

Graph-Based Deep Learning for Non-invasive Dementia Screening via Retinal Vascular Topology: A Proof-of-Concept Study

Sayantana Mukherjee¹, Anurag Shrivastava²

¹Research Scholar, Department of Computer Application, Eklavya University, Damoh, M.P.

²Supervisor, Department of Computer Application, Eklavya University, Damoh, M.P.

Abstract

Retinal fundus imaging provides a non-invasive, scalable window into systemic microvascular health, and growing evidence links retinal vascular architecture to cerebrovascular changes associated with cognitive decline and dementia. Conventional deep learning approaches process retinal images as Euclidean pixel grids and are not explicitly designed to capture the relational, graph-structured topology of the vascular network — information that may carry diagnostically relevant biomarker signals beyond local pixel features. This paper proposes a graph-based deep learning pipeline that converts retinal fundus images into structured graphs via Frangi vesselness filtering, morphological skeletonisation, and k-nearest-neighbour edge construction, enabling Graph Neural Networks (GNNs) to reason directly over vascular topology. The primary scientific contribution is the graph-based vascular topology modelling framework itself; the central empirical claim is that vascular topology constitutes an independent discriminative feature space beyond pixel-level information. Pipeline validity was confirmed on the DRIVE dataset (148 nodes, 416 undirected edges; image 21_training.tif). Cross-device portability was demonstrated on the STARE dataset (TopCon TRV-50, 35° FOV): five images processed without pipeline modification, yielding average 565 nodes and 1,985 edges per image (range: 82–936 nodes). Classification performance was evaluated on 3,296 retinal graphs from the APTOS 2019 dataset using diabetic retinopathy as a controlled proxy task for retinal microvascular pathology detection. A controlled ablation compared three architectures: a topology-free MLP baseline, a Graph Convolutional Network (GCN), and a Graph Attention Network (GAT). The GCN achieved the strongest performance (Balanced Accuracy 92.07%, AUC-ROC 0.9711), outperforming the topology-free MLP baseline (AUC 0.9420; $\Delta\text{AUC} = +0.0291$). Statistical significance testing was not performed; results are exploratory. These findings provide proof-of-concept support for graph-based retinal topology modelling as a methodological foundation for future non-invasive dementia screening.

Index Terms — Retinal imaging; graph neural networks; vascular topology; dementia screening; GCN; GAT; microvascular pathology; APTOS 2019; DRIVE; STARE; deep learning; cross-device validation.

1. Introduction

Dementia affects more than 55 million individuals globally, with Alzheimer's disease (AD) accounting for 60–80% of cases and vascular dementia constituting most of the remainder [1]. Current diagnostic pathways — structural MRI, amyloid PET imaging, cerebrospinal

10.48047/jocaaa.2024.33.02.54

fluid (CSF) tau and amyloid-beta assays, and comprehensive neuropsychological batteries — are resource-intensive, often invasive, and inaccessible in resource-limited healthcare settings [2]. These modalities are typically deployed after symptomatic presentation, foreclosing opportunities for early intervention on modifiable risk factors [3]. A non-invasive, scalable, primary-care-compatible screening tool capable of flagging at-risk individuals before clinical symptoms emerge represents a significant unmet need.

The retina offers a compelling candidate biomarker channel. Sharing a common developmental lineage and continuous microvascular architecture with the cerebral cortex, the retina provides a uniquely accessible window into central nervous system microvascular health. Retinal vascular changes — arteriolar narrowing, increased venular calibre, arteriovenous nicking, elevated tortuosity, and reduced fractal dimension — have been associated with poorer cognitive performance and elevated dementia risk across multiple cross-sectional studies [4, 5, 6]. Retinal fundus photography is rapid, inexpensive, non-invasive, and deliverable in primary care settings, making it an attractive platform for population-level cognitive risk screening if reliable computational biomarkers can be identified and validated.

A fundamental limitation of existing computational approaches is representational mismatch. Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs) — the dominant paradigms in retinal image analysis — represent images as regular Euclidean grids and are not explicitly designed to model the relational, graph-structured topology of the retinal vascular network: which vessel segments connect to which, at what branching angles, and with what calibre relationships. This topological structure is fundamentally non-Euclidean and may encode diagnostically relevant information that pixel-level feature learning cannot efficiently exploit. Prior CNN-based retinal biomarker studies implicitly aggregate topology through convolutional receptive fields, but provide no mechanism for explicit topological reasoning.

Graph Neural Networks (GNNs) provide a principled solution. By representing the vasculature as a graph — nodes at vessel junction points, edges encoding connectivity — GNNs perform message-passing aggregation directly over topological structure, enabling models to reason about relational vessel properties alongside local node features [7, 8]. GNN-based models of anatomical graph structures have demonstrated classification advantages over pixel-grid approaches for brain connectivity analysis and MCI classification [9, 10], motivating analogous investigation in retinal vascular topology.

This paper proposes a graph-based deep learning pipeline for retinal vascular topology modelling and presents the following contributions:

- A graph construction pipeline converting fundus images to PyTorch Geometric (PyG) Data objects via Frangi vesselness filtering [11], morphological skeletonisation [12], and k-nearest-neighbour (k-NN) edge construction, with 8-dimensional geometric node feature vectors.
- Pipeline validity confirmed on the DRIVE dataset (148 nodes, 416 undirected edges from image 21_training.tif) against independently annotated expert vessel ground truth.

10.48047/jocaaa.2024.33.02.54

- Cross-device portability demonstrated on the STARE dataset (TopCon TRV-50, 35° FOV, 700×605 px native resolution): five images processed with an identical, unmodified pipeline yielding average 565 nodes and 1,985 undirected edges per image.
- A controlled ablation study comparing three architectures — a topology-free MLP baseline, a GCN [7], and a GAT [8] — evaluated on 3,296 retinal graphs from APTOS 2019, isolating the classification contribution of vascular topology.
- Experimental evidence that vascular topology constitutes an independent discriminative feature space: GCN achieves $\Delta\text{AUC} = +0.0291$ over the topology-free MLP baseline (AUC 0.9711 vs. 0.9420), confirmed across two graph-structured architectures.
- A critical, pre-registered scope statement and a nine-point limitation assessment covering the proxy-task design, statistical testing, and cross-device validation scope.

Three explicit scope constraints govern interpretation. First, this study does not directly detect dementia. No publicly available large-scale retinal imaging dataset with accompanying dementia diagnoses exists for deep learning training at the time of writing. Diabetic retinopathy (DR) classification on APTOS 2019 is used as a controlled proxy task: DR involves retinal microvascular pathology mechanistically related — though not identical — to cerebrovascular disease [5]. Second, the contribution is the framework and the empirical topology-vs-no-topology finding, not the GCN or GAT architectures, which are established methods [7, 8]. Third, no direct CNN pixel-grid comparison is conducted, as the input representations are fundamentally incompatible; this is acknowledged as a limitation.

2. Related Work

2.1 Retinal Biomarkers and Cognitive Decline

Shared developmental lineage and continuous endothelial continuity between retinal and cerebral microvasculature provide the biological rationale for retinal biomarker research in neurodegeneration [6]. Epidemiological studies have associated reduced retinal arteriolar calibre, elevated venular calibre, increased arteriovenous ratio deviation, and greater tortuosity with cognitive decline and dementia incidence [4, 5]. Semi-automated tools such as the Singapore I Vessel Assessment (SIVA) platform have established normative calibre metrics and demonstrated associations with cardiovascular risk.

Critical limitations of this evidence must be stated. Retinal microvascular changes are not pathognomonic for dementia: similar changes occur in hypertension, type 2 diabetes mellitus, chronic kidney disease, and cardiovascular disease generally [13]. Failure to adjust for these comorbidities risks conflating a generic vascular disease signal with a dementia-specific one. Furthermore, virtually all published associations are cross-sectional; longitudinal evidence establishing that retinal vascular changes temporally

10.48047/jocaaa.2024.33.02.54

precede cognitive decline by a clinically meaningful interval remains sparse [4]. The relationship between retinal changes and specific dementia subtypes — AD, vascular dementia, Lewy body disease, frontotemporal dementia — is poorly characterised. These constraints inform our decision to position this study as a methodological proof-of-concept with a proxy task, rather than a clinical diagnostic validation.

2.2 Deep Learning in Retinal Image Analysis

CNNs have achieved strong performance across retinal classification benchmarks including DR grading, glaucoma screening, and age-related macular degeneration staging [14, 15]. The APTOS 2019 challenge and prior Kaggle DR competitions established competitive performance thresholds for the proxy task used in this study. Foundation models pre-trained at scale — most notably RETFound, trained on 1.6 million unlabelled retinal images — have demonstrated strong transfer performance across ocular and systemic disease prediction tasks [16].

The inductive bias of CNNs and ViTs towards local spatial features on a regular grid constitutes a principled limitation for vascular topology modelling. The relational connectivity between vessel segments and junction points is a global, graph-structured property. A CNN processing two fundus images with identical local vessel statistics but different branching topologies will assign similar intermediate representations to both images, as the convolutional operation has no explicit mechanism for representing inter-node relational structure. Whether this representational limitation translates to measurable performance deficits for microvascular pathology detection — and by how much — is the central empirical question this study addresses.

2.3 Graph Neural Networks in Medical Imaging

GNNs generalise deep learning to arbitrary graph-structured domains through message-passing: each node iteratively aggregates feature information from its neighbourhood, enabling models to capture both local node properties and global relational structure [7, 8]. The GCN [7] applies symmetric normalised aggregation with equal neighbourhood weights; the GAT [8] learns per-edge attention coefficients, enabling selective neighbour weighting. Both architectures are well-established and have been applied across medical domains including drug-target interaction prediction [17], brain connectivity modelling [9], population-level disease prediction [10], and retinal vessel segmentation [18].

In neuroimaging, graph-structured representations of functional brain connectivity — brain regions as nodes, functional correlations as edges — have consistently outperformed CNN baselines for MCI classification, providing proof-of-concept for the general principle that graph topology carries discriminative information beyond local features [9, 10]. The present study applies this principle to retinal vascular topology, a domain where explicit graph-based topology modelling has not previously been systematically evaluated for cognitive decline-related classification tasks.

3. Materials and Methods

3.1 Datasets

Three publicly available datasets were employed. All data access complied with the terms and conditions of the respective dataset providers. No new patient data were collected; accordingly, no institutional ethics approval was required. De-identification status is as provided by the original dataset curators.

3.1.1 APTOS 2019 Blindness Detection Dataset (Primary Classification Dataset)

The APTOS 2019 dataset [19] comprises 3,662 fundus photographs acquired at Aravind Eye Hospitals, India, graded on a five-class DR severity scale (Grade 0: no DR; Grades 1–4: mild to proliferative DR). For this study, images were binarised: Class 0 (no DR, Grade 0) and Class 1 (DR present, Grades 1–4). From the 3,662-image release, images yielding fewer than 10 extractable vessel junction nodes after preprocessing were excluded ($n = 366$), leaving 3,296 graphs, all of which produced valid vascular graphs. These were partitioned into training ($n = 2,490$, 75.5%), validation ($n = 366$, 11.1%), and test ($n = 440$, 13.4%) sets via stratified random split (random seed = 42). Test-set class distribution: no DR $n = 215$; DR $n = 225$.

Proxy task justification: DR involves retinal microvascular pathology — microaneurysms, capillary non-perfusion, neovascularisation, and haemorrhage — mechanistically linked through endothelial dysfunction and microvascular remodelling to the cerebrovascular changes associated with dementia [5]. APTOS 2019 provides a large-scale, standardised, publicly available dataset with clear binary pathology labels appropriate for a controlled methodological proof-of-concept. DR classification and dementia screening are not equivalent tasks; the proxy design is a stated limitation (see Section 5.5).

3.1.2 DRIVE Dataset (Pipeline Validity Confirmation)

The Digital Retinal Images for Vessel Extraction (DRIVE) dataset [20] contains 40 colour fundus photographs acquired with a Canon CR5 7-field non-mydratic camera at 45° FOV (565×584 px; TIFF images with GIF annotation masks) with vessel segmentation annotations from two independent human observers. The first observer's annotations (21_manual1.gif) served as reference for qualitative pipeline validation on image 21_training.tif. No classification training or testing was performed on DRIVE.

3.1.3 STARE Dataset (Cross-Device Pipeline Validation)

The STructured Analysis of the Retina (STARE) dataset [21] contains 397 retinal fundus images acquired with a TopCon TRV-50 camera at 35° FOV (700×605 px native resolution, PPM format), representing a categorically different imaging device from DRIVE. The Kaggle distribution (vidheeshnacode/stare-dataset) was used; no expert-annotated vessel segmentation masks were present in this distribution. Five representative images (im0001–im0005.ppm) were processed using the identical, unmodified pipeline to assess cross-device portability at the graph construction level. No classification training or testing was performed on STARE.

3.2 Graph Construction Pipeline

Each fundus image was converted to a PyG Data object through five sequential stages, implemented in Python 3.10 using scikit-image 0.21, OpenCV 4.8, and scikit-learn 1.3.

3.2.1 Preprocessing

Each image was resized to 512×512 pixels, and the green channel was extracted (optimal vessel-to-background contrast owing to the haemoglobin absorption peak at ≈540 nm) [20]. Contrast Limited Adaptive Histogram Equalisation (CLAHE; clip limit = 2.0; tile grid = 8×8) was applied to enhance local vessel contrast while suppressing noise amplification.

3.2.2 Vessel Enhancement and Segmentation

The Frangi vesselness filter [11] was applied across three scales ($\sigma \in \{1, 2, 3\}$ px) to produce a multi-scale vesselness probability map. Binary vessel segmentation was obtained by thresholding at 0.5, followed by morphological closing with a 2-pixel disk structuring element to fill small vessel-continuity gaps. The Frangi filter is derived from eigenvalue analysis of the local image Hessian, mathematically encoding multi-scale tubular structure — the expected morphology of retinal vessels — without requiring labelled training data. This makes it robust, interpretable, and independent of dataset-specific training, consistent with established retinal vessel analysis practice [22].

3.2.3 Skeletonisation

The binary vessel mask was thinned to a one-pixel-wide medial axis representation using morphological skeletonisation (Zhang–Suen algorithm [12]). Vessel junction points (skeleton pixels with three or more 8-connected neighbours) and vessel endpoints (skeleton pixels with exactly one neighbour) were identified as graph nodes.

3.2.4 Graph Construction

Edges were formed between node pairs within a 15-pixel Euclidean distance threshold via k-NN construction with $k = 4$ edges retained per node. Parameter justification: inter-junction distances in 512×512 fundus images correspond approximately to 10–20 pixels (anatomical capillary spacing); $k = 4$ reflects the predominant quadrifurcation branching pattern of retinal arterioles. Both parameters were applied identically to DRIVE and STARE without modification and confirmed to produce topologically coherent graphs (Section 4.5). Formal sensitivity analysis across (k , threshold) parameter space is acknowledged as a limitation and future work priority.

3.2.5 Node Feature Vector

Each node was assigned an 8-dimensional feature vector: (f1) normalised x-coordinate; (f2) normalised y-coordinate; (f3) local vessel pixel density in a 7×7 patch; (f4) patch vessel pixel count; (f5) node degree (number of k-NN edges); (f6) mean orientation of connected edges; (f7) local intensity contrast (patch standard deviation); (f8) normalised Euclidean distance to image centroid. This feature set captures geometric position, local vascular density, junction complexity, directional structure, and spatial context.

10.48047/jocaaa.2024.33.02.54

Table 1. *Cross-Device Pipeline Validation — DRIVE vs. STARE. Identical pipeline applied without modification. Results for STARE are from five sample images (im0001–im0005.ppm).*

Property	DRIVE (n = 1 image)	STARE (n = 5 images, avg)
Camera / system	Canon CR5 (7-field)	TopCon TRV-50
Field of view (FOV)	45°	35°
Native resolution	565 × 584 px	700 × 605 px
Pipeline input resolution	512 × 512 px	512 × 512 px
Vessel detected pixels	2,266	Varies by image quality
Graph nodes	148	565 avg (range: 82–936)
Graph edges (undirected)	416	1,985 avg (range: 233–3,410)
Node feature dimension	8	8 (identical)
Expert GT masks used	Qualitative QC	Not in Kaggle distribution
Pipeline code modified?	—	No
PyG sample object	Data(x=[148,8], edge_index=[2,832])	Data(x=[936,8], edge_index=[2,6820])

Note. STARE graph statistics are from five representative images only; cross-device portability at full dataset scale has not been systematically evaluated.

3.3 GNN Architectures

Three models were implemented and compared in a controlled ablation. All used identical optimiser, loss, batch construction, and early stopping configuration to ensure comparability.

3.3.1 MLP Baseline — Experiment A (No Topology)

A three-layer Multi-Layer Perceptron operating on global mean-pooled node features, with no message passing or edge information. This model establishes the performance achievable from node-level features alone, isolating the contribution of topology. Architecture: global mean pooling concatenated with global max pooling of 8-dimensional node features → 16-dim; Linear(16→128)→ReLU→Dropout(p=0.3)→Linear(128→64)→ReLU→Dropout(p=0.3)→Linear(64→2).

3.3.2 Graph Convolutional Network — Experiment B (Equal-Weight Aggregation)

A three-layer GCN [7] using the GCNConv operator performing symmetric normalised neighbourhood aggregation: $h_v^{(l+1)} = \sigma(\sum_{u \in N(v) \cup \{v\}} (d_v d_u)^{-1/2} W^{(l)} h_u^{(l)})$, where d_v denotes node degree and W is a learnable weight matrix. Architecture: GCNConv(8→512)→BatchNorm(512)→ReLU→Dropout(p=0.3); GCNConv(512→512)→BatchNorm(512)→ReLU→Dropout(p=0.3); GCNConv(512→64)→BatchNorm(64)→ReLU; global mean pooling concatenated with global max pooling → 128-dim; Linear(128→128)→ReLU→Dropout(p=0.3)→Linear(128→64)→ReLU→Dropout(p=0.3)→Linear(64→2).

3.3.3 Graph Attention Network — Experiment C (Learned Attention)

A three-layer GAT [8] with 8 attention heads using the GATConv operator: $h_v^{(l+1)} = \sigma(\sum_{u \in N(v) \cup \{v\}} \alpha_{vu} W^{(l)} h_u^{(l)})$, where $\alpha_{vu} = \text{softmax}_u(\text{LeakyReLU}(a^T [W h_v \parallel W h_u]))$. Architecture: GATConv(8→512, heads=8, concat=True, attn_dropout=0.3)→BatchNorm(512)→ReLU; GATConv(512→512, heads=8, concat=True, attn_dropout=0.3)→BatchNorm(512)→ReLU; GATConv(512→64, heads=1, concat=False, attn_dropout=0.3)→BatchNorm(64)→ReLU; global mean pooling concatenated with global max pooling → 128-dim; Linear(128→128)→ReLU→Dropout(p=0.3)→Linear(128→64)→ReLU→Dropout(p=0.3)→Linear(64→2).

3.4 Training Configuration and Evaluation

Optimiser: Adam (learning rate = 0.001; weight decay = 1×10^{-4}). Loss: cross-entropy with class weights inversely proportional to training set frequency (Class 0: 0.5104; Class 1: 0.4896) to address mild class imbalance. Early stopping: patience = 10 epochs on validation accuracy; best checkpoint saved. Maximum epochs: 50. Batch size: 32 graphs. Hardware: NVIDIA Tesla T4 × 2 (Kaggle Notebooks). Framework: PyTorch 2.x, PyTorch Geometric 2.x. Random seed: 42 (NumPy, PyTorch, Python random).

Primary evaluation metrics: overall test accuracy; balanced accuracy (arithmetic mean of per-class recall — preferred for mildly imbalanced evaluation); AUC-ROC; per-class precision, recall, and F1-score. AUC was computed using the one-vs-rest strategy.

Statistical testing note: No formal statistical significance tests (e.g., DeLong test for AUC comparison, bootstrap confidence intervals) were conducted. All model comparison results are descriptive and should be interpreted as exploratory. Formal significance testing is identified as a priority for future confirmatory work.

4. Results

4.1 Training Dynamics

GCN training converged earliest (early stopping at epoch 23; best validation accuracy 94.81%). GAT training stopped at epoch 19 (best validation accuracy 91.26% at epoch 9; validation loss showed mild instability of 0.26–0.63 after epoch 9). MLP training ran to the maximum of 40 epochs (best validation accuracy 91.80%). Early convergence of the GCN relative to both the GAT and MLP is consistent with its simpler inductive bias and lower parameter count.

4.2 Ablation Study — Main Classification Results

Table 2 presents the primary ablation results on the held-out test set ($n = 440$). GCN achieved the strongest performance: Balanced Accuracy 92.07%, AUC-ROC 0.9711. GAT achieved AUC-ROC 0.9504 (Balanced Accuracy 88.24%). The topology-free MLP baseline achieved AUC-ROC 0.9420 (Balanced Accuracy 88.12%). Both graph-structured models exceed the topology-free baseline (Table 3), confirming that graph topology carries discriminative information independent of node-level features. These differences are descriptive; no statistical significance tests were performed.

Table 2. *Ablation Study — Test Set Results ($n = 440$). Best-performing model (GCN) highlighted. Δ AUC computed relative to MLP baseline. No statistical significance testing was performed.*

Model	Test Acc. (%)	Balanced Acc. (%)	AUC-ROC	Δ AUC vs. MLP	Stop Epoch	Best Val Acc.
MLP — Topology-Free Baseline	88.18	88.12	0.9420	—	40	91.80%
GAT — Learned Attention (8 heads)	88.18	88.24	0.9504	+0.0084	19	91.26%
GCN — Equal-Weight Aggregation ★	92.05	92.07	0.9711	+0.0291	23	94.81%

★ GCN is the best-performing model on all metrics. Δ AUC differences are descriptive only; statistical significance was not tested.

Table 3. *Topology Contribution Analysis — Δ AUC relative to the topology-free MLP baseline.*

Comparison	Δ AUC	Interpretation
GCN vs. MLP (primary topology effect)	+0.0291	Graph topology provides independent discriminative information — confirmed
GAT vs. MLP (any graph)	+0.0084	Both graph architectures exceed topology-free

vs. no graph)		baseline
GCN vs. GAT (equal-weight vs. attention)	+0.0207 (GCN)	Diffuse vascular pathology favours uniform aggregation over selective attention (post-hoc)

4.3 GAT Per-Class Performance and Confusion Matrix

Table 4 presents per-class precision, recall, and F1-score for the GAT model; Table 5 presents the confusion matrix. GAT achieves higher sensitivity for no-DR images (recall 0.9070) than for DR images (recall 0.8578), suggesting mild systematic bias toward the majority class despite class-weighted training.

Table 4. *GAT Model — Per-Class Performance (Test Set, $n = 440$).*

Class	Precision	Recall (Sensitivity)	Specificity	F1-Score	Support
No DR (Grade 0)	0.8590	0.9070	0.8578	0.8824	215
DR (Grades 1–4)	0.9061	0.8578	0.9070	0.8813	225
Macro Average	0.8826	0.8824	0.8824	0.8818	440
Weighted Average	0.8831	0.8818	0.8818	0.8818	440

Table 5. *GAT Confusion Matrix (Test Set, $n = 440$).*

	Predicted: No DR	Predicted: DR
Actual: No DR ($n = 215$)	195 (TN)	20 (FP)
Actual: DR ($n = 225$)	32 (FN)	193 (TP)

4.4 GCN Performance

The GCN achieved Test Accuracy 92.05%, Balanced Accuracy 92.07%, and AUC-ROC 0.9711. The GCN converged at epoch 23, reaching best validation accuracy 94.81%. Full per-class breakdown was not separately tabulated for GCN, as the primary purpose of the per-class analysis (Table 4) was to characterise the GAT model's error patterns for comparison with the aggregate ablation. GCN's superiority over the GAT on aggregate metrics is interpreted in Section 5.2.

4.5 Cross-Dataset Pipeline Validation (DRIVE and STARE)

The graph construction pipeline was independently validated on two datasets representing different camera systems. No classification was performed; the validation quantifies graph topology statistics as evidence of pipeline validity and cross-device portability.

DRIVE Validation: Applied to image 21_training.tif against expert annotation 21_manual1.gif. Frangi filtering detected 2,266 vessel pixels; skeletonisation and junction extraction yielded 148 nodes; k-NN construction produced 416 undirected edges (832 directed). PyG object: Data($x=[148, 8]$, $edge_index=[2, 832]$, $y=[1]$). The resulting graph topology was qualitatively consistent with the expert-annotated vessel structure.

10.48047/jocaaa.2024.33.02.54

STARE Validation: Five STARE images (im0001–im0005.ppm) were processed using the identical, unmodified pipeline. Table 6 presents per-image results; Table 1 presents the cross-device comparison. The pipeline produced valid graphs for all five images. Table 1 is already presented in Section 3.2.

Table 6. STARE Dataset — Per-Image Graph Statistics ($n = 5$ images). Identical pipeline applied without modification.

Image File	Nodes	Edges (Undirected)	Node Feat. Dim	PyG Object Representation
im0001.ppm	936	3,410	8	Data(x=[936, 8], edge_index=[2, 6820])
im0002.ppm	845	3,171	8	Data(x=[845, 8], edge_index=[2, 6342])
im0003.ppm	486	1,499	8	Data(x=[486, 8], edge_index=[2, 2998])
im0004.ppm	82	233	8	Data(x=[82, 8], edge_index=[2, 466])
im0005.ppm	476	1,610	8	Data(x=[476, 8], edge_index=[2, 3220])
Average ($n = 5$)	565	1,985	8	Range: 82–936 nodes; 233–3,410 edges

Note. High within-dataset node count variance (82–936) reflects heterogeneous image quality across STARE images. Systematic quality filtering was not applied. Results should not be extrapolated to the full STARE dataset ($n = 397$).

The substantially higher graph density in STARE (avg 565 nodes / 1,985 edges) relative to DRIVE (148 nodes / 416 edges) is expected, reflecting STARE’s higher native resolution (700×605 vs. 565×584 px) and different optical characteristics. Critically, both datasets produced graphs within the same 8-dimensional node feature space, and no pipeline modification was required. This confirms that the pipeline is not camera-specific and does not require device-dependent recalibration — a necessary prerequisite for any clinical screening tool intended for deployment across heterogeneous imaging hardware.

5. Discussion

5.1 Principal Findings

This study produces three interconnected findings. First, and most importantly: graph-based modelling of retinal vascular topology yields a measurable improvement over topology-free classification. The GCN achieved $\Delta AUC = +0.0291$ over the MLP baseline using identical node features but without any message passing or edge information. This constitutes experimental evidence that vascular topology — the relational connectivity of

10.48047/jocaaa.2024.33.02.54

the retinal vascular network — carries independent discriminative information not captured by local node-level statistics. Vascular topology is an independent discriminative feature space. This is the central empirical claim of this paper and is supported by a controlled, pre-specified ablation design.

Second, cross-device validation on five STARE images confirms preliminary pipeline portability across fundus camera systems. The same pipeline code, applied without modification to TopCon TRV-50 images (35° FOV, 700×605 px) and Canon CR5 DRIVE images (45° FOV, 565×584 px), produces structurally valid graphs in the same 8-dimensional feature space. This device-agnostic behaviour is a necessary condition for clinical screening tools intended for deployment across heterogeneous hardware environments.

Third, the ablation demonstrates that both graph-structured architectures (GCN and GAT) outperform the topology-free baseline, confirming that the performance advantage derives from graph structure rather than from any specific architectural choice or implementation detail. Results are descriptive; no statistical significance tests were performed and all findings are exploratory.

5.2 Why GCN Outperforms GAT: A Post-Hoc Interpretation

The GCN outperforming the GAT ($\Delta\text{AUC} = +0.0207$) is theoretically interpretable, though the explanation is post-hoc and should be prospectively tested. GAT's learned attention concentrates information flow on selected neighbours, an inductive bias suited to tasks where specific relational configurations drive classification — such as brain connectivity, where hub nodes have disproportionate predictive value [9]. Retinal microvascular pathology in DR is spatially diffuse: microaneurysms, haemorrhages, and calibre changes are distributed across the retinal field without concentration at specific topological types of junction. Equal-weight neighbourhood aggregation over all connected nodes — the GCN approach — may more faithfully propagate this diffuse signal than selective attention, which may suppress informative but weakly-attended connections. This interpretation should be tested prospectively using attention map visualisation across retinal disease phenotypes with varying spatial distributions.

5.3 Proxy Task Validity and Dementia Screening Implications

DR is used as a proxy because no large-scale, publicly available retinal imaging dataset with accompanying dementia diagnoses exists for deep learning training. DR and dementia-associated retinal changes share biological substrates — endothelial dysfunction, capillary loss, arteriolar remodelling — but are distinct phenotypes. The classification signal discriminating DR from no-DR may partially reflect DR-specific pathology (microaneurysms, neovascularisation) rather than the subtler, diffuse calibre changes associated with cognitive decline.

The scientific contribution of this work is the framework and the empirical topology-vs-no-topology finding; not a claim about dementia detection performance. The $\Delta\text{AUC} = +0.0291$ establishes proof-of-concept that topology carries discriminative information in one controlled retinal microvascular pathology dataset. Applying this framework to dementia-

10.48047/jocaaa.2024.33.02.54

specific longitudinal retinal imaging cohorts — including the UK Biobank retinal imaging substudy ($n > 67,000$) and prospective ophthalmology-cognitive decline cohorts currently in recruitment — is the essential next step. The pipeline presented here is designed to be directly applicable to these datasets when they become available at sufficient scale.

5.4 CNN Comparison: Scope and Future Work

No direct comparison with CNN pixel-grid models was conducted. CNNs and GNNs operate on fundamentally incompatible input representations — pixel arrays versus graph objects — making head-to-head AUC comparison on the same data split methodologically problematic: both the input space and the discriminative features exploited differ. The MLP ablation design correctly isolates the topology contribution by holding node features constant and varying only the presence of graph structure. A methodologically sound CNN-GNN comparison would require a hybrid design — CNN-extracted patch embeddings used as node features for the GNN, compared against the same CNN used for pixel-grid classification — and is a clear future work direction. No claim that GNNs outperform CNNs for retinal classification can be made from these results.

5.5 Limitations

- Proxy task non-equivalence: DR classification is not equivalent to dementia screening. Biological validity of vascular topology features for dementia-specific retinal changes remains undemonstrated and is the primary open scientific question.
- Absence of statistical significance testing: No formal significance tests (DeLong test for AUC, bootstrap confidence intervals, McNemar test for accuracy) were performed. All model comparison results are descriptive and exploratory. This is a material methodological limitation.
- No external classification validation: All APTOS 2019 results derive from a single multi-clinic acquisition programme. Performance generalisation to images from different institutions, cameras, or population demographics is unknown.
- No direct CNN comparison: Incompatible input representations preclude a methodologically sound head-to-head comparison with CNN pixel-grid baselines. Topology advantage over pixel-based learning cannot be claimed from these results.
- STARE sample size: Only five STARE images were processed. Cross-device portability at full dataset scale ($n = 397$) and across broader device variation has not been established. The five-image validation confirms pipeline function but does not characterise cross-device performance statistically.
- STARE graph density heterogeneity: Wide node count variance (82–936 nodes across five images) indicates substantial image quality variation within the STARE sample. Systematic quality filtering was not applied; future cross-device studies should implement standardised quality inclusion criteria.

10.48047/jocaaa.2024.33.02.54

- Graph construction parameter sensitivity: $k = 4$ and distance threshold = 15 px were anatomically motivated and confirmed on DRIVE and STARE without code modification, but were not formally optimised via sensitivity analysis across the parameter space.
- Hand-crafted node features: The 8-dimensional geometric feature vector was designed by domain knowledge rather than learned. CNN patch embeddings as node features may provide richer representations and should be explored in future hybrid CNN-GNN designs.
- Cross-sectional design: APTOS 2019 is a cross-sectional dataset. Longitudinal topology trajectory tracking across disease progression cannot be performed and is essential for establishing temporal biomarker validity.

6. Conclusion

This paper presents a graph-based deep learning framework for retinal vascular topology modelling as a methodological foundation for non-invasive dementia screening. The primary contribution is the framework itself — converting fundus images into topologically structured graphs via Frangi filtering, morphological skeletonisation, and k -NN edge construction with 8-dimensional geometric node features — and the empirical demonstration that vascular topology constitutes an independent discriminative feature space.

Pipeline validity was confirmed on DRIVE (148 nodes, 416 edges from image 21_training.tif, qualitatively consistent with expert vessel annotation). Preliminary cross-device portability was demonstrated on STARE (TopCon TRV-50, 35° FOV): five images processed without pipeline modification, yielding average 565 nodes and 1,985 edges per image. This constitutes evidence of device-agnostic pipeline design at the graph construction level; it does not represent a claim of classification performance generalisation across imaging devices.

In a controlled ablation on 3,296 APTOS 2019 retinal graphs, the GCN achieved AUC-ROC 0.9711 and Balanced Accuracy 92.07%, outperforming both the GAT (AUC 0.9504) and the topology-free MLP baseline (AUC 0.9420; $\Delta\text{AUC} = +0.0291$). Both graph-structured models exceeded the topology-free baseline, confirming the topology effect across architectures. Statistical significance testing was not performed; all results are exploratory. GCN outperforming GAT is consistent with the hypothesis that diffuse microvascular pathology signal favours uniform neighbourhood aggregation over selective attention.

These findings are a proof-of-concept, not a clinical validation. No direct dementia detection claim is made. Applying this framework to longitudinal, dementia-specific retinal imaging cohorts — including the UK Biobank retinal imaging sub study and prospective ophthalmology-cognitive decline datasets currently in recruitment — is the essential and immediate next step. The framework is designed for direct applicability to those datasets when they reach sufficient scale for deep learning training.

Declarations

Ethics Approval and Consent to Participate

This study used exclusively publicly available, de-identified datasets (APTOS 2019, DRIVE, STARE) accessed in accordance with the terms and conditions of the respective dataset providers and challenge organisers. No new patient data were collected, and no patients were recruited. Accordingly, no institutional ethics board approval or patient consent was required for this study.

Consent for Publication

Not applicable. This study does not contain any individual person's data in any form.

Availability of Data and Materials

All datasets used in this study are publicly available:

- APTOS 2019 Blindness Detection: <https://www.kaggle.com/competitions/aptos2019-blindness-detection>
- DRIVE (Digital Retinal Images for Vessel Extraction): <https://drive.grand-challenge.org/>
- STARE (STructured Analysis of the Retina): <http://cecas.clemson.edu/~ahoover/stare/>

The graph construction pipeline code and PyTorch Geometric training scripts are available from the corresponding author upon reasonable request.

Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. Computational resources were provided by the Kaggle Notebooks platform (NVIDIA Tesla T4 $\times$ 2 GPUs, accessed free of charge).

Authors' Contributions

CRedit (Contributor Roles Taxonomy) statement:

- Conceptualisation: S.M.
- Methodology: S.M.
- Software and implementation: S.M.

- Formal analysis: S.M.
- Data curation: S.M.
- Writing — original draft: S.M.
- Writing — review and editing: S.M., A.S.
- Supervision: A.S.
- Project administration: A.S.

(S.M. = First Author; A.S. = Supervisor/Co-Author. Full names withheld for double-blind review.)

Acknowledgements

The authors acknowledge the APTOS 2019 challenge organisers and Aravind Eye Hospitals for making the dataset publicly available; the DRIVE and STARE dataset curators for providing publicly accessible retinal imaging benchmarks; and the Kaggle platform for providing free GPU computational resources. No AI writing assistants were used in the preparation of this manuscript.

References

- [1] M. Prince, A. Wimo, M. Guerchet, G. Ali, Y.-T. Wu, and M. Prina, “World Alzheimer Report 2015: The Global Impact of Dementia,” Alzheimer’s Disease International, London, 2016.
- [2] B. Dubois et al., “Preclinical Alzheimer’s disease: Definition, natural history, and diagnostic criteria,” *Alzheimer’s Dement.*, vol. 12, no. 3, pp. 292–323, 2016. doi: 10.1016/j.jalz.2016.02.002
- [3] C. R. Jack et al., “NIA-AA Research Framework: Toward a biological definition of Alzheimer’s disease,” *Alzheimer’s Dement.*, vol. 14, no. 4, pp. 535–562, 2018. doi: 10.1016/j.jalz.2018.02.018
- [4] S. McGrory et al., “The application of retinal fundus camera imaging in dementia: A systematic review,” *Alzheimer’s Dement. Diagn. Assess. Dis. Monit.*, vol. 6, no. 1, e12023, 2021. doi: 10.1002/dad2.12023
- [5] Y. Zhu et al., “Optical coherence tomography angiography reveals vascular changes in early stages of diabetic retinopathy,” *Br. J. Ophthalmol.*, vol. 103, no. 2, pp. 217–223, 2019. doi: 10.1136/bjophthalmol-2018-312041
- [6] N. Patton et al., “Retinal vascular image analysis as a potential screening tool for cerebrovascular disease: A rationale based on homology between cerebral and retinal microvasculatures,” *J. Anat.*, vol. 206, no. 4, pp. 319–348, 2006. doi: 10.1111/j.1469-7580.2005.00455.x

10.48047/jocaaa.2024.33.02.54

- [7] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in Proc. Int. Conf. Learn. Representations (ICLR), 2017. [Online]. Available: <https://arxiv.org/abs/1609.02907>
- [8] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," in Proc. Int. Conf. Learn. Representations (ICLR), 2018. [Online]. Available: <https://arxiv.org/abs/1710.10903>
- [9] S. Parisot et al., "Disease prediction using graph convolutional networks: Application to autism spectrum disorder and Alzheimer's disease," *Med. Image Anal.*, vol. 48, pp. 117–130, 2018. doi: 10.1016/j.media.2018.06.001
- [10] S. I. Ktena et al., "Distance metric learning using graph convolutional networks: Application to functional brain networks," in Proc. MICCAI 2017, LNCS vol. 10433, pp. 469–477, Springer, 2017. doi: 10.1007/978-3-319-66182-7_54
- [11] A. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever, "Multiscale vessel enhancement filtering," in Proc. MICCAI 1998, LNCS vol. 1496, pp. 130–137, Springer, 1998. doi: 10.1007/BFb0056195
- [12] T. Y. Zhang and C. Y. Suen, "A fast parallel algorithm for thinning digital patterns," *Commun. ACM*, vol. 27, no. 3, pp. 236–239, 1984. doi: 10.1145/357994.358023
- [13] J. M. Wardlaw, C. Smith, and M. Dichgans, "Mechanisms of sporadic cerebral small vessel disease: Insights from neuroimaging," *Lancet Neurol.*, vol. 12, no. 5, pp. 483–497, 2013. doi: 10.1016/S1474-4422(13)70060-7
- [14] G. Litjens et al., "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, 2017. doi: 10.1016/j.media.2017.07.005
- [15] J. H. Tan, H. Fujita, S. Sivaprasad et al., "Automated segmentation of exudates, haemorrhages, microaneurysms using single convolutional neural network," *Inf. Sci.*, vol. 420, pp. 66–76, 2017. doi: 10.1016/j.ins.2017.08.050
- [16] Y. Zhou et al., "A foundation model for generalizable disease detection from retinal images," *Nature*, vol. 622, pp. 156–163, 2023. doi: 10.1038/s41586-023-06555-x
- [17] M. Zitnik, M. Agrawal, and J. Leskovec, "Modeling polypharmacy side effects with graph convolutional networks," *Bioinformatics*, vol. 34, no. 13, pp. i457–i466, 2018. doi: 10.1093/bioinformatics/bty294
- [18] S. I. Ktena et al., "Metric learning with spectral graph convolutions on brain connectivity networks," *NeuroImage*, vol. 169, pp. 431–442, 2018. doi: 10.1016/j.neuroimage.2017.12.052
- [19] APTOS 2019 Blindness Detection Challenge, Kaggle, 2019. [Online]. Available: <https://www.kaggle.com/c/aptos2019-blindness-detection>

10.48047/jocaaa.2024.33.02.54

- [20] J. J. Staal, M. D. Abramoff, M. Niemeijer, M. A. Viergever, and B. van Ginneken, "Ridge-based vessel segmentation in color images of the retina," *IEEE Trans. Med. Imaging*, vol. 23, no. 4, pp. 501–509, 2004. doi: 10.1109/TMI.2004.825627
- [21] A. Hoover, V. Kouznetsova, and M. Goldbaum, "Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response," *IEEE Trans. Med. Imaging*, vol. 19, no. 3, pp. 203–210, 2000. doi: 10.1109/42.845178
- [22] J. V. B. Soares, J. J. G. Leandro, R. M. Cesar, H. F. Jelinek, and M. J. Cree, "Retinal vessel segmentation using the 2-D Gabor wavelet and supervised classification," *IEEE Trans. Med. Imaging*, vol. 25, no. 9, pp. 1214–1222, 2006. doi: 10.1109/TMI.2006.879967
- [23] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. ICML*, vol. 97, pp. 6105–6114, 2019. [Online]. Available: <https://arxiv.org/abs/1905.11946>
- [24] Alzheimer's Association, "2021 Alzheimer's disease facts and figures," *Alzheimer's Dement.*, vol. 17, no. 3, pp. 327–406, 2021. doi: 10.1002/alz.12328