Article

Not peer-reviewed version

Deep Learning for Automated Intracranial Hemorrhage Detection in CT Imaging: A Narrative Literature Review

Annika Svensson *

Posted Date: 18 March 2026

doi: 10.20944/preprints202603.1484.v1

Keywords: intracranial hemorrhage; deep learning; computed tomography; medical image analysis; subtype classification; segmentation; clinical AI

Article

Deep Learning for Automated Intracranial Hemorrhage Detection in CT Imaging: A Narrative Literature Review

Annika Svensson

Mid Sweden University; annikasvensson.msu@gmail.com

Abstract

Intracranial hemorrhage (ICH) is a time-sensitive neurological emergency in which rapid detection can influence triage, specialist consultation, and treatment planning. Recent deep learning research has substantially advanced automated ICH detection on non-contrast head computed tomography (CT), moving from slice-level convolutional neural networks to sequence-aware, three-dimensional, segmentation-driven, and workflow-integrated systems. This narrative literature review synthesizes representative studies on automated ICH detection, subtype classification, lesion segmentation, external validation, and clinical implementation. The review is anchored by Kar et al. (2024), which presented a deep learning framework for automated intracranial hemorrhage detection in medical image analysis, and situates that paper within a broader body of work spanning seminal benchmarks, clinically oriented validation studies, and recent systematic reviews. The literature shows that deep learning models can achieve strong diagnostic performance, especially when trained on large annotated datasets and designed to exploit volumetric context. However, real-world deployment remains limited by dataset shift, class imbalance, low-prevalence screening environments, limited prospective multicenter evaluation, false-positive burden, and insufficient interpretability and governance mechanisms. The review concludes that the next stage of research should prioritize robust external validation, clinically meaningful reporting, uncertainty estimation, workflow-aware evaluation, and hybrid pipelines that unify detection, subtype classification, segmentation, and outcome-oriented triage.

Keywords: intracranial hemorrhage; deep learning; computed tomography; medical image analysis; subtype classification; segmentation; clinical AI

---

1. Introduction

Intracranial hemorrhage includes several clinically important subtypes—subdural, epidural, intraparenchymal, intraventricular, and subarachnoid hemorrhage—that often require rapid recognition on non-contrast head CT. In acute settings, delays in interpretation can affect neurosurgical escalation, stroke triage, and intensive monitoring decisions. Because CT remains the front-line imaging modality for suspected acute hemorrhage, it has become a major target for computer-aided diagnosis and triage systems (Chilamkurthy et al., 2018; Kuo et al., 2019).

Research in this area has progressed quickly. Early work focused on binary or multi-label hemorrhage recognition using convolutional neural networks (CNNs), while later studies incorporated recurrent neural networks (RNNs), three-dimensional context, attention mechanisms, ensemble strategies, and dense segmentation architectures to improve sensitivity and subtype recognition (Ye et al., 2019; Sage et al., 2020; Wang et al., 2021; Alis et al., 2022). More recent work has shifted toward external validation, workflow integration, reader-performance analysis, and meta-analytic synthesis of model accuracy and efficiency (Hu et al., 2024; Del Gaizo et al., 2024; Nada et al., 2025; Karamian et al., 2025).

 doi:10.20944/preprints202603.1484.v1

The present review is related directly to Kar et al. (2024), who described an automated intracranial hemorrhage detection framework using deep learning in medical image analysis. Their paper reflects a continuing trend toward practical architectures that combine feature extraction, localization support, and classification to improve ICH identification. Building from that contribution, this review examines how the field has developed, which model families have been most influential, what evidence exists for real-world use, and where important research gaps remain.

2. Review Approach

This paper is a narrative literature review rather than a formal systematic review. The goal is to synthesize representative and influential studies relevant to automated ICH detection on head CT, with special attention to architecture design, subtype classification, segmentation, validation, and clinical deployment. Priority was given to peer-reviewed journal and conference publications that are frequently cited or clinically informative, along with a small number of very recent studies that illuminate current translational directions.

The reviewed literature can be grouped into five broad streams: (1) foundational deep learning models for hemorrhage detection, (2) sequence-aware and three-dimensional models for subtype classification, (3) segmentation and quantification studies, (4) external validation and real-world implementation papers, and (5) review and meta-analysis papers that summarize cumulative performance. This structure helps place Kar et al. (2024) within the evolving research landscape rather than treating it as an isolated technical contribution.

3. Evolution of Deep Learning Architectures for ICH Detection

Foundational work established that deep learning could detect hemorrhage on head CT with clinically meaningful performance. Chilamkurthy et al. (2018) demonstrated that deep learning algorithms could identify critical findings on head CT, including intracranial hemorrhage and its subtypes, thereby framing AI as a triage tool rather than merely an academic classifier. Kuo et al. (2019) further accelerated the field by showing expert-level performance for acute ICH detection using a fully convolutional approach, helping set a high benchmark for sensitivity and clinical relevance.

A key shift soon followed: researchers moved beyond isolated 2D slice classification toward models that better capture inter-slice context. Ye et al. (2019) proposed a three-dimensional joint convolutional-recurrent framework for precise diagnosis of ICH and its subtypes. Sage et al. (2020) used a combined CNN-RNN strategy for subtype detection in head CT, reinforcing the value of sequential information. Wang et al. (2021) then introduced a pipeline that explicitly mimicked radiologists' reading behavior by coupling a 2D CNN with sequence models, thereby balancing computational efficiency with scan-level contextual reasoning.

Subsequent studies refined these ideas with attention mechanisms, ensembling, and stronger annotations. Alis et al. (2022) used a joint CNN-RNN architecture with attention to improve identification and classification, while Kang et al. (2023) showed that weighted ensembles trained on strongly annotated CT data could outperform several individual deep models. These developments indicate a clear pattern in the literature: performance improves not only through deeper networks, but also through better exploitation of volumetric structure, better labels, and model combinations that reduce single-model failure modes.

Kar et al. (2024) belongs to this lineage of detection-oriented work. Although presented as a conference paper, it reflects an important design direction in the field: using deep architectures that support efficient feature extraction and practical decision support for acute CT interpretation. Its relevance lies less in claiming an entirely new paradigm and more in reinforcing the translational momentum of automated ICH detection research.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 March 2026 doi:10.20944/preprints202603.1484.v1

Table 1. Representative studies in automated intracranial hemorrhage detection and analysis.

Year	Study	Technical focus	Main contribution	Key limitation
2018	Chilamkurthy et al.	Emergency head CT triage	Established large-scale deep learning for urgent abnormality detection including ICH.	Broad abnormality scope rather than ICH-only optimization.
2019	Kuo et al.	Expert-level acute ICH detection	Set an influential benchmark for acute hemorrhage detection.	Limited discussion of deployment workflow.
2019	Ye et al.	3D CNN + RNN	Improved subtype diagnosis by modeling volumetric context.	Higher complexity and data demands.
2021	Wang et al.	2D CNN + sequence models	Mimicked radiologist reading order for scan-level classification.	Retrospective emphasis.
2021	Xu et al.	Dense U-Net segmentation	Enabled automated hemorrhage quantification and volume estimation.	Focused on segmentation more than workflow.
2022	Alis et al.	CNN-RNN with attention	Combined strong classification performance with interpretability support.	Performance drop in prospective use remained possible.
2023	Kang et al.	Strong labels + ensembles	Showed value of high-quality annotations and weighted ensembles.	Still dependent on curated data.
2024	Kar et al.	Detection-oriented deep learning framework	Conference contribution related to practical automated ICH detection.	Limited external validation and conference-scale scope.
2024	Del Gaizo et al.	National teleradiology deployment	Highlighted efficiency trade-offs in low-prevalence settings.	Operational results may depend on local workflow.
2025	Kang et al.	Reader-performance validation	Demonstrated improvement in non-expert reader performance.	Single-system evaluation.

4. Subtype Classification, Segmentation, and Quantification

Binary detection is clinically useful, but subtype classification and lesion delineation often matter more for management. Epidural hemorrhage, subdural hemorrhage, subarachnoid hemorrhage, intraparenchymal hemorrhage, and intraventricular hemorrhage differ in prognosis, urgency, and likely intervention. For that reason, multiple studies have targeted fine-grained subtype recognition rather than only presence-versus-absence detection (Ye et al., 2019; Sage et al., 2020; Wang et al., 2021).

Segmentation research addresses another clinically relevant need: hematoma localization and volume estimation. Xu et al. (2021) used a Dense U-Net framework for automatic segmentation and quantification of intracranial hemorrhage on CT, showing strong agreement with expert reference annotations. Sharrock et al. (2021) demonstrated that three-dimensional deep neural networks can segment intracerebral hemorrhage and associated ventricular extension with useful anatomical context. D'Angelo et al. (2024) reported strong diagnostic performance and marked time efficiency from a Dense U-Net-based pipeline. Hu et al. (2024) reached a similar conclusion in their meta-analysis, arguing that deep learning systems for detection and segmentation can approach expert performance while reducing processing time.

The literature suggests that segmentation adds value in at least three ways. First, it provides localization evidence that can make algorithmic decisions easier to verify visually. Second, it supports clinically meaningful measurements such as volume and mass effect. Third, it creates a bridge between diagnostic AI and downstream prognostic tasks, including hematoma expansion prediction and treatment planning. Thus, future systems related to Kar et al. (2024) would likely be stronger if they moved beyond detection alone and incorporated lesion-level outputs.

 doi:10.20944/preprints202603.1484.v1

5. External Validation and Workflow Integration

One of the most important themes in recent literature is the gap between retrospective benchmark performance and real-world utility. Several externally validated or workflow-integrated studies show that apparently strong algorithms may behave very differently in routine practice. Kundisch et al. (2021) evaluated a deep learning algorithm on emergency CTs, while Voter et al. (2021) emphasized diagnostic accuracy together with failure mode analysis. These papers underscore that performance metrics alone are insufficient; understanding when models fail is equally important.

Clinical validation studies provide a more nuanced picture. Wang et al. (2023) reported real-world validation of an AI-based CT hemorrhage detection tool, suggesting potential benefits for worklist prioritization. Choi et al. (2024) examined the effect of a deep learning-based interpretation aid on emergency decision-making and found that less experienced users may gain sensitivity, but also become more vulnerable to false positives. Del Gaizo et al. (2024) studied deployment in a national teleradiology program and showed that false-positive flags can lengthen interpretation time in low-prevalence settings, raising a critical operational issue: a model can be statistically accurate yet still reduce efficiency if its positive predictive value is low in practice.

Additional evidence from external validation continues this pattern. Nada et al. (2025) found satisfactory performance on a heterogeneous real-world dataset, but the need for analysis across patient risk factors and clinical settings remained evident. Kang et al. (2025) reported that deep learning assistance improved reader performance, especially among non-expert readers. Fang et al. (2025) further indicated that post-deployment surveillance and confidence monitoring may become standard requirements in high-stakes medical imaging AI. Together, these findings suggest that deployment success depends not only on area under the curve, but also on prevalence, calibration, user trust, review burden, uncertainty handling, and how the system fits radiology operations.

6. Evidence from Reviews and Meta-Analyses

Review papers increasingly show that the field has matured enough for higher-level synthesis. Hu et al. (2024) concluded that deep learning systems for ICH detection, segmentation, and quantification often perform comparably to experienced clinicians while reducing processing time. Karamian et al. (2025) reported promising pooled accuracy for deep learning-based ICH detection on non-contrast CT and noted the need for more prospective evidence. These review papers are important because they reveal recurrent limitations across individual studies.

Common issues include retrospective single-center designs, inconsistent ground-truth procedures, insufficient subgroup reporting, weak comparison to real clinical baselines, limited external validation, and selective reporting of favorable metrics. The broader literature therefore supports a cautious interpretation of strong headline results. While many studies report high sensitivity, specificity, or area under the curve, fewer investigate the practical consequences of false alarms, calibration drift, data heterogeneity, or clinically meaningful outcomes such as reduced turnaround time, earlier intervention, or improved patient prognosis.

7. Key Challenges and Research Gaps

Several challenges remain unresolved. The first is dataset shift. Models trained on one mix of scanners, reconstruction kernels, institutions, and labeling conventions may not generalize cleanly to another environment. This problem is visible in the growing number of external validation papers and in deployment-oriented studies that report performance degradation relative to development datasets (Wang et al., 2023; Del Gaizo et al., 2024; Nada et al., 2025).

The second challenge is annotation quality. Binary labels are easier to obtain, but subtype classification and segmentation require more expensive expert labeling. Kang et al. (2023) showed that strongly annotated datasets can substantially improve model performance. This finding implies that future work should invest in better curation pipelines, semi-supervised learning, and active-learning strategies rather than simply scaling network depth.

Preprints.org (www.preprints.org) | NOT PEER-REVIEWED | Posted: 18 March 2026 doi:10.20944/preprints202603.1484.v1

A third challenge is explainability and uncertainty. Heatmaps and segmentation overlays help, but they do not automatically guarantee trustworthy decisions. Reader-assistance studies imply that clinicians can be influenced by incorrect model outputs, especially when they are less experienced (Choi et al., 2024). This makes calibrated uncertainty estimation and failure detection especially important. Fang et al. (2025) addressed this issue by proposing real-time monitoring for ICH AI systems, indicating that post-deployment surveillance may become a standard requirement in high-stakes medical imaging AI.

A fourth challenge concerns evaluation design. Many papers emphasize classification metrics but underreport prevalence, positive predictive value in real-world settings, error taxonomy, subgroup robustness, or workflow burden. Del Gaizo et al. (2024) demonstrated why that matters: a tool with reasonable sensitivity and specificity can still generate operational inefficiency when deployed at scale in a low-prevalence environment. Future work should therefore report not only traditional machine-learning metrics, but also user-centered and system-level outcomes.

Finally, there remains a translation gap between detection and action. In practice, clinicians need more than a binary alert. They need reliable subtype identification, lesion localization, volume estimation, confidence information, and prioritization logic that fits emergency and neuroradiology workflows. This is a major opportunity for extending work such as Kar et al. (2024).

8. Future Research Agenda

The next phase of research should move toward integrated, clinically accountable systems. First, future models should combine detection, subtype classification, and segmentation in unified multi-task frameworks so that a single system can support both triage and structured reporting. Second, studies should prioritize multicenter prospective validation with transparent case prevalence and workflow analysis. Third, uncertainty estimation and continuous monitoring should become standard components of deployment, especially in high-volume environments where false positives may create downstream costs.

Fourth, more attention should be given to fairness, robustness, and data governance. Even when demographic bias is not the primary focus, models should be tested across age groups, scanner vendors, trauma versus non-trauma cohorts, and postoperative or artifact-heavy cases. Fifth, benchmark studies should be supplemented by reader studies and implementation analyses that measure whether AI actually improves decision quality, turnaround time, or patient care rather than only retrospective diagnostic accuracy.

For researchers building on Kar et al. (2024), a strong follow-up agenda would include external validation on heterogeneous head CT datasets, subtype-aware lesion localization, segmentation-guided explainability, and comparison against commercially deployed systems. Such work would position the original detection framework within the more clinically mature direction the literature is now taking.

9. Conclusions

Automated intracranial hemorrhage detection has evolved from promising CNN classifiers to increasingly sophisticated clinical AI systems that incorporate sequence modeling, segmentation, ensembling, reader support, and real-world validation. The literature consistently shows that deep learning can support rapid and accurate ICH recognition on head CT, and papers such as Kar et al. (2024) contribute to this growing translational body of work. At the same time, the field has learned that strong retrospective accuracy does not guarantee practical success. Generalizability, alert burden, explainability, annotation quality, and workflow fit all shape whether a model becomes clinically useful.

Overall, the most promising future direction is not a single better classifier, but a robust end-to-end decision-support ecosystem: one that detects hemorrhage, identifies subtype, localizes the lesion, quantifies burden, expresses uncertainty, and performs reliably across institutions. In that broader

 doi:10.20944/preprints202603.1484.v1

context, literature reviews such as this can help clarify how individual technical papers relate to the real requirements of deployable medical AI.

References

Agrawal, D., Poonamallee, L., Joshi, S., & Bahel, V. (2023). Automated intracranial hemorrhage detection in traumatic brain injury using 3D CNN. Journal of Neurosciences in Rural Practice, 14, 615–621. https://doi.org/10.25259/JNRP_172_2023

Alis, D., Alis, C., Yergin, M., Topel, C., Asmakutlu, O., Bagcilar, O., Senli, Y. D., Ustundag, A., Salt, V., Dogan, S. N., et al. (2022). A joint convolutional-recurrent neural network with an attention mechanism for detecting intracranial hemorrhage on noncontrast head CT. Scientific Reports, 12, 2084. https://doi.org/10.1038/s41598-022-05872-x

Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N. G., Venugopal, V. K., Mahajan, V., Rao, P., Warier, P., & others. (2018). Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study. The Lancet, 392(10162), 2388–2396.

Choi, S. Y., et al. (2024). Impact of a deep learning-based brain CT interpretation algorithm on clinical decision-making for intracranial hemorrhage in the emergency department. Scientific Reports, 14, 21929.

Cortés-Ferre, L., Gutiérrez-Naranjo, M. A., Egea-Guerrero, J. J., Pérez-Sánchez, S., & Balcerzyk, M. (2023). Deep learning applied to intracranial hemorrhage detection. Journal of Imaging, 9(2), 37. https://doi.org/10.3390/jimaging9020037

Del Gaizo, A. J., et al. (2024). Deep learning to detect intracranial hemorrhage in a national teleradiology program and the impact on interpretation time. Radiology: Artificial Intelligence, 6(4), e240067.

D'Angelo, T., Bucolo, G. M., Kamareddine, T., Yel, I., Koch, V., Gruenewald, L. D., Martin, S., Alizadeh, L. S., Mazziotti, S., Blandino, A., Vogl, T. J., & Booz, C. (2024). Accuracy and time efficiency of a novel deep learning algorithm for intracranial hemorrhage detection in CT scans. La Radiologia Medica, 129(10), 1499–1506. https://doi.org/10.1007/s11547-024-01867-y

Fang, Z., Johnston, A., Cheuy, L., et al. (2025). Automated real-time assessment of intracranial hemorrhage detection AI using an ensembled monitoring model (EMM). NPJ Digital Medicine, 8(1), 608. https://doi.org/10.1038/s41746-025-02007-0

Hu, P., Yan, T., Xiao, B., et al. (2024). Deep learning-assisted detection and segmentation of intracranial hemorrhage in noncontrast computed tomography scans of acute stroke patients: A systematic review and meta-analysis. International Journal of Surgery, 110, 3839–3847.

Kar, N. K., Jana, S., Rahman, A., Ashokrao, P. R., & Mangai, R. A. (2024, July). Automated intracranial hemorrhage detection using deep learning in medical image analysis. In 2024 International Conference on Data Science and Network Security (ICDSNS) (pp. 1–6). IEEE. DOI: 10.1109/ICDSNS62112.2024.10691276

Kang, D.-W., Park, G.-H., Ryu, W.-S., Schellingerhout, D., Kim, M., Kim, Y. S., Park, C.-Y., Lee, K.-J., Han, M.-K., Jeong, H.-G., & Kim, D.-E. (2023). Strengthening deep-learning models for intracranial hemorrhage detection: Strongly annotated computed tomography images and model ensembles. Frontiers in Neurology, 14, 1321964. https://doi.org/10.3389/fneur.2023.1321964

Kang, D. W., Kim, M., Park, G. H., Kim, Y. S., Han, M. K., Lee, M., Kim, D., Ryu, W. S., & Jeong, H. G. (2025). Deep learning-assisted detection of intracranial hemorrhage: Validation and impact on reader performance. Neuroradiology, 67(6), 1511–1519. https://doi.org/10.1007/s00234-025-03560-x

Karamian, A., et al. (2025). Diagnostic accuracy of deep learning for intracranial hemorrhage detection in non-contrast computed tomography: A systematic review and meta-analysis. Journal of Clinical Medicine, 14(7), 2377.

Kundisch, A., Hönning, A., Mutze, S., Kreissl, L., Spohn, F., Lemcke, J., et al. (2021). Deep learning algorithm in detecting intracranial hemorrhages on emergency computed tomographies. PLoS ONE, 16(11), e0260560. https://doi.org/10.1371/journal.pone.0260560

Kuo, W., Häne, C., Mukherjee, P., Malik, J., & Yuh, E. L. (2019). Expert-level detection of acute intracranial hemorrhage on head computed tomography using deep learning. Proceedings of the National Academy of Sciences, 116(45), 22737–22745. https://doi.org/10.1073/pnas.1908021116

doi:10.20944/preprints202603.1484.v1

Nada, A., Sayed, A. A., Hamouda, M., Tantawi, M., Khan, A., Alt, A., Hassanein, H., Sevim, B. C., Altes, T., & Gaballah, A. (2025). External validation and performance analysis of a deep learning-based model for the detection of intracranial hemorrhage. The Neuroradiology Journal, 38(3), 312–321. https://doi.org/10.1177/19714009241303078

Sage, A., Badura, P., Szymura, J., & van Rijn, R. R. (2020). Intracranial hemorrhage detection in head CT using combined CNN-RNN. Applied Sciences, 10(21), 7577. https://doi.org/10.3390/app10217577

Sharrock, M. F., Mould, W. A., Ali, H., Hildreth, M., Miller, D. E., Qian, G., Sun, Y., Hanley, D. F., & Carass, A. (2021). 3D deep neural network segmentation of intracerebral hemorrhage: Development and validation for clinical trials. Stroke, 52(6), 1984–1992.

Voter, A. F., Meram, E., Garrett, J. W., & Yu, J.-P. J. (2021). Diagnostic accuracy and failure mode analysis of a deep learning algorithm for the detection of intracranial hemorrhage. Journal of the American College of Radiology, 18(8), 1143–1152. https://doi.org/10.1016/j.jacr.2021.03.005

Wang, D., Jin, R., Shieh, C.-C., Ng, A. Y., Pham, H., Dugal, T., Barnett, M., Winoto, L., Wang, C., & Barnett, Y. (2023). Real world validation of an AI-based CT hemorrhage detection tool. Frontiers in Neurology, 14, 1177723. https://doi.org/10.3389/fneur.2023.1177723

N. Kumar Kar, S. Jana, A. Rahman, P. Rahul Ashokrao, I. G and R. Alarmelu Mangai, "Automated Intracranial Hemorrhage Detection Using Deep Learning in Medical Image Analysis," 2024 International Conference on Data Science and Network Security (ICDSNS), Tiptur, India, 2024, pp. 1-6, doi: 10.1109/ICDSNS62112.2024.10691276.

Wang, H.-C., Wang, S.-C., Xiao, F., Ho, U.-C., Lee, C.-H., Yan, J.-L., Chen, Y.-F., & Ko, L.-W. (2025). Development of a clinically applicable deep learning system based on sparse training data to accurately detect acute intracranial hemorrhage from non-enhanced head computed tomography. Neurologia Medico-Chirurgica, 65(3), 103–112. https://doi.org/10.2176/jns-nmc.2024-0163

Wang, X., Shen, T., Yang, S., Lan, J., Xu, Y., Wang, M., et al. (2021). A deep learning algorithm for automatic detection and classification of acute intracranial hemorrhages in head CT scans. NeuroImage: Clinical, 32, 102785. https://doi.org/10.1016/j.nicl.2021.102785

Xu, J., Zhang, R., Zhou, Z., Wu, C., Gong, Q., Zhang, H., Wu, S., Wu, G., Deng, Y., Xia, C., & Ma, J. (2021). Deep network for the automatic segmentation and quantification of intracranial hemorrhage on CT. Frontiers in Neuroscience, 14, 541817. https://doi.org/10.3389/fnins.2020.541817

Ye, H., Gao, F., Yin, Y., Guo, D., Zhao, P., Lu, Y., Bai, J., Cao, K., Song, Q., Wang, X., Chen, W., Guo, X., & Xia, J. (2019). Precise diagnosis of intracranial hemorrhage and subtypes using a three-dimensional joint convolutional and recurrent neural network. European Radiology, 29(11), 6191–6201. https://doi.org/10.1007/s00330-019-06163-2