SUPPLEMENTARY INFORMATION FOR:
FORVA AND GCM-CLIP: A MILLION-SCALE MULTIMODAL DATASET AND REPRESENTATION LEARNING FRAMEWORK FOR VIRTUAL AUTOPSY

Jikai Mao
maojikai@zju.edu.cn

Nanze Du
dunanze@hotmail.com

Lyu Tu
leo_lyu@yeah.net

Li Hao
lihao@cifs.gov.cn

Yi Shen
sevensevensysy@163.com

Liang Shen
xiaoqiao2006@139.com

Junjun Guo
865408638@qq.com

Jing Cai*
caijing@zjjcxy.cn

Supplementary Note 1 Prompt Design for Data Annotation and Information Extraction

Supplementary Note 1.1 Generative Description Prompt

1. Prompts are suitable for providing inputs to generative vision-language large models to enable the accurate and precise description of virtual anatomical images.

Supplementary Note 2 Prompt Design for Data Annotation and Information Extraction

Supplementary Note 2.1 Generative Description Prompt

1. Prompts are suitable for providing inputs to generative vision-language large models to enable the accurate and precise description of virtual anatomical images.

```
"""
You are an expert forensic radiologist.  The provided Virtual Autopsy slice
shows a post-mortem CT image with a known context of '{cause_of_death}' and
a specific target symptom.

Your task is to explicitly decouple your radiological report into two
hierarchical sections:

(1) 'discovery':  Provide a detailed, multi-level semantic description
containing exactly TWO sentences.  Sentence 1 (Positive):  Affirm
the presence of the symptom with fine-grained attributes (e.g.,
'Contrast-enhanced chest CT scan shows an intimal flap in the aorta,
diagnosing an aortic dissection.').  Sentence 2 (Negative):  Affirm the
healthy, normal state without this symptom (e.g., 'Contrast-enhanced chest
CT scan reveals a normal aorta with no signs of intimal flap, dissection,
or other pathological lesions.').
```

```
(2) 'impression': This section provides the diagnostic cues and MUST
strictly contain the explicit semantic anchors as a nested structure:
- 'location': Identify the main explicit anatomical anchor. Choose from
the 8 core regions: [Head, Neck, Chest, Abdomen, Pelvis, Unknown, Upper
Leg, Lower Leg].
- 'disease': Identify the primary forensic pathology or injury typology.
You MUST select exactly ONE category from the following 40 predefined
options: [Healthy, Laryngeal edema, Putrefactive gas bubbles, Soft tissue
hematoma, Fractures, Aortic rupture, Gastric content reflux, Incised
wounds, Skin burns, Intimal tear, Blood sedimentation, Airway burns,
Hypostatic pulmonary edema, Brain herniation, Visceral autolysis, Airway
and digestive tract fluid, Hemopericardium, Internal bleeding, Pulmonary
edema, Skull fracture, Internal organ injury, Pericardial effusion,
Hemopneumothorax, Ballistic hemorrhage, Hyoid bone fracture, Visceral
perforation, High cervical fracture-dislocation, Neck hematoma, Fracture
along the wound track, Aortic dissection, Tissue necrosis, Coronary artery
calcification, Cerebral edema, Hyperdense lesion in brain parenchyma,
Intracranial hemorrhage, Pulmonary embolism, Sinus fluid, Airway foreign
body, Periaortic hematoma, Coronary artery occlusion].
- 'text': Provide concise diagnostic cues summarizing the pathological
localization and findings.

Respond ONLY in a valid JSON format with two main keys: "discovery" and
"impression".
Example:
{
"discovery": "⟨value⟩",
"impression": {
"location": "⟨value⟩",
"disease": "⟨value⟩",
"text": "⟨value⟩"
}
}
"""
```

Supplementary Note 2.2 Information Extraction and Verification Prompt

2. Prompts are used to extract fine-grained injury typologies from generated imaging reports and strictly verify their consistency, acting as a quality assurance gate to eliminate semantic hallucinations or logical contradictions.

```
"""
You are a professional forensic medical information extraction and
verification expert. Your task is to analyze the input virtual autopsy
text, extract key pathological information, and strictly VERIFY it against
predefined forensic taxonomies. Output in pure JSON format.

Output Requirements
(1) Output in pure JSON format without any markdown blocks or additional
text.
(2) The JSON structure must reflect the hierarchical nature of the report,
containing the following fields:
- verified_impression: A nested object containing the explicit semantic
anchors: - disease: SINGLE most prominent injury/disease type (Must
be exactly from the 40 predefined categories). - location: SINGLE most
prominent anatomical location (Must be exactly from the 8 core regions). -
concise_description: Key evidence description.
- verification_status: "Pass" or "Fail".
- error_reason: Leave empty if "Pass". If "Fail", state the logical
contradiction (e.g., "Text says normal aorta, but disease claims Aortic
```

```
dissection").

Verification & Extraction Principles
1.  Disease Category Verification (from Impression):
- Identify the SINGLE most prominent forensic pathology from the text.
- You MUST map it to exactly ONE of the following 40 categories:  [Healthy,
Laryngeal edema, Putrefactive gas bubbles, Soft tissue hematoma, Fractures,
Aortic rupture, Gastric content reflux, Incised wounds, Skin burns, Intimal
tear, Blood sedimentation, Airway burns, Hypostatic pulmonary edema,
Brain herniation, Visceral autolysis, Airway and digestive tract fluid,
Hemopericardium, Internal bleeding, Pulmonary edema, Skull fracture,
Internal organ injury, Pericardial effusion, Hemopneumothorax, Ballistic
hemorrhage, Hyoid bone fracture, Visceral perforation, High cervical
fracture-dislocation, Neck hematoma, Fracture along the wound track, Aortic
dissection, Tissue necrosis, Coronary artery calcification, Cerebral edema,
Hyperdense lesion in brain parenchyma, Intracranial hemorrhage, Pulmonary
embolism, Sinus fluid, Airway foreign body, Periaortic hematoma, Coronary
artery occlusion].
- If the text describes a completely normal state, select "Healthy".

2.  Anatomical Location Verification (from Impression):
- Map the primary location to exactly ONE of the 8 core regions:  [Head,
Neck, Chest, Abdomen, Pelvis, Unknown, Upper Leg, Lower Leg].
- If multiple are involved, choose the one with the primary lethal lesion.

3.  Key Description Extraction:
- Summarize core forensic/imaging features in 1-2 sentences.
- Retain explicit anchors (e.g., fracture size, fluid density, exact tissue
involved).

Processing Workflow
1.  Locate the forensic findings in the text, specifically identifying the
'impression' cues.
2.  Extract the explicit semantic anchors (disease and location) and map
them to the strict taxonomies.
3.  Consistency Check (Crucial):  Verify if the text logically supports the
mapped disease.  If the text asserts the absence of a lesion but the mapped
disease is a positive injury, mark verification_status as "Fail".
4.  Synthesize the concise description.

Now please analyze and verify the following forensic text description:
"""
```

Supplementary Table S1. Standardized Virtual Autopsy Dataset Acquisition and Reconstruction Protocols. Detailed scanning parameters (kV, mAs, pitch) and reconstruction window settings (Window Level/Width) applied across core anatomical regions to ensure high-fidelity imaging for the dataset.

Region	Parameter Type	Specific Parameters
Head and Neck	Scanning Params	120 kV, 380 mAs, Collimator 64Œ0.625 mm, Slice 0.625 mm, Pitch 0.4
	Reconstruction - Brain Window	WL: 40, WW: 85
	Reconstruction - Bone Window	WL: 800, WW: 2000
	Reconstruction - Soft Tissue Window	WL: 45, WW: 300
Thorax and Abdomen	Scanning Params	120 kV, 227 mAs (227-400), Collimator 64Œ0.625 mm, Slice 1.0 mm, Pitch 0.6
	Reconstruction - Lung Window	WL: -600, WW: 1500
	Reconstruction - Mediastinum Window	WL: 40, WW: 400
	Reconstruction - Bone Window	WL: 600, WW: 2000
Pelvis and Lower Limbs	Scanning Params	120 kV, 250-400 mAs, Collimator 64Œ0.625 mm, Slice 1.0 mm, Pitch 0.6
	Reconstruction - Bone Window	WL: 600, WW: 2000
	Reconstruction - Soft Tissue Window	WL: 50, WW: 350

Supplementary Table S2. Statistical Composition of the Large-Scale Pre-training Dataset. The table details the distribution of 1,257,349 image-text pairs across diverse causes of death, serving as the foundational data for GCM-CLIP pre-training.

Cause of Death	Number of Text-Image Pairs
traffic_accident	331,595
high_fall_death	216,985
drowning	200,701
sudden_death	97,610
mechanical_asphyxia	78,415
sharp_force_injury_death	71,495
blunt_trauma_death	65,536
coronary_heart_disease_death	48,810
natural_death	32,114
mechanical_injury_death	25,670
aortic_rupture_death	24,815
cerebral_hemorrhage_death	23,836
electrocution	18,761
burn	11,190
aspiration_infarction	5,220
gunshot_wound_death	4,596
	Total: 1,257,349

Supplementary Table S3. Statistical Summary of the Fine-Grained Annotated Dataset for Model Fine-Tuning and Evaluation. This table details the hierarchical distribution of 203,478 image-text pairs across 8 anatomical locations, 40 disease-level classes, and 98 instance-level subtypes, serving as the benchmark for few-shot classification and performance validation.

Location	Disease-level	Instance-level (Full Name)	Instance-level (Abbreviation)	Number of Image-Text Pairs
Abdomen	Blood Sed	Abdominal Mass	Abd Mass	507
Abdomen	Gastric Reflux	Aspiration	Aspiration	1,521
Abdomen	Aortic Rupture	Contrast Extravasation	Contrast Extrav	247
Abdomen	Blood Sed	Complex Cystic and Solid Mass	Cx Mass	201
Abdomen	Gastric Reflux	Esophageal Air-Fluid Level	Esoph AFL	442
Abdomen	Gastric Reflux	Esoph Hyperdense	Esoph Hyperdense	1,416
Abdomen	Gastric Reflux	Gastroesophageal Reflux	GER	21,488
Abdomen	Putrefactive Gas	Gas-containing abscess	Gas Abscess	274
Abdomen	Internal Hemo	Hematoma	Hematoma	3,565
Abdomen	Internal Hemo	Hemopertioneum	Hemopertioneum	5,120
Abdomen	Blood Sed	Hyperdense Abdominal Lesion	Hyperdense Abd Lesion	283
Abdomen	Blood Sed	Hyperdense Abdominal Sedimentation	Hyperdense Abd Sed	281
Abdomen	Blood Sed	Hyperdense Sedimentation	Hyperdense Sed	511
Abdomen	Internal Hemo	Intra-abdominal hemorrhage	IA Hemo	359
Abdomen	Putrefactive Gas	Intracranial Gas	IC Gas	2,418
Abdomen	Blood Sed	Intracranial Hemorrhage	ICH	2,305
Abdomen	Internal Hemo	Intraperitoneal hemorrhage	IPHemo	292
Abdomen	Organ Inj	Hepatic hemorrhage	Liver Hemo	359
Abdomen	Organ Inj	Hepatic injury	Liver Inj	2,765
Abdomen	Organ Inj	Hepatic laceration	Liver Lac	2,348
Abdomen	Gastric Reflux	Oropharyngeal foreign material	OP FB	218
Abdomen	Gastric Reflux	Oropharyngeal Reflux	OP Reflux	2,811
Abdomen	Internal Hemo	Pelvic Hemorrhage	Pelvic Hemo	2,234
Abdomen	Gastric Reflux	Pharyngeal Gastric Content	Pharyn Gastric Cont	258
Abdomen	Gastric Reflux	Pharyngeal Reflux	Pharyn Reflux	1,704
Abdomen	Putrefactive Gas	Pneumatosis	Pneumatosis	2,363
Abdomen	Putrefactive Gas	Pneumocephalus	Pneumocephalus	4,025
Abdomen	Putrefactive Gas	Pneumoperitoneum	Pneumoperit	3,060
Abdomen	Gastric Reflux	Posterior Pharyngeal Hyperdensity	Post Pharyn Hyperdense	1,172
Abdomen	Autolysis	Post-mortem changes	Postmortem	8,416
Abdomen	Internal Hemo	Retroperitoneal hemorrhage	RP Hemo	1,240
Abdomen	Blood Sed	Subarachnoid Hemorrhage	SAH	776
Abdomen	Blood Sed	Subdural Hematoma	SDH	238
Abdomen	Blood Sed	Soft Tissue Calcification	ST Calc	293
Abdomen	Putrefactive Gas	Soft Tissue Emphysema	ST Emphysema	15,071
Abdomen	Blood Sed	Soft Tissue Hyperdensity	ST Hyperdense	564
Abdomen	Blood Sed	Soft Tissue Mass	ST Mass	1,067
Abdomen	Organ Inj	Splenic injury	Splenic Inj	4,486
Abdomen	Putrefactive Gas	Gas Emphysema	SubQ Emphys	1,820
Abdomen	Blood Sed	Target Lesion	Target Lesion	417
Chest	Pulm Edema	Cardiogenic pulmonary edema	Cardio Edema	498
Chest	Pericard Eff	Cardiac silhouette enlargement due to pericardial effusion	Cardiomegaly (PE)	204
Chest	Hypostatic Edema	Dependent Opacities	Dependent Opac	1,108

Supplementary Table S3. Continued: Statistical Summary of the Fine-Grained Annotated Dataset

Location	Disease-level	Instance-level (Full Name)	Instance-level (Abbreviation)	Number of Image-Text Pairs
Chest	Pulm Edema	Interstitial and Ground-Glass Opacities	Interst & GGO	912
Chest	Pulm Edema	Interstitial Edema	Interst Edema	945
Chest	Pulm Edema	Interstitial and Alveolar Edema	Interst/Alv Edema	1,525
Chest	Hypostatic Edema	Lower lobe predominant pulmonary edema	LL Edema	1,096
Chest	Hypostatic Edema	Lower lung zone predominant ground-glass opacities	LL GGO	486
Chest	Hypostatic Edema	Lower Lung Zone Opacities	LL Opac	1,680
Chest	Hemopericard	Pericardial Effusion	Pericard Eff	1,001
Chest	Pericard Eff	Simple pericardial effusion	Simple Pericard Eff	237
Chest	Aortic Dissect	Stanford Type A	Stanford A	2,925
Chest	HPTX	Tension Hemopneumothorax	Tension HPTX	7,221
Chest	Pericard Eff	Normal	nl	614
Head	CE	Cerebral edema	CE	818
Head	CE	Cytotoxic Edema	Cyto Edema	844
Head	CE	Diffuse Cerebral Edema	DCE	851
Head	Ballistic Hemo	Intraparenchymal Hemorrhage	IPH	309
Head	ICH	Intraparenchymal Hemorrhage	IPH	15,244
Head	ICH	Intraparenchymal hemorrhage with mass effect	IPH-ME	10,049
Head	ICH	Intraparenchymal hemorrhage with mass effect and midline shift	IPH-ME/MS	3,023
Head	Skull Fx	Linear Skull Fracture	Linear Skull Fx	3,088
Head	Sinus Fluid	Maxillary Sinus Opacification	Max Sinus Opac	2,251
Head	Sinus Fluid	Maxillary sinusitis	Max Sinusitis	2,291
Head	Incised Wnd	Soft Tissue Injury	ST Injury	677
Head	Sinus Fluid	Sinus fluid	Sinus Fluid	4,859
Head	Sinus Fluid	Paranasal Sinus Opacification	Sinus Opac	8,710
Head	ICH	Sphenoid wing hemorrhage	Sphenoid Hemo	1,394
Head	Brain Hern	Subfalcine Herniation	Subfalcine Hern	2,322
Head	Incised Wnd	Normal	nl	614
Head	Burns	Normal Head CT	nl Head CT	781
Neck	High C-spine Fx/Lux	Atlanto-occipital dislocation	AOD	460
Neck	Airway/Digest Fluid	Fluid in the airway and digestive tract	Airway/Digest Fluid	301
Neck	Neck Hema	Anterior Neck Hematoma	Ant Neck Hematoma	1,882
Neck	Airway/Digest Fluid	Bowel Obstruction	Bowel Obstr	297
Neck	Airway/Digest Fluid	Cervical Fluid Accumulation	C-spine Fluid	218
Neck	Intimal Tear	Carotid artery intimal tear	Carotid Tear	202
Neck	ST Hematoma	Heterogeneous soft tissue density with irregular margins	Hetero ST Density	247
Neck	Hyoid Fx	Hyoid Bone Cortical Disruption	Hyoid Fx-Disrupt	678
Neck	Hyoid Fx	Hyoid bone fracture with discontinuity and irregularity	Hyoid Fx-Irreg	484

Supplementary Table S3. Continued: Statistical Summary of the Fine-Grained Annotated Dataset

Location	Disease-level	Instance-level (Full Name)	Instance-level (Abbreviation)	Number of Image-Text Pairs
Neck	Hyoid Fx	Hyoid Bone Fracture with Discontinuity and Misalignment	Hyoid Fx-Mal	216
Neck	ST Hematoma	Hyperdense Soft Tissue Collection	Hyperdense ST Coll	732
Neck	ST Hematoma	Hyperdense Soft Tissue Hematoma	Hyperdense ST Hema	675
Neck	ST Hematoma	Hyperdense Soft Tissue Lesion	Hyperdense ST Lesion	897
Neck	ST Hematoma	Intramuscular Hematoma	IM Hematoma	543
Neck	Larynx Edema	Laryngeal Soft Tissue Edema	Larynx Edema	229
Neck	Airway/Digest Fluid	Retropharyngeal Fluid Collection	Retropharyn Fluid Coll	227
Neck	Neck Hema	Soft Tissue Hematoma	ST Hematoma	3,888
Pelvis	Fx	Bilateral iliac wing fractures	Bilat Iliac Fx	816
Pelvis	Fx	Cervical Spine Fracture	C-spine Fx	670
Pelvis	Fx	Comminuted Fracture	Comminuted Fx	3,422
Pelvis	Fx	Iliac Wing Fractures	Iliac Fx	246
Pelvis	Fx	Left iliac wing fracture	L Iliac Fx	331
Pelvis	Fx	Pelvic Ring Fracture	Pelvic Ring Fx	6,082
Pelvis	Fx	Rib fractures	Rib Fx	353
Pelvis	Fx	Skull Fracture	Skull Fx	836
Pelvis	Fx	Transverse Fracture	Trans Fx	480
Pelvis	Fx	Transverse Tibial Fracture	Trans Tib Fx	554
			Total	**203,478**

Supplementary Table S4. Distribution of the ForVA Test Set. This table details the composition of 4,096 image-text pairs from our self-constructed forensic dataset. To optimize space, the categories are split into two columns.

Disease-level	Pairs	Disease-level	Pairs
Healthy	124	Laryngeal edema	104
Putrefactive gas bubbles	111	Soft tissue hematoma	104
Fractures	110	Aortic rupture	104
Gastric content reflux	109	Incised wounds	104
Skin burns	108	Intimal tear	103
Blood sedimentation	107	Airway burns	103
Hypostatic pulmonary edema	107	Brain herniation	103
Visceral autolysis	107	Airway and digestive tract fluid	103
Hemopericardium	106	Internal bleeding	103
Pulmonary edema	106	Skull fracture	103
Internal organ injury	106	Pericardial effusion	103
Hemopneumothorax	106	Ballistic hemorrhage	102
Hyoid bone fracture	106	Visceral perforation	102
High cervical fracture-dislocation	106	Neck hematoma	102
Fracture along the wound track	106	Aortic dissection	102
Tissue necrosis	105	Coronary artery calcification	102
Cerebral edema	105	Hyperdense lesion in brain parenchyma	86
Intracranial hemorrhage	105	Pulmonary embolism	80
Sinus fluid	105	Airway foreign body	72
Periaortic hematoma	104	Coronary artery occlusion	62
		Total Image-Text Pairs: 4,096	

Supplementary Table S5. Composition of the External Zero-Shot Test Set: MIMIC-CXR. A subset of 1,024 chest CT image-text pairs strictly filtered from the MIMIC-CXR dataset using "CT" keywords. This set evaluates the model's capability to transfer forensic knowledge to routine clinical chest imaging without fine-tuning.

Disease-level	**Number of Image-Text Pairs**
Pleural effusion and pleural lesions	65
Unknown cause and unclassified lesions	64
Cardiovascular system lesions	58
Atelectasis and reduction of lung volume	57
Pulmonary tumor lesions	55
Pulmonary infectious lesions	54
Benign nodules and cysts in the lungs	53
Emphysema and COPD	53
Chest wall and bone lesions	52
Interstitial lung disease	52
Mediastinal lesions	51
Normal and other rare lesions	51
Postoperative/radiation therapy changes	51
Diaphragmatic abnormalities and diaphragmatic hernia	51
Bronchopathy	51
Calcifying lesions in the lungs	51
Abdominal related chest involvement lesions	50
Thyroid lesions	41
Pulmonary hemorrhage and contusion	40
Pulmonary embolism and pulmonary vascular disease	24
	Total: 1,024

Supplementary Table S6. Composition of the External Zero-Shot Test Set: PMC-OA. A curated subset of 4,096 image-text pairs derived from biomedical literature in the PMC-OA dataset. Screened for chest CT modalities, this set assesses the model's generalization across diverse, literature-sourced medical imagery.

Disease Level	Near-Equal Number
Other/Complex and Rare Diseases	205
Nodule and Mass	205
Vascular Abnormalities	205
Ground-Glass Opacity/Consolidation	205
Trauma/Fracture	205
Cyst	205
Lymphadenectasis	205
Diffuse Bone Destruction	205
Obstruction/Dilatation	205
Hemorrhage/Hematoma	205
Infarction/Ischemia	205
Calcification	204
Pneumoperitoneum/Free Gas	204
Effusion	204
Necrosis/Liquefaction	204
Fibrosis/Scar	205
Osteoproliferation/Sclerosis	205
Fat Density	205
Edema/Exudation	205
Cavity	205
	Total:4,096

Supplementary Table S7. Composition of the External Zero-Shot Test Set: ROCOv2. A selection of 4,096 multimodal radiology images and captions from the ROCOv2 dataset. This subset challenges the model with varied biomedical imaging contexts to rigorously test its zero-shot cross-domain adaptability.

Disease Level	Number of Image-Text Pairs
Nodule_Mass	205
Necrosis_Liquefaction	205
Trauma_Fracture	205
OtherCategories	205
Obstruction_Dilatation	205
Cyst	205
Hemorrhage_Hematoma	205
GroundGlassOpacity_Consolidation	205
Effusion	205
Pneumoperitoneum_FreeGas	205
Edema_Exudation	205
VascularAbnormalities	205
Calcification	204
Cavity_Sinus	205
Infarction_Ischemia	205
DiffuseBoneDestruction	205
Fibrosis_Scar	205
FatDensity	204
BoneHyperplasia_Sclerosis	204
Lymphadenopathy	204
	Total:4,096

Supplementary Table S8. Statistical Distribution of the Dedicated Object Detection Dataset. This table details the sample composition for the six key forensic lesion categories (e.g., Pulmonary Edema, Hyoid Fracture) used in the downstream object detection task. These expert-annotated samples serve as a fine-grained benchmark for evaluating the model's localization generalization.

Lesion Category	**Number of Samples**
Sinus Fluid	379
Pulm Edema	4,190
HPTX	291
Gastric Reflux	1,572
Hyoid FX	20
Airway/Digest Fluid	2,778
	Total: 9,230

Supplementary Table S9. Comparative Evaluation of Zero-Shot Classification Performance on the ForVA Benchmark. This table presents a quantitative comparison of GCM-CLIP against state-of-the-art baseline models across three hierarchical tasks: anatomical localization (`Loc`), disease identification (`Dis`), and joint fine-grained classification (`Dis_Loc`). The substantial performance gains (e.g., +0.099 in Loc@1 and +0.051 in Dis@1 vs. BiomedCLIP) validate the efficacy of the *Explicit-Implicit Multi-task Supervision* in overcoming forensic image noise and establishing robust cross-modal semantic mappings.

Model	Dis_Loc@1	Dis_Loc@5	Dis@1	Dis@5	Loc@1	Loc@5
PubMedCLIP	0.0175	0.1552	0.0856	0.2421	0.2614	0.4960
BiomedCLIP	0.0993	0.2519	0.2028	0.3981	0.5546	0.9187
GLoRIA	0.0769	0.2356	0.1530	0.2844	0.4328	0.5827
MGCA	0.0759	0.2788	0.1530	0.3505	0.3413	0.5507
medsiglip	0.0478	0.2458	0.1696	0.3442	0.2417	0.4006
GCM-CLIP	**0.1220**	**0.2858**	**0.2536**	**0.4169**	**0.6530**	**0.9502**

Supplementary Table S10. Quantitative Assessment of Cross-Modal Semantic Alignment via Retrieval Tasks. This table reports the Recall@K metrics (R@1, R@5, R@10) for bidirectional retrieval tasks. GCM-CLIP consistently outperforms state-of-the-art baselines across all metrics, achieving substantial gains (e.g., $\sim$0.08–0.10 improvement in Image-to-Text accuracy over BiomedCLIP). These results validate the effectiveness of the *Generalized Category Mining (GCM)* mechanism in disentangling complex forensic semantics and enhancing feature discrimination through explicit-implicit supervision.

Model	image_to text_R@1	image_to text_R@5	image_to text_R@10	text_to image_R@1	text_to image_R@5	text_to image_R@10
PubMedCLIP	0.0752	0.2763	0.4345	0.0905	0.2995	0.4440
BiomedCLIP	0.2431	0.5459	0.6823	0.2500	0.5407	0.6796
GLoRIA	0.1728	0.4870	0.6591	0.1679	0.4960	0.6630
MGCA	0.2294	0.5761	0.7089	0.2448	0.5722	0.7177
medsiglip	0.2731	0.5937	0.7487	0.2675	0.6032	0.7448
GCM-CLIP	**0.3230**	**0.6425**	**0.7675**	**0.3134**	**0.6413**	**0.7702**

Supplementary Table S11. Assessment of Data Efficiency and Generalization Capabilities for Unseen Diseases. This table compares model performance across varying training data fractions. GCM-CLIP demonstrates a breakthrough in data efficiency, addressing the scarcity of labeled samples for rare forensic pathologies. Notably, in the extreme scenario using only 1% of labeled data, GCM-CLIP achieves an accuracy of 0.5429 on the Instance-level (Unseen Disease) task, significantly outperforming the baseline BiomedCLIP trained with 100% full annotations (0.4860).

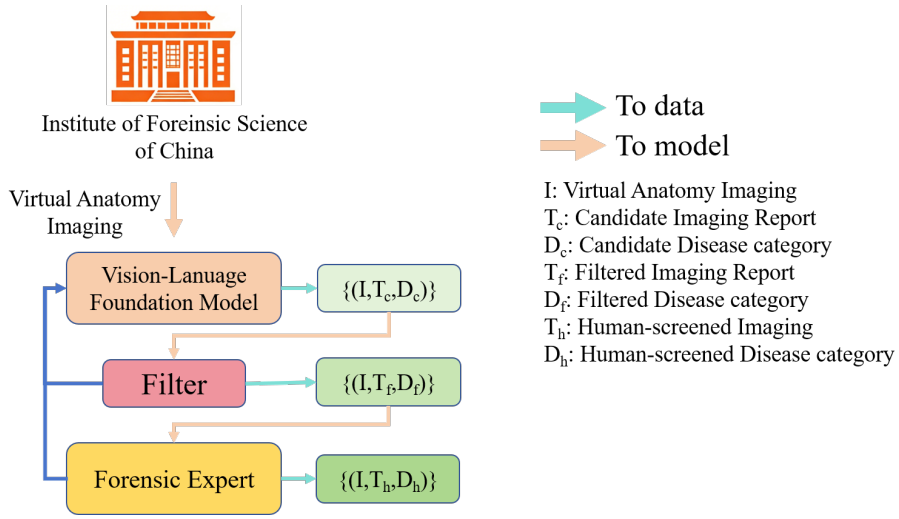
Model	Location			Disease-level			Instance-level (Unseen Disease)		
	1%	10%	100%	1%	10%	100%	1%	10%	100%
PubMedCLIP	0.4240	0.4223	0.4035	0.0512	0.0324	0.0324	0.5716	0.5826	0.5875
BiomedCLIP	0.6149	0.6928	0.7627	0.3584	0.6167	0.6564	0.4528	0.4770	0.4860
GLoRIA	0.6169	0.6237	0.6171	0.1084	0.1103	0.0949	0.5071	0.5394	0.5421
MGCA	0.7470	0.7487	0.7460	0.4519	0.5173	0.5466	0.5281	0.5575	0.6027
medsiglip	0.6264	0.6037	0.6076	0.5141	0.6420	0.6743	0.5271	**0.5971**	0.5980
GCM-CLIP	**0.7580**	**0.7688**	**0.7793**	**0.5976**	**0.7155**	**0.7238**	**0.5429**	0.5849	**0.6215**

Supplementary Table S12. Quantitative evaluation of multi-granularity semantic disentanglement. The Normalized Mutual Information (NMI) scores measure the correlation between the mined implicit clusters and explicit forensic labels. Note the significant increase in pathological alignment from Set A to Set B, while anatomical correlation remains consistently low.

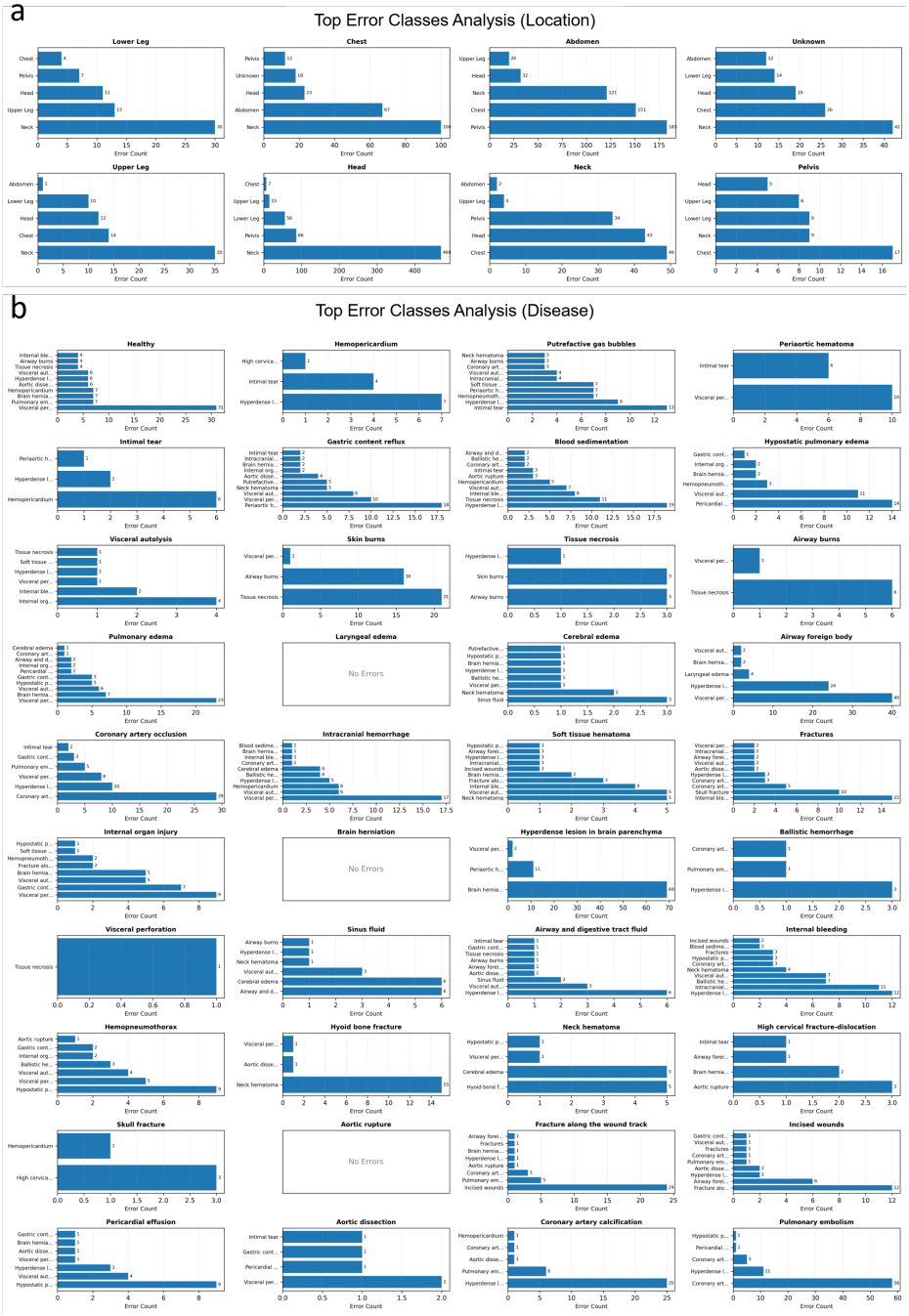
		Semantic Correlation (NMI)		
Cluster Set	**Semantic Granularity**	**Disease**	**Location**	**Δ Gap**
Set A	Coarse-grained	0.6730	0.3152	0.3578
Set B	Fine-grained	0.8298	0.3774	0.4524

Supplementary Table S13. Ablation Study Quantifying the Contribution of Core GCM Components. This table presents a systematic evaluation of model performance under different configurations of Explicit Supervision (Exp. Sup.), Implicit Supervision (Imp. Sup.), and the GOS module. The results demonstrate that while Explicit Supervision establishes foundational alignment, its combination with Implicit Supervision creates a powerful synergistic effect, driving a substantial performance leap on the domain-specific ForVA dataset (Loc@1 increases to 0.6530). The final inclusion of GOS further enhances generalization, achieving peak accuracy across all four testing benchmarks.

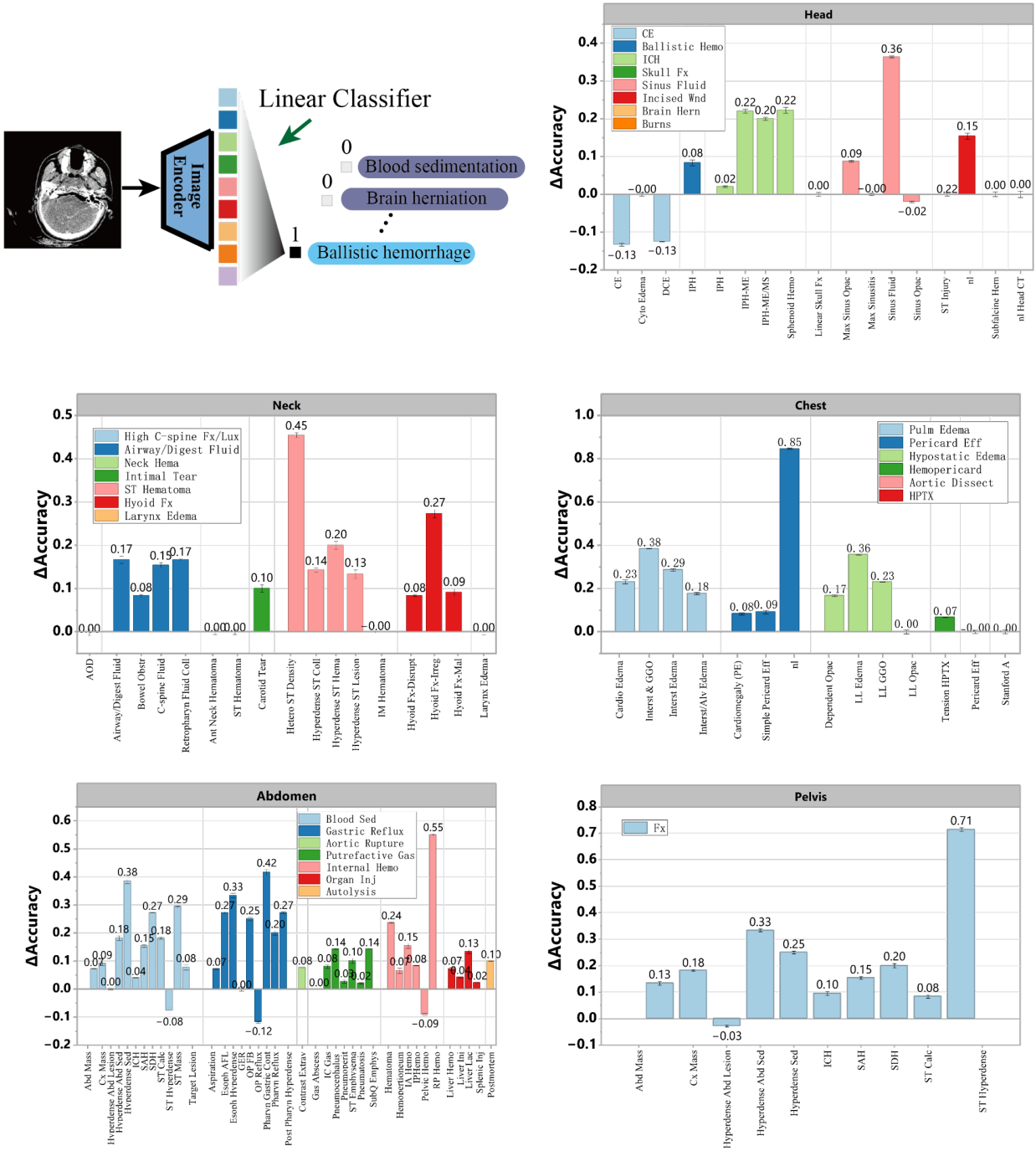
Exp. Sup.	Imp. Sup.	GOS	roco		mimic-cxr		pmc-oa		ForVA	
			Loc@1	Cat@1	Loc@1	Cat@1	Loc@1	Cat@1	Loc@1	Cat@1
			0.3452	0.1306	0.1333	0.1499	0.3491	0.1042	0.5544	0.2026
✓			0.4223	0.1440	0.3828	0.1508	0.4008	0.1093	0.5897	0.2290
	✓		0.3264	0.1396	0.0976	0.1440	0.3750	0.0991	0.3960	0.1291
✓	✓		0.5854	0.1702	0.5642	0.1528	0.4194	0.1179	0.6227	0.2368
✓	✓	✓	**0.6073**	**0.1707**	**0.5790**	**0.1632**	**0.5086**	**0.2157**	**0.6530**	**0.2536**



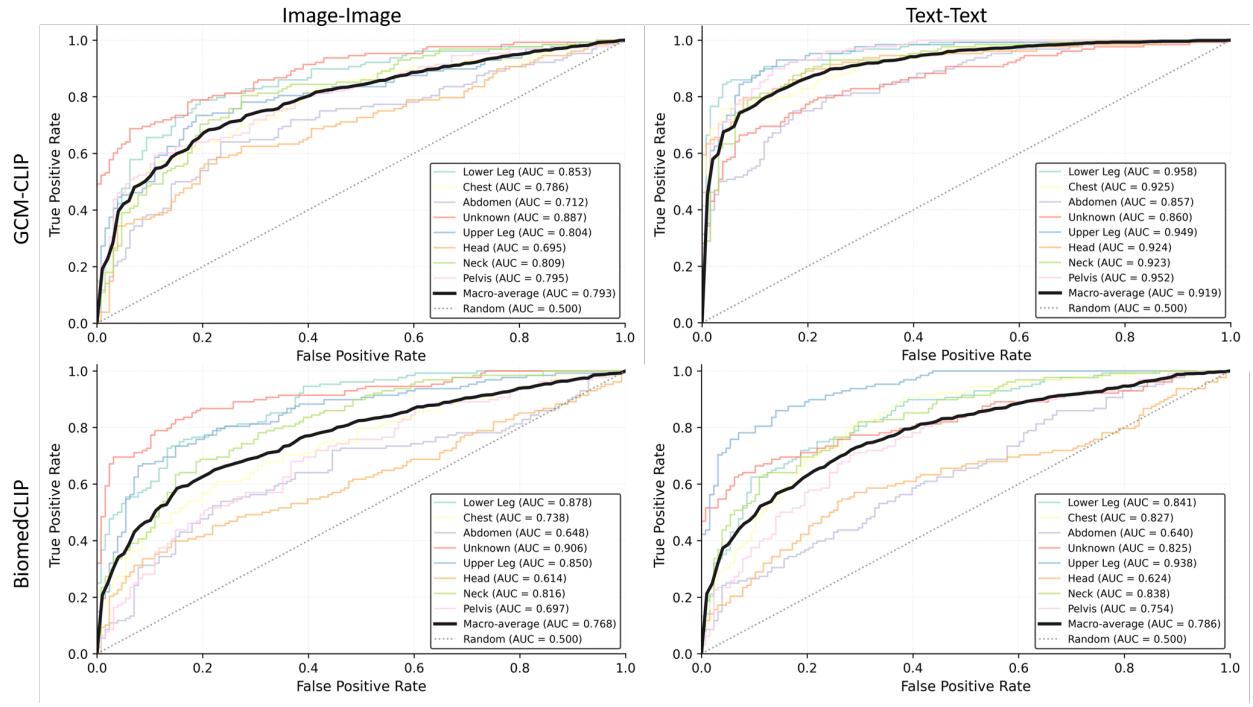
Supplementary Figure S1. Iterative data curation pipeline for forensic virtual anatomy. Raw virtual anatomy imaging (I) sourced from the Institute of Forensic Science of China undergoes a coarse-to-fine annotation process. Initially, a Vision-Language Foundation Model generates candidate imaging reports (T_c) and disease categories (D_c). To ensure evidentiary reliability, these outputs are processed by an automated Filter to produce a refined subset (T_f, D_f), followed by rigorous screening by Forensic Experts to establish the final human-validated ground truth (T_h, D_h). Crucially, a feedback mechanism (blue arrows) leverages verification outcomes from both the filtering and expert screening stages to iteratively refine the foundation model. Teal arrows denote the generation of data outputs, while peach arrows indicate the sequential processing flow.



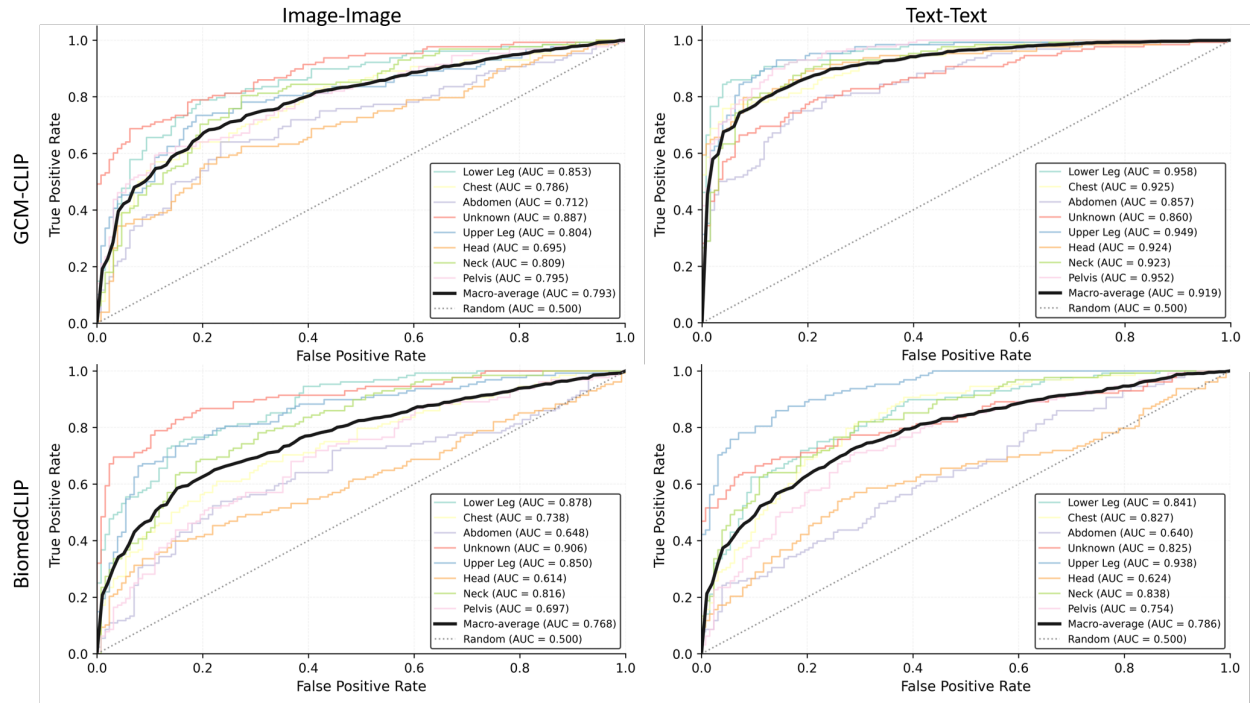
Supplementary Figure S2. Error Class Distribution of GCM-CLIP. (a) Bar charts showing the frequency of mispredicted anatomical locations when the model fails to identify the true site. (b) Bar charts showing the distribution of mispredicted injury categories, with "No Errors" denoting categories without misclassifications. Our analysis reveals that the model's error distribution adheres to the forensic "injury complex" principle: misclassifications tend to involve pathologically related adjacent categories rather than random noise. For instance, Intracranial hemorrhage is often misclassified as Cerebral edema or Skull fracture, reflecting the pathological cascade of skull fracture causing traumatic hemorrhage, which in turn induces secondary edema. Similarly, Intimal tear is frequently misclassified as Hemopericardium, aligning with the clinical pathology that aortic intimal tears often precede rupture into the pericardial cavity. These findings demonstrate that GCM-CLIP implicitly learns the pathological associations between injuries.



Supplementary Figure S3. Evaluation of Fine-Grained Representation Capabilities via Linear Probing. The schematic (top left) illustrates the linear probing protocol used to assess the discriminative quality of frozen image features. The bar charts present the differential accuracy (ΔAccuracy), calculated as the performance of GCM-CLIP minus the baseline BiomedCLIP, across five core anatomical regions (Head, Neck, Chest, Abdomen, and Pelvis). Positive values indicate superior performance of GCM-CLIP. The results demonstrate significant gains in identifying subtle and abstract forensic pathological subtypes. Notably, in the **Head** region, GCM-CLIP achieves a performance boost of 0.22 for Intracerebral hemorrhage with Mass Effect (IPH-ME) and 0.22 for Sphenoid Sinus Hemorrhage (Sphenoid Hemo)—a critical indicator for skull base fractures. In the **Abdomen** region, the model exhibits a substantial gain of 0.55 for Retroperitoneal Hematoma (RP Hemo), validating its robustness in detecting deep-seated traumatic features and complex injuries often compromised by post-mortem artifacts.



Supplementary Figure S4. Qualitative Visualization of Coarse-Grained Semantic Disentanglement via Sankey Diagrams. This diagram illustrates the mapping dynamics between the macroscopic coarse clusters generated by the GCM mechanism and explicit forensic labels (anatomical locations and lesion categories). The non-bijective flows reveal that these coarse clusters broadly group macroscopic lesion mechanisms and "pathological families" rather than merely memorizing static explicit anchors. This visualization qualitatively confirms the GCM mechanism's effectiveness in isolating ontological pathological features from redundant anatomical background noise at a macroscopic level.



Supplementary Figure S5. ROC Curve Analysis for Anatomical Region Identification. This figure illustrates the Receiver Operating Characteristic (ROC) curves evaluating the model's performance in discriminating across eight core anatomical regions (e.g., Head, Neck, Chest, Abdomen, and Pelvis). The tight clustering of curves in the upper-left quadrant demonstrates GCM-CLIP's exceptional robustness in capturing macroscopic anatomical features. This high discriminative power confirms that the GCM mechanism effectively filters out post-mortem artifacts and background noise, ensuring stable and precise anatomical localization as a prerequisite for downstream pathological diagnostic tasks.