**PUNE VIDYARTHI GRIHA'S COLLEGE OF ENGINEERING, TECHNOLOGY & MANAGEMENT, PUNE**
Department of Artificial Intelligence & Data Science · Project Review-II · AY 2026–27

# AI-Powered Non-Destructive Reading & Text Recovery from Closed Ancient Books and Scrolls

Detailed project report: feasibility, scope, requirements, design, validation and plan



| Team | Roll No. | Module ownership |
|---|---|---|
| **Divesh Adivarekar** | 4008 | M2 Segmentation, M6 GUI |
| **Pranav Gugale** | 3078 | M3 ARAP flattening, M5 OCR & translation |
| **Rohan Gorde** | 3022 | M1 Ingestion, M6 GUI |
| **Samarth Sangolgi** | 3085 | M4 Ink detection, M5 OCR & translation |
| **Internal guide** | | Prof. Ravindra Mule |





*A carbonised scroll: unopenable, text intact inside.*

## Summary

Millions of bound books and rolled scrolls are fused, brittle or carbonised. Opening them destroys them, and leaving them closed keeps their contents unread. Micro-CT can see inside, but the raw volume is a warped, fused block of voxels in which the ink is almost invisible.

This project builds a **modular six-module AI pipeline** that virtually unwraps such artifacts from pre-acquired CT or X-ray phase-contrast volumes: it isolates each page, flattens it, detects the ink, reads the text and reports a **confidence score for every character**. Nothing uncertain is presented as certain.

# 1. Problem statement and motivation

**Problem.** Develop an AI-powered system that can accurately segment, flatten and reconstruct hidden text from closed books and scrolls using only pre-acquired volumetric CT / X-ray scans, with no physical opening, unrolling or destructive intervention.

The task breaks into four hard sub-problems:

| Sub-problem | Why it is hard | Our answer |
|---|---|---|
| **Fused layers** | Adjacent pages touch at sub-voxel distance; naive segmentation jumps between sheets. | FastSAM3D-style 3D segmentation with surface-normal tracking (M2). |
| **Warped geometry** | Pages are crumpled 3D manifolds that must unfold without distorting letters. | As-Rigid-As-Possible flattening (M3). |

| Sub-problem | Why it is hard | Our answer |
|---|---|---|
| **Invisible ink** | Carbon ink has almost the same X-ray density as carbonised papyrus. | 3D CNN / TimeSformer ensembles on local voxel windows (M4). |
| **Damaged text** | Recovered characters are broken, and guesses can mislead scholars. | Per-token log-likelihood; low-confidence output flagged (M5). |




*Scale of the heritage at risk.*

### Why it matters

- **Lost literature.** About 1,800 carbonised Herculaneum scrolls remain unread.
- **Decaying archives.** Millions of acidic 19th-century volumes are becoming too brittle to open.
- **Research value.** Pushes 3D surface tracking, zero-shot volumetric segmentation and sparse transformers on terabyte data.
- **Transfer.** The same core applies to medical pathology, industrial NDT and PCB inspection.

# 2. Aim and objectives

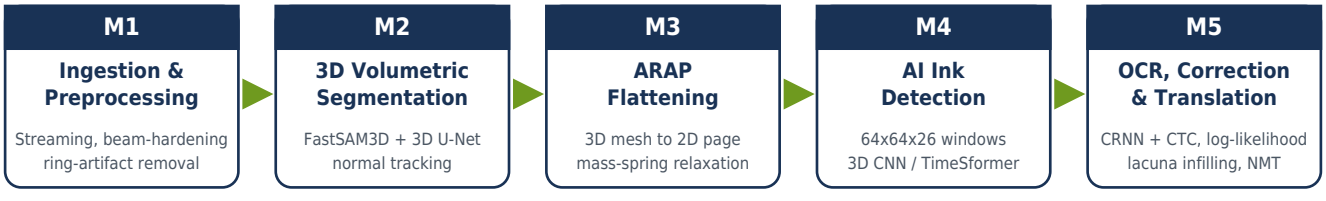
**Aim.** Architect an end-to-end AI platform that converts raw volumetric CT scans of closed books and scrolls into flattened, readable, confidence-scored and translated digital text.

| # | Objective | Measure |
|---|---|---|
| **O1** | Data preprocessing: remove beam-hardening and ring artifacts on TB-scale volumes by out-of-core streaming | Handles a 50 GB test volume within RAM |
| **O2** | Zero-shot segmentation of tightly bound pages without layer-jumping | Dice > 0.90 |
| **O3** | Isometric ARAP flattening that preserves character shape | Edge-length distortion within tolerance |
| **O4** | Ink detection for low-Z carbon ink with 3D CNN / TimeSformer ensembles | Ink-versus-substrate separation |
| **O5** | CRNN OCR and Ithaca-style lacuna infilling with confidence scoring | CER / WER and per-token log-likelihood |
| **O6** | Researcher GUI to load, monitor and export | All FR-6.x met |

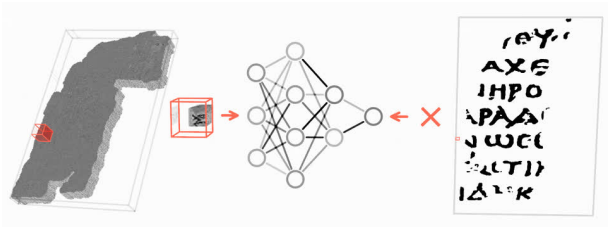
# 3. System architecture

Five processing phases run in sequence, each producing a documented data product for the next. A sixth module, the GUI, sits across all of them. Scanning itself is outside the project; the system consumes reconstructed volumes in OME-Zarr, HDF5 or 16-bit TIFF.

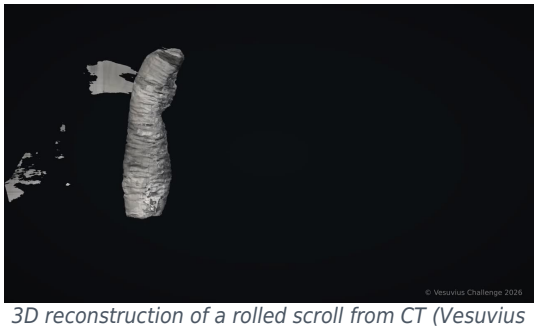
*Closed book / scroll → external micro-CT / X-ray scan → reconstructed volume (10–15 TB) [out of scope]*




**M6 Researcher GUI (cross-cutting): load scan · monitor phases · before/after previews · confidence review · export text**



*Ink recovery: a 3D window of the scan is passed to a neural network, which outputs the letters.*




*3D reconstruction of a rolled scroll from CT (Vesuvius Challenge).*

### Data flow (DFD)

At context level the researcher selects a scan from the CT repository and receives a confidence-scored translated transcript. At level 1, processes 1.0 to 5.0 map one-to-one onto modules M1 to M5. Three data stores are used: **D1** raw and corrected volumes, **D2** segmented meshes and flattened page maps, **D3** model weights, benchmarks and metric logs. Module 6 reads status from all five and writes the export to D3.

# 4. Modules in detail

| Module | What it does and how | Technology | Owner |
|---|---|---|---|
| **M1 Ingestion & Preprocessing** | Streams terabyte volumes in chunks; iterative polynomial correction removes beam-hardening “cupping”; polar-coordinate filtering suppresses ring artifacts. | Zarr / HDF5, NumPy, scikit-image, ASTRA | Rohan Gorde |
| **M2 3D Segmentation** | Builds point clouds at each sheet centre, groups them into page surfaces with a zero/few-shot FastSAM3D-style model and 3D U-Net, tracks surface normals to avoid layer-jumping. | FastSAM3D, 3D U-Net, Open3D, PyTorch | Divesh Adivarekar |
| **M3 ARAP Flattening** | Converts each page surface to a mesh and minimises ARAP energy with mass-spring relaxation; keeps a 2D-to-3D map for traceability. | ARAP, Bullet Physics, Open3D | Pranav Gugale |
| **M4 AI Ink Detection** | Feeds a 64×64×26 voxel window per pixel into a 3D CNN / TimeSformer ensemble; outputs a per-pixel ink probability for carbon and metallic inks. | 3D CNN, TimeSformer, CUDA | Samarth Sangolgi |
| **M5 OCR & Translation** | CRNN + CTC transcribes with per-character log-likelihood; flags low scores; masked language model fills gaps; neural MT outputs modern languages. | CRNN, masked LM, NMT | Pranav Gugale, Samarth Sangolgi |
| **M6 Researcher GUI** | Loads scans, shows phase progress and before/after previews, highlights uncertain text, exports annotated output. | Desktop GUI, local IPC | Rohan Gorde, Divesh Adivarekar |

# 5. Techniques and mathematical models

| Model | Expression | Role in the project |
|---|---|---|
| **Beer–Lambert attenuation** | $I = I_0 \cdot \exp(-\mu x)$ | Governs X-ray loss through the artifact; basis of ink–substrate contrast. |
| **Photoelectric absorption** | $P \propto Z^n / E^3$ ($n \approx 4$–5) | Explains why high-Z metallic ink is visible and low-Z carbon ink is not. |
| **ARAP energy** | $E = \Sigma\, w_i \, \|(p_i - p_j) - R_i(p_i^0 - p_j^0)\|^2$ | Flattens curved pages with minimal shear and stretch. |
| **Dice coefficient** | $2\|A \cap B\| / (\|A\| + \|B\|)$ | Acceptance gate for segmentation; target > 0.90. |
| **Token log-likelihood** | $\log P(\text{token}_t \mid \text{context})$ | Per-character confidence for flagging, instead of guessing. |

# 6. Software requirements (IEEE 830 structure)

The SRS defines **22 functional** and **8 non-functional** requirements, each labelled, prioritised and mapped to a test. Functional requirements are labelled FR-<module>.<n>; M1 to M5 are High priority and M6 is Medium.

| Group | Key requirements |
|---|---|
| **Performance** | NFR-1 about 8 ms per volumetric patch on an A100-class GPU · NFR-2 10–15 TB volumes processed out-of-core · NFR-3 Dice ≥ 0.90 · NFR-4 low-latency GUI progress |
| **Safety** | NFR-5 operates only on pre-acquired data and never touches the artifact · NFR-6 low-confidence output is never shown as definitive |
| **Security** | NFR-7 access limited to authorised users · NFR-8 raw volumes never sent to third-party services without explicit configuration |
| **Quality** | Modular (each phase swappable) · Reliable (graceful on malformed scans) · Reusable (segmentation and flattening) · Testable (Dice, IoU, CER/WER) |

**In scope:** processing of pre-acquired scans, artifact correction, 3D segmentation, ARAP flattening, ink detection, OCR with lacuna infilling and translation, researcher GUI, quantitative validation on EduceLab-Scrolls and ASTRA phantoms. **Out of scope:** scanner hardware, physical conservation, production cloud deployment, provenance verification, and medical / NDT / PCB use (future scope only).

**Open decisions (TBD):** target translation languages; final OCR confidence threshold; client-server protocol; access-control mechanism. None blocks the critical path of M1 to M4.

# 7. Feasibility, cost and risk

| Dimension | Verdict | Reasoning |
|---|---|---|
| **Technical** | Feasible | Every phase builds on published, reproducible work; the hard part is the Dice gate on fused layers. |
| **Economic** | Feasible | Open-source stack and public data; low cost on the institute GPU cluster. |
| **Schedule** | Feasible | 12 months, 4 phases; about 2,300 person-hours at ~12 h/week per member. |
| **Operational** | Feasible | Offline cluster pipeline with a simple archivist GUI; no uptime obligation. |

| Cost head | Basis | Cloud scenario | Institute scenario |
|---|---|---|---|
| **GPU compute** | ~600 A100-hours | ₹1,50,000 | ₹0 (cluster) |
| **Storage** | ~20 TB, 12 months | ₹60,000 | ₹10,000 |
| **Scan data / software** | Public benchmarks, open source | ₹0 | ₹0 |
| **Publication** | 1 paper + registration | ₹25,000 | ₹25,000 |

| Cost head | Basis | Cloud scenario | Institute scenario |
|---|---|---|---|
| **Contingency** | 15% buffer | ₹35,000 | ₹5,000 |
| **Total** | | ≈ ₹2,70,000 | ≈ ₹40,000 |

| Risk | P / I | Mitigation |
|---|---|---|
| **Layer-jumping; Dice stays below 0.90** | H / H | Point-cloud tracking with manual seeds; normal-consistency loss. |
| **Carbon ink undetectable in absorption CT** | M / H | Train on phase-contrast subsets; validate on metallic ink first. |
| **Hallucinated text in gap filling** | M / H | Hard-gate on log-likelihood (NFR-6). |
| **GPU cluster time unavailable** | M / H | Patch-wise training at reduced resolution; cloud reserve. |
| **Terabyte I/O bottleneck** | M / M | Chunked Zarr streaming; develop on cropped sub-volumes. |
| **Phase 2 slips** | M / H | Build M4 / M5 in parallel on synthetic pages; descope to scrolls only. |
| **No real manuscript scans** | H / M | ASTRA phantoms plus public benchmark scans. |
| **Member unavailable** | L / M | Documented interfaces; backup owner per module. |

# 8. Project plan and validation

| Phase | Months | Deliverable |
|---|---|---|
| **1** | 1–3 | Cluster setup, ASTRA synthetic CT, ingestion and artifact correction (M1) · MS1 at month 3 |
| **2** | 4–7 | FastSAM3D segmentation, layer tracking, ARAP flattening (M2, M3) · MS2 month 5, MS3 month 7 |
| **3** | 8–10 | Ink inference, CRNN OCR, scoring, translation (M4, M5) · MS4 month 9 |
| **4** | 11–12 | GUI, benchmark validation, Docker packaging, paper (M6) · MS5 month 11, MS6 month 12 |

| Test | Requirement | Description | Criterion | Status |
|---|---|---|---|---|
| **T1** | FR-1.1 | Out-of-core load of 50 GB Zarr volume | Within host RAM | **Passed** |
| **T2** | FR-1.2/1.3 | Beam-hardening and ring correction on phantom | Cupping flattened | In progress |
| **T3** | FR-2.2/2.4 | Segmentation on fused layers | Dice > 0.90 | In progress |
| **T4** | FR-2.3 | Layer-jump test, 40 µm gap | Zero cross-layer merges | Planned |
| **T5** | FR-3.2/3.3 | ARAP distortion on glyph grid | Within tolerance | Planned |
| **T6** | FR-4.2/4.4 | Carbon and metallic ink detection | Ink separated | Planned |
| **T7** | FR-5.1/5.5 | OCR accuracy and confidence | CER / WER logged | Planned |
| **T8** | FR-5.2, NFR-6 | Low-confidence flagging | 100% flagged | Planned |

# 9. Outcomes, applications and future scope



*Sealed historical volumes in archives worldwide.*

## Direct outcomes

- Working six-module pipeline from raw voxels to translated text.
- Dice > 0.90 and reported CER / WER on public benchmarks.
- Confidence-annotated transcripts for archivists.
- Researcher GUI, documentation and a submittable paper.

## Applications and future scope

- Mass digitisation of acidic-paper holdings.
- Medical tumour and organ segmentation.
- Industrial non-destructive testing of composites.
- PCB and microchip inspection for tampering.

# 10. Comparison with existing systems

| Criterion | Spine-cutting / flatbed | ThaumatoAnakalyptor | Generic 2D SAM + OCR | This project |
|---|---|---|---|---|
| **Artifact safety** | Destroys binding | Non-destructive | n/a | Non-destructive |
| **Artifacts** | Books, after cutting | Rolled scrolls | Flat images | Books and scrolls |
| **Automation** | Manual | Semi-manual | Partial | End-to-end |
| **3D segmentation** | None | Hand-tuned point clouds | 2D masks | FastSAM3D + 3D U-Net |
| **Confidence** | n/a | Limited | Binary CER | Per-token log-likelihood |
| **Reuse** | None | Scroll-specific | Not volumetric | Modular |

# References

[1] Y. Assael et al., “Restoring and attributing ancient texts using Ithaca,” *Nature*, vol. 603, no. 7900, pp. 280–283, 2022.

[2] A. Kirillov et al., “Segment Anything,” in *Proc. IEEE/CVF ICCV*, 2023, pp. 4015–4026.

[3] W. Schilliger et al., “ThaumatoAnakalyptor: Automatic Scroll Segmentation Pipeline for CT Scans,” 2023.

[4] Vesuvius Challenge Team, “3D Sliding-Window TimeSformer Neural Network for Carbon Ink Detection,” Technical Report, 2023.

[5] Y. Tang et al., “Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis,” in *Proc. IEEE/CVF CVPR*, 2022.

[6] Y. Wang et al., “FastSAM3D: An Efficient Segment Anything Model for 3D Medical Images,” in *Proc. MICCAI*, 2024.

[7] Z. Cen et al., “Segment Anything in 3D with NeRFs (SA3D),” in *NeurIPS*, 2024.

[8] X. Yin et al., “SAI3D: Segment Any Instance in 3D Scenes,” in *Proc. IEEE/CVF CVPR*, 2024.

[9] IEEE, “IEEE Recommended Practice for Software Requirements Specifications,” IEEE Std 830-1998, 1998.