images.

We therefore retrospectively analyzed adults with clinically suspected acute ICH who underwent handheld transcranial sonography and NCCT during the same period, developed a binary deep learning model for acute ICH based on TCS, and evaluated its diagnostic performance in the test set using NCCT as the reference standard. Sonologists with different levels of experience independently interpreted the images under blinded conditions, and paired analyses were used to compare the model with the sonologists. Our aims were to assess the feasibility and potential adjunctive diagnostic value of a TCS-based deep learning model for acute ICH in adults and to define its clinical role relative to sonologists with different levels of experience.

Methods

Study Participants

The study was approved by the Ethics Committee of Chinese People's Liberation Army General Hospital (approval number S2023-176-01), which waived the requirement for informed consent because of the retrospective design and the use of de-identified data.

Of 435 consecutively enrolled patients with suspected intracranial hemorrhage, 40 were excluded (23 because of poor transcranial acoustic windows and 17 because of incomplete clinical or imaging data), leaving 395 participants for analysis. NCCT findings served as the diagnostic reference standard for all cases.

Ultrasound Examination

All examinations were performed by a sonologist with more than 10 years of experience in ultrasound according to a standardized scanning protocol. Participants were examined in the supine position using a handheld color Doppler ultrasound system (TE AIR; Mindray, Shenzhen, China) with a transcranial phased-array probe (i3P; 1.0–2.5 MHz). High-penetration grayscale imaging was used with an examination depth of 16–18 cm. The temporal window served as the acoustic window, with the probe placed over the preauricular temporal region and its long axis parallel to the orbitomeatal line; intracranial structures below, at, and above the level of the midbrain were scanned continuously by adjusting the probe position and angle [9, 12]. Static and dynamic TCS images were acquired during the examination for subsequent frame extraction, model analysis, and human interpretation.

Dataset Construction and Partitioning

Static frames were extracted from the dynamic images at a rate of 1 frame per second and, after format conversion, de-identification, and size normalization, were included in the analysis. Image preprocessing and dataset partitioning followed the reporting requirements of the CLAIM 2024 and STARD-AI checklists [15, 16]. All data were partitioned at the participant level, with participants randomly allocated to training and test sets at a ratio of 8:2. During partitioning, all images from a given participant were confined to a single subset to avoid participant-level data leakage. The training set was used for iterative updating of model parameters, and the test set for performance evaluation. A random seed of 42 was used to ensure reproducibility of the entire experiment.

Deep Learning Model

Model development was performed in Python 3.11 with PyTorch 2.3.0 and CUDA 12.1 on an RTX 4090 graphics processing unit (24 GB). TCS images of 260 × 260 pixels were used as input, and data augmentation was implemented by random cropping, flipping, rotation, and brightness and contrast perturbation. The backbone was ImageNet-pretrained EfficientNet-B2 [13]. A layerwise fine-tuning strategy was adopted: the first three backbone layers were frozen, and only the deeper feature layers and a custom binary classification head were trained. The classification head comprised multilevel dropout,

fully connected layers, batch normalization, and rectified linear unit (ReLU) activation, and output the predicted probability of hemorrhage (Fig. 1).

Training used cross-entropy loss and the AdamW optimizer [17] with an initial learning rate of 3×10^{-4}, weight decay of 0.02, cosine annealing learning rate scheduling, a batch size of 16, and 20 epochs. Before training, an internal validation set was split from the training set at a ratio of 4:1, and the weights with the highest validation accuracy were saved; the test set remained completely independent throughout training and model selection. Multiple frames from the same participant were treated as a "bag," and the softmax predicted probabilities of the individual frames were fused by mean pooling to obtain a participant-level prediction [18]; a threshold of 0.5 defined a positive result. Gradient-weighted class activation mapping (Grad-CAM) [19] was used to generate heat maps from the final convolutional layer of the backbone; these were overlaid on the original images to visualize the regions on which the model focused.

Human Image Interpretation

All test set images were organized at the participant level, stripped of participant identity and diagnostic information, randomly ordered, renumbered R001–R078, and interpreted by six sonologists. According to years of ultrasound experience, the sonologists were divided into 3 groups of 2 sonologists each: junior (< 5 years), mid-level (5–10 years), and senior (> 10 years). This panel spanned a sufficiently broad range of experience and exceeded the minimum of 5 recommended for multireader diagnostic accuracy studies [20]. All sonologists received uniform training and calibration before interpretation and remained blinded to participants' clinical data, one another's interpretations, deep learning model predictions, and the NCCT reference standard. Positive interpretation was based on previous studies showing that acute intracranial hematomas appear markedly hyperechoic on grayscale images [21, 22].

Statistical Analysis

Statistical analyses were performed with R 4.3.1 and Python 3.11. Normally distributed continuous variables are presented as mean ± SD and non-normally distributed variables as median (interquartile range [IQR]); between-group comparisons used the independent-samples t test and the Mann–Whitney U test, respectively. Categorical variables are presented as counts and percentages [n (%)], and between-group comparisons used the χ^2 test or the Fisher exact test. When at least 20% of the cells in a contingency table had an expected count below 5, a Monte Carlo conditional permutation exact test (200,000 iterations) was used instead. Using NCCT findings as the reference standard, we calculated the model AUC, accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The 95% confidence intervals (CIs) for accuracy, sensitivity, specificity, PPV, and NPV were calculated with the Clopper–Pearson exact method, and the 95% CI for the AUC with the DeLong method. Two-sided exact McNemar tests were used to compare the AI model with sonologists of different experience levels, and agreement was assessed with Cohen's κ, with 95% CIs obtained by the

percentile method from 2,000 bootstrap resamples. All tests were two-sided, and P < 0.05 was considered statistically significant.

Use of Artificial Intelligence Tools

WorkBuddy, an artificial intelligence–based writing assistant, was used to translate the original Chinese draft into English and to assist with language polishing and proofreading. It was not used for study design, data collection, image acquisition or interpretation, statistical analysis, or generation of any figure or table. All scientific content was conceived, analyzed, and verified by the authors, who take full responsibility for the integrity and accuracy of the work. This tool does not meet the ICMJE criteria for authorship and is therefore not listed as an author.

Results

Baseline Characteristics of the Dataset

A total of 395 participants were included, with a mean age of 60.74 ± 16.85 years, and 1,734 cranial ultrasound images were extracted. There were 201 participants without hemorrhage and 194 with hemorrhage. According to the predefined partitioning scheme, the training set comprised 317 participants (156 with intracranial hemorrhage and 161 without) with 1,391 images, and the test set comprised 78 participants with 343 images. Baseline characteristics are presented in Table 1. There were no statistically significant differences between the training and test sets in age, sex, time from symptom onset to examination, admission Glasgow Coma Scale (GCS) score, or history of hypertension (all $P > 0.05$). Among participants with hemorrhage, hemorrhage type and hemorrhage volume also did not differ significantly between the two sets (all $P > 0.05$).

Table 1
Baseline Characteristics of the Study Participants

Characteristic	Overall (n = 395)	Training Set (n = 317)	Test Set (n = 78)	*P*
Age, y, mean ± SD	60.74 ± 16.85	60.73 ± 17.03	60.81 ± 16.22	0.969
Male sex, n (%)	295 (74.7)	240 (75.7)	55 (70.5)	0.344
Time from onset to examination, h, median (IQR)	23.03 (12.6–42.5)	22.10 (11.8–41.9)	25.04 (15.8–45.7)	0.182
Admission GCS score, median (IQR)	13 (7–15)	12 (7–15)	15 (7–15)	0.197
History of hypertension, n (%)	233 (59.0)	189 (59.6)	44 (56.4)	0.605
ICH confirmed on NCCT, n (%)	194 (49.1)	156 (49.2)	38 (48.7)	—
Hemorrhage subgroup characteristics*	n = 194	n = 156	n = 38	
Hemorrhage type, n (%)				> 0.99
Intraparenchymal hemorrhage	135 (69.6)	109 (69.9)	26 (68.4)	
Subarachnoid hemorrhage	17 (8.8)	14 (9.0)	3 (7.9)	
Subdural hematoma	14 (7.2)	11 (7.1)	3 (7.9)	
Intraventricular hemorrhage	11 (5.7)	9 (5.8)	2 (5.3)	
Brainstem hemorrhage	13 (6.7)	10 (6.4)	3 (7.9)	
Epidural hematoma	4 (2.1)	3 (1.9)	1 (2.6)	
Hemorrhage volume*, mL, median (IQR)	17.22 (6.62–37.79)	18.71 (6.49–37.74)	16.14 (7.46–41.03)	0.934

Data are presented as mean ± SD, median (interquartile range), or counts (percentages). *These rows apply to participants with NCCT-confirmed intracranial hemorrhage, with denominators of 194, 156, and 38 in the three columns. GCS indicates Glasgow Coma Scale; ICH, intracranial hemorrhage; IQR, interquartile range; and NCCT, noncontrast computed tomography.

Diagnostic Performance of the Deep Learning Model in the Test Set

The analysis included 78 participants, of whom 38 (48.72%) had NCCT-confirmed intracranial hemorrhage and 40 (51.28%) did not. The model achieved an AUC of 0.956 (95% CI, 0.892–0.996), accuracy 91.03% (95% CI, 82.38%–96.32%), sensitivity 89.47% (95% CI, 75.20%–97.06%), specificity 92.50% (95% CI, 79.61%–98.43%), PPV 91.89% (95% CI, 78.09%–98.30%), NPV 90.24% (95% CI, 76.87%–97.28%), F1 score 0.907 (95% CI, 0.826–0.966), and precision–recall AUC (PR-AUC) 0.960 (95% CI, 0.897–0.997); the receiver operating characteristic curve is shown in Fig. 2A. The confusion matrix

showed that the model correctly identified intracranial hemorrhage in 34 participants and correctly excluded it in 37, with 3 false-positive and 4 false-negative results (Fig. 2B).

Comparison With Sonologists of Different Experience Levels

Among the 78 participants in the test set, the diagnostic performance of each sonologist group and the corresponding 95% CIs are presented in Table 2. Sensitivity for the senior, mid-level, and junior sonologist groups was 94.74%, 82.89%, and 69.74%; specificity was 88.75%, 75.00%, and 62.50%; and accuracy was 91.67%, 78.85%, and 66.03%. Model sensitivity and specificity were both significantly higher than those of the junior sonologist group ($P = 0.007$ and $P = 0.001$) and did not differ significantly from those of the senior or mid-level groups (all $P > 0.05$) (Table 3). Model accuracy did not differ significantly from that of the senior group ($P > 0.05$) but was significantly higher than that of the mid-level and junior groups ($P = 0.008$ and $P < 0.001$). Agreement between experience groups, within groups, and with the AI model is presented in Table 4.

Table 2

Participant-Level Diagnostic Performance of the Deep Learning Model and Sonologists by Experience

Method	Sensitivity	Specificity	PPV	NPV	Accuracy
AI model	89.47 (75.20–97.06)	92.50 (79.61–98.43)	91.89 (78.09–98.30)	90.24 (76.87–97.28)	91.03 (82.38–96.32)
Senior A	94.74 (82.25–99.36)	90.00 (76.34–97.21)	90.00 (76.34–97.21)	94.74 (82.25–99.36)	92.31 (84.01–97.12)
Senior B	94.74 (82.25–99.36)	87.50 (73.20–95.81)	87.80 (73.80–95.92)	94.59 (81.81–99.34)	91.03 (82.38–96.32)
Senior group (> 10 y)	94.74 (87.07–98.55)	88.75 (79.72–94.72)	88.89 (79.95–94.79)	94.67 (86.90–98.53)	91.67 (86.17–95.49)
Mid-level C	86.84 (71.91–95.59)	72.50 (56.11–85.40)	75.00 (59.66–86.81)	85.29 (68.94–95.05)	79.49 (68.84–87.80)
Mid-level D	78.95 (62.68–90.45)	77.50 (61.55–89.16)	76.92 (60.67–88.87)	79.49 (63.54–90.70)	78.21 (67.41–86.76)
Mid-level group (5–10 y)	82.89 (72.53–90.57)	75.00 (64.06–84.01)	75.90 (65.27–84.62)	82.19 (71.47–90.16)	78.85 (71.59–84.97)
Junior E	68.42 (51.35–82.50)	65.00 (48.32–79.37)	65.00 (48.32–79.37)	68.42 (51.35–82.50)	66.67 (55.08–76.94)
Junior F	71.05 (54.10–84.58)	60.00 (43.33–75.14)	62.79 (46.73–77.02)	68.57 (50.71–83.15)	65.38 (53.76–75.80)
Junior group (< 5 y)	69.74 (58.13–79.75)	62.50 (50.96–73.08)	63.86 (52.57–74.12)	68.49 (56.56–78.87)	66.03 (58.02–73.41)

All values are percentages, with 95% confidence intervals in parentheses. A and B, C and D, and E and F denote senior (> 10 years), mid-level (5–10 years), and junior (< 5 years) sonologists, respectively. AI indicates the deep learning model; NPV, negative predictive value; and PPV, positive predictive value; and y, years.

Table 3
Comparison of Diagnostic Performance Between the AI Model and Sonologists by Experience

Comparison	***P* (accuracy)**	***P* (sensitivity)**	***P* (specificity)**
AI model vs senior A	1.000	0.688	1.000
AI model vs senior B	1.000	0.500	0.688
AI model vs senior group	0.861	0.168	0.529
AI model vs mid-level C	0.022	1.000	0.008
AI model vs mid-level D	0.021	0.289	0.070
AI model vs mid-level group	0.008	0.278	0.058
AI model vs junior E	< 0.001	0.008	0.003
AI model vs junior F	< 0.001	0.035	< 0.001
AI model vs junior group	< 0.001	0.007	0.001

n = 78 (38 with and 40 without intracranial hemorrhage). *P* values are from two-sided exact McNemar tests comparing the AI model with each sonologist on the same participants.

Table 4
Agreement Between the AI Model and Sonologists of Different Experience Levels

Comparison	κ	95% CI	Strength of Agreement
Senior A	0.667	0.493–0.819	Substantial
Senior B	0.795	0.651–0.922	Substantial
Mid-level C	0.669	0.494–0.821	Substantial
Mid-level D	0.590	0.398–0.767	Moderate
Junior E	0.462	0.257–0.644	Moderate
Junior F	0.337	0.126–0.522	Fair
Senior group (A–B)	0.769	0.615–0.897	Substantial
Mid-level group (C–D)	0.667	0.490–0.821	Substantial
Junior group (E–F)	0.460	0.253–0.642	Moderate
Senior vs mid-level	0.653	0.519–0.769	Substantial
Senior vs junior	0.383	0.231–0.538	Fair
Mid-level vs junior	0.434	0.287–0.576	Moderate

Agreement was assessed with Cohen's κ (95% CIs by percentile bootstrap with 2,000 resamples) and graded per Landis and Koch (0.81–1.00, almost perfect; 0.61–0.80, substantial; 0.41–0.60, moderate; 0.21–0.40, fair; ≤0.20, poor). CI indicates confidence interval.

Calibration and Clinical Utility

The calibration curve (Fig. 3A) showed close agreement between the predicted probabilities and the observed event frequencies, with only minor deviation at low predicted probabilities, indicating good calibration. Decision curve analysis (Fig. 3B) showed that the net benefit of the model was higher than that of the "treat all" and "treat none" reference strategies across a wide range of threshold probabilities, indicating potential clinical utility within that range.

Visual Interpretation of Model Decisions

Grad-CAM visualization was used to display the regions on which the model focused (Fig. 4). In the true-positive case (A), model attention was concentrated on the hyperechoic hemorrhage in the cerebral hemisphere contralateral to the probe, closely corresponding to the location of the hemorrhage. In the false-positive case (B), attention was concentrated on a hyperechoic artifact of the contralateral petrous temporal bone, which was misinterpreted as hemorrhage. The true-negative case (C) was a physiologic hyperechoic calcification that the model correctly classified as no hemorrhage. In the false-negative case (D), attention was concentrated on the deep contralateral brain parenchyma, and a small contralateral subdural hemorrhage was missed.

Discussion

In this study, we established a binary deep learning model for acute intracranial hemorrhage using handheld transcranial sonography, with NCCT as the reference standard, and compared the performance of the model with that of sonologists of different experience levels in the same test set. The model achieved an AUC of 0.956, accuracy of 91.03%, sensitivity of 89.47%, and specificity of 92.50%. Model sensitivity and specificity were comparable to those of the senior and mid-level sonologist groups and significantly better than those of the junior group, whereas model accuracy was comparable to that of the senior group and significantly higher than that of the mid-level and junior groups.

Handheld transcranial sonography is portable, permits bedside examination, involves no ionizing radiation, and allows repeated dynamic assessment at the point of care. The model developed in this study achieved an AUC of 0.956 for the diagnosis of acute intracranial hemorrhage. Previous studies have confirmed that transcranial sonography can effectively identify intracranial hemorrhage [23]; a systematic review evaluating transcranial ultrasonography for the diagnosis of hemorrhagic stroke reported a pooled AUC of 0.953 [6]; and meta-analyses of AI models for intracranial hemorrhage on NCCT images have consistently reported pooled AUCs of 0.94–0.96 [10, 11]. The AUC of our model is comparable to these results, suggesting that deep learning can stably extract features of acute intracranial hemorrhage from transcranial sonography images and providing additional evidence supporting AI-assisted TCS for the diagnosis of acute ICH. Through automated interpretation, the model achieved overall diagnostic performance comparable to that of the sonologists. In prehospital care, bedside diagnosis, and resource-limited settings, the combination of handheld transcranial sonography and AI is expected to reduce dependence on operator experience and the subjective variability associated with human interpretation, thereby expanding the clinical applicability of this technique and providing a feasible approach for the rapid screening and evaluation of acute intracranial hemorrhage.

Our results also indicate that the performance of handheld transcranial sonography for diagnosing acute intracranial hemorrhage varies markedly with sonologist experience, declining as experience decreases, with a higher risk of missed and erroneous diagnoses among junior sonologists; this finding is consistent with the inherent operator dependence of transcranial sonography [24, 25]. The sensitivity and specificity of our AI model were significantly better than those of the junior sonologist group and did not differ significantly from those of the senior and mid-level groups, while model accuracy was comparable to that of the senior group and significantly higher than that of the mid-level and junior groups. Together with the agreement analysis, these findings indicate that the AI model can provide stable automated interpretation, compensate for the limited experience of junior sonologists, and reduce the subjective variability of human interpretation [26–28]. In prehospital care, bedside diagnosis, and resource-limited settings, combining AI with portable transcranial sonography may provide a reliable adjunctive tool for the rapid screening of acute intracranial hemorrhage.

Grad-CAM visualization showed that, in true-positive cases, the high-response regions of the model were concentrated mainly on the abnormal hyperechoic hemorrhage and the surrounding tissue. For

hyperechoic artifacts and physiologic calcifications that are difficult for some sonologists to distinguish [29, 30], the model was often able to discriminate correctly, although misclassifications still occurred; some small extra-axial subdural hematomas were missed, which we attribute to the small number of such cases, irregular lesion morphology, and insufficient model training, and which also explains why sensitivity was lower than specificity in this study. The calibration curve indicated good calibration, and decision curve analysis showed that the net benefit of the model was higher than that of the “treat all” and “treat none” strategies, suggesting that using model-predicted probabilities to support clinical decision making yields a positive net benefit over a wide range of thresholds and therefore has potential clinical value.

This study has several limitations. First, the sample size was limited, and the model supports only binary classification without lesion segmentation; it therefore cannot quantify hemorrhage extent, grade hemorrhage severity, or distinguish hemorrhage subtypes. Second, this was a single-center retrospective study without external validation, so the stability and generalizability of the results require further confirmation in larger multicenter independent prospective cohorts. Third, we compared the independent diagnostic performance of the AI model with that of sonologists of different experience levels but did not evaluate whether sonologist performance improves under an AI-assisted reading mode.

Conclusions

The deep learning model based on transcranial sonography images demonstrated good discrimination of acute intracranial hemorrhage in adults. Its diagnostic accuracy was comparable to that of senior sonologists and significantly higher than that of mid-level and junior sonologists; its sensitivity and specificity were comparable to those of senior and mid-level sonologists and significantly higher than those of junior sonologists. The model may serve as an adjunctive screening tool in prehospital care, bedside diagnosis, and resource-limited settings, although its performance in larger multicenter populations requires further validation.

Abbreviations

AI
artificial intelligence
AUC
area under the receiver operating characteristic curve
CI
confidence interval
CLAIM
Checklist for Artificial Intelligence in Medical Imaging
Grad-CAM
gradient-weighted class activation mapping
ICH

intracranial hemorrhage
IQR
interquartile range
NCCT
noncontrast computed tomography
NPV
negative predictive value
PPV
positive predictive value
ROC
receiver operating characteristic
SD
standard deviation
STARD
Standards for Reporting of Diagnostic Accuracy
TCS
transcranial sonography.

Declarations

Ethics approval and consent to participate

This retrospective study was approved by the Ethics Committee of Chinese People's Liberation Army General Hospital (approval number S2023-176-01), which waived the requirement for informed consent because of the retrospective design and the use of fully de-identified data. All procedures were performed in accordance with the Declaration of Helsinki.

Consent for publication

Not applicable. All ultrasound images and clinical data were fully de-identified before analysis, and no individual person's data are presented in any form that could permit identification.

Availability of data and materials

The datasets generated and/or analyzed during the current study are not publicly available due to patient privacy and ethical restrictions, but are available from the corresponding author on reasonable request, subject to approval by the institutional ethics committee.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported by the Independent Research Project for the National Natural Science Foundation of China (grant nos. 82371972 and 82572227).

Authors' contributions

J.G. conceived and designed the study, acquired clinical data, and drafted the manuscript. X.Q. developed the deep learning model and performed the statistical analysis. Y.L. and X.L. contributed to the statistical analysis. X.Z., Y.F., N.F., R.N., C.M. and X.C. acquired ultrasound data and participated in image interpretation. L.Z. and Z.W. supervised the study and critically revised the manuscript for important intellectual content. All authors reviewed and approved the final manuscript and agree to be accountable for all aspects of the work.

Acknowledgements

Special thanks to the participants who contributed ultrasound data for this study.

References

1. Greenberg SM, Ziai WC, Cordonnier C, et al. 2022 guideline for the management of patients with spontaneous intracerebral hemorrhage: a guideline from the American Heart Association/American Stroke Association. Stroke. 2022;53(7):e282–361. 10.1161/STR.0000000000000407.
2. Steiner T, Purrucker JC, Aguiar de Sousa D, et al. European Stroke Organisation (ESO) and European Association of Neurosurgical Societies (EANS) guideline on stroke due to spontaneous intracerebral haemorrhage. Eur Stroke J. 2025;10(4):1007–86. 10.1177/23969873251340815.
3. Seiffge DJ, Fandler-Höfler S, Du Y, et al. Intracerebral haemorrhage—mechanisms, diagnosis and prospects for treatment and prevention. Nat Rev Neurol. 2024;20(12):708–23. 10.1038/s41582-024-01035-w.
4. Meyer-Wiethe K, Sallustio F, Kern R. Diagnosis of intracerebral hemorrhage with transcranial ultrasound. Cerebrovasc Dis. 2009;27(Suppl 2):40–7. 10.1159/000203125.
5. Kukulska-Pawluczuk B, Książkiewicz B, Nowaczewska M. Imaging of spontaneous intracerebral hemorrhages by means of transcranial color-coded sonography. Eur J Radiol. 2012;81(6):1253–8. 10.1016/j.ejrad.2011.02.066.
6. Pinto-Villalba R, Paz A, Arellano J, et al. Transcranial ultrasonography in the detection of cerebrovascular accident: a systematic review and bivariate random-effects meta-analysis. Front Neurol. 2026;17:1726046. 10.3389/fneur.2026.1726046.

7. Allen BC, Kapoor S, Anzalone A, et al. Transcranial ultrasonography to detect intracranial pathology: a systematic review and meta-analysis. J Neuroimaging. 2023;33(3):333–58. 10.1111/jon.13087.
8. Fernandes RCL, Santos RP. Methodological standards for transcranial sonography of brain parenchyma. Rev Bras Neurol. 2025;61(1):12–8. 10.46979/rbn.v61i1.68004.
9. Miceli G, Basso MG, Cocciola E, Tuttolomondo A. Artificial intelligence in the diagnostic use of transcranial Doppler and sonography: a scoping review of current applications and future directions. Bioeng (Basel). 2025;12(7):681. 10.3390/bioengineering12070681.
10. Hu P, Yan T, Xiao B, et al. Deep learning-assisted detection and segmentation of intracranial hemorrhage in noncontrast computed tomography scans of acute stroke patients: a systematic review and meta-analysis. Int J Surg. 2024;110(6):3839–47. 10.1097/JS9.0000000000001266.
11. Karamian A, Seifi A. Diagnostic accuracy of deep learning for intracranial hemorrhage detection in non-contrast brain CT scans: a systematic review and meta-analysis. J Clin Med. 2025;14(7):2377. 10.3390/jcm14072377.
12. Simonetti R, Canals P, Gonzalez Riveros JD, et al. Feasibility of an AI-assisted transcranial duplex sonography protocol for early detection of intracerebral haemorrhage: the HYPER-AI-SCAN single-centre prospective study. BMJ Open. 2025;15(11):e102903. 10.1136/bmjopen-2025-102903.
13. Tan M, Le QV. EfficientNet: rethinking model scaling for convolutional neural networks. In: Proceedings of the 36th International Conference on Machine Learning. 2019;97:6105–6114.
14. Choi YH, Park JY, Lee SY, et al. Diagnostic performance of real-time artificial intelligence using deep learning analysis of endoscopic ultrasound videos for gallbladder polypoid lesions. Sci Rep. 2026;16(1):189. 10.1038/s41598-025-29179-9.
15. Tejani AS, Klontzas ME, Gatti AA, et al. Checklist for Artificial Intelligence in Medical Imaging (CLAIM): 2024 update. Radiol Artif Intell. 2024;6(4):e240300. 10.1148/ryai.240300.
16. Sounderajah V, Guni A, Liu X, STARD-AI Steering Committee, et al. The STARD-AI reporting guideline for diagnostic accuracy studies using artificial intelligence. Nat Med. 2025;31(10):3283–9. 10.1038/s41591-025-03953-8.
17. Loshchilov I, Hutter F. Decoupled weight decay regularization. 7th International Conference on Learning Representations (ICLR). 2019. arXiv:1711.05101.
18. Ilse M, Tomczak JM, Welling M. Attention-based deep multiple instance learning. In: Dy J, Krause A, eds. Proceedings of the 35th International Conference on Machine Learning. Vol 80. Proceedings of Machine Learning Research. PMLR; 2018:2127–2136.
19. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017:618–626. 10.1109/ICCV.2017.74
20. Obuchowski NA, Bullen J. Multireader diagnostic accuracy imaging studies: fundamentals of design and analysis. Radiology. 2022;303(1):26–34. 10.1148/radiol.211593.
21. Mäurer M, Shambal S, Berg D, et al. Differentiation between intracerebral hemorrhage and ischemic stroke by transcranial color-coded duplex-sonography. Stroke. 1998;29(12):2563–7.

10.1161/01.STR.29.12.2563.

22. Woydt M, Greiner K, Perez J, et al. Transcranial duplex-sonography in intracranial hemorrhage: evaluation of transcranial duplex-sonography in the diagnosis of spontaneous and traumatic intracerebral hemorrhage. Zentralblatt für Neurochirurgie. 1996;57(3):129–35.
23. Dinsmore M, Venkatraghavan L. Clinical applications of point-of-care ultrasound in brain injury: a narrative review. Anaesthesia. 2022;77(Suppl 1):69–77. 10.1111/anae.15604.
24. Kaffes M, Damian M, Davis SM, et al. Systematic review on noncomputed tomography-based technologies for prehospital stroke diagnosis. J Am Heart Assoc. 2026;15(17):e045588. 10.1161/JAHA.125.045588.
25. Bhuiyan MR, Deb S, Mitchell RA, Teddy PJ, Drummond KJ. The effect of formal training on the clinical utility of transcranial Doppler ultrasound monitoring in patients with aneurysmal subarachnoid haemorrhage. J Clin Neurosci. 2012;19(9):1255–60. 10.1016/j.jocn.2012.02.001.
26. Reschke P, Gruenewald LD, Koch V, et al. AI assistance improves radiology resident reader performance in CT diagnosis of intracranial hemorrhage. Radiol Med. 2026;131(2):280–91. 10.1007/s11547-025-02119-3.
27. Mabit L, Lepoittevin M, Valls M, et al. Real-life performance of a commercially available AI tool for post-traumatic intracranial hemorrhage detection on CT scans: a supportive tool. J Clin Med. 2025;14(13):4403. 10.3390/jcm14134403.
28. Forookhi A, Laschena L, Pecoraro M, et al. Bridging the experience gap in prostate multiparametric magnetic resonance imaging using artificial intelligence: a prospective multi-reader comparison study on inter-reader agreement in PI-RADS v2.1, image quality and reporting time between novice and expert readers. Eur J Radiol. 2023;161:110749. 10.1016/j.ejrad.2023.110749.
29. Kapoor S, Offnick A, Allen B, et al. Brain topography on adult ultrasound images: techniques, interpretation, and image library. J Neuroimaging. 2022;32(6):1013–26. 10.1111/jon.13031.
30. Sarwal A, Patel Y, D'Agostino R Jr, et al. Exploratory study to assess feasibility of intracerebral hemorrhage detection by point of care cranial ultrasound. Ultrasound J. 2022;14(1):40. 10.1186/s13089-022-00289-z.

Figures

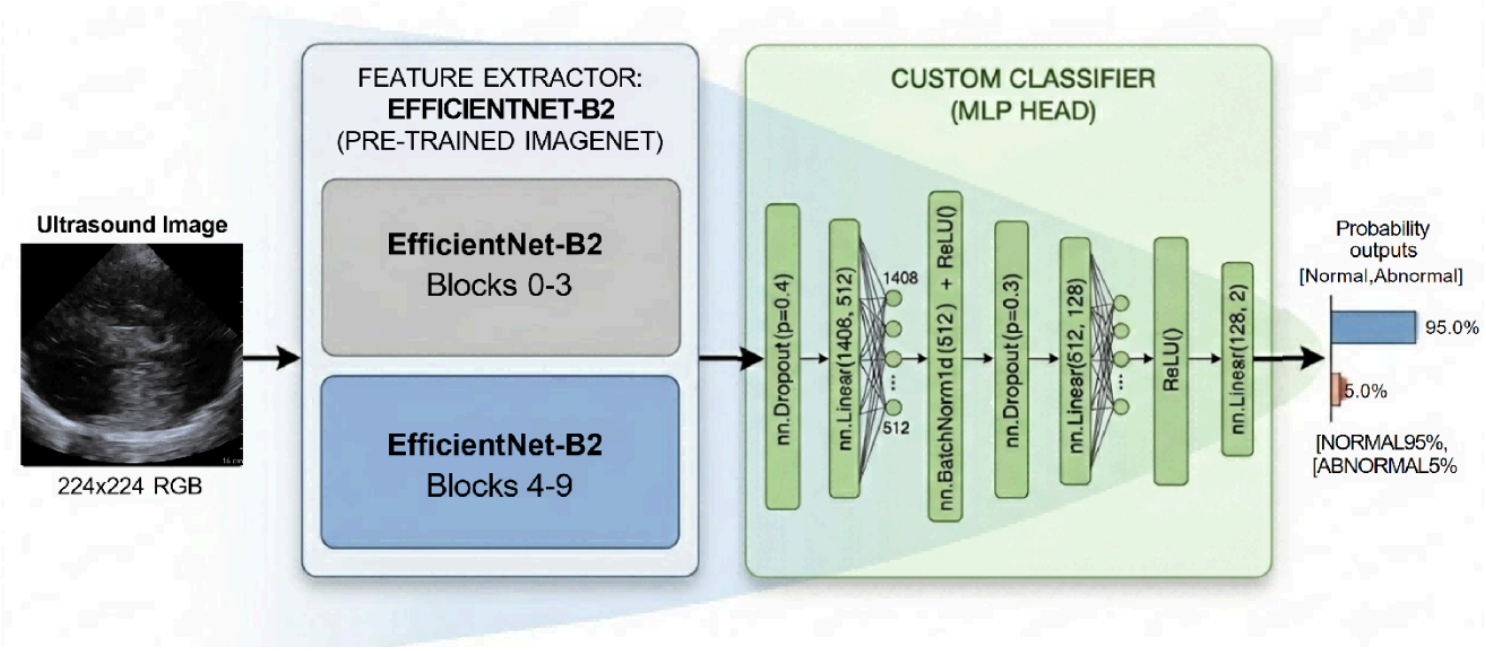


Figure 1

Architecture of the enhanced transfer learning model for ultrasound image classification based on EfficientNet-B2.

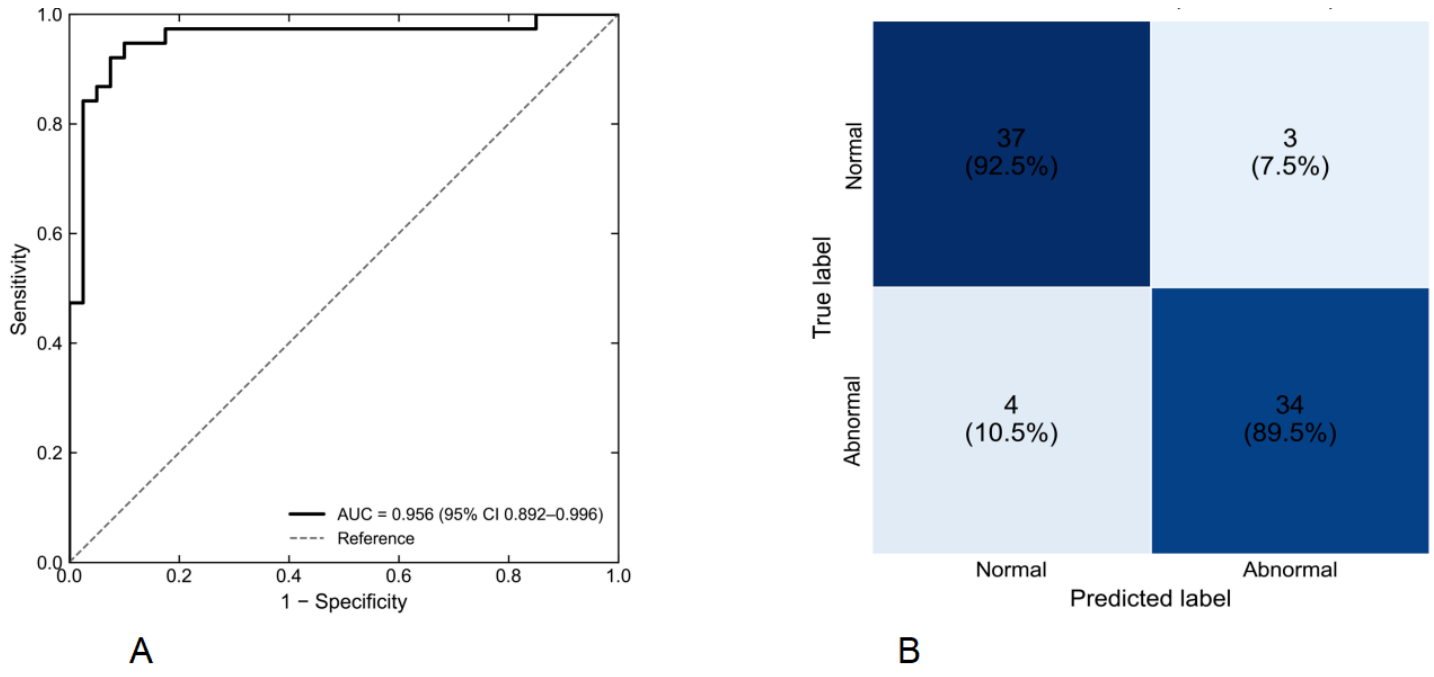


Figure 2

Receiver operating characteristic (ROC) curve and confusion matrix of the deep learning model. A, ROC curve. B, Confusion matrix.

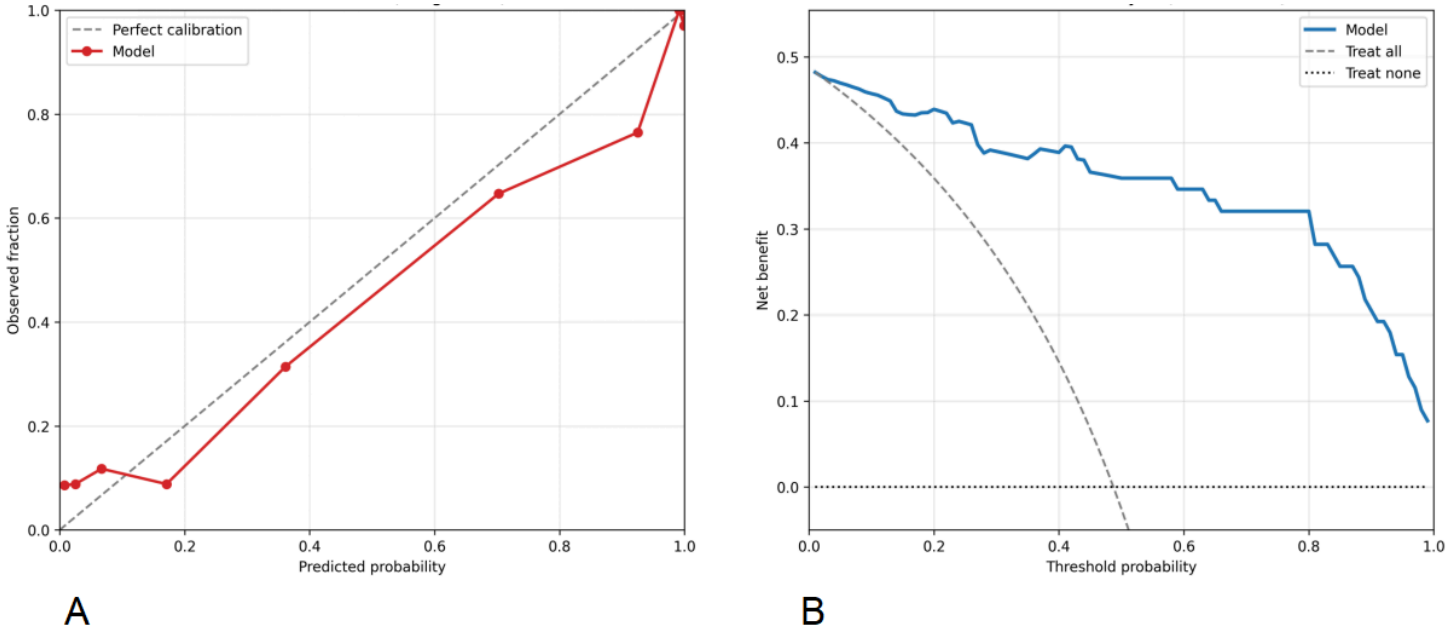


Figure 3

Calibration curve (A) and decision curve analysis (B).

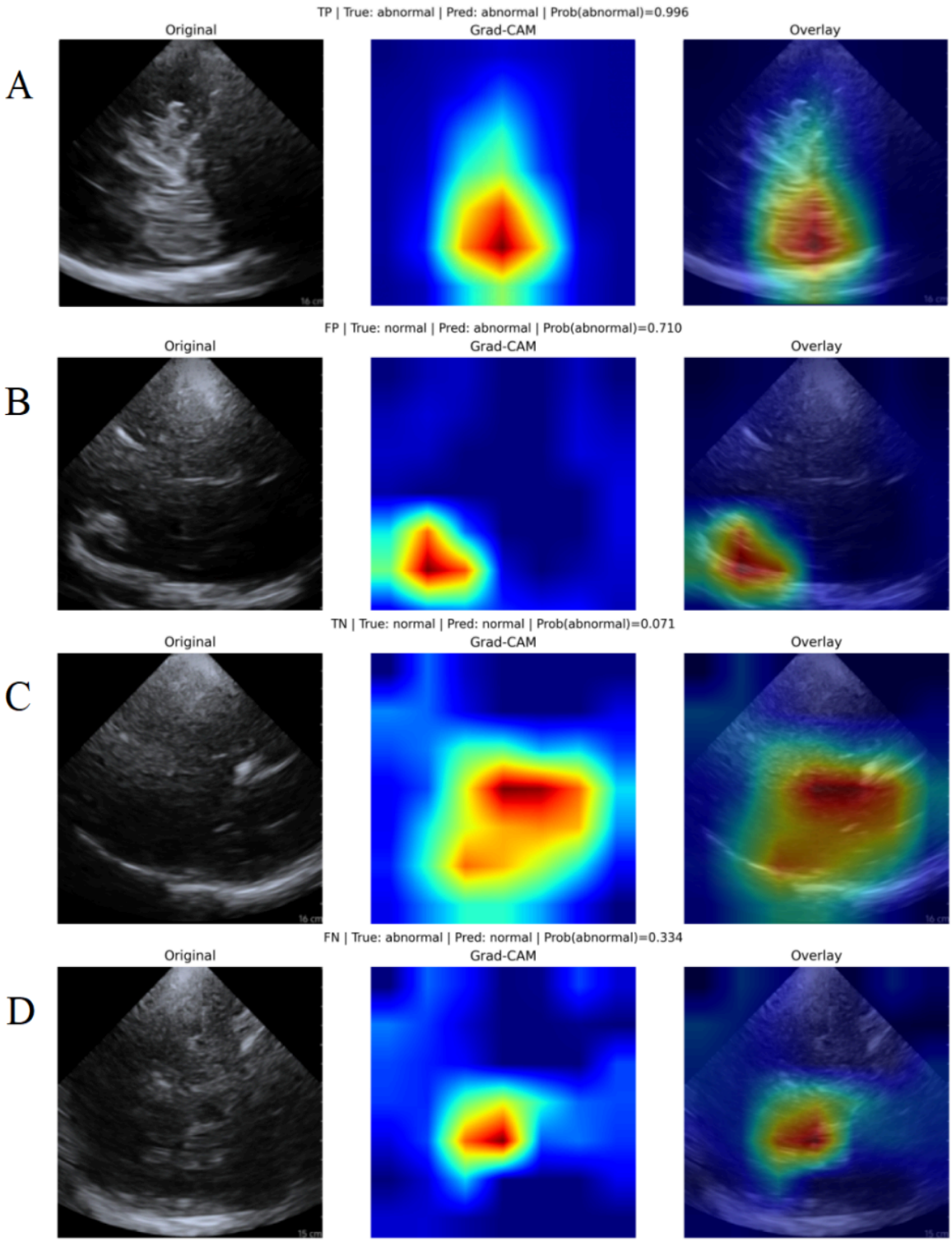


Figure 4

Grad-CAM visualizations of representative cases. Each triplet shows the original transcranial sonography image (left), Grad-CAM heat map (middle), and overlay (right). A, True positive: hemorrhage in the cerebral hemisphere contralateral to the probe. B, False positive: hyperechoic artifact of the contralateral petrous temporal bone. C, True negative: physiologic hyperechoic calcification. D, False negative: small contralateral subdural hemorrhage.