Retinal Fundus Multi-Disease Image Classification Using Ensembled CNNs with Transformer Head

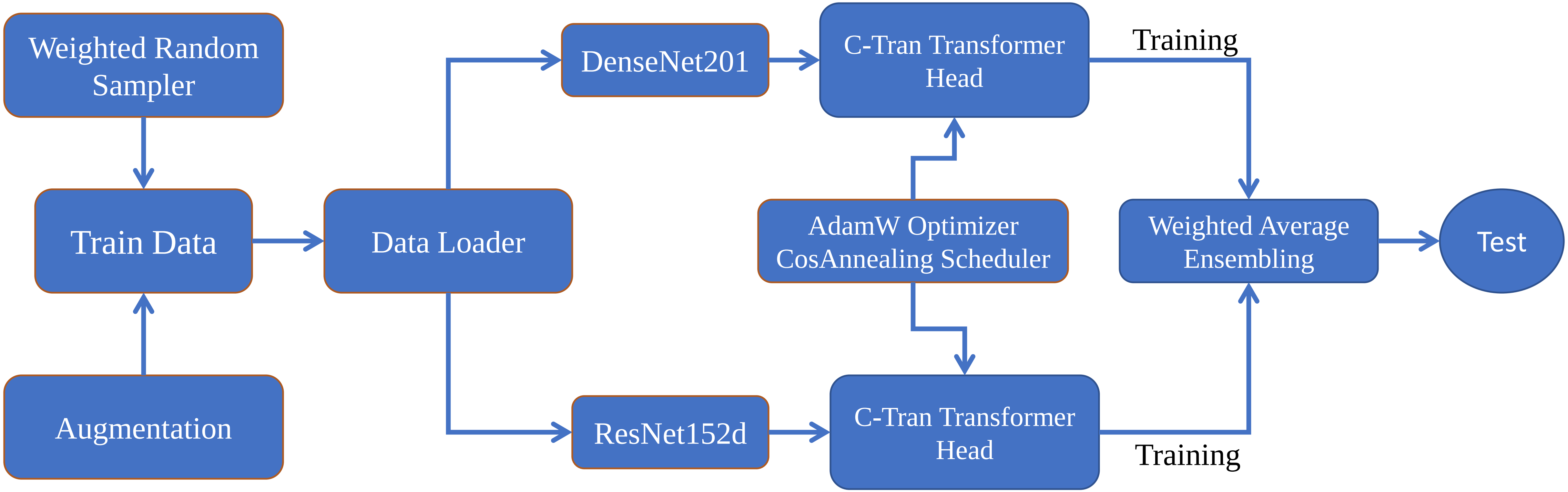
- ***End Term Presentation***

- Saksham Agarwal (2011144)
- Deependra Singh (2011058)

Comparison between our model and the baseline model

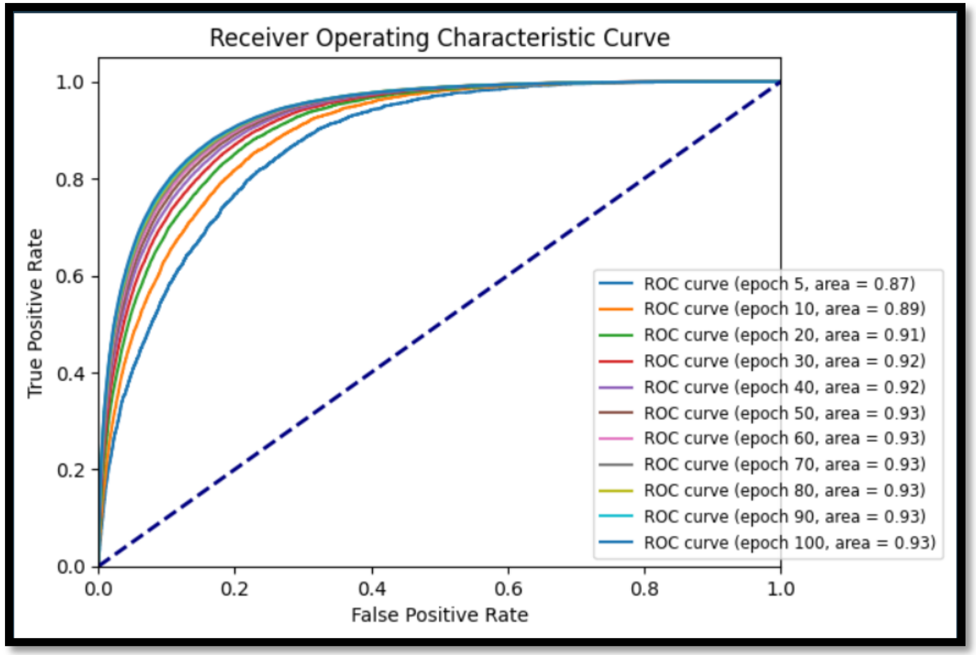
Baseline Model	Our Model
Sampler: LP ROS 10%	Sampler: Weighted Random Sampler
Backbone: DenseNet161	Backbone: DenseNet201 and ResNet152d
Architecture: CNN backbone with a C-Tran transformer Head	Architecture: Two C-Tran models with separate CNN backbones ensembled together
Ensembling: None	Ensembling: Weighted Averaging Ensemble
Best Result: Model Score = 0.900	Best Result: Model Score = 0.917

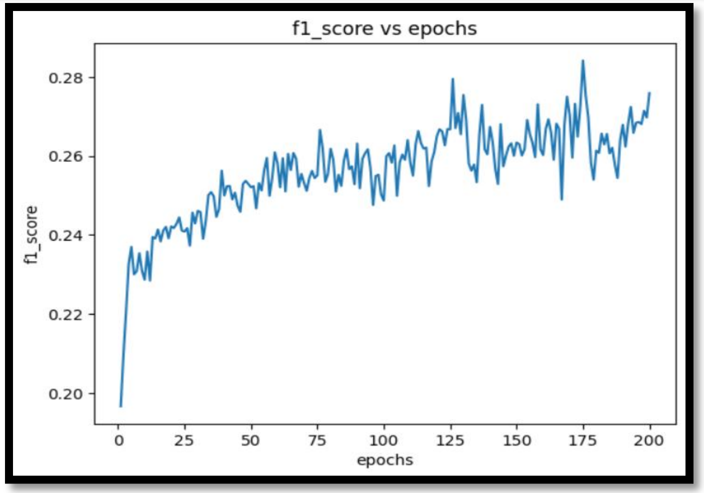
Model architecture



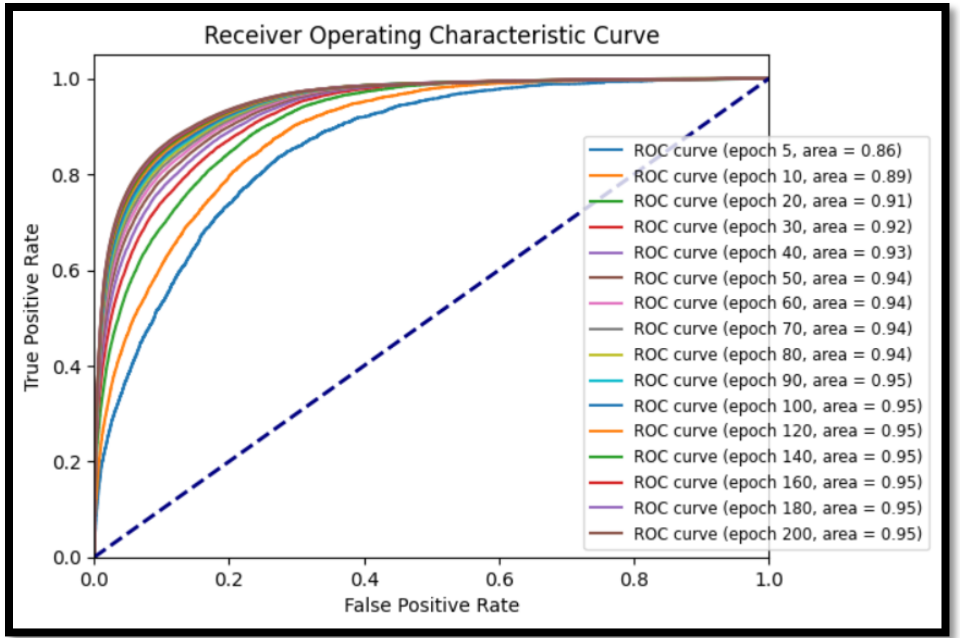
Review of Prior Methods and Their Efficacy

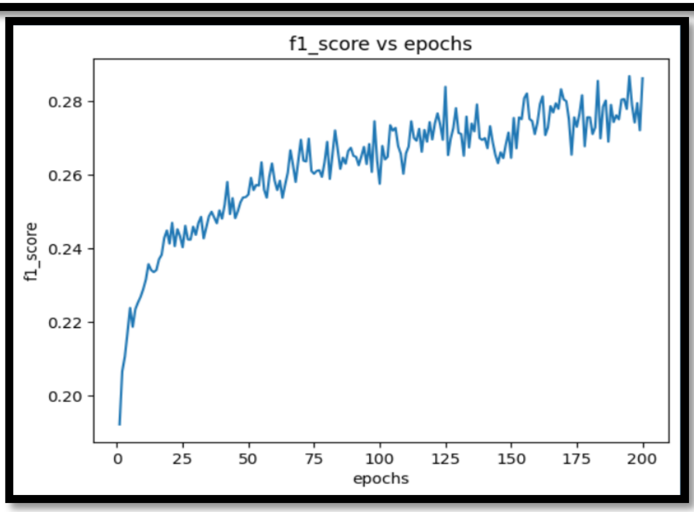
DenseNet121





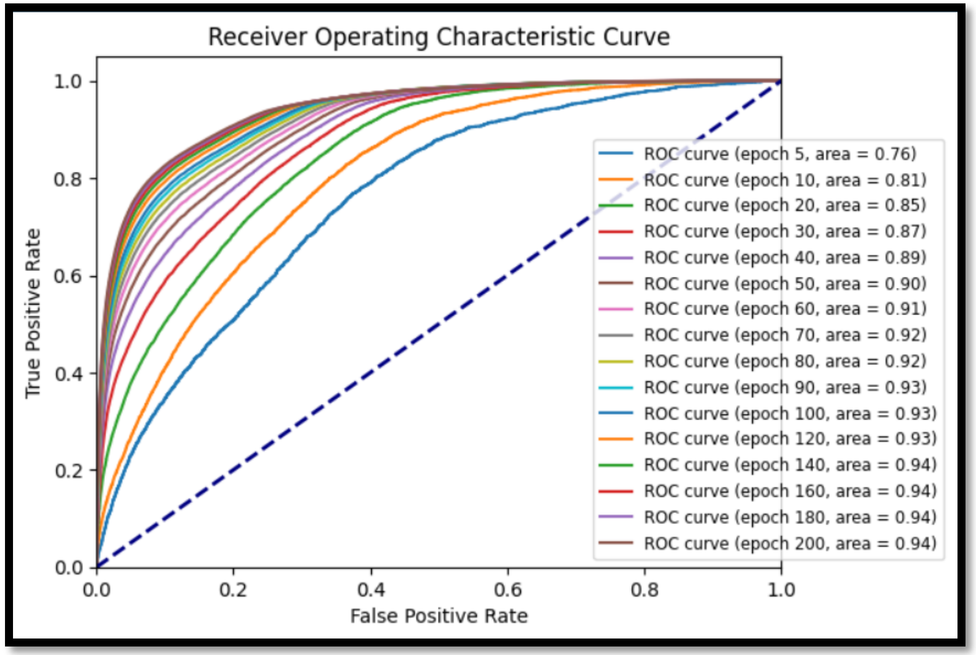
DenseNet201

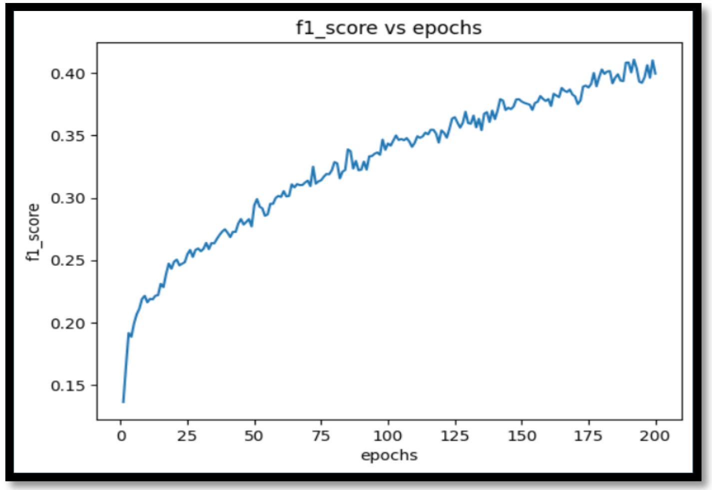




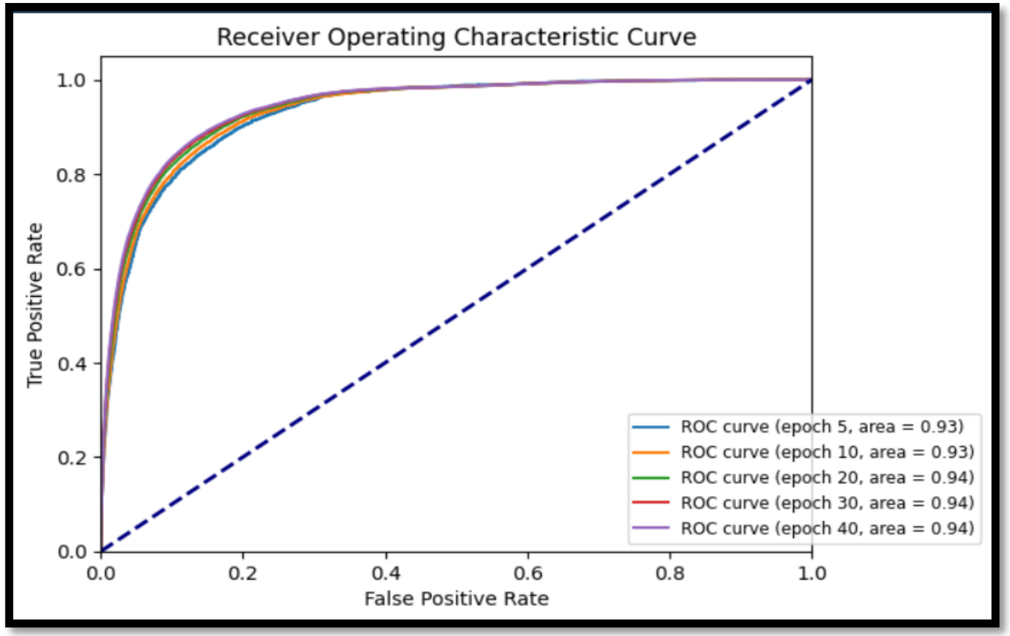
Review of Prior Methods and Their Efficacy

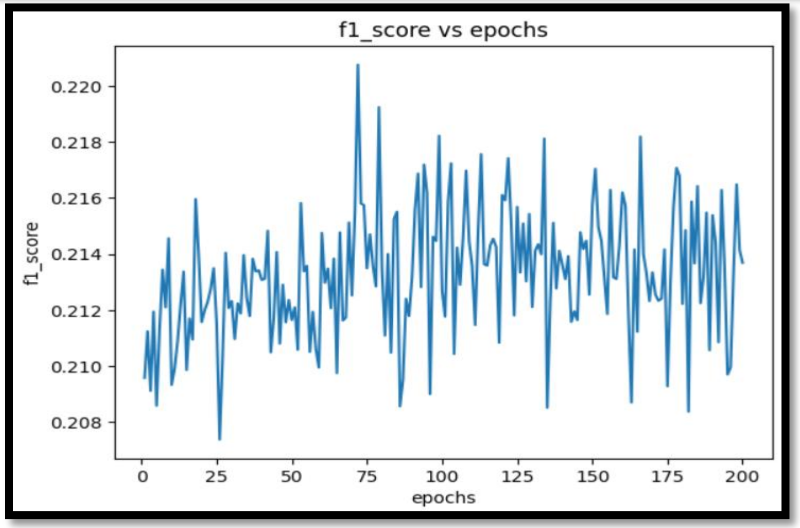
ResNet152d





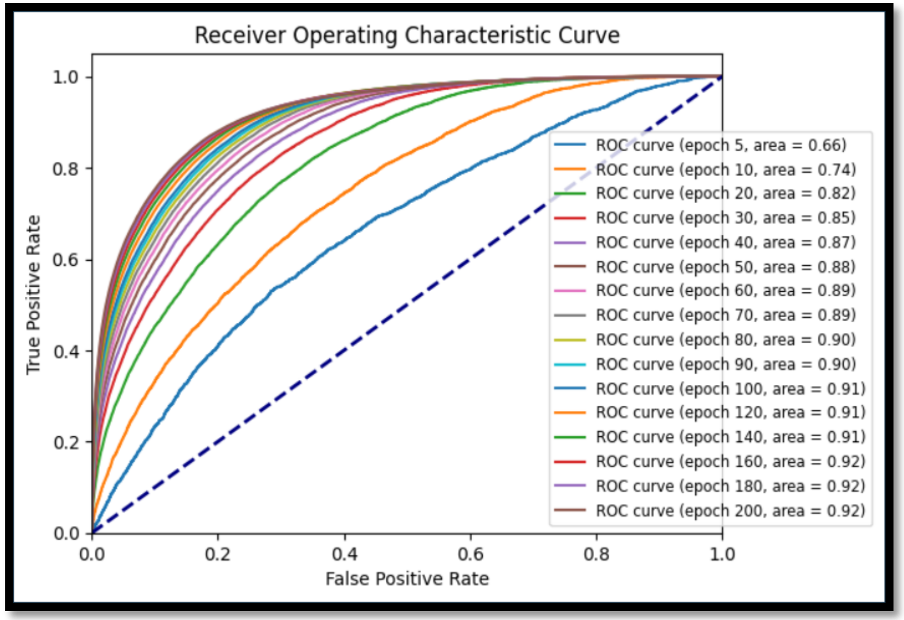
EfficientNetV2-S

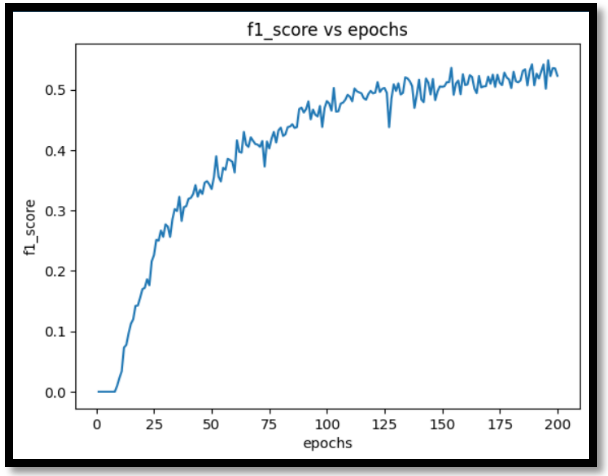


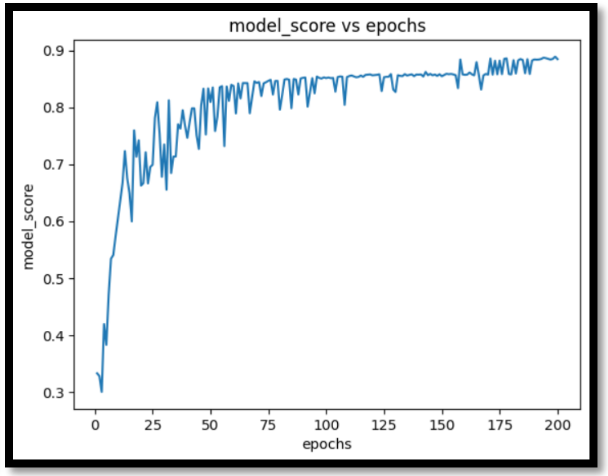


C-Tran Without Label Embeddings

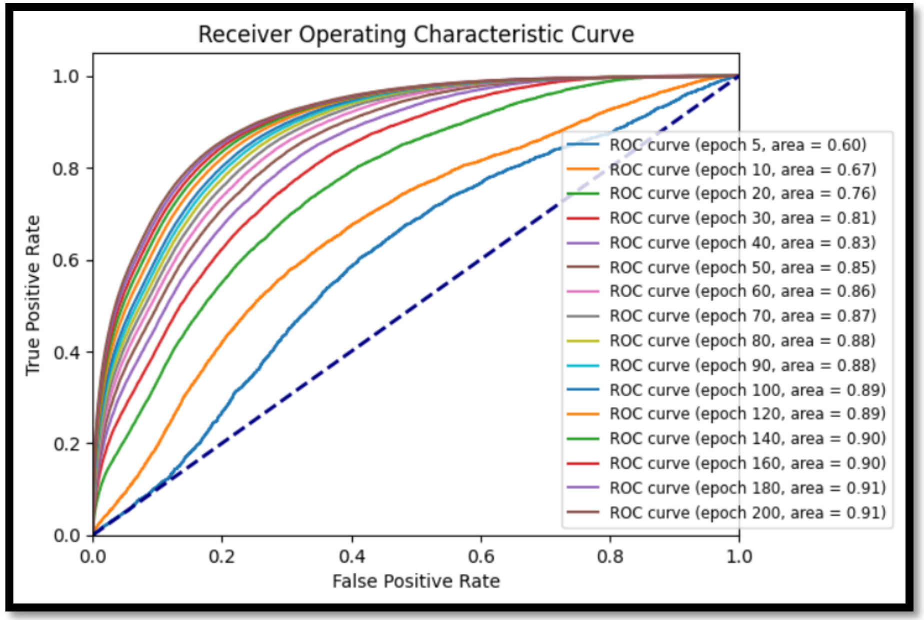
Backbone: DenseNet201

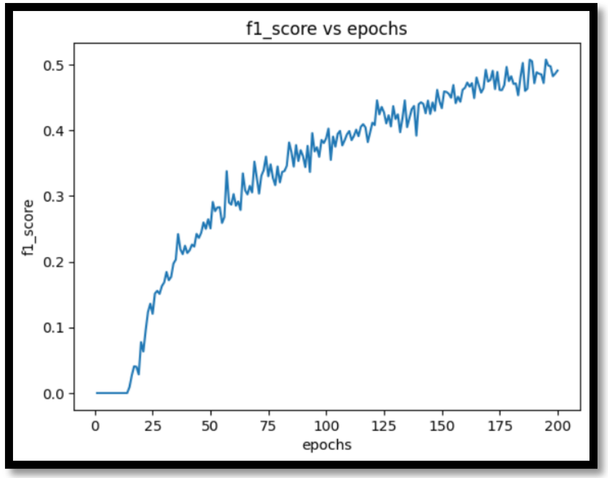


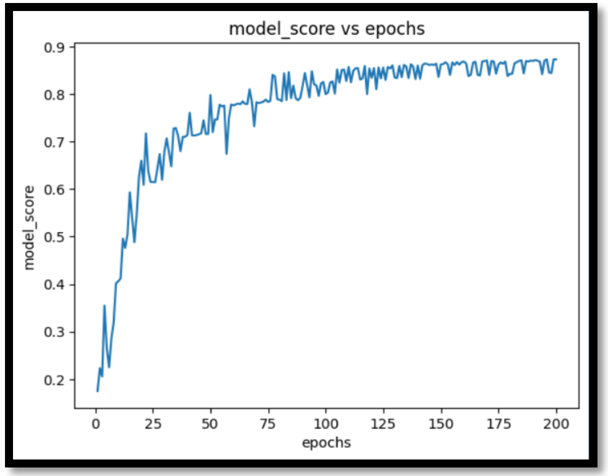




Backbone: ResNet152d

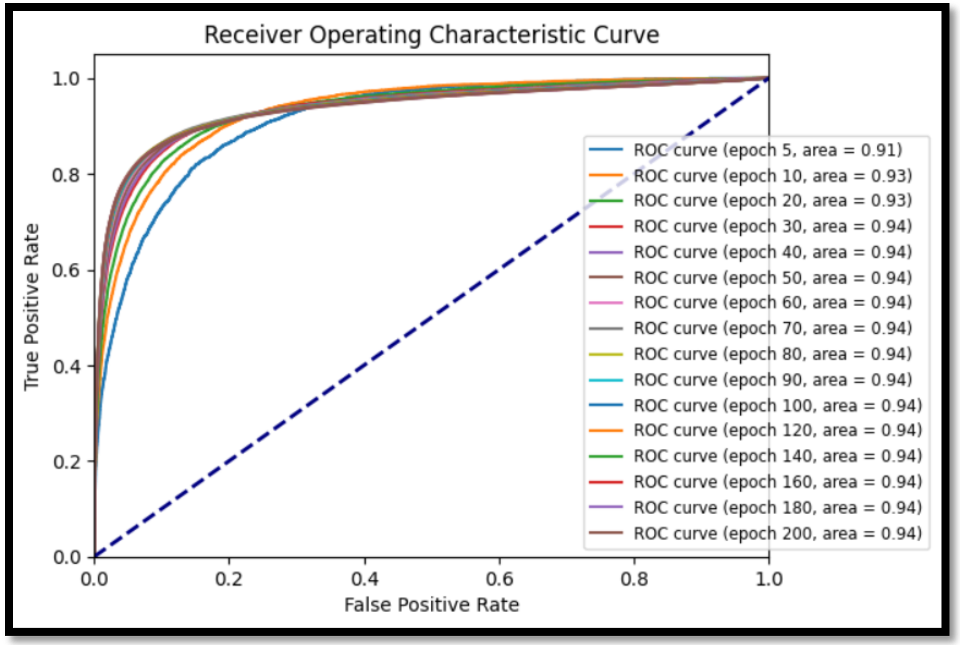


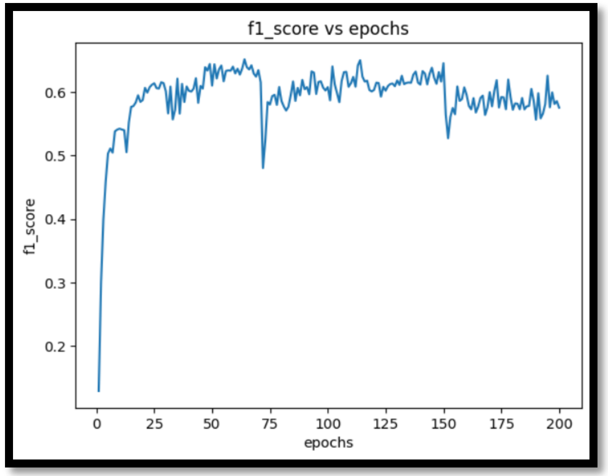


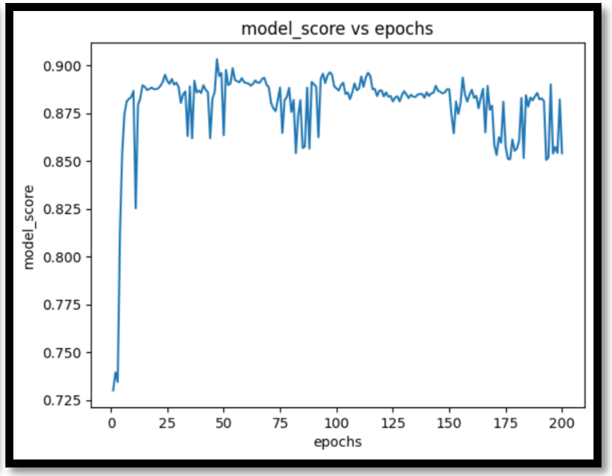


C-Tran With Label Embeddings

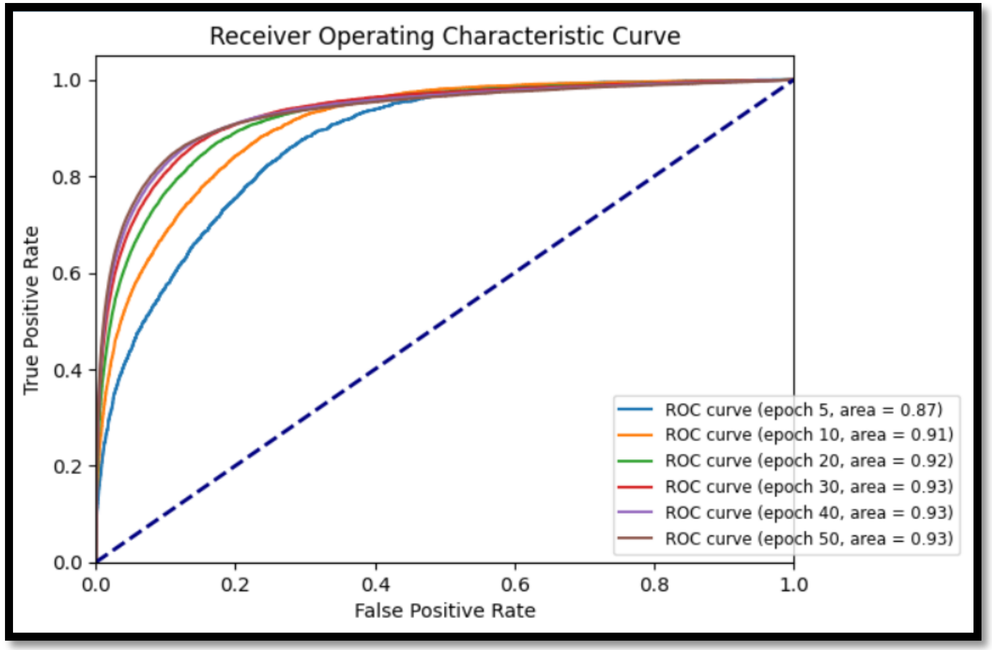
Backbone: DenseNet201

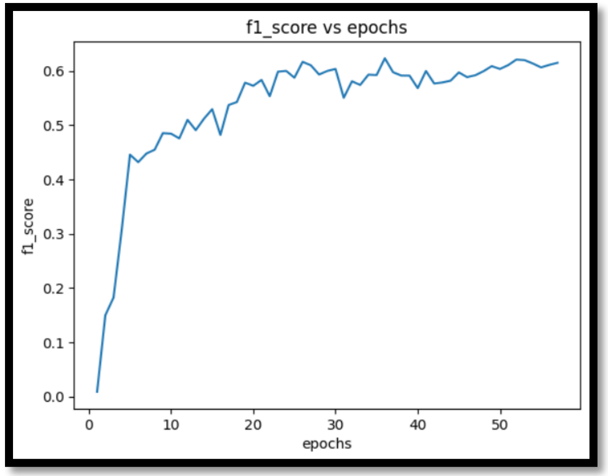


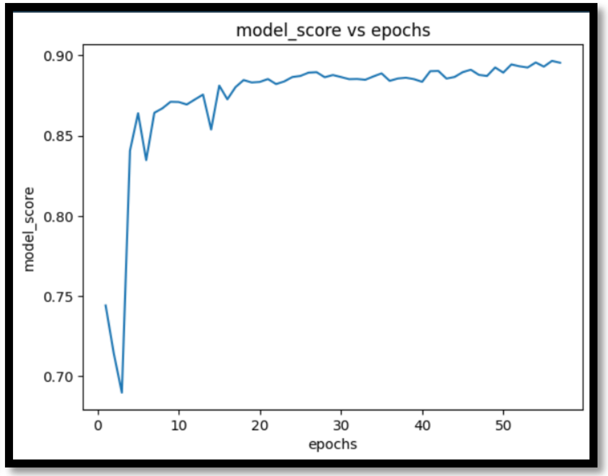




Backbone: ResNet152d

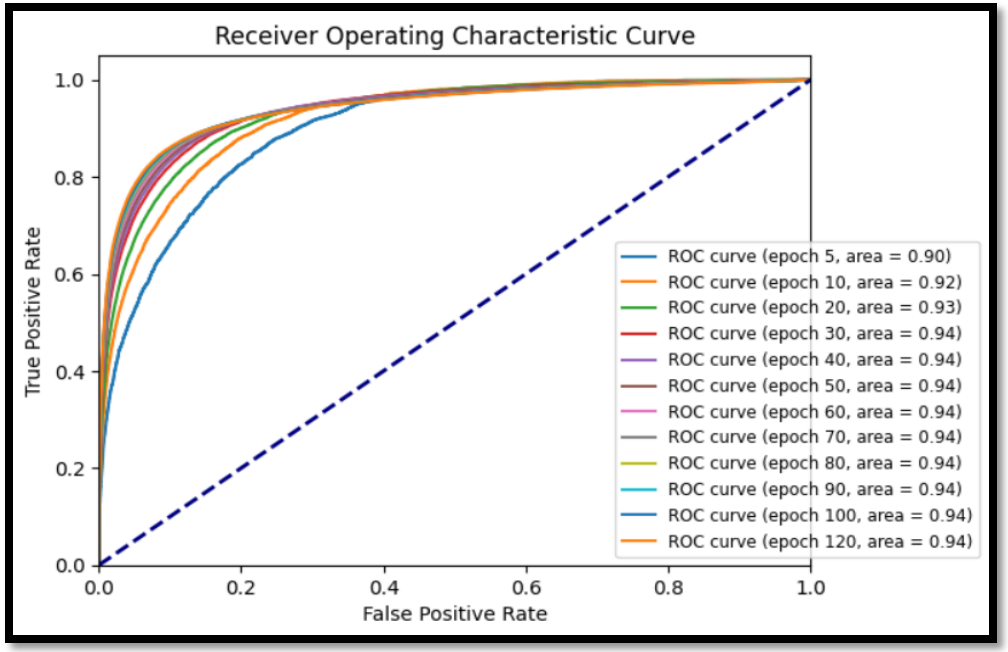


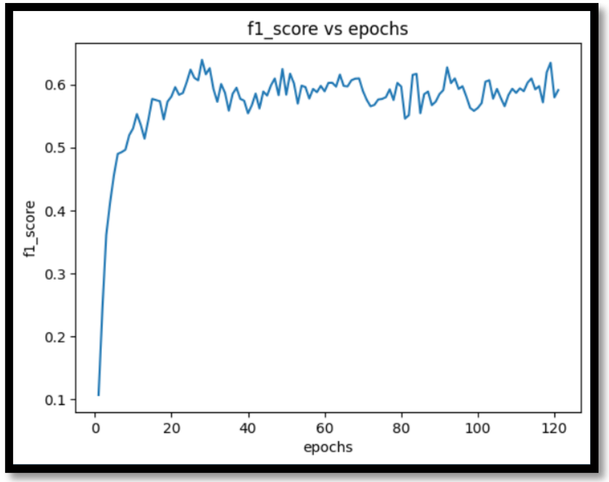


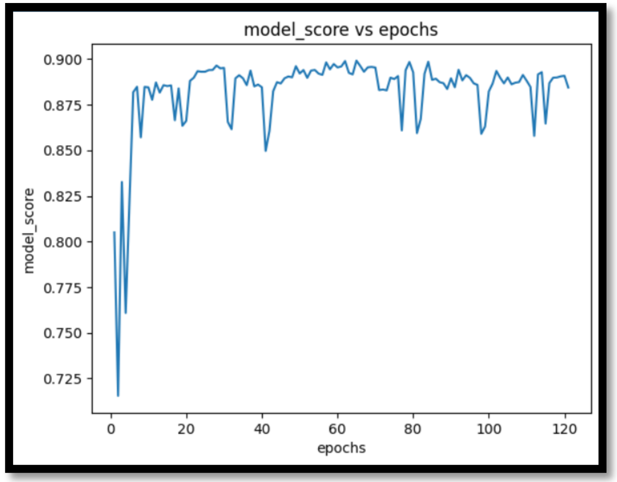


C-Tran Ensemble Models

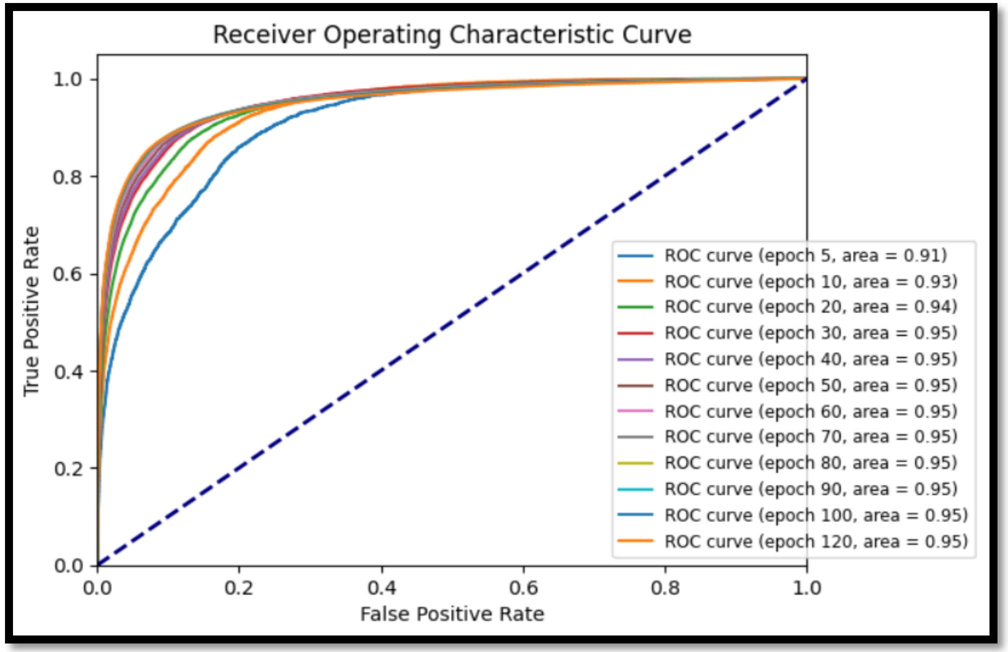
Strong: RN152d, Weak: DN201

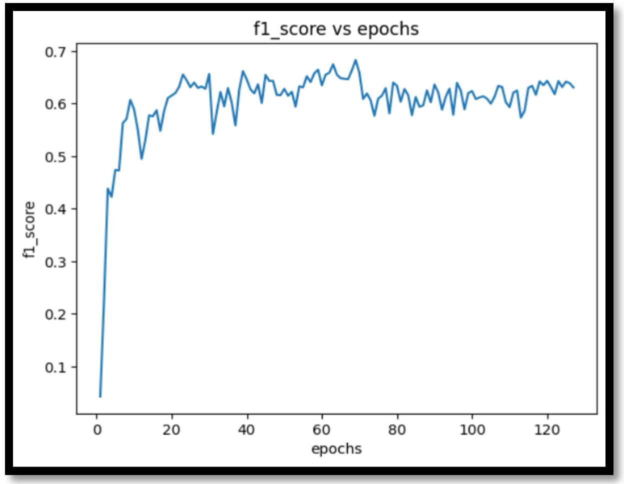


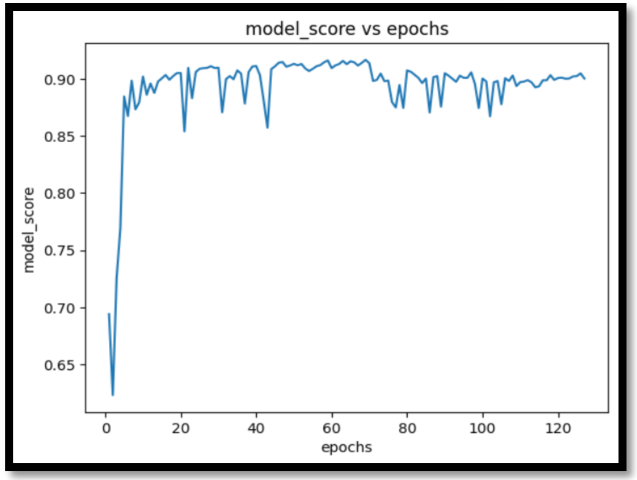




Strong: DN201, Weak: RN152d







Results

Model	Loss	ML mAP	ML F1	ML AUC	ML Score	Bin AUC	Bin F1	Model Score
DenseNet121	2.121	-	0.276	0.951	-	-	-	-
DenseNet201	1.692	-	0.272	0.959	-	-	-	-
ResNet152d	1.815	-	0.41	0.947	-	-	-	-
EfficientNetV2-S	1.881	-	0.214	0.958	-	-	-	-
C-Tran (without LE):								
DenseNet201	0.245	0.623	0.535	0.932	0.778	1	0	0.888
ResNet152d	0.177	0.572	0.486	0.921	0.7464	1	0	0.873
C-Tran (with LE)								
DenseNet201	0.344	0.691	0.639	0.922	0.807	1	0.667	**0.903**
ResNet152d	0.073	0.658	0.604	0.931	0.795	1	1	0.897
Mured	-	0.573	0.685	0.962	0.824	0.976	0.824	**0.9**
Ensemble Models:								
S: RN, W: DN	0.373	0.669	0.597	0.928	0.798	1	1	0.899
S: DN, W: RN	0.246	0.712	0.683	0.954	0.833	1	1	**0.917**

Label	Precision	Recall	f1	AUC
DR	0.35	0.3333	0	0.9474
NORMAL	0.6833	0.6667	1	1
MH	0.6833	0.6667	1	1
ODC	0.6833	0.6667	1	1
TSLN	0.6833	0.6667	1	1
ARMD	0.6833	0.6667	1	1
DN	0.6833	0.6667	1	1
MYA	0.6833	0.6667	1	1
BRVO	0.6833	0.6667	1	1
ODP	0.6833	0.6667	1	1
CRVO	0.6833	0.6667	1	1
CNV	0.6833	0.6667	1	1
RS	0.6833	0.6667	1	1
ODE	0.6833	0.6667	1	1
LS	0.6833	0.6667	1	1
CSR	0.6833	0.6667	1	1
HTR	0.35	0.3333	0	0.8947
ASR	0.525	0.5	0	1
CRS	0.6833	0.6667	1	1
OTHER	0.7	0.5	0.6667	1

Comparison with the Mured Model [1]

- The ensemble model surpasses the Mured model in 5 out of 7 metrics including the model score.
- When comparing the ResNet152d and DenseNet201 models, we find that the former converged faster while the later gave a better results.
- DenseNet converged to its maximum value at 76th , while ResNet at the 45th epoch.
- On combining the two, we see that for the weak DenseNet and strong ResNet model, with weights 0.52 and 0.48 respectively, the model converged at 25th epoch while the score got lowered.
- On the other hand, on using the strong DenseNet and weak ResNet model, we find that it converges quicker than DenseNet alone and gives a better score.
- Compared to the Mured Model with two labels with 0.9 AUC, we only have one label with AUC less than 0.9.

Future Plans

- Applying and comparing different sampling techniques like LP-ROS, ML-ROS, SMOTE-ENN, MLMO and Cascade Sampling.
- Optimizing the augmenters.
- Optimizing the weights of ensemble model.
- Increasing the dataset by combining the MURED dataset with the RFMID 2.0 dataset.
- Using Label Mask Training if necessary.

References:

[1] Rodriguez, M. A., AlMarzouqi, H., Liatsis, P. (2022). Multi-label Retinal Disease Classification Using Transformers. IEEE Journal of Biomedical and Health Informatics.

[2] Lanchantin, J., Wang, T., Ordonez, V., & Qi, Y. (2021). General multi-label image classification with transformers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 16478-16488).