Name: Yikai Chen

1. What we set out to build (very brief)

Our proposal targets five-grade DR classification by combining (i) lesion-aware attention on CNN features, (ii) an explicit vessel-tree pathway, and (iii) quad-modal fusion of local CNN, global ViT, and vascular cues, trained with imbalance-aware objectives and strong augmentation. The rough 8-week schedule—literature/data → baselines → lesion attention → vessel module → fusion → full training/validation—is summarized in the timeline table on page 2 of the proposal.

2. Experimental Setup

2.1 Datasets

APTOS-2019 (complete so far), RFMiD-2020, Messidor, and IDRiD (in progress). Labels mapped to 5 DR grades.

Availability of results. At the time of writing, completed validation metrics are available for APTOS-2019 only. All other dataset results will be finalized in the final report.

2.2 Pre-processing and Augmentation

Images are resized to 448×448. Training uses random crop, horizontal/vertical flip, small rotations (≤15°), affine transforms, color jitter, perspective warp, and random erasing; validation uses a deterministic resize. A severe-DR-focused augmentation profile is available but not required for the main experiments. Normalization follows ImageNet statistics.

2.3 Optimization Details

Optimizer: AdamW; weight decay 1e-4; gradient clipping (L2 ≤ 1.0).

LR schedule: ReduceLROnPlateau on validation kappa (patience = 5).

Loss: cross-entropy with label smoothing $\varepsilon = 0.1$ by default; focal/class-balanced focal loss available for sensitivity analyses.

Batch size. We used 12 images per GPU. With 5×RTX-4090 and DataParallel, the effective batch size per step was 60 (the loader batch size was set to args.bs × #GPUs, and DataParallel split it evenly across devices).

2.4 Evaluation Metrics

We report quadratic weighted kappa (K)—the community's primary DR metric—together with accuracy (A), macro/weighted F1, and ROC-AUC (one-vs-rest, weighted). Our code computes confusion matrices and per-class F1 for auditability.

Training uses five NVIDIA RTX 4090 GPUs with 5-fold cross-validation and 30 epochs per fold.

3. What is implemented now

Model & training code. The current training script implements the planned hybrid model and adds several engineering safeguards. Major components include:

3.1 Lesion Aware Attention on high level ResNet 50 features, with channel, image driven lesion, and severe DR attention branches.

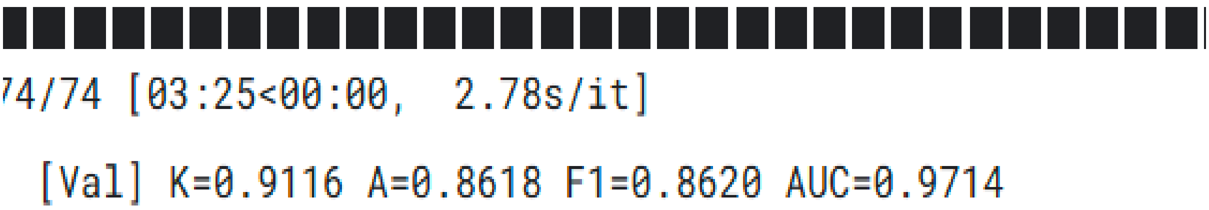
3.2 VesselTree pathway: green-channel vessel mask (OpenCV with automatic PIL fallback), encoded via a lightweight CNN + KAN layer into a compact embedding.

3.3 Global encoders (ViT-S/L) with KAN-MLP replacements and an MoE-style gate; FPN-like reduction of ResNet multi-stage features pre-fusion.

3.4 Fusion & head. Quad modal concatenation (ResNet+FPN, fused ViT, vessel features, and a small conv fusion of vessel maps), followed by KAN layers and a light graph refinement head (EnhancedKANGCN).

3.5 Training pipeline. AdamW, ReduceLROnPlateau, gradient clipping, label smoothing / focal / class-balanced focal, mixup, strong aug, multi-GPU support; automatic caps for very-large datasets during early iteration; confusion-matrix & per-class F1 export.

4. Result



We present KAN-VesselGraph-DR, a clinically grounded DR grading framework that combines vessel-aware reasoning, multi-scale lesion fusion, and an ordinal refinement head within a hybrid CNN–Transformer–KAN architecture. On **APTOS-2019**, the system achieves **K = 0.9116, A = 0.8618, F1 = 0.8620, AUC = 0.9714** on validation. Pending other dataset and external validations will finalize the comparative picture, but these results already indicate that explicitly modeling vascular structure and progression yields practical gains and interpretability for DR screening.

5. Difficulties encountered & how we addressed them

5.1 Class imbalance & rare-class collapse. Severe DR (grade 3) under-performed (<0.1 F1 initially).

Mitigations: class-balanced focal loss with explicit higher weights for rare/weak classes; a more aggressive “severe-case” augmentation path; optional mixup; label smoothing for stability.

5.2 Multi-GPU reliability. DistributedDataParallel caused local port/NCCL issues.

Mitigation: switch default to DataParallel (script still supports DDP) and use safer backends/timeouts when DDP is required.

5.3 Vessel extraction environment variance. OpenCV may be unavailable in some environments.

Mitigation: automatic PIL-based fallback with contrast enhancement and edge cues; consistent tensor conversion guards.

5.4 Feature-shape mismatches in fusion. FPN and multi-branch concatenation occasionally mis-aligned channel dims.

Mitigations: rigorous dimension checks, on-the-fly adjustment of the first fusion KAN layer (mil_fc1) when concat dims change, and protective fallbacks to zeros to avoid crashes while iterating.

5.5 Numerical stability. Sporadic NaN/Inf in logits during early ablations.

Mitigations: gradient clipping, conservative learning rates, layer-norms in KAN blocks, and safe fallbacks when invalid logits are detected.

5.6 Memory/perf constraints. KAN layers can be heavy.

Mitigations: smaller grid_size, chunked processing in KAN ops, environment limits on intra-op threads, pooled features (FPN), and selective parameter freezing in early ViT blocks.

6. How the proposal has changed in light of progress

6.1 Scope prioritization. VesselTree + lesion-aware attention + multi-scale ViT fusion are fully implemented and trainable; the STLE (spatio-temporal lesion evolution) idea remains a stretch goal.

6.2 Training strategy. Beyond the original emphasis (CB-Focal + mixup + strong aug), we now support a triad of losses (label smoothing / focal / class-balanced focal) with ReduceLROnPlateau + gradient clipping by default.

6.3 Engineering choices. Standardize on DataParallel as the default; keep DDP as an option; add OpenCV→PIL fallback to ensure portability.

6.4 Timeline update (ahead of plan). Compared with the initial 8-week plan, we have already completed items slated through fusion + early full-training. This frees up additional time to (a) extend validation across datasets, (b) run deeper ablations, and (c) polish deployment-quality training utilities.

7. Next steps

7.1 Stabilize validation on the other full training set(RFMiD2020 dataset; IDRiD datasets.) and run 5-fold with consistent seeds to report Kappa/Weighted-F1/AUC/AUG.

7.2 Lift grade-3 performance tune class weights, introduce mild ordinal constraints or label smoothing tuned per class, and extend the “severe-case” augmentation policy.

7.3 Ablations (i) with/without vessel pathway, (ii) lesion attention variants, (iii) ViT-S vs ViT-L contributions, (iv) KAN vs Linear MLP in ViT.

7.4 (Stretch) implement the **STLE** idea from the draft if time permits (e.g., a simple ordinal/transition prior embedded via a small CRF-like head).

8. Brief takeaway

We have implemented the core architecture (lesion attention + vessel pathway + multi-scale ViTs + fusion) and a stable training pipeline with imbalance-aware losses. Early signals show strong performance on majority classes but severe DR remains challenging, which we are directly addressing through loss weighting, augmentation, and fusion refinements before running full, cross-validated evaluations.