

Hybrid CNN-Vision Transformer for Rabbit Gastric Dilation Classification

First AI for Rabbit Diagnostics • 5-Fold Ensemble • 89.4% F1 • Clinical-Grade Uncertainty

Markus Clauss¹, Kerstin Müller², Francesca Del Chicca¹, Marcus Clauss³, Henning Richter¹

¹Diagnostic Imaging Research Unit (DIRU), Clinic for Diagnostic Imaging, Vetsuisse Faculty, University of Zurich, Switzerland • ²Small Animal Clinic, Department of Veterinary Medicine, Freie Universität Berlin, Germany • ³Clinic for Zoo Animals, Exotic Pets and Wildlife, Vetsuisse Faculty, University of Zurich, Switzerland



Universität
Zürich^{UZH}



THE PROBLEM

Gastric Dilation

- Life-threatening emergency
- Rapid diagnosis critical
- X-ray interpretation challenging
- Limited expertise available
- No automated tools exist

RESEARCH GAP

0/422

Rabbit AI
Papers

0/34

Reported
Hybrid in Vet

Based on systematic reviews: AI for rabbit imaging entirely absent (0/422 veterinary AI publications). Hybrid CNN-Transformer architectures remain extremely limited in veterinary imaging, with no hybrid architectures reported for veterinary diagnostic radiology.

First Reported Hybrid for Vet Diagnostics

CLINICAL EXAMPLES

Figure 1: Representative cases. **A:** Non-dilated stomach (normal). **B:** Dilated stomach (gastric dilation). Left column: laterolateral projection, right column: ventrodorsal projection.

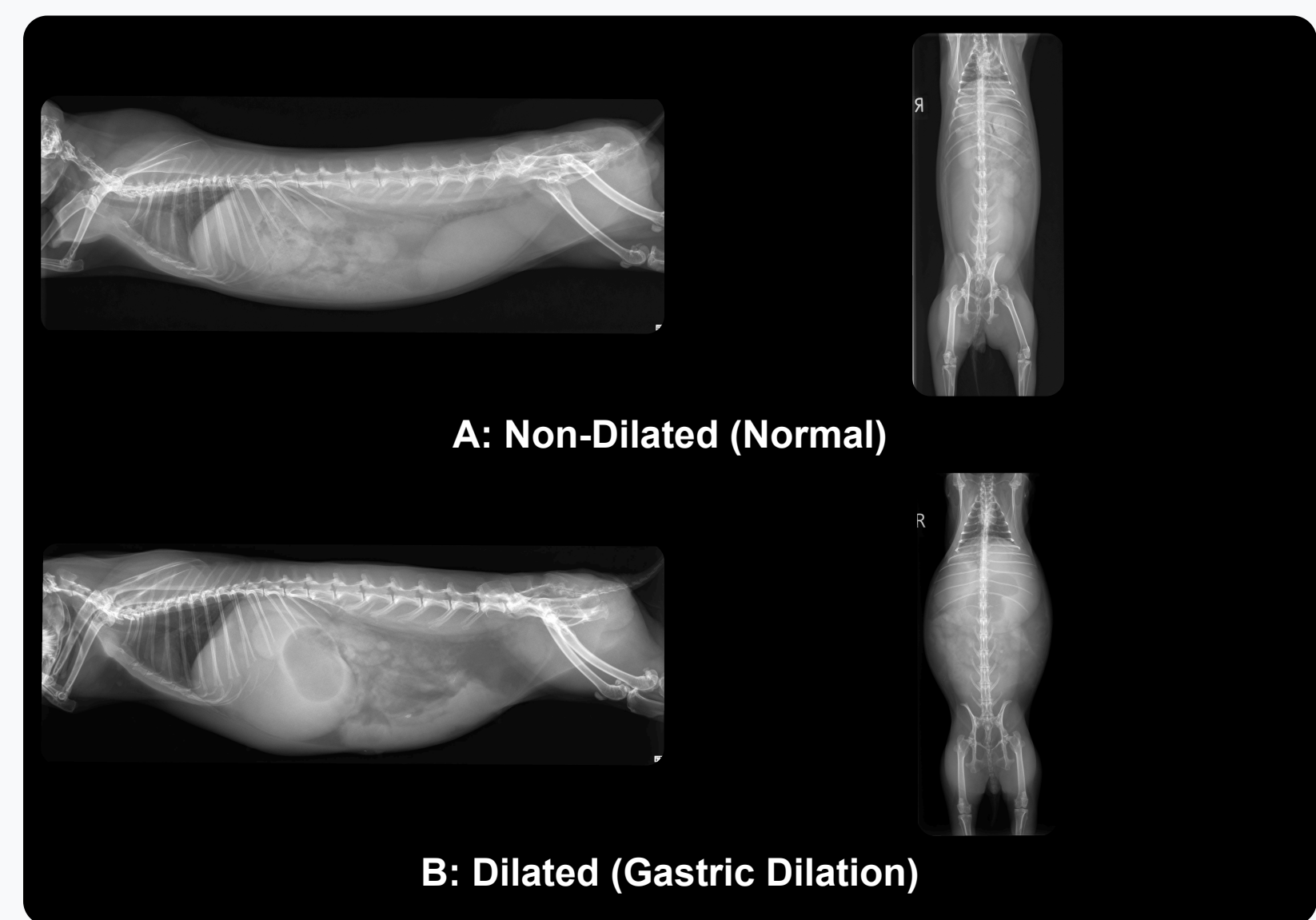
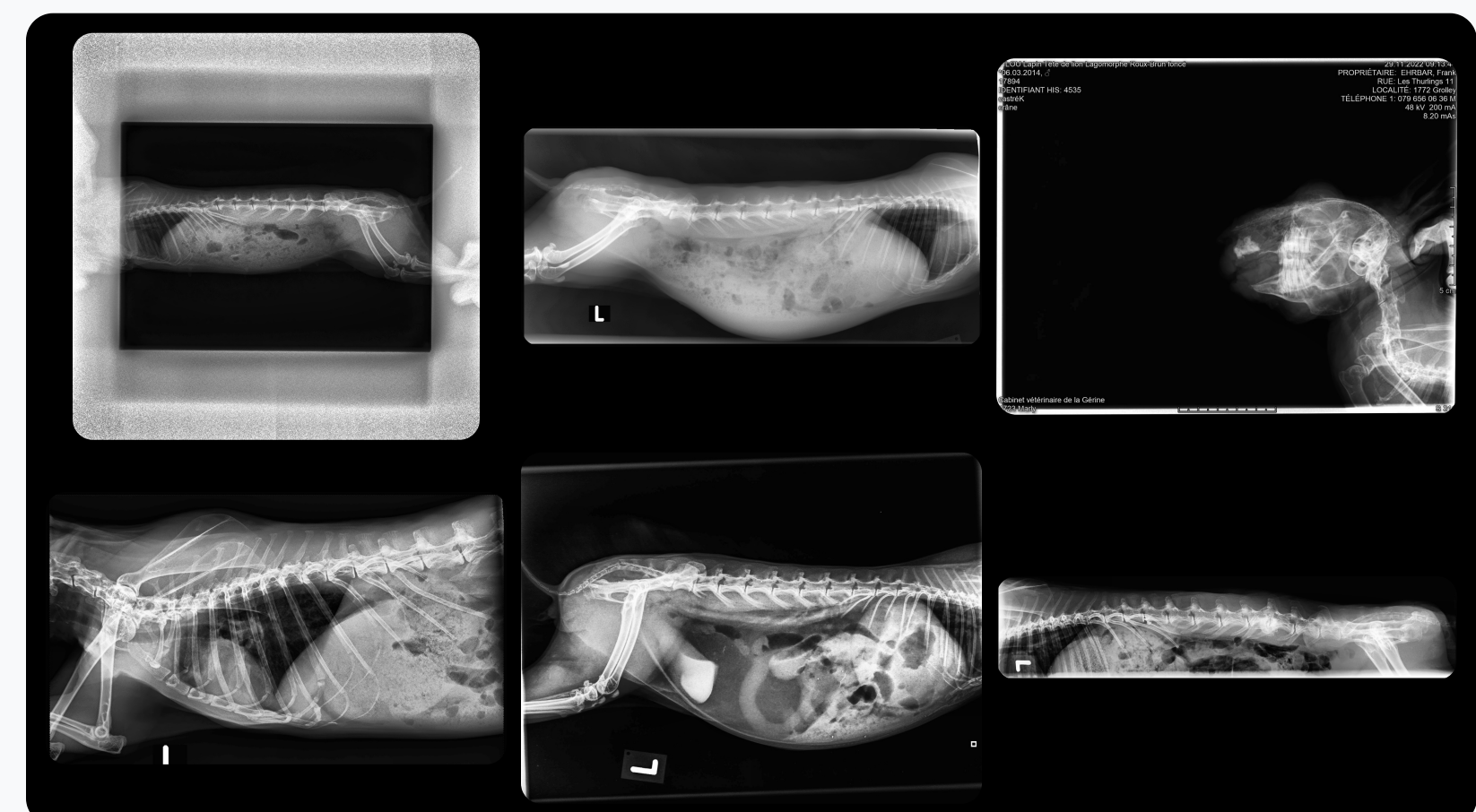


Figure 2: Outlier detection examples. Quality issues identified through manual review (n=17) and Isolation Forest algorithm (lateral: n=22, ventral: n=14). Outliers removed before training to ensure data quality.



DATASET

679 Multi-Institutional Images

ND = Non-Dilated (healthy stomach) • D = Dilated (gastric dilation)

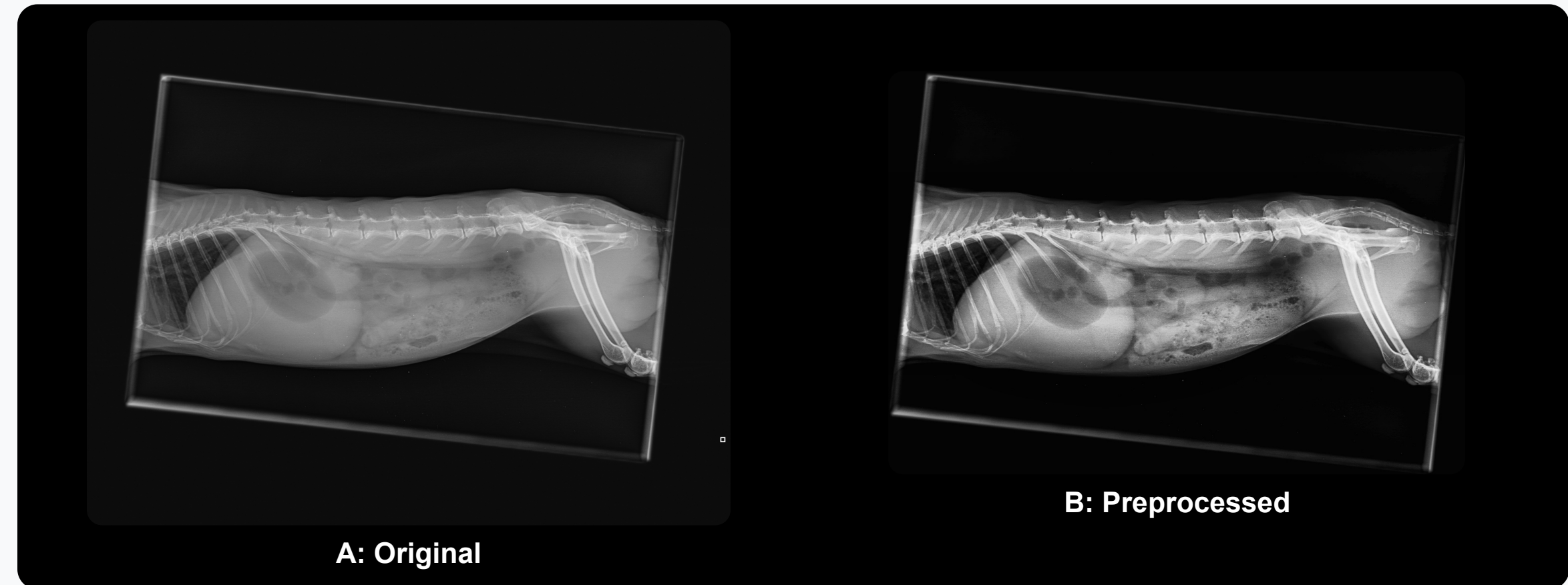
- Laterolateral: 371 (ND:211, D:160)
- Ventrodorsal: 308 (ND:167, D:141)
- Zurich 38%, Berlin 62%
- 11-year span (2014-2025)

External Validation:

- 60 images (July 2025)
- 11-month temporal separation
- 31 laterolateral, 29 ventrodorsal

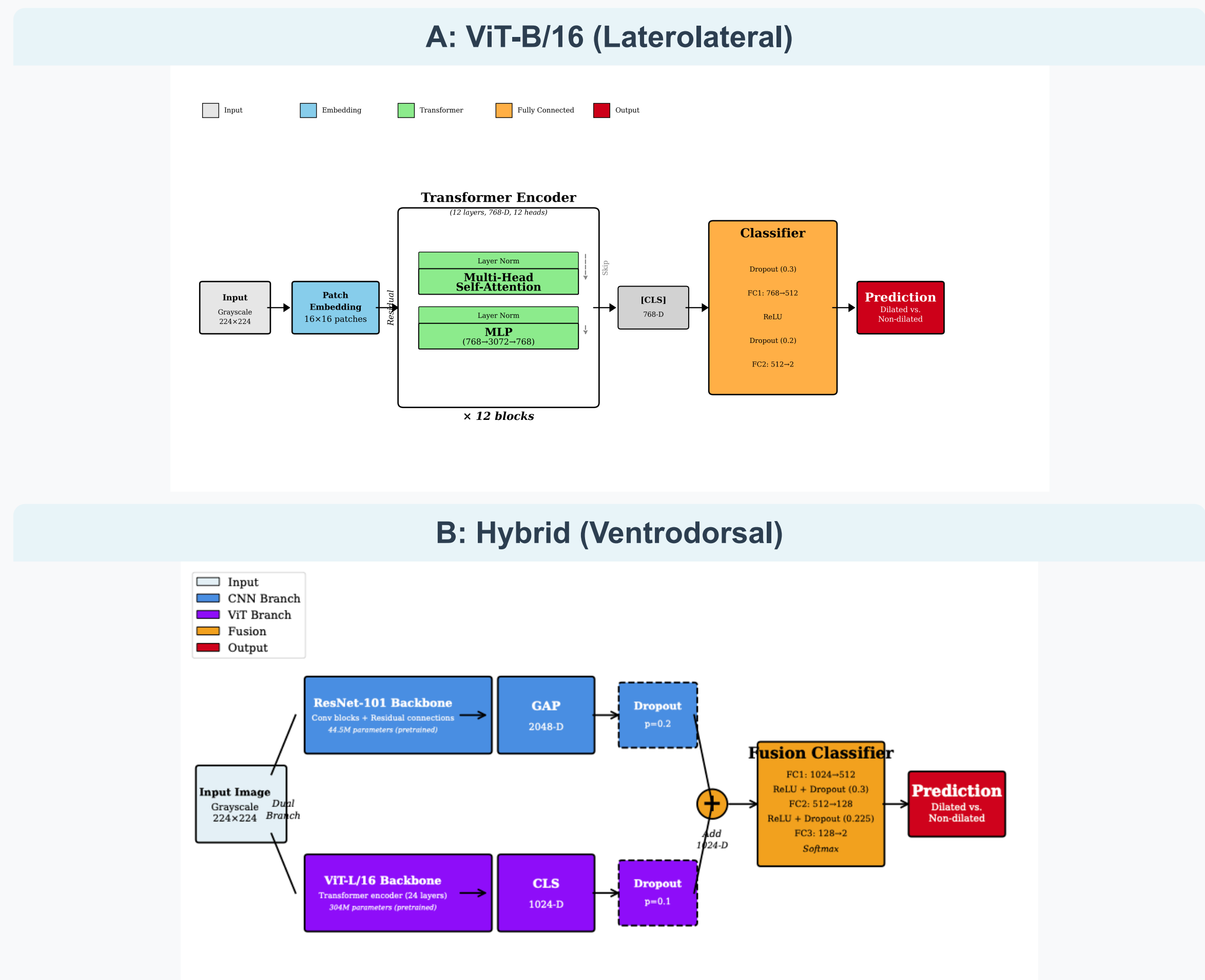
PREPROCESSING

Figure 3: Preprocessing pipeline transforms raw DICOM images. **A:** Original DICOM with min-max normalization. **B:** After windowing (WL=400, WW=1500), CLAHE enhancement (clip=0.03), and automatic cropping.



DEPLOYED ARCHITECTURES

Figure 4: Deployed architectures in production 5-fold ensemble system. **A:** ViT-B/16 (Vision Transformer, 86M parameters, deployed for laterolateral projection). **B:** Hybrid Late Fusion (CNN-Transformer fusion, 349M parameters, deployed for ventrodorsal projection). Both pretrained on ImageNet and fine-tuned on rabbit gastric dilation radiographs.



5-FOLD CV RESULTS

CV = Cross-Validation • F1 = F1 Score (harmonic mean of precision and recall) • Ext. = External validation cohort • Gap = Performance difference between test and external validation

Arch.	Test F1	Ext. F1	Gap
Laterolateral (n=371, ext n=31)			
ViT-B/16	91.62%	87.26%	4.36%
Hybrid	91.46%	87.19%	4.27%
ViT-L/16	89.80%	88.35%	1.45%
ResNet-101	92.39%	85.74%	6.66%
Ventrodorsal (n=308, ext n=29)			
Hybrid	87.93%	86.16%	1.77%
ResNet-101	90.16%	79.11%	11.05%
ViT-B/16	87.27%	82.20%	5.07%
ViT-L/16	86.81%	81.99%	4.81%

Key Findings:

- Laterolateral:** All architectures perform equally well (0.44% performance range between best and worst model - no significant differences)
- Ventrodorsal:** Hybrid architecture outperforms pure CNN by +2.76% (Cohen's d=0.78, medium effect size)
- Generalization:** Hybrid shows exceptional stability with only 1.77% performance drop on external validation

EXPERT VALIDATION

213 Blinded Cases (Radiographic Assessment Only)

42%

Expert-Model
Agreement*

79.6%

Expert
Consistency

***Expert-Model Agreement:** Cases where board-certified radiologist (radiographic assessment only, no clinical history) agreed with model prediction over ground truth labels.

- Suggests potential systematic annotation differences
- 2.6% diagnostically ambiguous cases
- Model may perform 3-5% better than metrics suggest

CLINICAL PERFORMANCE

92.3%

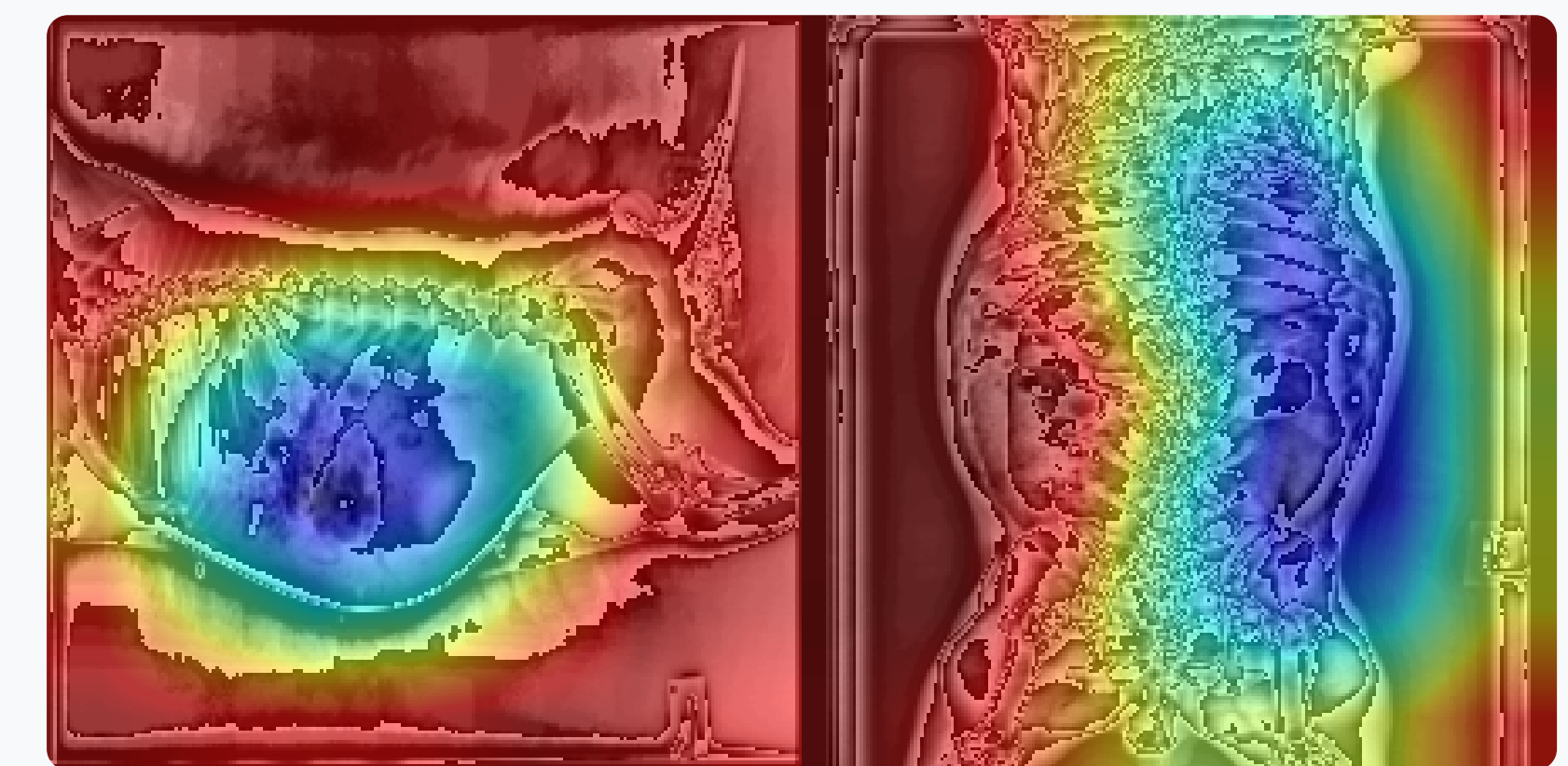
Sensitivity (Hybrid)

Sensitivity = Ability to correctly identify dilated cases (true positives) • Specificity = Ability to correctly identify non-dilated cases (true negatives)

- 87-100% sensitivity range
- 81-94% specificity range
- Minimal false negatives
- Emergency triage suitable:** Reliable enough for clinical decision support in urgent cases

GRAD-CAM VISUALIZATION

Figure 5: Grad-CAM (Gradient-weighted Class Activation Mapping) visualizations showing model attention for gastric dilation detection. Representative case with laterolateral (left) and ventrodorsal (right) projections. Blue regions indicate high model attention (focused on gastric area), red regions indicate low attention.



DEPLOYED PLATFORM

5-Fold Ensemble System

- ViT-B/16 (laterolateral, 89.38%)
- Hybrid (ventrodorsal, 87.03%)
- Uncertainty: $\sigma < 0.05$ typical
- 95% bootstrap CIs
- Inference: <500ms
- Clinical action guidance

Try the Demo



zivsfdiru01.uzh.ch

IMPACT

Novel Scientific Finding:

Projection-dependent architectural requirements challenge the notion of universal architectural superiority in medical imaging AI

- Laterolateral projections:** Simple anatomy with clear gastric boundaries enables architectural equivalence (0.44% performance range)
- Ventrodorsal projections:** Complex anatomical overlap requires hybrid architecture for superior performance (+2.76%, exceptional 1.77% generalization gap)
- Clinical translation:** First validated AI system for rabbit diagnostics, deployed with real-time uncertainty quantification
- Methodological innovation:** Bidirectional quality assurance (expert validation identified 42% potential ground truth errors)