Acta Neurochirurgica (2026) 168:200
https://doi.org/10.1007/s00701-026-06954-9

RESEARCH

An agentic AI framework for integrated decision support and surgical planning in intracerebral hemorrhage

Eugen Kochuiev[3] · Vladyslav Kaliuzhka[1,2] · Mykyta Markevych[2] · Zoia Kochuieva[3] · Volodymyr Piatykop[1,2]

Received: 21 February 2026 / Accepted: 9 June 2026
© The Author(s) 2026

Abstract

Background Intracerebral hemorrhage (ICH) remains associated with high mortality and treatment variability. Current workflows rely on fragmented imaging interpretation and operator-dependent surgical planning.

Methods The objective was to develop and validate an agentic artificial intelligence (AI) framework integrating automated imaging analysis, guideline-based reasoning, and trajectory optimization for ICH treatment. Fifty consecutive computed tomography (CT) and computed tomography angiography (CTA) datasets from patients with spontaneous ICH were retrospectively analyzed. The system performed multi-class anatomical segmentation of skin, skull, brain, ventricles, and hematoma, followed by volumetric quantification and JavaScript Object Notation (JSON) based structured encoding of imaging biomarkers. A knowledge-based module incorporating international ICH guidelines generated risk stratification and treatment recommendations. When evacuation was indicated, an automated trajectory modeling module proposed a patient-specific minimally invasive surgical corridor.

Results Overall agreement between AI-generated and expert treatment recommendations was 82% (41/50 cases), with substantial agreement beyond chance (Cohen's $\kappa = 0.71$). Discrepancies occurred primarily in borderline surgical indication scenarios. In evacuation candidates, the automated planner generated feasible trajectories in all 50 cases. Median angular deviation between AI-generated and expert-defined trajectories was 7.6°, interquartile range (IQR) 5.1–9.8°. AI-generated trajectories demonstrated equal or greater safety margins relative to expert planning in the majority of cases. End-to-end processing has a potential to substantially reduce simulated decision-support time compared with manual workflow.

Conclusion The proposed agentic AI framework enables structured, explainable, and workflow-integrated decision support for ICH management. This system may reduce operator variability and enhance precision in minimally invasive evacuation planning.

Keywords Automatic planning · Artificial intelligence · Deep learning · Intracerebral hemorrhage · Agentic AI

Abbreviations

AI	Artificial intelligence
CT	Computed tomography
CTA	Computed tomography angiography
ICH	Intracerebral hemorrhage
IQR	Interquartile range
AI	Artificial intelligence
CI	Confidence interval
CT	Computed tomography
CTA	Computed tomography angiography

✉ Vladyslav Kaliuzhka
vy.kaliuzhka@knmu.edu.ua

Eugen Kochuiev
eugen@kochuev.com

Mykyta Markevych
mykyta.markevych@gmail.com

Zoia Kochuieva
aliseiko@gmail.com

Volodymyr Piatykop
vo.pyatikop@knmu.edu.ua

[1] Department of Neurology and Neurosurgery, Kharkiv National Medical University, Kharkiv, Ukraine

[2] Neurosurgery Department, Communal Non-Commercial Enterprise of the Kharkiv Regional Council "Regional Clinical Hospital", Kharkiv, Ukraine

[3] National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine

Published online: 23 June 2026

DICOM	Digital Imaging and Communications in Medicine
EVD	External ventricular drainage
HU	Hounsfield unit
ICH	Intracerebral hemorrhage
JSON	JavaScript Object Notation
GCS	Glasgow Coma Scale
LLM	Large language model
MLS	Midline shift
mRS	Modified Rankin Scale
NIHSS	National Institutes of Health Stroke Scale
nnU-Net	No-new-U-Net
PCA	Principal component analysis
SD	Standard deviation

Introduction

Intracerebral hemorrhage (ICH) remains one of the most devastating forms of stroke, constituting a neurosurgical emergency associated with high mortality, severe long-term disability, and substantial socioeconomic burden. Contemporary epidemiological data report a 30-day case fatality rate of 38.2% (95% confidence interval (CI) 35.1%–41.5%), with 52.7% of 12-month survivors remaining severely functionally dependent (modified Rankin Scale score ≥ 4). The global economic impact is equally profound, with direct medical costs exceeding US$15.6 billion annually in high-income economies alone [1].

Despite decades of technical refinement in microsurgical and minimally invasive evacuation techniques, persistent controversy remains regarding optimal surgical indications [6]. Recent trials such as Early MiNimally-invasive Removal of IntraCerebral Hemorrhage (ENRICH) trial (NCT02880878) and Minimally Invasive Surgeries for spontaneous hypertensive IntraCerebral Hemorrhage (MISICH) trial (NCT02811614) have demonstrated potential functional benefits of early minimally invasive evacuation in selected patients, yet uncertainty persists regarding patient selection, volumetric thresholds, hematoma morphology, and trajectory optimization [7, 11]. The pathoanatomical complexity of ICH – including multilobulated architecture, septations, and deep extension into basal ganglia or thalamic structures – further complicates surgical planning and risk assessment.

A critical determinant of treatment strategy is hematoma volume estimation; however, variability in segmentation techniques, operator experience, and imaging interpretation contributes to inconsistent volumetric assessment associated potential time loss and the cognitive burden placed on surgeons. Prior semiautomated workflows using platforms such as 3D Slicer have demonstrated feasibility, but they require technical expertise and manual interaction that limit scalability and reproducibility – particularly in resource-limited settings where the majority of global ICH cases are treated [11]. Also manual planning in emergency settings often occurs under extreme time pressure, which exacerbates inter-operator variability.

Emerging artificial intelligence (AI) systems have achieved high performance in automated hemorrhage detection and segmentation, with reported diagnostic accuracies exceeding 94% [3]. Nevertheless, most existing AI applications remain confined to image analysis tasks [9]. Their role in procedural guidance, surgical trajectory optimization, or integrated clinical reasoning remains limited [12]. To date, no system provides end-to-end decision support that integrates volumetric analysis, anatomical safety assessment, trajectory planning, and risk stratification into a unified, automated framework.

The concept of agentic AI – defined as autonomous, goal-directed systems capable of multistep reasoning, adaptive planning, and coordinated task execution – offers a paradigm shift beyond passive segmentation models. Rather than functioning as isolated prediction tools, agentic systems can orchestrate sequential modules, incorporate domain-specific constraints, and generate actionable surgical strategies.

In this study, we present and validate an agentic AI framework for ICH management that functions as an augmented intelligence system, integrating automated hematoma segmentation, image reorientation, safety zone delineation, risk stratification, and trajectory optimization into a unified clinical decision support platform, supporting neurosurgeons rather than replacing clinical decision-making. Our objective is to develop an end-to-end reasoning architecture that reduces operator dependence, standardizes preoperative planning, and advances precision neurosurgical care for patients with intracerebral hemorrhage.

Materials and methods

This retrospective imaging study included 50 cases with spontaneous intracerebral hemorrhage (ICH) who underwent noncontrast computed tomography (CT) and CT angiography (CTA) at a Communal Non-Commercial Enterprise of the Kharkiv Regional Council "Regional Clinical Hospital" (Kharkiv, Ukraine) during 2025. The local Ethics and Bioethics Committee approved study. All imaging data and patient personal data were anonymized prior to analysis. CT and CTA examinations were performed using two clinical CT systems (SOMATOM X.ceed scanner (Siemens Healthineers, Germany), and Optima CT540 scanner (GE Healthcare, USA).

Inclusion criteria were adults (≥ 18 years) with spontaneous supratentorial and infratentorial ICH, available CTA at admission, complete Digital Imaging and Communications in Medicine (DICOM) imaging dataset suitable for 3D

reconstruction, slice thickness ≤ 2.5 mm. Exclusion criteria were traumatic intracranial hemorrhage, postoperative or secondary hemorrhage, severe motion artifacts impairing segmentation.

Clinical variables include age, National Institutes of Health Stroke Scale (NIHSS) score, Glasgow Coma Scale (GCS) score, Modified Rankin Scale (mRS) score, anticoagulant use, blood pressure at admission, time from symptom onset.

Clinical outcomes: actual management performed, discharge disposition. Clinical data were de-identified prior to analysis.

Radiological characteristics

DICOM datasets (CT and CTA) were uploaded through a secure web-based interface. DICOM parsing and metadata extraction were performed using the pydicom library, enabling validation of acquisition parameters and consistency across studies. The preprocessing pipeline included CT/CTA co-registration, followed by isotropic resampling to $1 \times 1 \times 1$ mm^3 voxel resolution using trilinear interpolation, selected to preserve spatial continuity and minimize boundary artifacts relevant for hematoma delineation. To avoid information loss associated with global downsampling, volumes were processed in a foreground-aware manner with automatic cropping of non-informative regions, enabling preservation of native spatial detail during subsequent analysis. All images were reoriented to a standard anatomical coordinate system (RAS convention), and intensity values were normalized using a brain window of − 100 to 200 Hounsfield units (HU).

AI agent architecture

The proposed agentic framework consists of three tightly coupled modules: (1) multi-class anatomical segmentation with quantitative biomarker extraction, (2) guideline-driven clinical reasoning, and (3) patient-specific trajectory optimization (Fig. 1). The architecture is designed to be end-to-end, traceable, and modular – each stage produces structured outputs that are consumed by the next stage, enabling reproducible decision-making and surgical planning. Everything is implemented within a custom web-based platform (eXtraVision LLC, Ukraine) integrating all three modules into a unified pipeline. The system provides secure data storage and built-in anonymization of patient information, ensuring compliance with data protection requirements. This architecture enables simultaneous processing, seamless data flow between modules, and standardized outputs, supporting traceable and reproducible results across different users and clinical settings.

Orchestration is implemented on LangGraph (LangChain), a framework for stateful, multi-actor LLM workflows. A typed shared-state object is updated by each node and read by the controller, which selects the next node or tool at runtime based on intermediate results. This design deliberately combines LLM-based reasoning with deterministic execution: the LLM is used where flexible, context-sensitive interpretation is required (e.g., synthesizing imaging biomarkers, clinical features, and guideline evidence into a recommendation), while quantitative steps — volumetric and morphometric computation, risk-score calculation (ICH score, FUNC score), EVD-indication checking, and trajectory optimization — are encapsulated as deterministic, validated tools with structured inputs and outputs. The LLM invokes these tools through a ReAct-style loop and consumes their outputs, but does not reproduce their computations, ensuring that numerical results are reproducible and auditable across runs and users.

The execution graph is case-specific. Cases meeting conservative-management criteria terminate without entering the surgical-planning subgraph; cases with intraventricular extension or hydrocephalus trigger an EVD-evaluation branch; cases meeting evacuation criteria enter the trajectory-planning subgraph with parameters (target, entry-corridor constraints, eloquent-structure and vascular avoidance masks) configured dynamically from prior module outputs. Each structured output is schema-validated and checked against clinical-plausibility constraints (volume consistency, laterality concordance, guideline-citation grounding) before being passed downstream; on

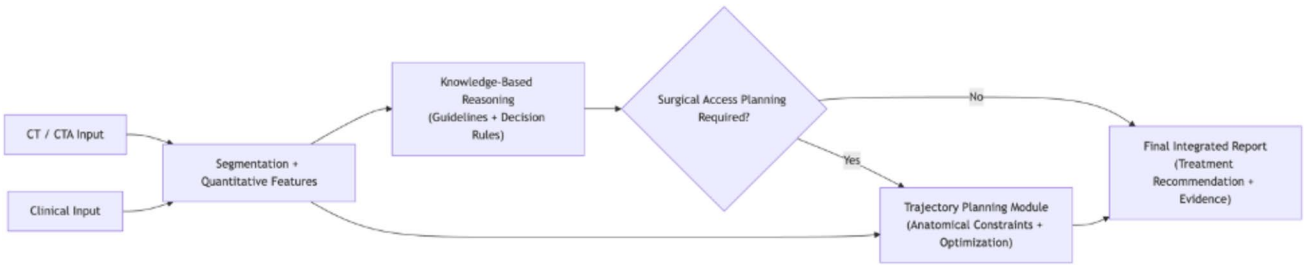


Fig. 1 Framework high-level architecture

failure, the relevant sub-agent is re-invoked with corrective context.

The reasoning LLM is deployed within Microsoft Azure AI Foundry (formerly Azure OpenAI Service) under an enterprise tenant with zero data retention, regional data residency, encryption in transit and at rest, and role-based access control; patient inputs are anonymized at ingestion prior to any LLM invocation.

CT segmentation and quantitative features extraction module

Automated multistructure segmentation constituted the perception layer of the proposed agentic AI framework; however, segmentation itself was not the primary objective. Its role was to enable reproducible extraction of quantitative imaging biomarkers required as structured inputs for the downstream reasoning module. A customized 3D encoder–decoder convolutional neural network derived from the no-new-U-Net (nnU-Net) paradigm was trained for volumetric CT/CTA data to segment skin, skull, brain parenchyma, ventricular system, intracerebral hematoma, and intracranial vascular structures (CTA when available). All segmentations were resampled to 1-mm isotropic resolution to ensure geometric consistency across cases.

From these patient-specific segmentations, the system automatically derived a panel of quantitative imaging features aligned with clinically actionable decision variables in ICH. Hematoma characterization included total volume calculated by voxel summation, centroid coordinates in patient space, laterality relative to the segmented mid-sagittal plane, anatomical compartment involvement based on overlap with segmented brain structures, maximum axial and craniocaudal diameters, geometric descriptors such as sphericity, surface-to-volume ratio, and elongation index. Intraventricular extension was quantified directly from spatial intersection between hematoma and ventricular masks and encoded both as a binary variable and as absolute intraventricular hemorrhage volume.

To estimate mass effect and secondary injury, the framework computed midline shift (MLS) in mm, as displacement of the septum pellucidum from the skull-defined midline, total ventricular volume and linear ventricular indices to detect hydrocephalus, and minimum Euclidean distances between the hematoma and critical structures including the ventricular system, cortical surface, and major vascular structures in CTA cases. Perihematomal edema was not explicitly segmented and was therefore not incorporated as an independent volumetric input variable in the current system, reflecting the intended scope of the framework, which is focused on hyper-acute decision-making.

All imaging-derived measurements, together with structured clinical variables such as age, Glasgow Coma Scale, NIHSS, blood pressure, anticoagulation status, and onset-to-scan timing, were serialized into a predefined JSON schema comprising metadata, clinical parameters, quantitative imaging features, and derived logical indicators. JSON (JavaScript Object Notation) is a lightweight, text-based, language-independent data interchange format used to store and transport structured data. This JSON-based interface functions as the "nervous system" of the agentic framework, translating heterogeneous raw imaging and clinical data into structured, machine-readable clinical insights that can be consistently propagated across modules. The large language model (chatGPT 5.1, OpenAI) is the primary consumer of this representation. Rather than receiving free-text radiological summaries, the model operates exclusively on explicit numerical variables and structured flags. This design constrains reasoning to verifiable inputs, improves auditability, and reduces hallucination risk by grounding the LLM's decision process in deterministic quantitative features extracted directly from segmentation outputs.

Knowledge-based reasoning module

To ensure clinical correctness and traceability, the system is built around an indexed guideline knowledge base rather than free-form language generation. Contemporary recommendations for spontaneous intracerebral hemorrhage (American Heart Association, American Stroke Association, European Stroke Organisation, and locally adopted protocols) are ingested from source documents using structure-preserving extraction, enabling retention of headings, tables, and conditional statements [4, 5]. The extracted content is normalized and stored with rich metadata (guideline source, version, section, recommendation type), allowing the downstream decision-support layer to constrain reasoning to verifiable, up-to-date guideline passages.

At inference time, the agent forms a query from the case-specific structured feature set (clinical variables + quantitative imaging biomarkers) and retrieves evidence via hybrid search (lexical + semantic). Retrieved candidates are re-ranked to prioritize guideline fragments that contain explicit thresholds, contraindications, and location-specific surgical criteria. This retrieval-grounded approach is designed to reduce omission of critical exceptions and to ensure that every proposed management pathway is supported by directly referenced guideline text.

Decision support is implemented as a hierarchical pipeline in which deterministic, data-driven components precede language generation. First, the imaging module produces quantitative biomarkers derived from segmentation and measurement (e.g., hematoma volume, dimensions, anatomical compartment/localization, ventricular involvement), which are combined with clinician-provided variables (e.g., age, neurological status, time from onset, anticoagulation/physiologic

parameters). These structured inputs drive computable guideline constructs and rule-based "gates" (e.g., volumetric thresholds, location-dependent considerations, EVD criteria, and standardized risk score components such as ICH score). The LLM layer then operates primarily as an orchestration and explanation module: it synthesizes the retrieved evidence with the structured features to generate one of four predefined management pathways – conservative management, minimally invasive evacuation, open surgical evacuation (craniotomy/decompressive craniectomy as appropriate), or external ventricular drainage – together with a structured quantitative case summary and guideline-referenced justification.

Trajectory planning module

When surgical evacuation is indicated, this module performs fully automated trajectory modelling to select a single best catheter corridor for minimally invasive hematoma access. The module operates on the segmented hematoma, skull/skin surface, ventricular system, venous sinuses, and CTA-derived vascular structures.

Trajectories are generated as straight-line paths from skull/skin entry points to intralesional target points. Each trajectory candidate is first evaluated against non-negotiable exclusion criteria: major vessel intersection (CTA-based), ventricular intersection, venous sinus intersection. Any candidate that intersects these masks is rejected and removed from further consideration.

For candidates that pass hard rejection, we additionally compute proximity-based penalties to discourage trajectories that pass too close to critical structures even if they do not intersect them. Let d_{vessel}, d_{vent} and d_{sinus} be the minimum distances (mm) from the trajectory to each structure. A soft penalty is computed as a monotonic decreasing function of distance, e.g.:

$$P(T)= \omega_{vessel}\frac{1}{d_{vessel}(T)+\varepsilon} + \omega_{vent}\frac{1}{d_{vent}(T)+\varepsilon} + \omega_{sinus}\frac{1}{d_{sinus}(T)+\varepsilon}$$

where:

ω_{vessel}, ω_{vent} and $\omega_{sinus} \geq 0$ are structure-specific weighting coefficients that were empirically selected based on two considerations: structure segmentation quality and structure diffusivity; vessels were assigned a lower penalty because they are more spatially diffuse, whereas ventricles and venous sinuses were assigned higher penalties because they are more robustly segmented and clinically important to avoid;

$\varepsilon > 0$ is a small positive constant introduced for numerical stability and regularization.

Trajectory function is defined as following:

$$T(s) = E + s\boldsymbol{u}, s \in [0, L]$$

where:

$E \in \mathbb{R}^3$	is the skull/skin entry point;
$\boldsymbol{u}$	is a unitl direction vector;
L	is the trajectory length;
s	is the scalar path parameter representing depth along the trajectory.

Each surviving trajectory is scored primarily by how much hematoma it traverses, operationalized as the number of hematoma voxels intersected along the path. As the trajectory is sampled through voxel space, we have:

$$V_{\mathcal{H}}(T) = \sum\nolimits_{s\in T} 1_{\mathcal{H}}(s)$$

where $1_{\mathcal{H}}(s)$ is 1 when the sampled point lies inside the hematoma mask. Maximizing $V_{\mathcal{H}}(T)$ directly implements the objective that the catheter should traverse a longer path within the hematoma cavity.

To reduce unnecessary tissue transgression, we also minimize the distance from the entry point to the first hematoma intersection:

$$D_{skin\rightarrow hema}(T) = \| E - H_{entry} \|$$

So, for each candidate trajectory T that passes hard constraints, we compute a single composite score:

$$Score(T) = \alpha \cdot V_{\mathcal{H}}(T) - \beta \cdot D_{skin\rightarrow hema}(T) - \gamma \cdot P(T)$$

where:

$V_{\mathcal{H}}(T)$	is hematoma voxel traversal (maximize);
$D_{skin\rightarrow hema}(T)$	represents extracerebral path length (minimize);
$P(T)$	stands for proximity-based soft risk (minimize);
α, β, γ	are tunable weights that were selected empirically rather than formally optimized.

Trajectory optimization is formulated as a constrained geometric search problem in 3D voxel space:

$$T^* = \underset{T}{argmax}\ Score(T)$$

The objective is to maximize intralesional catheter traversal while minimizing extracerebral path length and penalizing proximity to critical structures, subject to strict non-intersection constraints with major vessels (CTA), ventricles, and venous sinuses.

The search is initialized using the intrinsic geometry of the hematoma. The dominant spatial orientation of the clot is estimated via principal component analysis (PCA) of the hematoma voxel distribution. The first principal component defines the major longitudinal axis of the cavity, which serves as a physiologically meaningful prior for trajectory

alignment. Because optimal drainage efficiency is associated with catheter alignment along the central longitudinal axis of the hematoma, candidate trajectories are first generated along both directions of this principal axis.

From these axis-aligned initializations, the algorithm performs constrained angular exploration in a local neighborhood of the principal direction. Direction vectors are perturbed within a bounded angular cone, generating a discrete set of candidate trajectories that are projected from potential skull entry points toward intralesional targets. Each candidate trajectory is evaluated in voxel space.

Trajectories that intersect segmented major vessels, ventricles, or venous sinuses are immediately discarded.

All weights (α, β, γ and ω_{vessel}, ω_{vent}, ω_{sinus}) were set by iterative manual tuning on a representative development subset stratified by hematoma location and volume, with top-ranked trajectories visually inspected in 3D until they were judged anatomically plausible. The ordering $\alpha > \beta > \gamma$ reflects the planning hierarchy: intralesional traversal is the primary objective, extracerebral path length is secondary, and the soft penalty only breaks ties among candidates already filtered by hard constraints. Structure-specific weights follow segmentation reliability and clinical consequence of injury (lower for diffuse vessels, higher for robustly segmented ventricles and sinuses). The values used in all reported experiments were $\alpha = 3.0$, $\beta = 2.0$, $\gamma = 1.0$, $\omega_{vessel} = 1.0$, $\omega_{vent} = 2.0$, $\omega_{sinus} = 2.5$, and $\varepsilon = 1.0$ mm, these are heuristic settings, not formally optimized.

Reference standard

A three-member expert panel (2 board-certified neurosurgeons and 1 neuroradiologist), blinded to AI outputs, independently reviewed all cases. For each case, experts determined: hematoma parameters (volume, localization), optimal management class, treatment strategy, surgical approach type (if applicable), proposed surgical trajectory (if applicable). Discrepancies were resolved by consensus. The consensus decision was considered as a reference standard.

Statistical analysis and ethics

Statistical analysis was performed at the patient level ($N = 50$). Categorical variables are presented as counts and percentages. Continuous variables are reported as mean ± standard deviation (SD) when normally distributed, or median and interquartile range (IQR) when distribution is non-normal. The primary endpoint was agreement between AI-generated and expert treatment recommendations (four-category management classification). Agreement beyond chance was assessed using Cohen's kappa (κ) coefficient with 95% confidence intervals (CI). Observed agreement and expected agreement by chance (Pe) were calculated. Agreement between AI-derived and reference hematoma volumes was assessed using Pearson correlation coefficient (r). Correlation was evaluated using a two-sided test. Additionally, agreement between AI-derived and reference hematoma volumes was evaluated using Bland–Altman analysis, with calculation of mean bias and 95% limits of agreement. For trajectory comparison, angular deviation between AI-generated and expert-defined vectors was calculated. As angular deviation did not follow a normal distribution, results are reported as median and IQR. All statistical tests were two-sided. A p-value < 0.05 was considered statistically significant.

The study was approved by the institutional Ethics Committee (protocol No. 63/1 dated 15.07.2025) and conducted in accordance with the Declaration of Helsinki. Given the retrospective design, the requirement for written informed consent was waived by the Ethics Committee.

Results

A total of 50 patients with spontaneous intracerebral hemorrhage were included in the analysis. The mean age of the cohort was 65 years (range 49–83 years), with a slight predominance of male patients. Deep hemorrhages involving the basal ganglia were the most common localization, followed by lobar and thalamic hemorrhages. Cerebellar hemorrhages accounted for a smaller proportion of cases (Table 1).

According to the ICH score classification, patients were distributed across the full severity spectrum (0–5), with the majority falling into intermediate-risk categories. Higher ICH scores were associated with lower GCS scores and larger hematoma volumes.

Retrospective analysis of the data showed that conservative management was applied in 29 patients (58%), surgical hematoma evacuation in 12 patients (24%), decompressive craniectomy in 7 patients (14%), and external ventricular drainage alone in 2 patients (4%).

Functional outcome at discharge, assessed using mRS, demonstrated favorable outcome (mRS 0–2) in approximately one-third of patients, while nearly half of the cohort experienced severe disability or death (mRS 4–6). Poor outcomes were predominantly observed in patients with massive hemorrhage and high ICH scores.

CT segmentation and quantitative features extraction

Model-derived hematoma volumes demonstrated robust agreement with reference (ground truth) segmentation across the entire volumetric spectrum. An excellent Pearson's

Table 1 Demographic and baseline clinical characteristics of the study population

Parameter	Value
Age (years)	
Mean ± SD	65.1 ± 9.2
Range	49–83
Sex, *n* (%)	
Male	29 (58%)
Female	21 (42%)
Admission Neurological Status	
NIHSS score, mean ± SD	14.8 ± 6.3
GCS score, mean ± SD	10.9 ± 3.2
ICH Score	
Mean ± SD	2.1 ± 1.4
Range	0–5
Functional Outcome at Discharge (mRS) score, *n* (%)	
0–2 (Favorable)	14 (28%)
3 (Moderate disability)	12 (24%)
4–6 (Poor outcome)	24 (48%)

Table 2 Quantitative imaging characteristics of intracerebral hemorrhage

Parameter	Value
Hematoma volume (mL)	
Mean ± SD	41.6 ± 28.4
Median (IQR)	34 (18–48)
Range	10–120
Massive ICH (≥ 80 mL), *n* (%)	10 (20%)
Anatomical location, *n* (%)	
Lobar	18 (36%)
Deep (basal ganglia/thalamus)	22 (44%)
Cerebellar	8 (16%)
Brainstem	2 (4%)
Laterality, *n* (%)	
Left	27 (54%)
Right	23 (46%)
Linear dimensions (mm)	
Maximum axial diameter	52.3 ± 14.8
Orthogonal diameter	38.7 ± 11.2
Craniocaudal extent	41.5 ± 12.6
CT attenuation (HU)	
Mean density	61.4 ± 7.9
Density SD	11.2 ± 4.3
10–90 percentile range	22.6 ± 6.1
Shape descriptors	
Sphericity	0.71 ± 0.12
Surface-to-volume ratio	0.42 ± 0.09
Elongation index	1.38 ± 0.24
Convexity ratio	0.83 ± 0.08

correlation was observed (r = 0.91), with minimal volumetric deviation along the line of identity, indicating accurate proportional scaling and preservation of case-to-case volumetric variability.

The reported parameters demonstrate substantial variability in hematoma size, geometry, and internal density, reflecting the heterogeneous anatomical phenotypes of intracerebral hemorrhage. Overall, these imaging features provide an objective characterization of hemorrhage morphology and complexity, supporting their potential role in, treatment selection, and outcome prediction (Table 2).

To further assess agreement beyond correlation, Bland–Altman analysis was performed (Fig. 2). The analysis demonstrated a small positive bias of 0.81 mL, indicating minimal systematic overestimation. The standard deviation of differences was 13.05 mL, yielding 95% limits of agreement from − 24.77 mL to 26.39 mL, supporting good agreement across clinically relevant volume ranges.

The mean absolute error was 9.96 mL, confirming overall quantitative accuracy. Error remained stable across the volumetric spectrum, including in larger hematomas (> 80 mL), where the mean absolute error was 11.78 mL. Population-level characteristics, including mean and dispersion, were comparable between model-derived and reference measurements, supporting the quantitative fidelity and clinical validity of the automated volumetric assessment.

Representative axial non-contrast CT slices demonstrating automated multistructure segmentation generated by the customized 3D nnU-Net model. (**A**) Brain parenchyma and ventricular structures are segmented, with the ventricular system highlighted in blue and additional anatomical masks shown for contextual reference. (**B**) Intracerebral hematoma segmentation (red) with concurrent ventricular segmentation (blue) and skull mask (pink), illustrating spatial relationships and enabling volumetric and radiomic feature extraction. The panel also displays automatically computed quantitative parameters, including voxel count, CT attenuation statistics, and total hematoma volume (Fig. 3).

Overall, the model demonstrated strong performance for deep and lobar hemorrhages and somewhat reduced — but clinically acceptable — performance for infratentorial cases, consistent with the well-documented challenges of segmenting lesions in the posterior fossa (smaller lesion volumes, anatomical complexity, and beam-hardening artifacts). These results are in agreement with previously reported performance metrics [14], supporting the suitability of the model for hematoma volume estimation within the proposed pipeline (Table 3).

Knowledge-based reasoning

The LLM demonstrated a slightly lower tendency toward conservative recommendations, consistent with

guideline-based threshold enforcement (Table 3). The overall agreement between LLM-generated recommendations and expert decisions was 82% (41/50 cases). To account for category imbalance and chance agreement, Cohen's kappa coefficient was calculated. The expected agreement by chance (Pe) was 0.379, yielding:

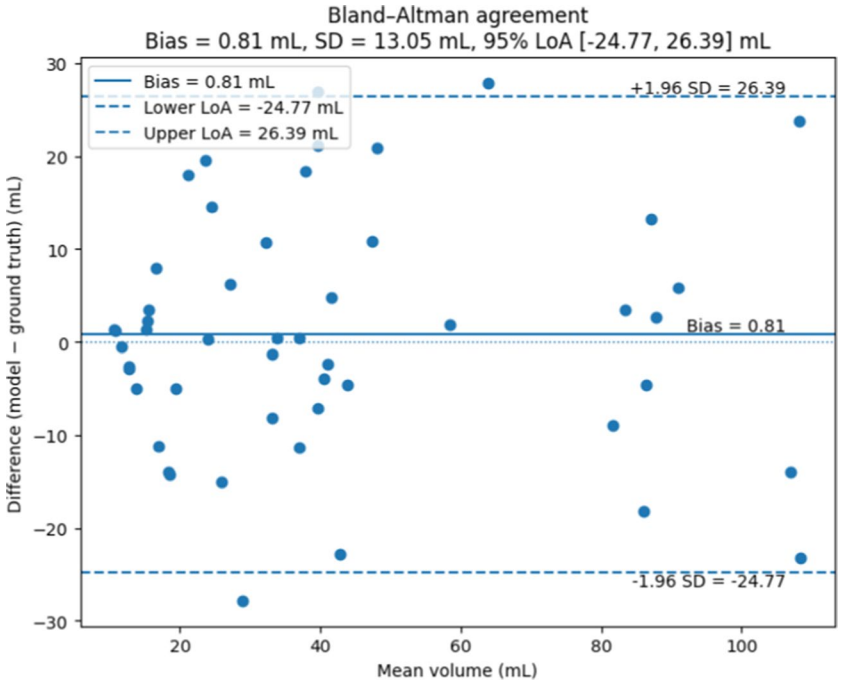


Fig. 2 Bland–altman agreement between ground truth and model-derived hematoma volumes

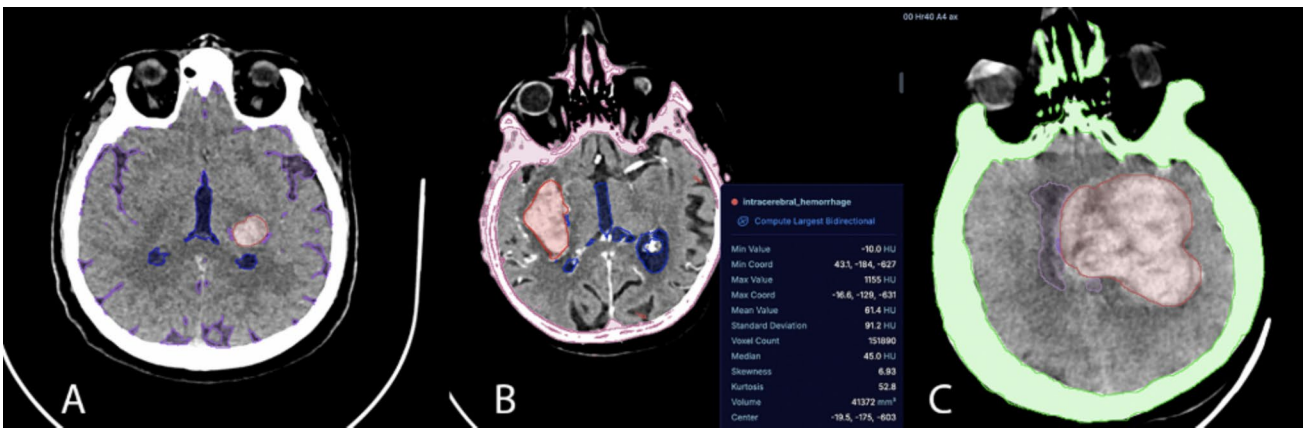


Fig. 3 Hematoma segmentation using nnU-Net

Table 3 Segmentation model performance on the test dataset

ICH Subtype	*n*	Dice Score (95% CI)	Sensitivity (95% CI)	PPV (95% CI)
Deep	22	0.89 (0.84–0.92)	0.87 (0.81–0.92)	0.92 (0.86–0.96)
Lobar	18	0.85 (0.78–0.90)	0.82 (0.73–0.89),	0.90 (0.83–0.95)
Infratentorial (Cerebellar & Brainstem)	10	0.53 (0.28–0.67)	0.39 (0.18–0.62)	0.93 (0.78–0.98)

$$\kappa = \frac{0.82 - 0.379}{1 - 0.379} = \frac{0.441}{0.621} = 0.71$$

The corresponding 95% confidence interval for κ was 0.54–0.88. This indicates substantial agreement between the LLM (Fig. 4) and expert clinical judgment (Table 4).

The system integrates key patient parameters (age, sex, GCS, clinical dynamics, and vital signs) to generate a standardized diagnosis of intracerebral hemorrhage, calculate the ICH score with associated risk stratification, and provide guideline-based treatment recommendations, including surgical and conservative management options.

Trajectory modelling

The trajectory planning module was evaluated on an independent test dataset of 50 CT scans with intracerebral hemorrhage (Fig. 5). For each case, expert neurosurgeons defined clinically optimal evacuation trajectories. The automated trajectory was compared against expert-defined references using geometric deviation and objective safety metrics.

Trajectory direction agreement was quantified using angular deviation between the automated and expert direction vectors:

Table 4 Concordance of human and agentic AI decision-making in intracerebral hemorrhage management

Treatment Strategy	Expert *n* (%)	LLM *n* (%)
Conservative management	29 (58%)	23 (46%)
Surgical hematoma evacuation	13 (26%)	17 (34%)
Decompressive craniectomy	6 (12%)	8 (16%)
External ventricular drainage (EVD) only	2 (4%)	2 (4%)
Total	50 (100%)	50 (100%)

$$\Theta = \arccos(\boldsymbol{u}_{exp} \cdot \boldsymbol{u}_{alg})$$

Across the 50 cases, the median angular deviation was 7.6° (IQR 5.1–9.8). In 43 of 50 cases (86%), angular deviation was below 10°, indicating close alignment with expert-defined surgical corridors.

All automated trajectories satisfied predefined hard safety constraints. No trajectory intersected major vessels (CTA-based), the ventricular system, or venous sinuses in 50/50 cases (100%). Minimum Euclidean distances from the automated trajectory to critical structures were computed. These findings indicate that the automated system achieved close geometric agreement with expert planning while providing consistently larger anatomical safety margins.

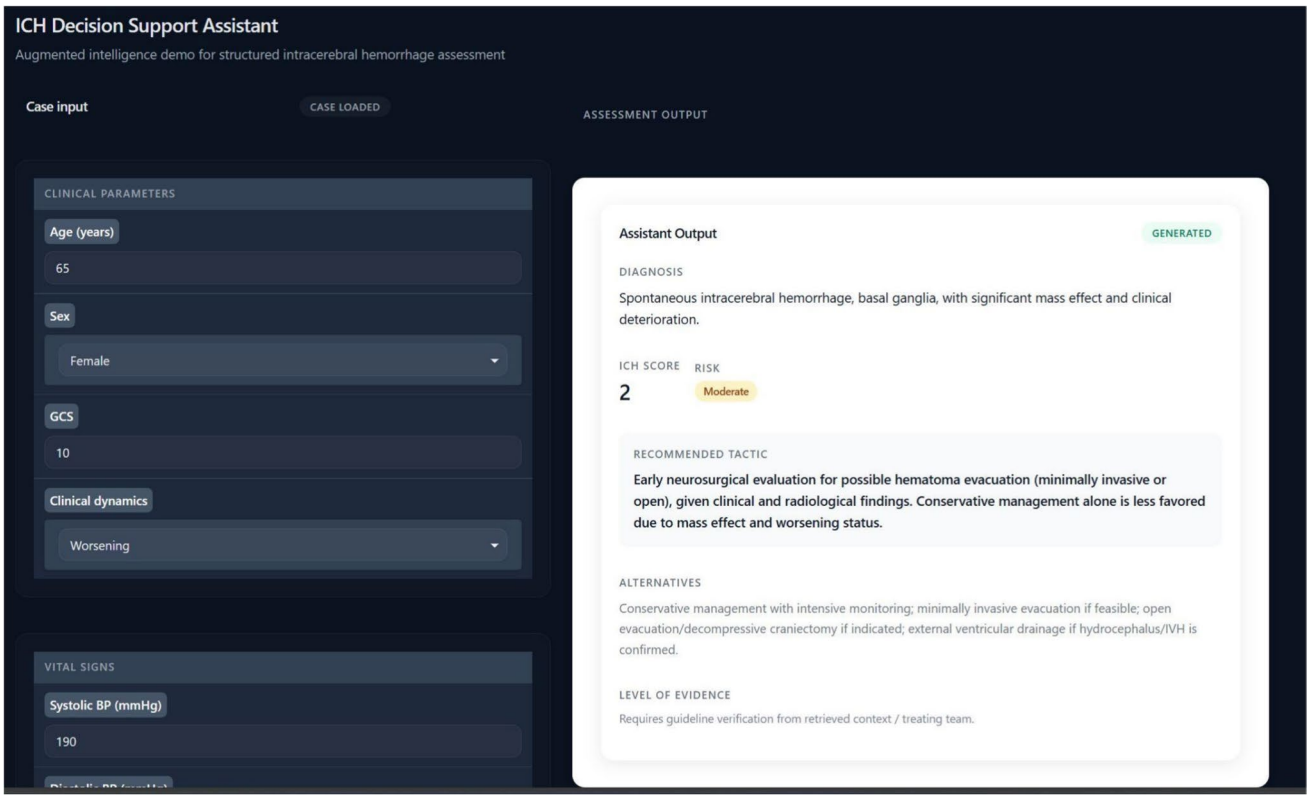


Fig. 4 Interface of the ICH decision support assistant with automated assessment output

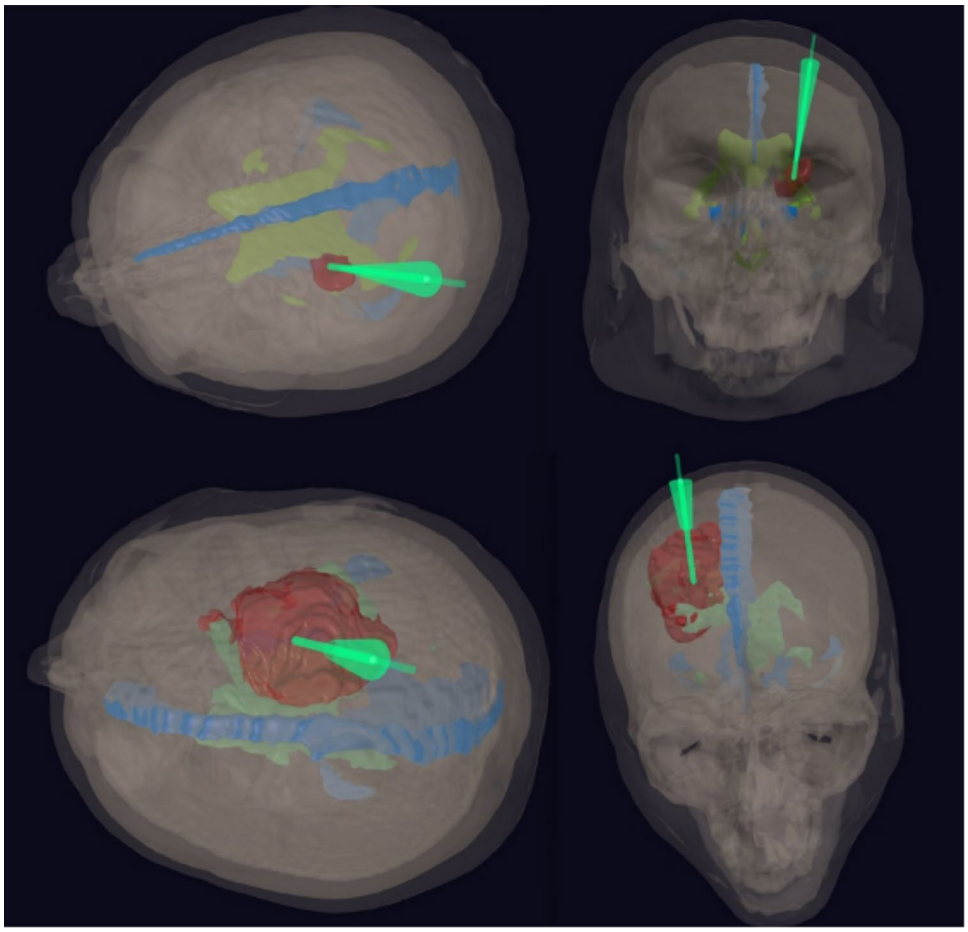


Fig. 5 Automated trajectory modeling for minimally invasive hematoma evacuation

Three-dimensional reconstructions of patient-specific cranial anatomy and intracranial structures generated from segmented CT/CTA data. The hematoma (red), ventricular system (blue), major vascular structures (green), and surrounding brain parenchyma (transparent) are visualized within a unified spatial framework. The automatically generated surgical trajectory (green line) is optimized to align with the principal long axis of the hematoma while minimizing transgression of eloquent tissue and proximity to critical vascular and ventricular structures.

Discussion

We present a fully automated framework that delivers end-to-end integration of multimodal imaging analysis, volumetric hematoma characterization, guideline-based classification, and personalized surgical trajectory planning. From raw CT/CTA acquisition to structured clinical recommendations, all analytical stages are executed within a unified agentic pipeline. This seamless integration enables rapid, reproducible, and context-aware decision support tailored to acute neurovascular workflows, where time efficiency and precision are critical.

Unlike conventional AI systems that primarily focus on anatomical segmentation, the present framework incorporates agentic reasoning mechanisms that synthesize anatomical, clinical, and contextual variables into higher-order inferences [15, 17]. Rather than limiting analysis to voxel-level predictions, the system transitions from descriptive imaging outputs to prescriptive clinical support [9]. This evolution from image interpretation to structured therapeutic guidance represents a substantive conceptual advance in neurosurgical AI applications [13].

By embedding standardized analytical workflows and guideline-driven logic, the framework reduces reliance on individual operator experience [8]. Automated data synthesis and structured recommendations mitigate variability introduced by subjective interpretation, especially in time-sensitive emergency settings. This promotes more consistent clinical reasoning across diverse practitioners and institutions reducing potential time loss and the cognitive burden placed on surgeons. In this context, the framework should be interpreted as augmented intelligence,

aimed at reducing cognitive load, standardizing analysis, and improving reproducibility in time-critical neurosurgical workflows.

All analytical outputs are structured through standardized JSON encoding, enabling potential interoperability with hospital information systems, neuronavigation platforms, and external clinical decision-support tools. This structured representation transforms raw imaging and clinical data into LLM-friendly, audit-ready outputs suitable for workflow automation and longitudinal quality tracking. To prevent "LLM-only" behavior, outputs are constrained through structured response schemas and post-generation validation. The system verifies completeness of required fields, internal consistency between measured biomarkers and the selected pathway, and the presence of explicit guideline grounding for each key recommendation. When critical inputs are missing or uncertain, the agent returns an "insufficient data" state with an explicit list of required variables and assumptions rather than extrapolating. Final treatment decisions remain under the authority of the attending neurosurgeon; the agent is intended to improve speed, consistency, and documentation quality through evidence-linked, reproducible decision support.

Trajectory planning is executed through automated optimization algorithms that account for vascular anatomy, ventricular avoidance, hematoma morphology, and alignment with the longitudinal axis of the hemorrhage cavity. This multi-constraint optimization ensures consistent generation of safe, efficient, and anatomically informed surgical corridors. A lightweight intraoperative interface provides surgeons with intuitive access to, trajectory visualizations, and structured recommendations without disrupting workflow dynamics. An important safety finding is the 100% avoidance of major vessels, ventricular system, and venous sinuses in all generated trajectories. This highlights the system's inherent "vascular awareness," achieved through explicit integration of segmented vascular anatomy and hard exclusion constraints during trajectory optimization. Such deterministic enforcement of safety rules has the potential to reduce human planning errors, particularly in emergency settings where fatigue, time pressure, and cognitive overload may affect decision-making. By consistently identifying and avoiding critical structures, the system provides an additional layer of safety and standardization in minimally invasive trajectory planning.

Most disagreements occurred in borderline surgical cases, where the LLM recommended more aggressive surgical treatment or minimally invasive management in patients selected by surgeons for open surgical intervention [12, 16]. While the proposed agentic framework incorporates established clinical and radiological criteria for treatment selection, it does not explicitly model ethical considerations related to long-term quality of life, functional independence, or patient-centered values. Decisions in severe neurological conditions often require careful balancing of survival benefit and anticipated disability, as well as consideration of patient and family preferences. These aspects remain under the responsibility of the treating team and are not intended to be replaced by the system. Future work may explore the integration of patient-reported outcomes and ethical decision-support components to further enhance individualized care. Importantly, no cases demonstrated extreme discordance (e.g., LLM recommending conservative management in cases selected for decompressive craniectomy with mass effect). The observed 82% agreement between AI-generated and expert decisions further supports the system's clinical reliability. This level of concordance suggests that the framework performs at an expert level in standard cases, while discrepancies are primarily confined to borderline scenarios where even human decision-making is inherently variable. These findings reinforce the potential of the system to provide consistent, guideline-aligned support in routine clinical situations while maintaining transparency in more complex cases.

A key conceptual strength of the proposed framework is the transition from purely *descriptive* image analysis to *prescriptive* clinical decision support. Importantly, the system effectively overcomes the traditional "black box" limitation of conventional AI by translating pixel-level imaging data into discrete, structured clinical facts encoded in a standardized JSON format.

These machine-readable variables (e.g., hematoma volume, location, ventricular involvement, midline shift) are subsequently processed through deterministic, rule-based logic aligned with established clinical guidelines. The large language model operates on this structured representation rather than raw images, serving primarily as an orchestration and explanation layer rather than an opaque decision-maker.

This architecture ensures transparency, traceability, and auditability at each step of the pipeline. Critically, by standardizing data interpretation and enforcing consistent decision pathways, the framework directly addresses one of the central challenges in neurosurgery—inter-operator variability [2]. By reducing dependence on individual experience and subjective interpretation, the system promotes reproducible, guideline-consistent decisions, particularly in high-pressure and emergency settings. Our work shifts the paradigm in neurosurgical decision support by moving from isolated image analysis toward integrated, end-to-end clinical reasoning systems. By combining automated imaging interpretation, structured data representation, guideline-based reasoning, and trajectory optimization within a unified framework, it contributes to the foundational framework of computational neurosurgery.

Limitations

Despite promising results, several limitations exist. The current evaluation is primarily retrospective and based on simulated planning scenarios. Prospective clinical validation is required to determine the real-world impact on surgical accuracy, complication rates, and functional outcomes, in alignment with established reporting frameworks for AI-based clinical decision support systems [4, 10]. The limited cohort size restricts statistical power and may affect generalizability. Perihematomal edema was not explicitly segmented or incorporated as an independent quantitative variable. This was a deliberate design choice, as the system is primarily focused on the hyper-acute phase of intracerebral hemorrhage with focus on hematoma characteristics (volume, location, mass effect) rather than edema evolution. As edema may become clinically relevant future iterations of the framework will incorporate automated edema segmentation and temporal modeling to extend applicability beyond the hyper-acute setting.

Importantly, this work does not aim to advance segmentation methodology itself. Instead, segmentation is treated as an upstream component that enables extraction of structured quantitative imaging biomarkers, which serve as inputs for the downstream agentic reasoning module. The primary focus of the proposed framework is the end-to-end decision-support pipeline, where a large language model operates on these structured features to generate treatment recommendations and invoke a tool to recommend optimal surgical trajectory. Consequently, segmentation performance is considered in the context of its utility for reliable quantitative representation rather than as a standalone objective.

Larger, multicenter datasets are necessary to confirm robustness across heterogeneous imaging protocols and patient populations. Finally, reliance on single-center data introduces potential bias related to local practice patterns and infrastructure. Broader validation across diverse healthcare systems will be essential before widespread clinical adoption. The trajectory scoring weights were selected by manual tuning rather than formal optimization, which introduces several biases: operator preference of those performing the tuning, coupling to upstream segmentation behavior (since structure-specific weights partly compensate for segmentation reliability), single-center bias in the development subset, and modelling bias from the additive linear form of the score. Future work will address these through multicenter expert-annotated reference trajectories, formal optimization, and per-weight sensitivity analyses.

This work contributes to the development of intelligent, agent-based decision-support systems for the management of intracerebral hemorrhage. By integrating automated imaging analysis, clinical risk stratification, and individualized trajectory planning, the proposed framework provides a comprehensive platform for supporting complex surgical decision-making. Beyond intracerebral hemorrhage, similar agentic architectures may be extended to other acute neurovascular and neurosurgical conditions requiring rapid, data-driven intervention. Future integration with robotic and augmented navigation systems may enable semi-autonomous execution of optimized surgical strategies under neurosurgeon supervision. Continued refinement of this framework may facilitate patient-specific treatment planning, minimizing iatrogenic injury while maximizing evacuation efficiency and neurological preservation, thereby improving procedural safety and long-term clinical outcomes.

Conclusion

This agentic AI framework provides a structured, explainable, and workflow-integrated augmented intelligence platform for ICH treatment planning. By combining imaging analysis, guideline reasoning, and trajectory optimization, the system represents a step toward intelligent neurosurgical decision support.

Acknowledgements The authors would like to thank Eugen Sodin, Arsenii Halaida and Oleksii Samoilenko for their assistance with segmentation and data extraction from CT and CTA datasets. We also thank Prof. Oleksii Kalmykov for his support with manuscript proofreading and language editing.

Author contributions Eugen Kochuiev contributed to the development of core algorithms and system architecture and participated in writing, reviewing, and editing the manuscript. Vladyslav Kaliuzhka provided neurosurgical expertise, contributed to the conceptualization and methodology of the study, and supervised the clinical aspects of the research. Mykyta Markevych performed data analysis and contributed to data curation, investigation, and manuscript review and editing. Zoia Kochuieva designed and implemented the LLM-based agentic decision-support module. olodymyr Piatykop contributed to clinical interpretation and expert evaluation of the cases. All authors read and approved the final version of the manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability The datasets generated and analyzed during the current study are not publicly available due to institutional patient privacy protocols but are available from the corresponding author on reasonable request.

Code availability The proprietary algorithms and custom 3D encoder-decoder convolutional neural network derived from the nnU-Net paradigm utilized in this study are the intellectual property of eXtra Vision LLC. Further technical details are available from the corresponding author upon reasonable request.

Declarations

Ethics approval This retrospective study involving human participants was in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki Declaration

and its later amendments or comparable ethical standards. The study was approved by the institutional Ethics and Bioethics Committee of Kharkiv National Medical University and Communal Non-Commercial Enterprise of the Kharkiv Regional Council "Regional Clinical Hospital" (protocol No. 63/1 dated 15.07.2025).

Consent to participate Given the retrospective design of the study, the requirement for written informed consent was waived by the institutional Ethics and Bioethics Committee of Kharkiv National Medical University and Communal Non-Commercial Enterprise of the Kharkiv Regional Council "Regional Clinical Hospital". All imaging data and patient personal data were anonymized prior to analysis.

Consent for publication Not applicable, as all imaging data and clinical parameters were fully de-identified and the manuscript contains no potentially identifiable patient information.

Competing interests Vladyslav Kaliuzhka is the Chief Executive Officer and Founder, Mykyta Markevych is the Chief Operating Officer and Co-Founder, and Eugen Kochuiev is the Chief Technical Officer of eXtra Vision LLC. These roles were held during the conduct of the study. The company is involved in the development of technologies related to the subject matter of this research. All other authors declare that they have no competing financial or non-financial interests.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

References

1. An SJ, Kim TJ, Yoon BW (2017) Epidemiology, risk factors, and clinical features of intracerebral hemorrhage: an update. Nat Rev Neurol. https://doi.org/10.1038/nrneurol.2016.178
2. Brügge E, Ricchizzi S, Arenbeck M, Keller MN, Schur L, Stummer W, Holling M, Lu MH, Darici D (2024) Large language models improve clinical decision making of medical students through patient simulation and structured feedback: a randomized controlled trial. BMC Med Educ 24(1):1391. https://doi.org/10.1186/s12909-024-06399-7
3. Chilamkurthy S, Ghosh R, Tanamala S et al (2018) Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. Lancet. https://doi.org/10.1016/S0140-6736(18)31645-3
4. DECIDE-AI Steering Committee (2022) DECIDE-AI: reporting guideline for AI-based decision support systems in healthcare. Nat Med. https://doi.org/10.1038/s41591-022-01772-5
5. Greenberg SM, Ziai WC, Cordonnier C et al (2022) Guideline for the management of patients with spontaneous intracerebral hemorrhage. Stroke. https://doi.org/10.1161/STR.0000000000000407
6. Hanley DF, Lane K, McBee N et al (2017) Thrombolytic removal of intraventricular haemorrhage in treatment of severe stroke (CLEAR III): a randomised, multicentre, multiregion, placebo-controlled trial. Lancet. https://doi.org/10.1016/S0140-6736(17)30195-7
7. Hanley DF, Thompson RE, Rosenblum M, Thompson CB, Thompson G et al (2019) Minimally invasive surgery with thrombolysis in intracerebral haemorrhage evacuation (MISTIE III): a randomised, controlled, open-label, blinded endpoint phase 3 trial. Lancet. https://doi.org/10.1016/S0140-6736(19)30195-3
8. Ho CM, Guan S, Mok PK, Lam CH, Ho WY, Mak CH, Qin H, Wong AKC, Hui V (2025) Development and validation of a large language model-powered chatbot for neurosurgery: mixed methods study on enhancing perioperative patient education. J Med Internet Res 27:e74299. https://doi.org/10.2196/74299
9. Iqbal U, Tanweer A, Rahmanti AR, Greenfield D, Lee LT, Li YJ (2025May 7) Impact of large language model (ChatGPT) in healthcare: an umbrella review and evidence synthesis. J Biomed Sci 32(1):45. https://doi.org/10.1186/s12929-025-01131-z.PMID:40335969;PMCID:PMC12057020
10. Patil A, Serrato P, Chisvo N, Arnaout O, See PA, Huang KT (2024Nov 23) Large language models in neurosurgery: a systematic review and meta-analysis. Acta Neurochir (Wien) 166(1):475. https://doi.org/10.1007/s00701-024-06372-9
11. Pradilla G et al (2024) Trial of early minimally invasive removal of intracerebral hemorrhage. N Engl J Med. https://doi.org/10.1056/NEJMoa2308440
12. Qureshi AI, Palesch YY, Barsan WG et al (2016) Intensive blood-pressure lowering in patients with acute cerebral hemorrhage. N Engl J Med. https://doi.org/10.1056/NEJMoa1603460
13. Salehi S, Singh Y, Horst KK, Hathaway QA, Erickson BJ (2025) Agentic AI and large language models in radiology: opportunities and hallucination challenges. Bioengineering (Basel) 12(12):1303. https://doi.org/10.3390/bioengineering12121303
14. Sharrock MF, Mould WA, Ali et al (2021) 3D deep neural network segmentation of intracerebral hemorrhage: development and validation for clinical trials. Neuroinformatics 415:403–415. https://doi.org/10.1007/s12021-020-09493-5
15. Soun JE, Chow DS, Nagamine M et al (2021) Artificial intelligence and acute stroke imaging. AJNR Am J Neuroradiol. https://doi.org/10.3174/ajnr.A6883
16. Sprigg N, Flaherty K, Appleton JP et al (2018) Tranexamic acid for hyperacute primary intracerebral haemorrhage (TICH-2): an international randomised, placebo-controlled, phase 3 superiority trial. Lancet. https://doi.org/10.1016/S0140-6736(18)31033-X
17. Yu N et al (2022) A robust deep learning segmentation method for hematoma volumetric detection in intracerebral hemorrhage. Stroke. https://doi.org/10.1161/STROKEAHA.120.03224

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.