Supplementary Materials

Table S1. Details of the unsupervised learning dataset (Reconstruction of images)

Unsupervised learning (D_{u1}) - Reconstruction of fundus images	
Unlabeled (40%)†	
Glaucoma	Normal
drishtiGS_021 - drishtiGS_032, drishtiGS_034, drishtiGS_038 - drishtiGS_040, drishtiGS_043 - drishtiGS_045, drishtiGS_048 - drishtiGS_056	drishtiGS_035 - drishtiGS_037, drishtiGS_041 - drishtiGS_042, drishtiGS_046 - drishtiGS_047, drishtiGS_057, drishtiGS_072, drishtiGS_078, drishtiGS_085, drishtiGS_089
Im0517_g_ORIGA - Im0583_g_ORIGA	Im0097_ORIGA - Im0289_ORIGA

† ORIGA and Drishti- GS datasets were used.

Table S2. Details of the pseudo-labeling dataset (Prediction of pseudo-labels)

Pseudo-labeling (D_{u1} or S_1)- Prediction of pseudo-labels	
Unlabeled (40%)†	
Glaucoma	Normal
drishtiGS_058 - drishtiGS_071, drishtiGS_073 - drishtiGS_077, drishtiGS_079 - drishtiGS_088	drishtiGS_090 - drishtiGS_101
Im0584_g_ORIGA - Im0650_g_ORIGA	Im0290_ORIGA - Im0482_ORIGA

† ORIGA and Drishti- GS datasets were used.

Table S3. Details of the supervised learning dataset (Glaucoma detection)

Supervised learning (D_L)- Glaucoma detection	
labeled (20%)†	
Glaucoma	Normal
drishtiGS_001 - drishtiGS_006, drishtiGS_010 - drishtiGS_012, drishtiGS_014 - drishtiGS_016, drishtiGS_019 - drishtiGS_020	drishtiGS_007 - drishtiGS_009, drishtiGS_013, drishtiGS_017 - drishtiGS_018, drishtiGS_033
Im0483_g_ORIGA - Im0516_g_ORIGA	Im0001_ORIGA - Im0096_ORIGA

† ORIGA and Drishti- GS datasets were used.

Table S4. Details of the training set for glaucoma detection via supervised learning

Training set†	
Glaucoma	Normal
drishtiGS_001 -drishtiGS_006, drishtiGS_010 -drishtiGS_012, drishtiGS_014 -drishtiGS_015	drishtiGS_007 -drishtiGS_0010, drishtiGS_013, drishtiGS_017
Im0483_g_ORIGA - Im0509_g_ORIGA	Im0001_ORIGA - Im0076_ORIGA

†ORIGA and Drishti- GS datasets were used.

Table S5. Details of the testing set for glaucoma detection

Testing set†	
Glaucoma	Normal
drishtiGS_016, drishtiGS_019 -drishtiGS_020	drishtiGS_018, drishtiGS_033
Im0510_g_ORIGA - Im0516_g_ORIGA	Im0077_ORIGA -Im0096_ORIGA

† ORIGA and Drishti- GS datasets were used.

- Table S1-S5 above shows the complete list of image identifiers. These tables cover the unsupervised (for image reconstruction), pseudo-label generation, and supervised learning (for glaucoma detection) training and test partitions for glaucoma detection.

- The inclusion of image identifier ensures that each image is uniquely identified, allowing for verifiable confirmation that there is no overlap between partitions.

Dataset Partitioning Procedure

Input:

- Two labeled datasets: **Drishti-GS** and **ORIGA**
- Each dataset includes images labeled as **Glaucoma** or **Normal**

Objective:

To manually split the data for various tasks:

- 40% for **Unsupervised Learning**
- 40% for **Pseudo-Label Generation**
- 20% for **Supervised Learning**, further divided into:
    - 80% **Training**
    - 20% **Testing**

Procedure:

1. **Level-1 consideration**
   For each dataset (**Drishti-GS** and **ORIGA**):
   - Separate the images into two groups based on their labels:
     - **Glaucoma**
     - **Normal**
2. **Calculate Class-Based Split Sizes**
   For each class (Glaucoma and Normal) in each dataset:
   - N1 = 40% of the total number of images for **Unsupervised Learning from both datasets combined**
   - N2 = 40% of the total number of images for **Pseudo-Label Generation from both datasets combined**
   - N3 = 20% of the total number of images for **Supervised Learning from both datasets combined**
3. **Manual Selection**
   - From each class, manually select the required number of images for each purpose by
   - Separate N1 number of images for **Unsupervised Learning**
   - Separate N2 number of images for **Pseudo-Label Generation**
   - Separate N3 number of images for **Supervised Learning**
   - Ensure that the selected images are evenly distributed between the classes (**Glaucoma** and **Normal)** as well as datasets within each split.
4. **Supervised Learning Sub-Split**
   - From the N3 number of images in the **Supervised Set**, split it into:
     - **80% for Training**
     - **20% for Testing**
5. **Final Partitioning**
   - Combine the corresponding partitions from Drishti-GS and ORIGA to form:
     - **Unsupervised Set**
     - **Pseudo Label Set**
     - **Supervised Training Set**
     - **Supervised Test Set**
6. **Verification**
   - Confirm there is no overlap between any of the partitions, to ensure uniqueness.

All image identifiers for each partition are included in the Supplementary Materials (Tables S1-S5). This procedure ensures that partitions are mutually exclusive.

Table S6. Summary of Training and Evaluation Setups for the Proposed Method

Experiments	Encoder Architecture	Unsupervised pre-training (D_{u1})	Pseudo-labelling (D_{u2}/L_1)	Supervised training ($D_L/E_{labeled}$),	Test set	Type of Evaluation
Core Architecture	Vanilla AE / CAE / DAE	Fundus images described in Table 1	-	D_L - Fundus images described in Table 4	Fundus images described in Table 5	Internal evaluation (Fundus images described in Table 5)
Proposed method	Vanilla AE / CAE / DAE	Fundus images described in Table 1	D_{u2} or S_1- Fundus images described in Table 2	$E_{labeled}$ – small, enhanced dataset	Fundus images described in Table 5	Internal evaluation (Fundus images described in Table 5)
	Vanilla AE / CAE / DAE	Fundus images described in Table 1	L_1- Complete fundus images in the following datasets: HRF, EyePACS – AIROGS, FIVES, MSHF, LES-AV, G1020, CRFO-V4 (Table 5 in the main manuscript)	$E_{labeled}$ – Large, enhanced dataset	Fundus images described in Table 5	Internal evaluation (Fundus images described in Table 5)
	CAE	Fundus images described in Table 1	L_1- Complete fundus images in the following datasets: HRF, EyePACS – AIROGS, FIVES, MSHF, LES-AV, G1020, CRFO-V4 (Table 5 in the main manuscript)	$E_{labeled}$ – Large, enhanced dataset	Fundus images described in Table 5	Internal evaluation (leave-one-out dataset) (ORIGA /Drishti-GS described in Table 5)
	CAE	Fundus images	L_1- Complete fundus images in	$E_{labeled}$ – Large	CRFO-v4 dataset	External (leave-one-out

	described in Table 1	the following datasets: HRF, EyePACS – AIROGS, FIVES, MSHF, LES-AV, G1020 (Table 5 in the main manuscript except CRFO-v4)	enhanced dataset (except CRFO-v4)		dataset) evaluation
CAE	Fundus images described in Table 1	L_1- Complete fundus images in the following datasets: HRF, EyePACS – AIROGS, FIVES, MSHF, LES-AV, G1020, CRFO-V4 (Table 5 in the main manuscript)	$E_{labeled}$ – Large enhanced dataset	Papilla dataset	External (fully unseen) evaluation

Note. All the supervised training and classification for glaucoma detection are conducted using SVM classifier.