Faculty of Computing
Sabaragamuwa University of Sri Lanka

ICARC

Deep Learning-Driven Retinal Vessel Segmentation Using a Transformer-Based Architecture: A Performance Evaluation

Yves Byiringiro[1], Jean De Dieu Niyonteze[2],Liliane Tuyishime[1],
Mediatrice Dusenge[3], Josue Uzigusenga[4], Benny Uhoranishema[2]

[1] Martin J. Whitman School of Management, Syracuse University Syracuse, NY, 13244, United States

[2] Goizueta Business School Emory University Atlanta, GA 30322, United States

[3] African Centre of Excellence in Data Science, University of Rwanda Kigali, Rwanda

[4] College of Sciences and Technology, University of Rwanda Kigali, Rwanda

ICARC 2026

Sabaragamuwa University
of Sri Lanka

Presenting Author : Jean De Dieu Niyonteze

Motivation and Problem Statement

- Retinal vessel segmentation automatically extracts blood vessels from fundus images to accurately represent the retinal vascular network.
- Retinal vasculature is a critical biomarker for detecting and monitoring diabetic retinopathy, glaucoma, hypertension, and cardiovascular disease.
- Segmentation remains difficult: small and thin vessels, low vessel–background contrast, uneven illumination, and imaging noise degrade performance.
- CNN-based models (U-Net, ResU-Net, Dense U-Net, Attention U-Net) capture local features well but have limited receptive fields and weak explicit global context modeling.
- These limitations hinder accurate representation of complex, thin, and tortuous vessels, especially under heterogeneous imaging conditions.
- Transformer-based models have emerged to address these challenges by modeling long-range dependencies and global contextual relationships via self-attention.

Study Aim and Contributions

This study aims to systematically evaluate the effectiveness of the SegFormer transformer architecture for retinal vessel segmentation across heterogeneous retinal fundus datasets.

Key Contributions

- A comprehensive multi-dataset benchmark for transformer-based retinal vessel segmentation.
- Demonstration of robust segmentation of both major and fine vessels.
- Evidence of SegFormer’s scalability and cross-dataset generalization potential.

Related Work

CNN-Based Retinal Vessel Segmentation

- Since 2020, CNN-based models have dominated retinal vessel segmentation, with architectures such as SA U-Net, Dense U-Net, and lightweight U-Net variants.
- SA U-Net reports Dice > 0.80 and accuracy > 0.95 on DRIVE, CHASE_DB1, STARE, and HRF, illustrating strong performance on standard benchmarks.
- Lightweight U-Net designs achieve competitive Dice scores (≈0.69–0.79) on DRIVE, CHASE_DB1, and HRF, with accuracies up to ≈0.97, while reducing computational cost.
- Enhanced loss functions and attention mechanisms improve segmentation on DRIVE and CHASE_DB1 but yield limited gains on HRF due to high variation in contrast and illumination.

Overall, CNNs effectively model local spatial structure yet struggle to simultaneously capture global context and fine-scale vessels in heterogeneous datasets.

Related Works

Transformer-Based Retinal Vessel Segmentation

- Transformer-based architectures excel at retinal vessel segmentation by capturing long-range dependencies and global context, improving detection of thin vessels and vessel continuity over CNNs.
- SegFormer achieves strong performance across multiple datasets while being smaller, faster, and more efficient; its lightweight version, SegFormer-B0, further reduces model size with minimal performance loss.
- Other transformer-based and hybrid CNN-transformer approaches enhance generalization on heterogeneous and low-contrast retinal images, motivating the use of SegFormer in this study.

MATERIALS AND METHODS

Datasets and Preprocessing Pipeline

- Evaluated SegFormer on eight public retinal datasets:

  DRIVE, STARE, CHASE_DB1, HRF, RETA Benchmark, AV-DRIVE, LES-AV, IOSTAR.
- Diverse in resolution, illumination, and annotation quality.
- Images resized to 512×512 pixels, normalized, and enhanced with CLAHE.
- AV-DRIVE artery/vein annotations merged into single binary vessel mask.
- Ground truth masks converted to binary vessel maps for consistency.
- Unified preprocessing ensured fair comparison across datasets.

SegFormer Architecture and Training Setup

SegFormer chosen for dense prediction due to:

- MiT encoder captures long-range dependencies via self-attention
- Preserves fine spatial details with overlapping patch embeddings
- Lightweight MLP decoder aggregates multi-scale features efficiently

Training setup:

- Adam optimizer, learning rate 1×10^{-4}.
- Batch size 4, max 100 epochs on an NVIDIA GeForce RTX 4070 GPU.
- Combined BCE + Dice loss.

Training Strategy, CV, and Validation Strategy / Inference on New Data

Five-fold cross-validation on each dataset, best validation Dice checkpoint retained.

Evaluation metrics: Dice score, accuracy, specificity, precision, recall, F1-score.

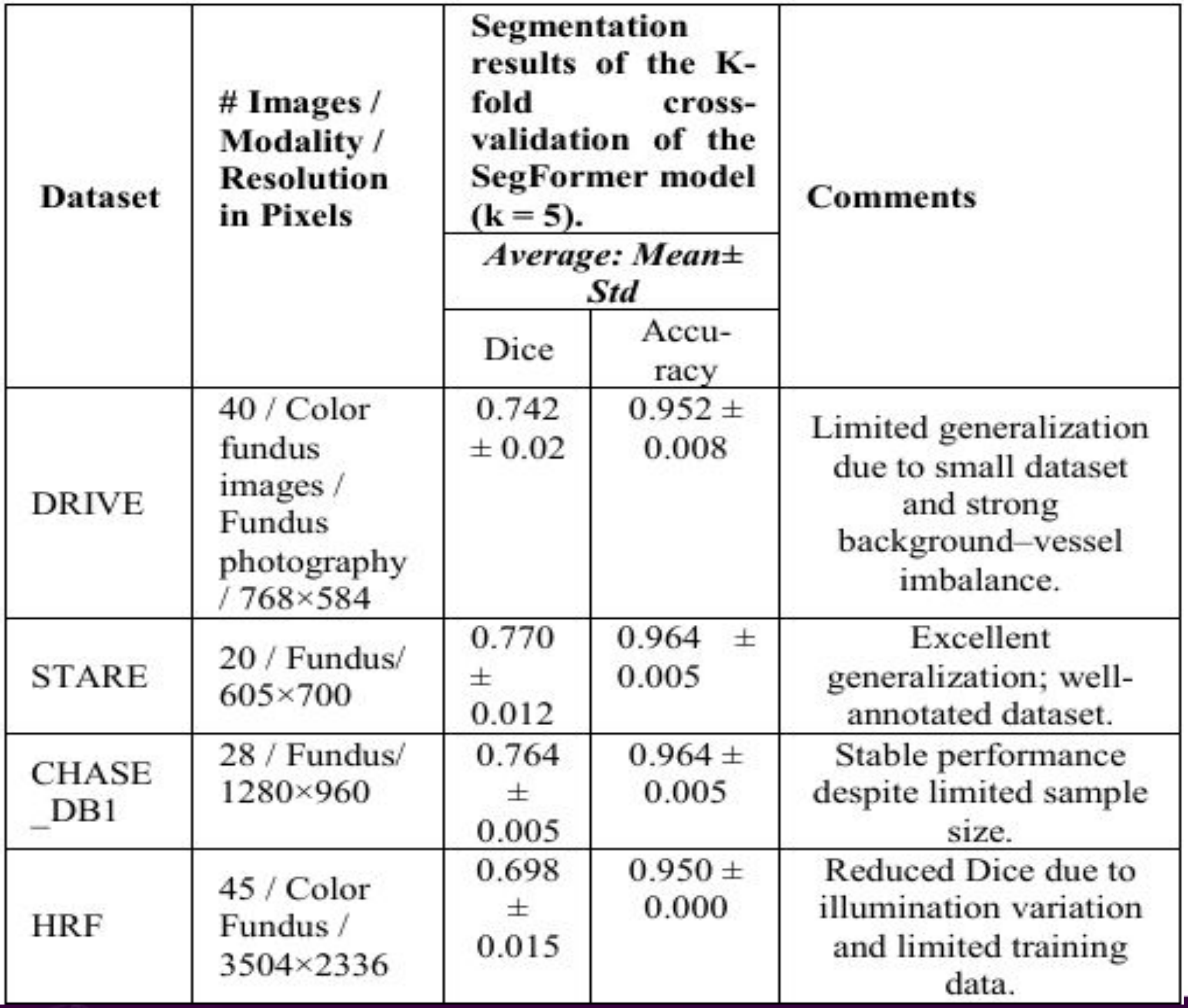
Validation strategy:

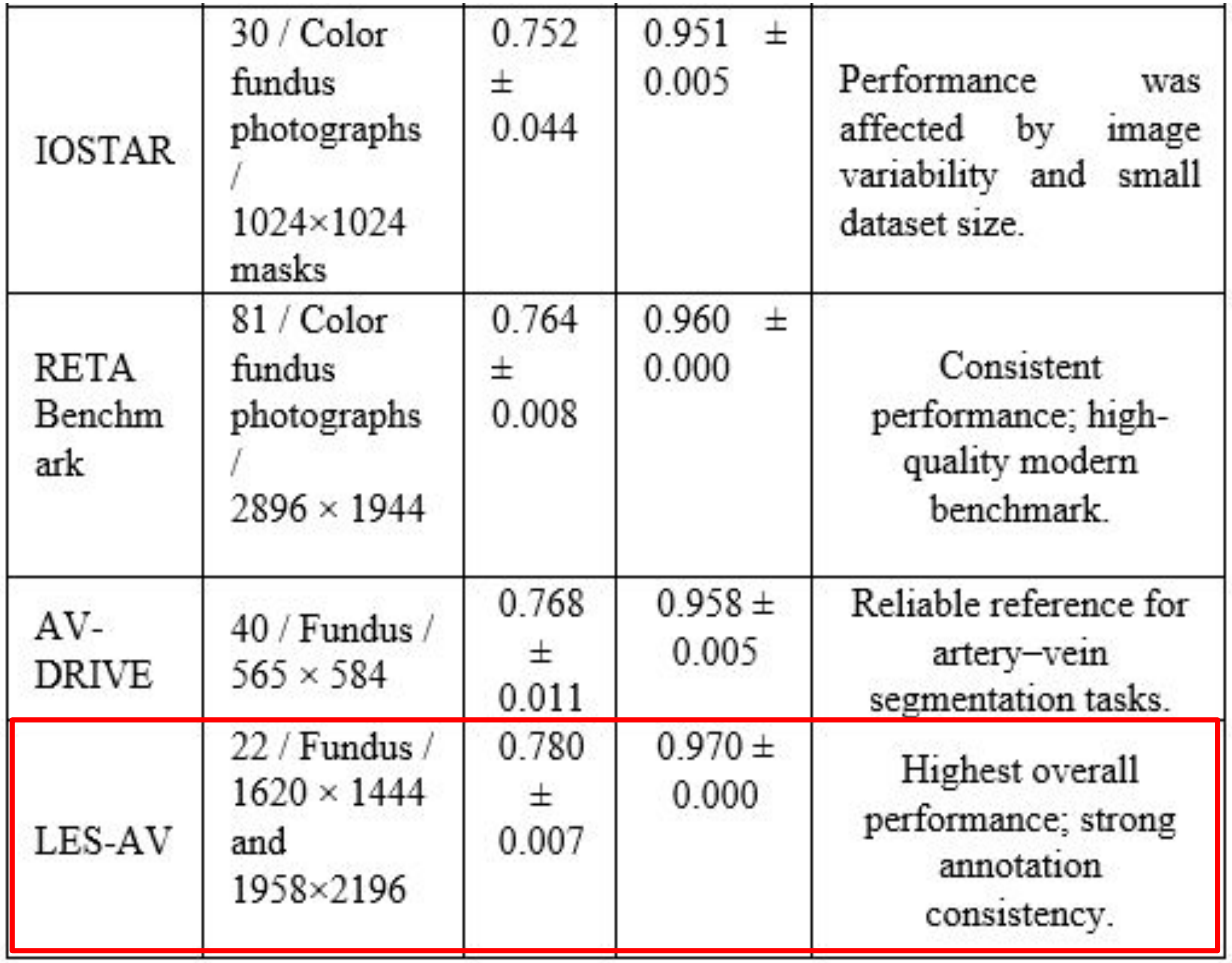
- Per-image metrics averaged for each dataset.
- Best-performing dataset was used for final training.
- Final model applied to external datasets without fine-tuning.

.

RESULTS AND DISCUSSION

Performance evaluation of the SegFormer model using five-fold cross-validation across multiple retinal vessel segmentation datasets





Dataset	# Images / Modality / Resolution in Pixels	Segmentation results of the K-fold cross-validation of the SegFormer model (k = 5).		Comments
		Average: Mean± Std		
		Dice	Accuracy	
DRIVE	40 / Color fundus images / Fundus photography / 768×584	0.742 ± 0.02	0.952 ± 0.008	Limited generalization due to small dataset and strong background–vessel imbalance.
STARE	20 / Fundus/ 605×700	0.770 ± 0.012	0.964 ± 0.005	Excellent generalization; well-annotated dataset.
CHASE _DB1	28 / Fundus/ 1280×960	0.764 ± 0.005	0.964 ± 0.005	Stable performance despite limited sample size.
HRF	45 / Color Fundus / 3504×2336	0.698 ± 0.015	0.950 ± 0.000	Reduced Dice due to illumination variation and limited training data.
IOSTAR	30 / Color fundus photographs / 1024×1024 masks	0.752 ± 0.044	0.951 ± 0.005	Performance was affected by image variability and small dataset size.
RETA Benchmark	81 / Color fundus photographs / 2896 × 1944	0.764 ± 0.008	0.960 ± 0.000	Consistent performance; high-quality modern benchmark.
AV-DRIVE	40 / Fundus / 565 × 584	0.768 ± 0.011	0.958 ± 0.005	Reliable reference for artery–vein segmentation tasks.
LES-AV	22 / Fundus / 1620 × 1444 and 1958×2196	0.780 ± 0.007	0.970 ± 0.000	Highest overall performance; strong annotation consistency.

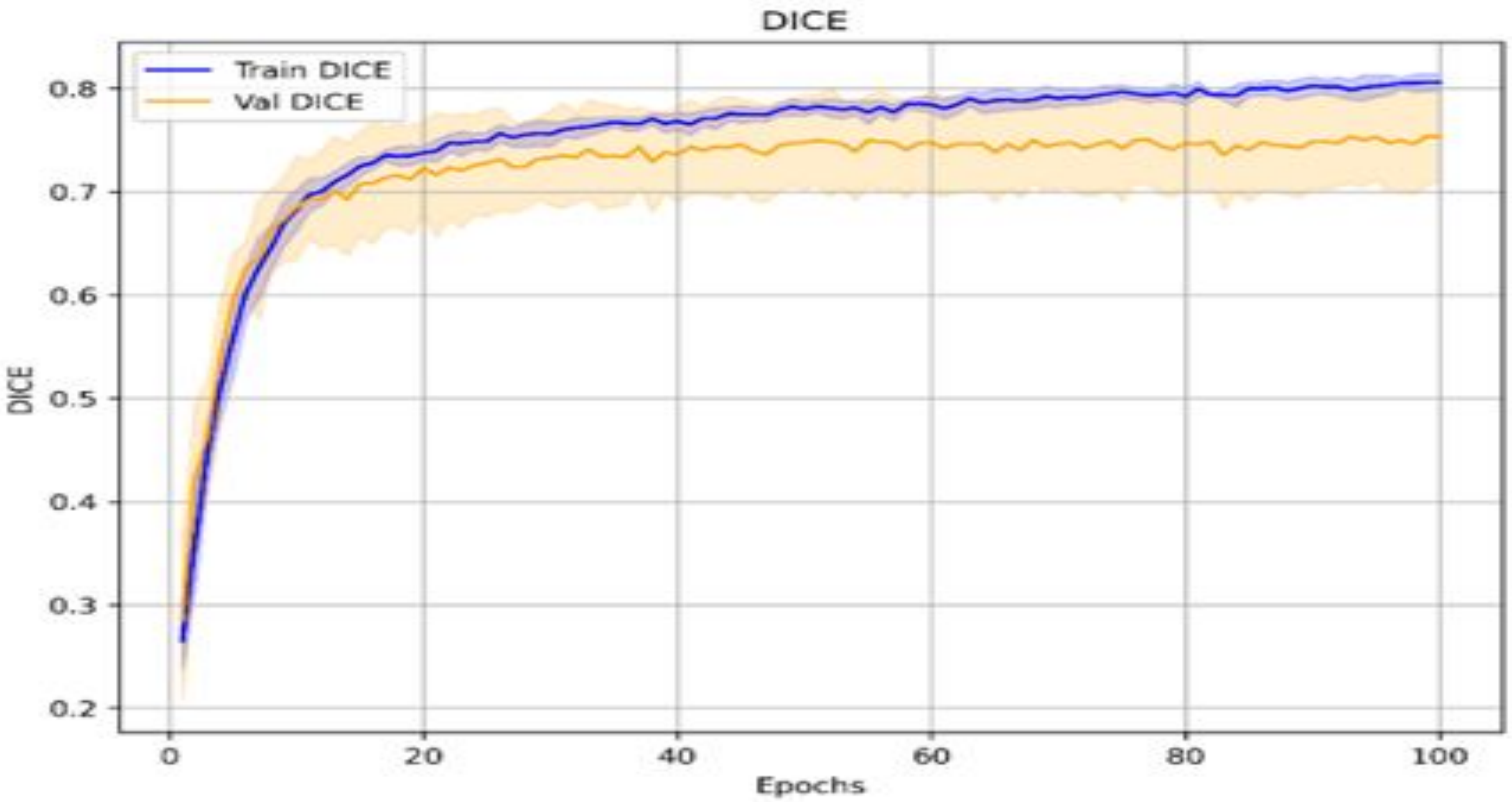
Cross-Validation Results

Dice scores: 0.70 – 0.82

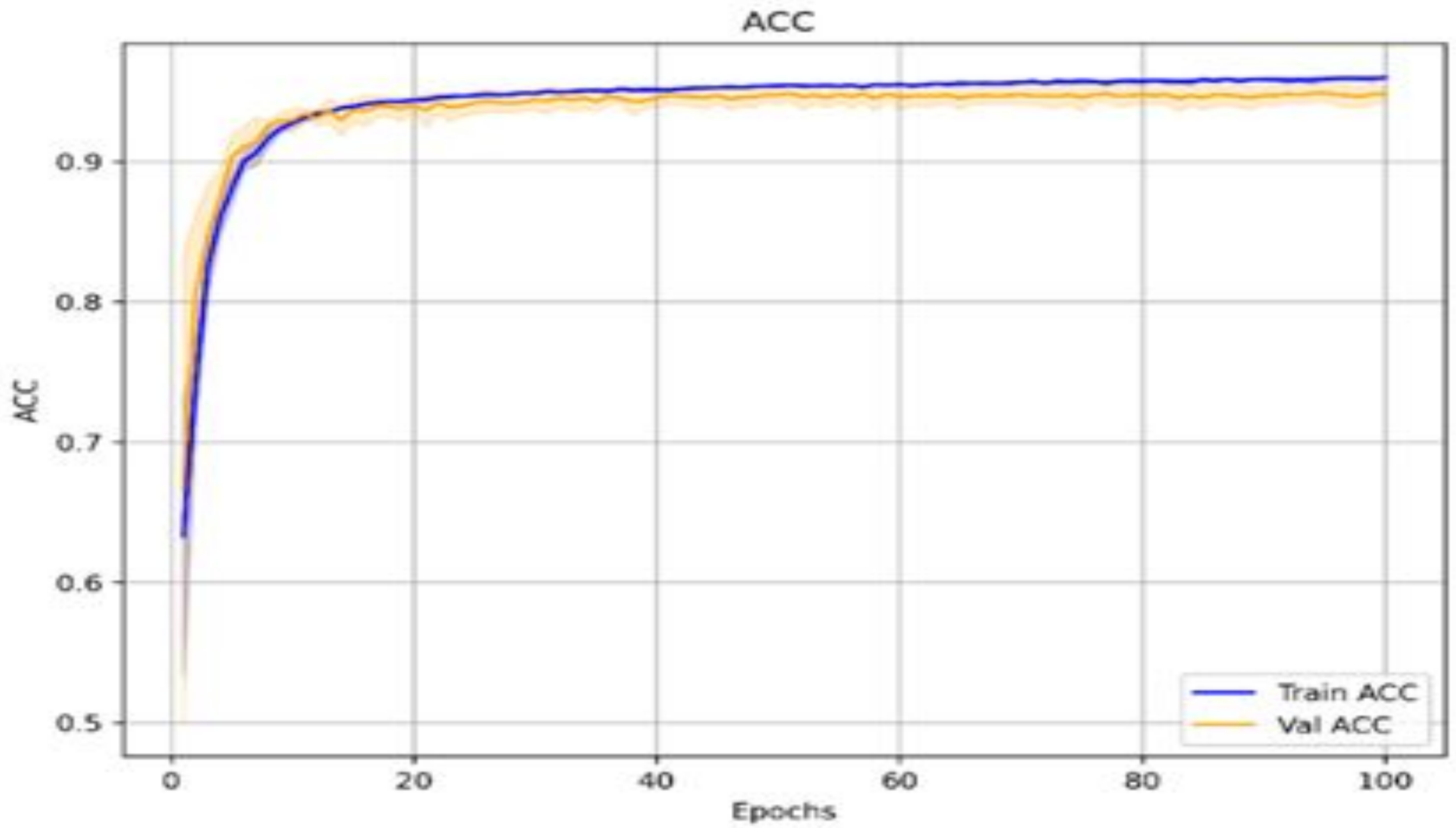
Accuracy consistently above 0.95

Stable performance across heterogeneous datasets

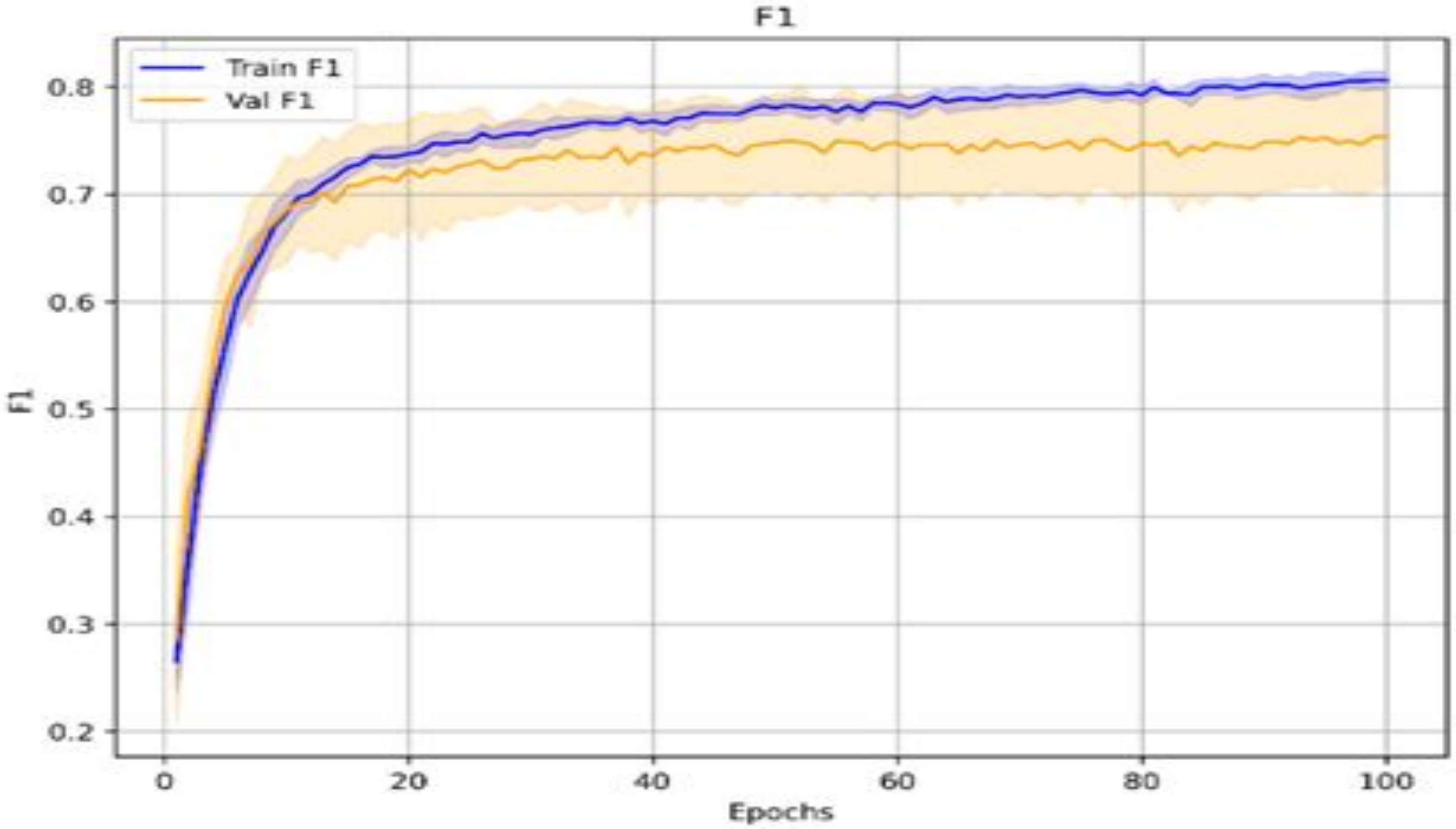
Training and validation dice curves of the SegFormer model on the LES-AV dataset obtained from five-fold cross-validation



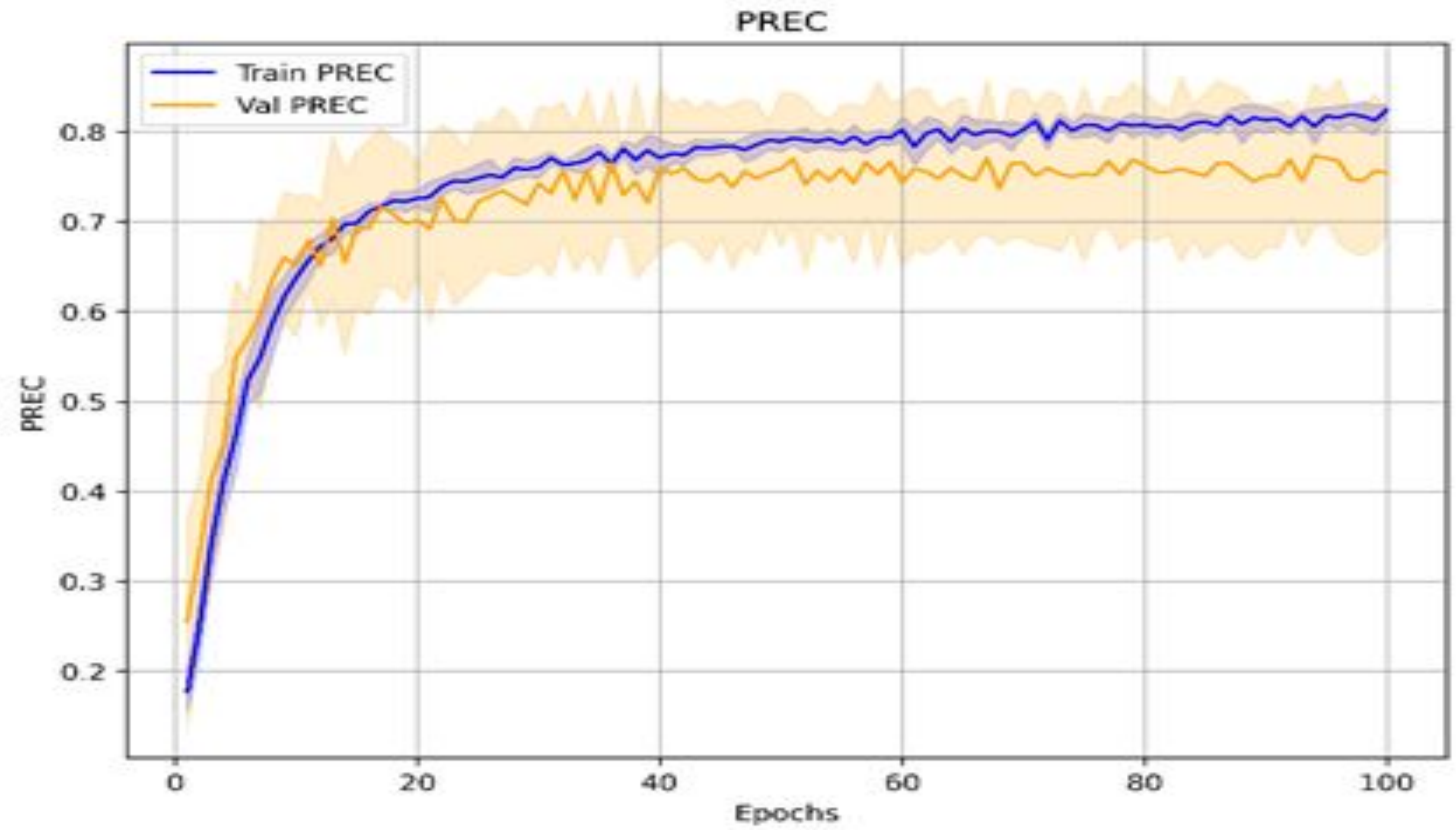
Training and validation accuracy curves of the SegFormer model on the LES-AV dataset obtained from five-fold cross-validation



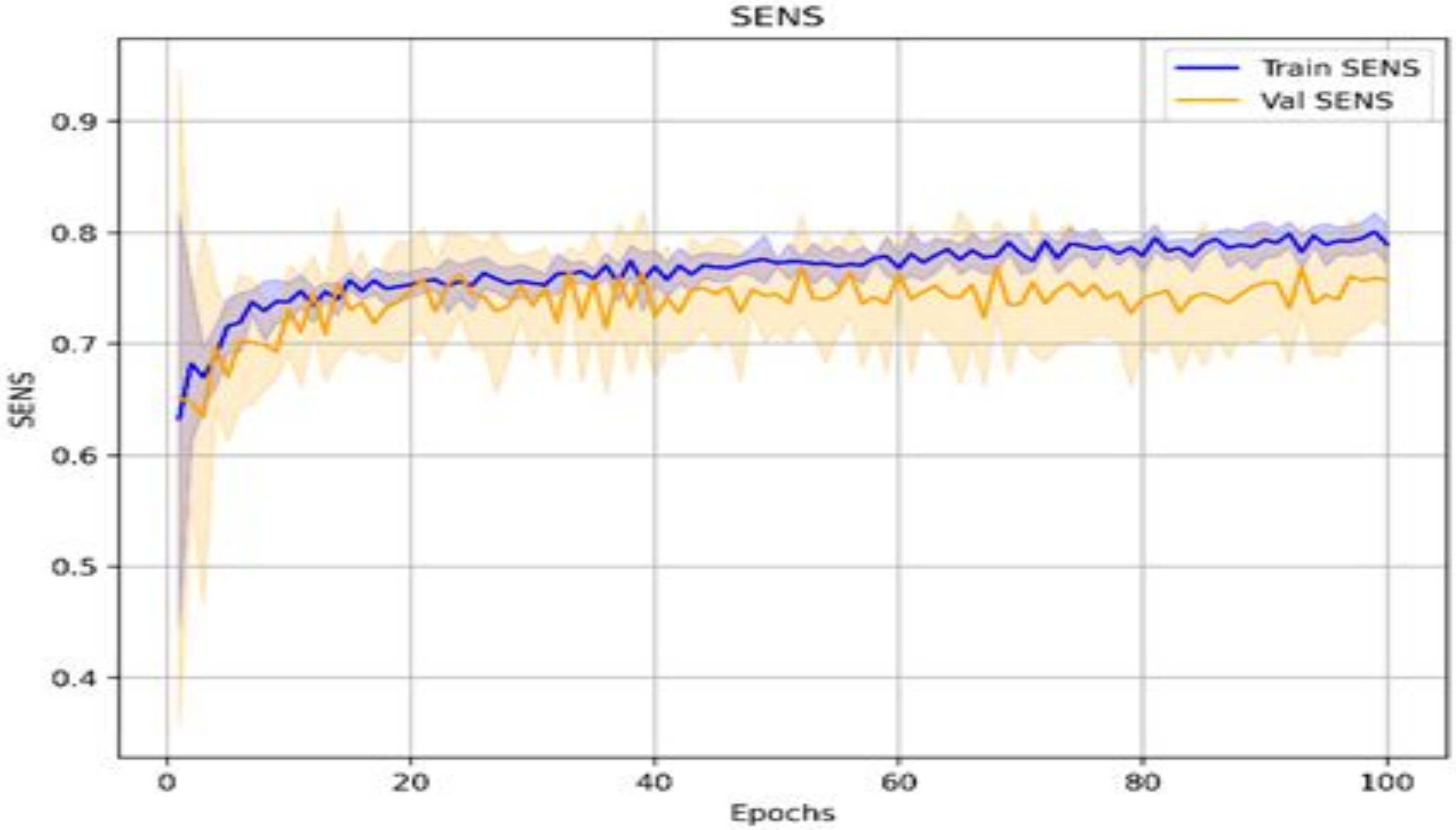
Training and validation F1-score curves of the SegFormer model on the LES-AV dataset obtained from five-fold cross-validation



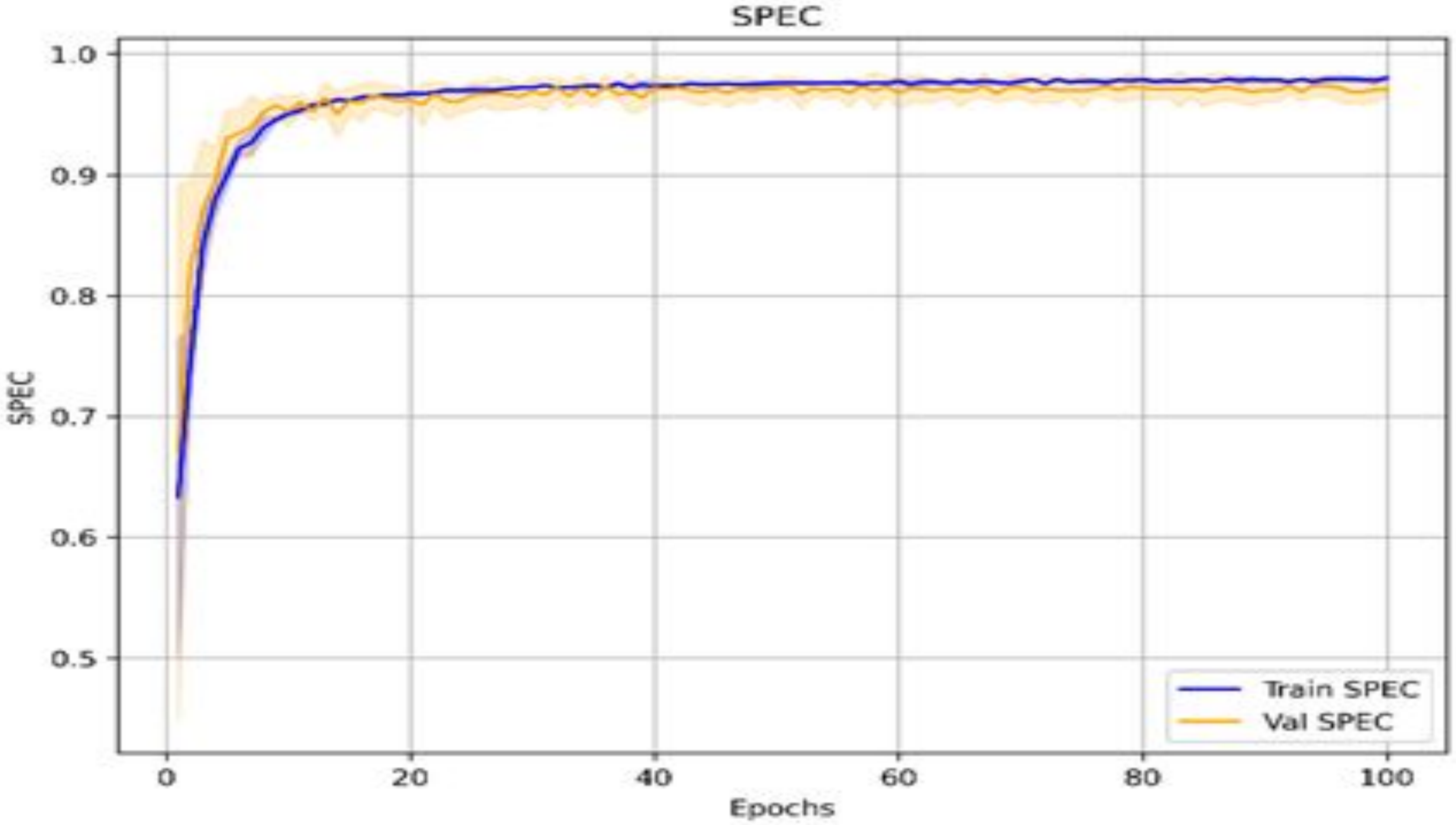
Training and validation precision curves of the SegFormer model on the LES-AV dataset obtained from five-fold cross-validation



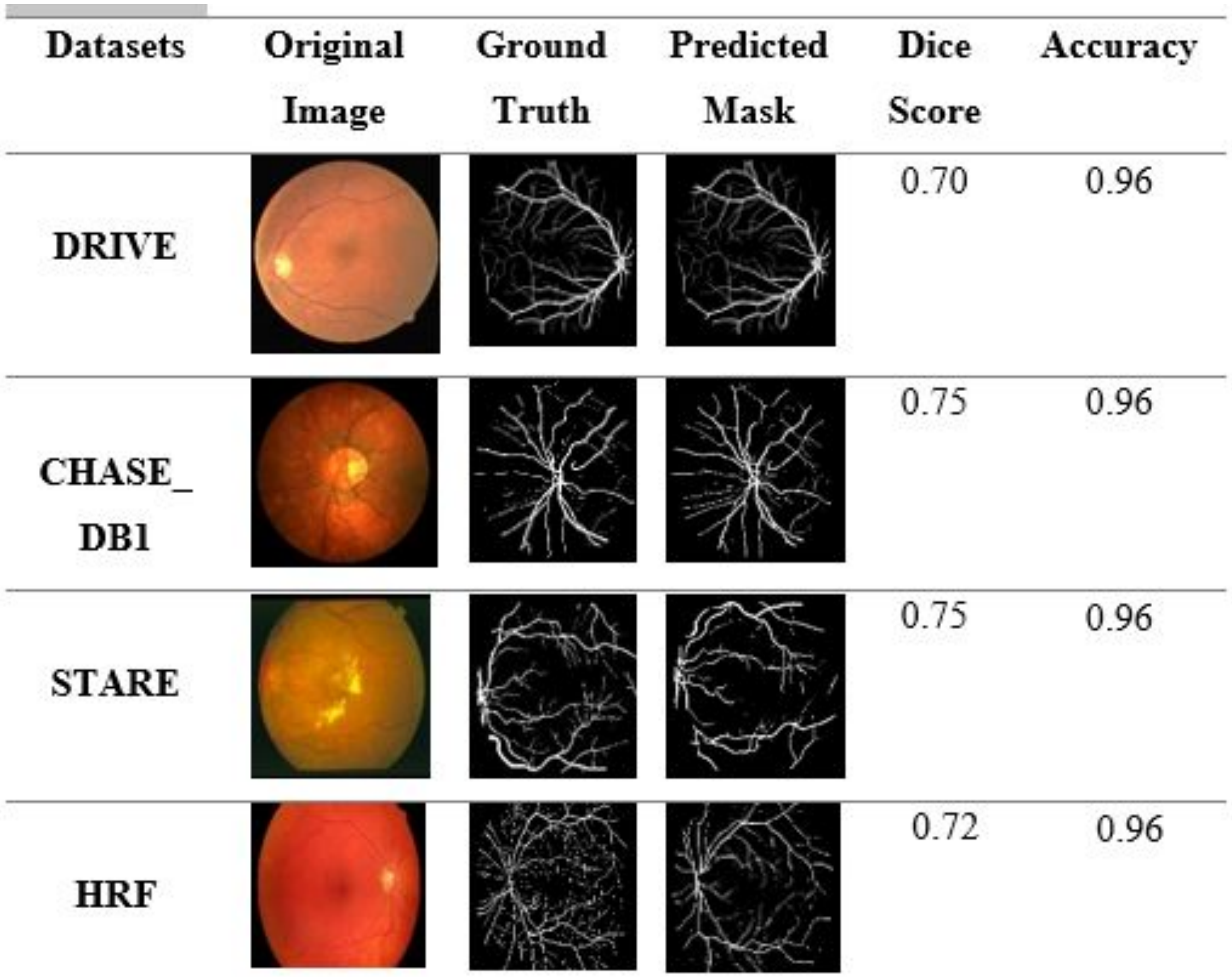
Training and validation sensitivity curves of the SegFormer model on the LES-AV dataset obtained from five-fold cross-validation

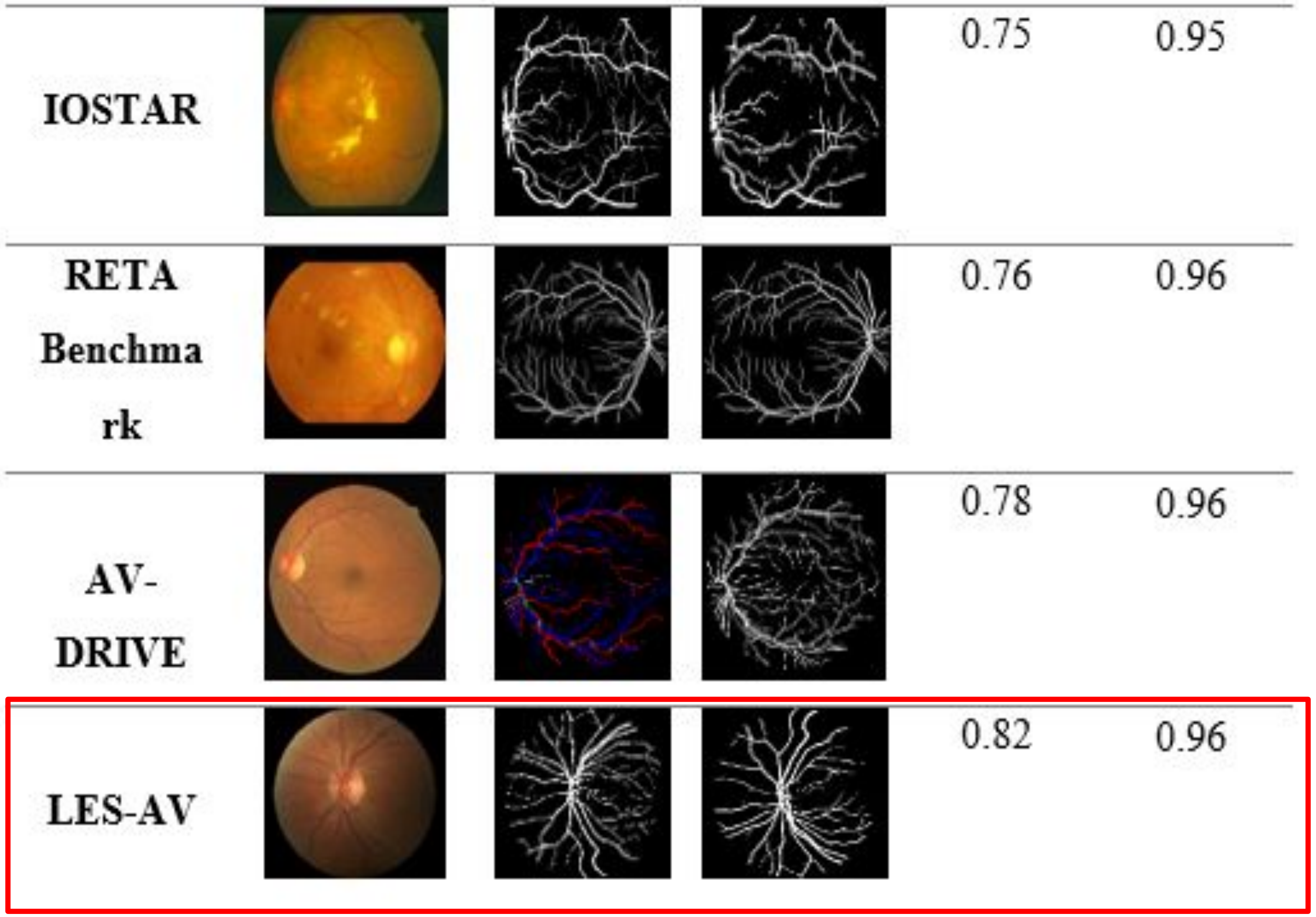


Training and validation specificity curves of the SegFormer model on the LES-AV dataset obtained from five-fold cross-validation



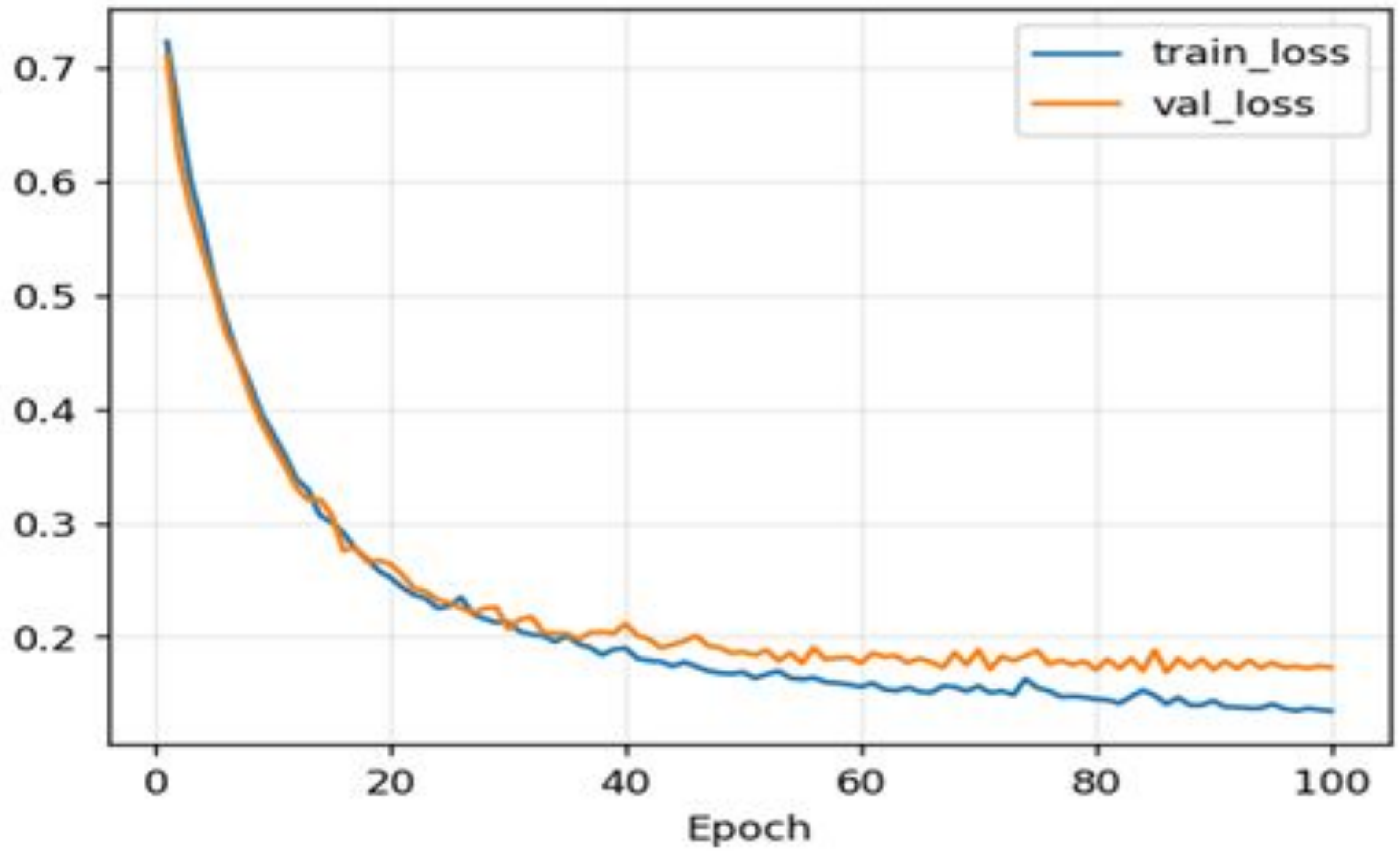
QUALITATIVE SEGMENTATION EXAMPLES FROM THE LES-AV–TRAINED SEGFORMER MODEL ACROSS ALL DATASETS, INCLUDING ORIGINAL IMAGES, GROUND-TRUTH MASKS, PREDICTED MASKS, DICE SCORES, AND ACCURACY.



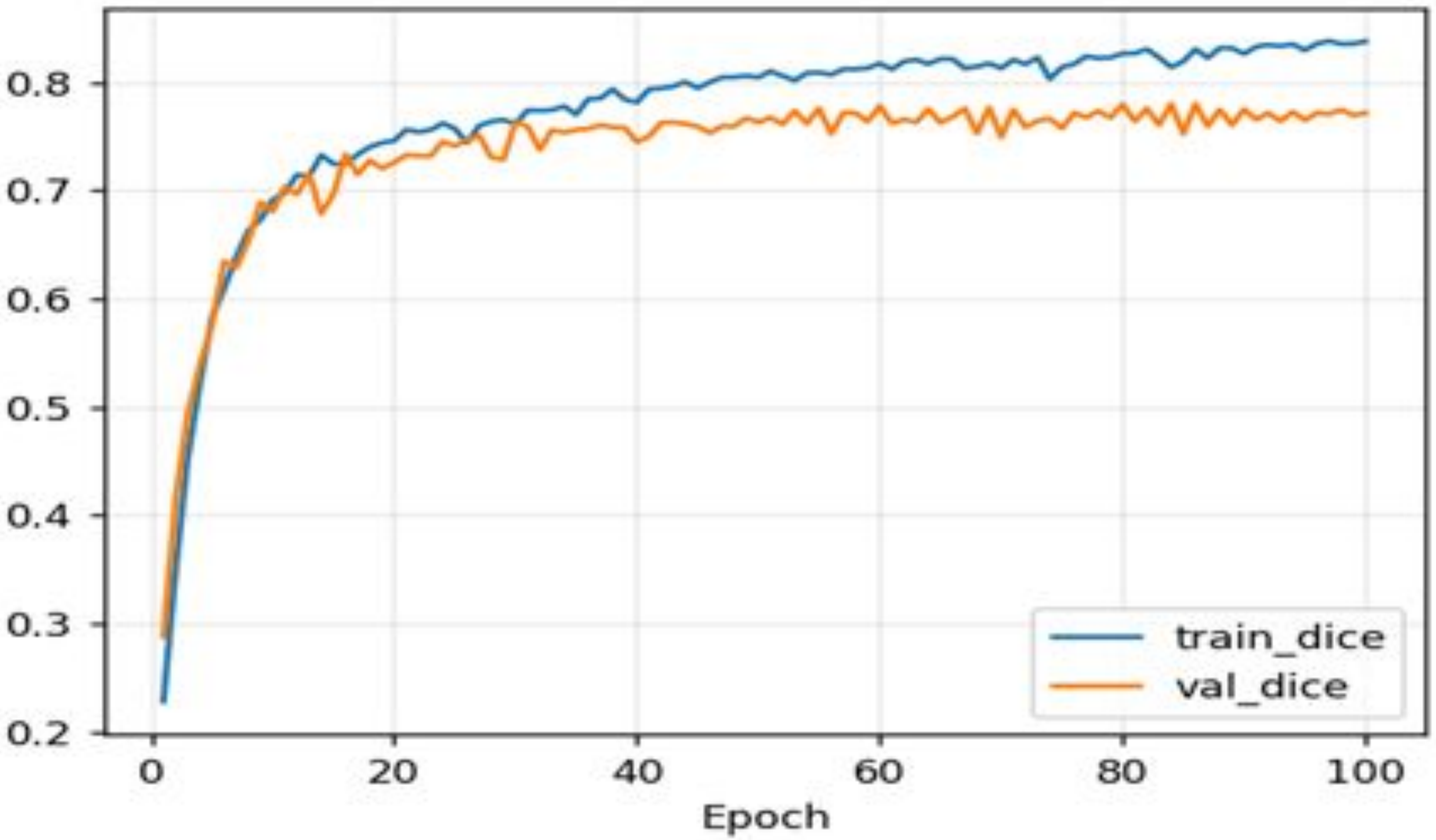


Datasets	Original Image	Ground Truth	Predicted Mask	Dice Score	Accuracy
DRIVE				0.70	0.96
CHASE_DB1				0.75	0.96
STARE				0.75	0.96
HRF				0.72	0.96
IOSTAR				0.75	0.95
RETA Benchmark				0.76	0.96
AV-DRIVE				0.78	0.96
LES-AV				0.82	0.96

Training and validation loss curves of the SegFormer model on the LES-AV dataset



Training and validation dice curves of the SegFormer model on the LES-AV dataset



Results Overview — Cross-Validation

- SegFormer evaluated on eight public retinal datasets using five-fold cross-validation
- Stable performance across heterogeneous imaging conditions
- Dice scores 0.70–0.82 with accuracy consistently > 0.95
- Best performance on LES-AV, with moderate scores on HRF and IOSTAR due to data variability

Conclusion & Future Work

Conclusion

- SegFormer demonstrates robust and generalizable retinal vessel segmentation
- Transformer-based models outperform CNNs in cross-dataset consistency

Future Work

- Evaluation of advanced transformer architectures
- Improved modeling of thin and low-contrast vessels
- Real-time clinical deployment analysis

👋

Q & A

👋

THANK
YOU!