**NIPGBOARD KIRA EXPERIMENT REPORT**

**ANTIKA DAS- 22/8/2020**

**DATA PREPARATION**: For this NIPGBoard Kira experiment , I have taken two categories of data -“balanced” and “unbalanced”.

**KIRA PARAMETERS:**

* This experiment has been done over three different **epochs (10,35,50)** for both balanced and unbalanced data to check the performance of kira with PCA,UMAP,T-SNE etc…
* 15% of all TN and 15% of all FN dataset will randomly be connected as a graph for positive pairwise constraints and rest of the data will be connected by transitivity.

TN-TN🡪positive constraints

FN-FN🡪 positive Constraints

TN-FN🡪 Negative Constraints

**Observations**:

* The positive pairing seems fine while number of negative pairs are way less than the positive pairs, which does not match the expected number of negative pairs. As an example, suppose we have 1000 FN and 600 TN, SO 15% of 1000 would be 150 and 15% of 600 is 90, So minimum number of negative pairs should be 90 at least(TN-FN) . But according to the experiment it is very less.
* T-SNE seems to perform better than Umap in making clusters, however still T-SNE fails to make well separated clusters for both unbalanced and balanced dataset (In all the clusters a good amount of TN and FN were mixed) .
* For unbalanced dataset T-SNE and Umap are working better with higher number of epoch( 50).
* For balanced dataset even though T-SNE shows some promising results but Umap (with different number of neighbors) doesn’t seem to help much even with higher number of epoch.
* There is a good room for improvement if we further verify the performance with different perplexity and learning rate and iterations in case of TSNE and different neighbors for U-map.
* I have also run the kira algorithm with 10% of whole TN dataset and 10% of whole FN dataset, but also there is no significant improvement.

\*\*\*\*After a good amount of experiments what I observed is, higher epochs are performing better clustering in case of T-SNE(of course other parameters eg. learning rate, iteration, perplexity of T-SNE matters too) for both balanced and unbalanced data. Standard perplexity and iterations being used in all the experiments are 30, and >2000 accordingly which showed promising result. While for U-map not much considerable improvement. Overall, the algorithm is almost working same with both unbalanced and balanced data.

**Experimental results with visualization:**

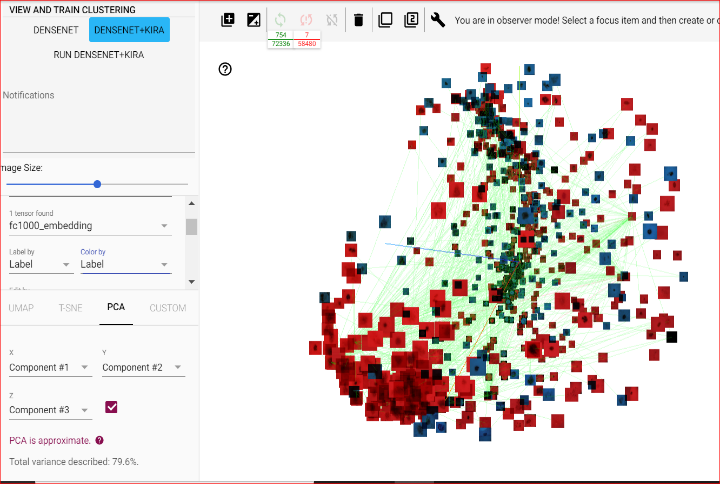
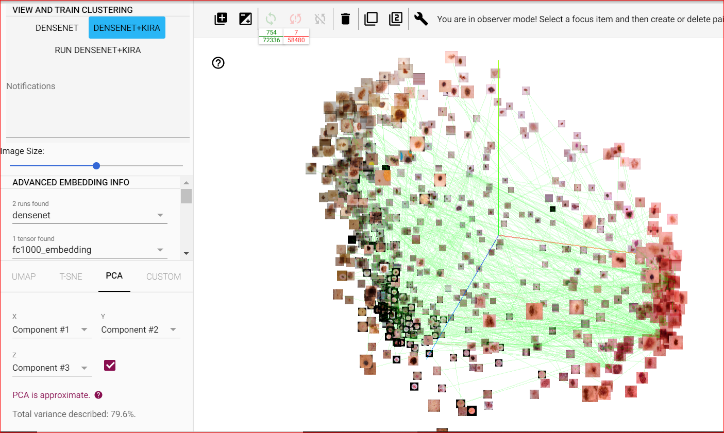
**Experiment 1:**

*Category* : Unbalanced [250 FN + 500 TN🡪 training ( 1:2)**]**

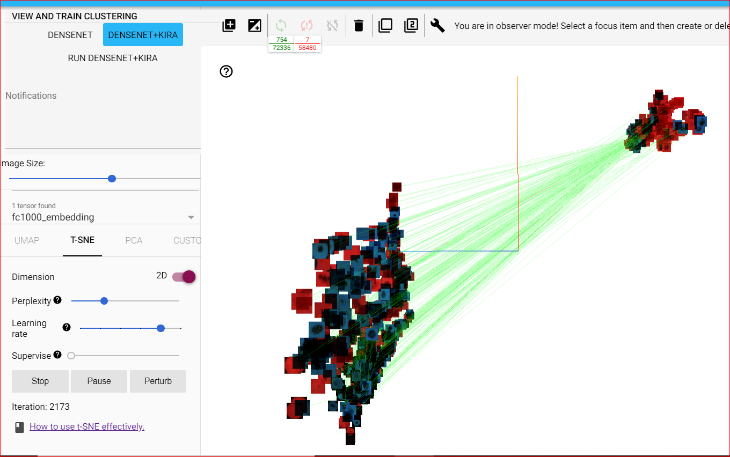
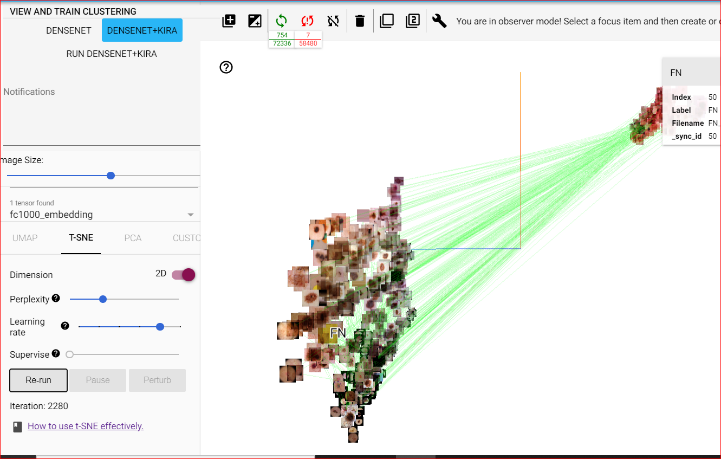
*Kira Parameters* : ***epoch*** 10,

***Pairwise constraints***: 15 % of TN and FN (connected graph)

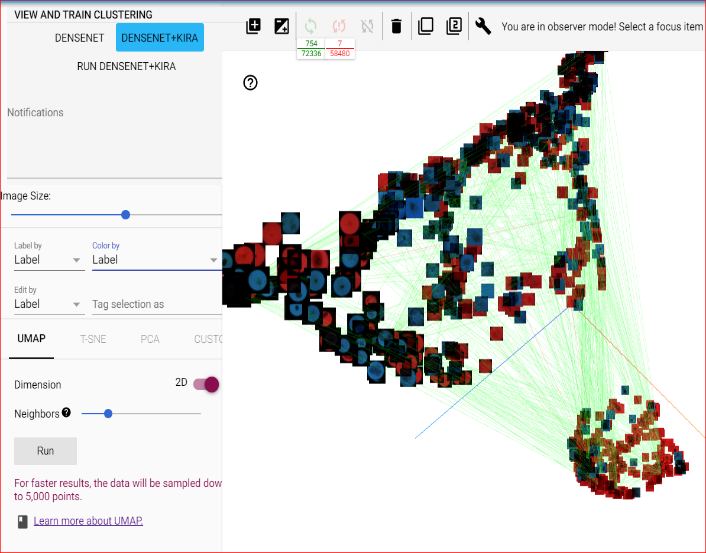
**PCA**



**T-SNE: perplexity**



**Umap: Neighbors- default (15)**



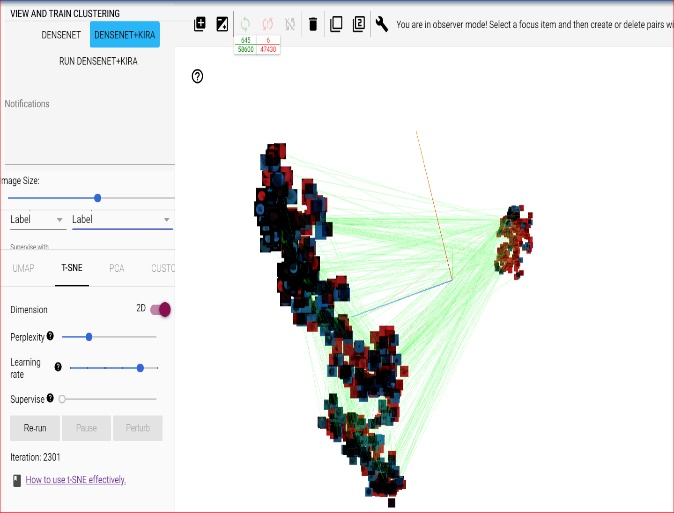
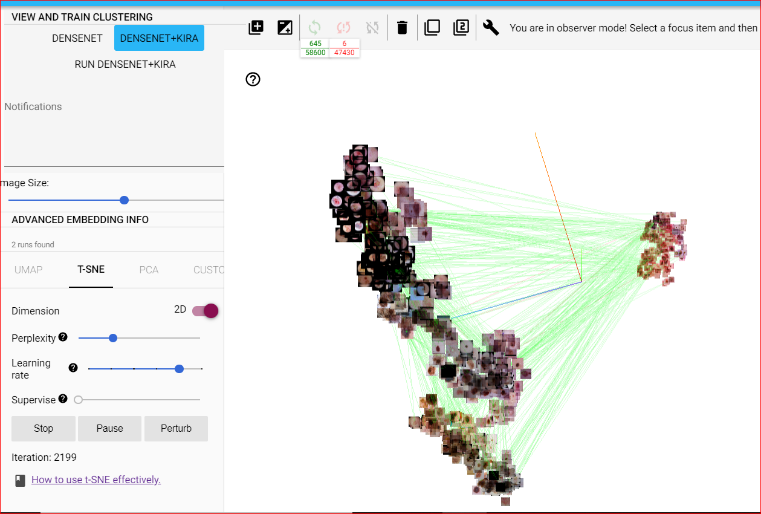
**Experiment 2:**

*Category* : Unbalanced [250 FN + 500 TN🡪 training ( 1:2)**]**

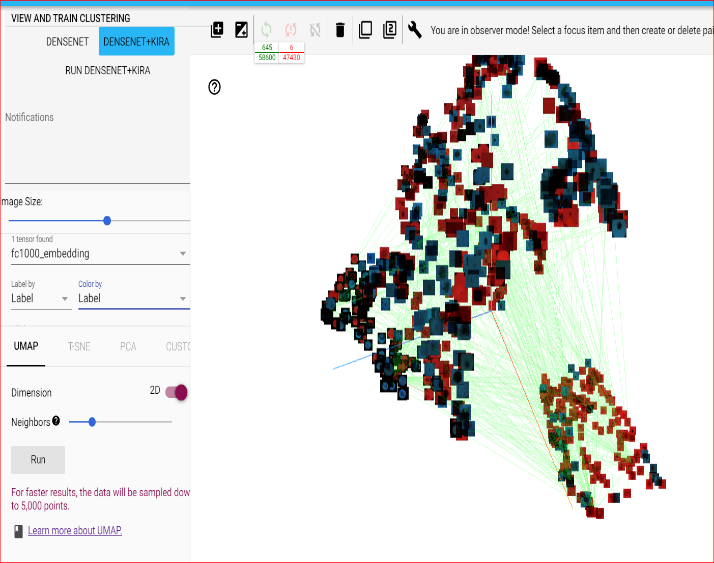
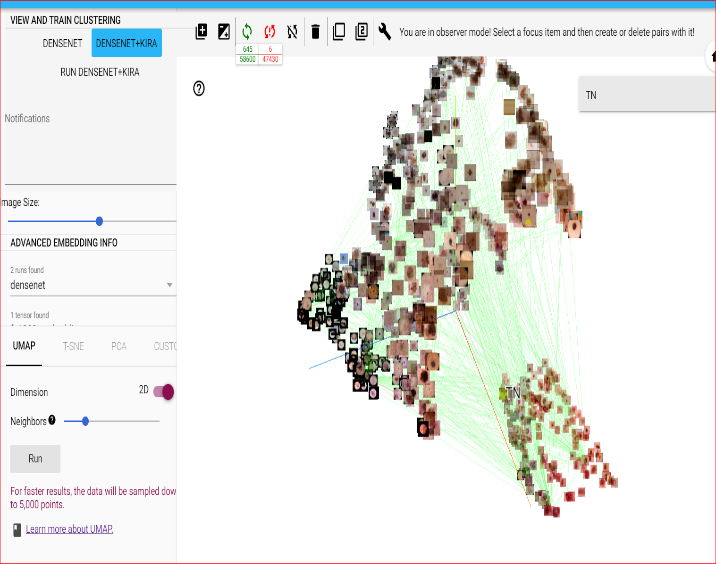
*Kira Parameters* : epoch 35,

Pairwise constraints: 15 % of TN and FN (connected graph)

**T-SNE**



**Umap:Neighbors default (15)**



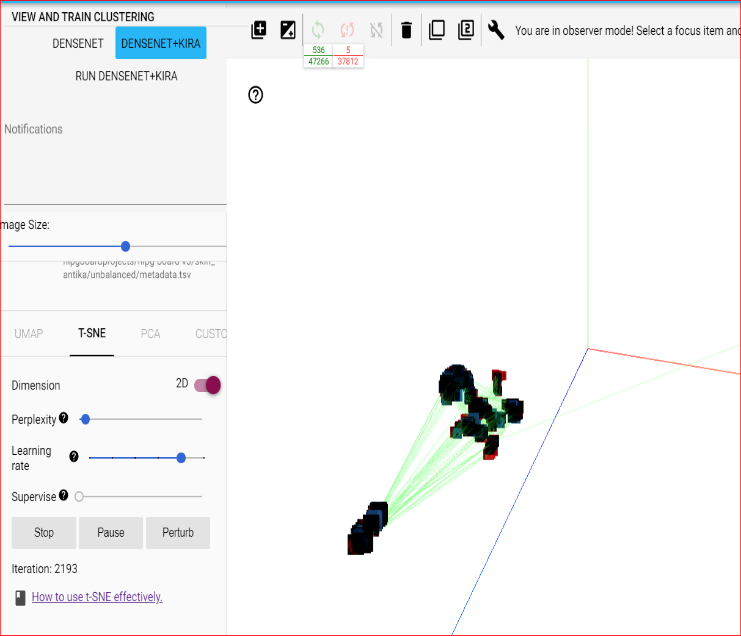
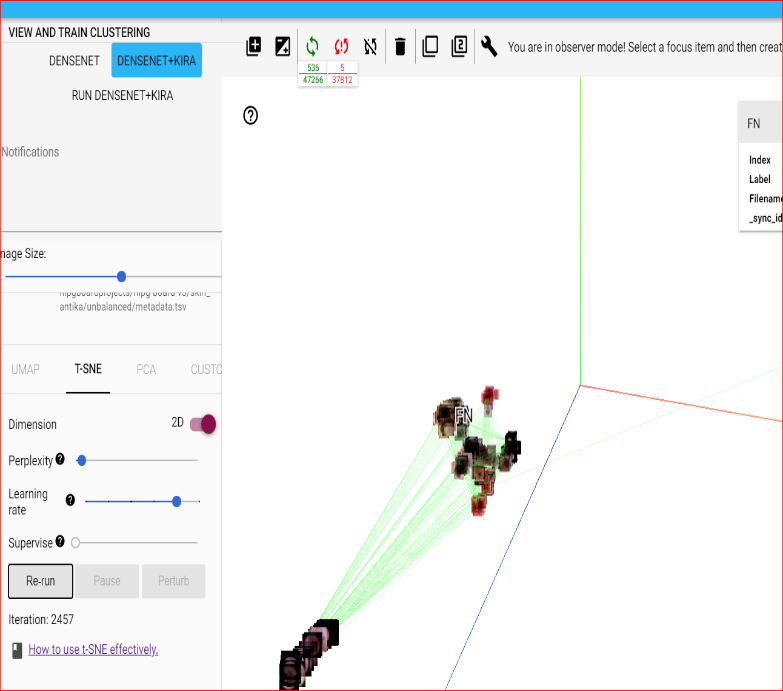
**Experiment 3.**

*Category* : Unbalanced [250 FN + 500 TN🡪 training ( 1:2)**]**

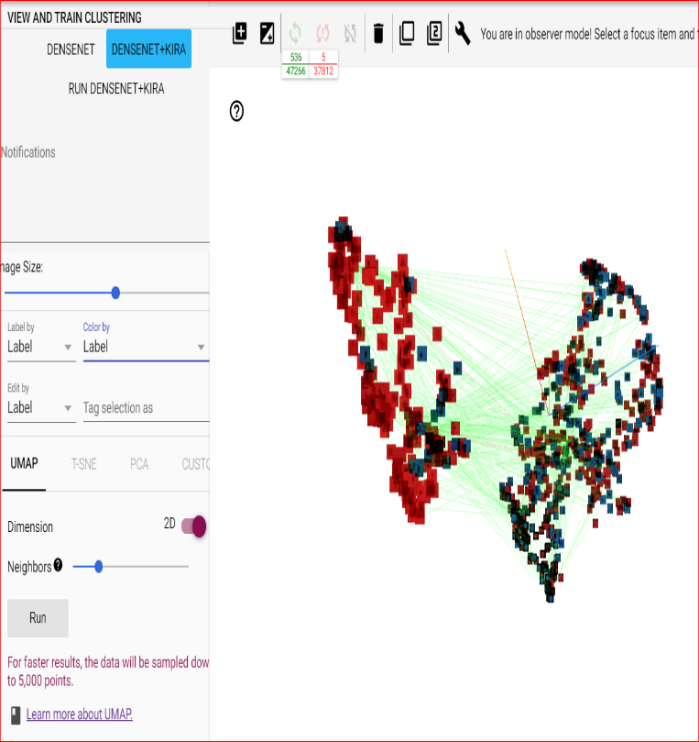
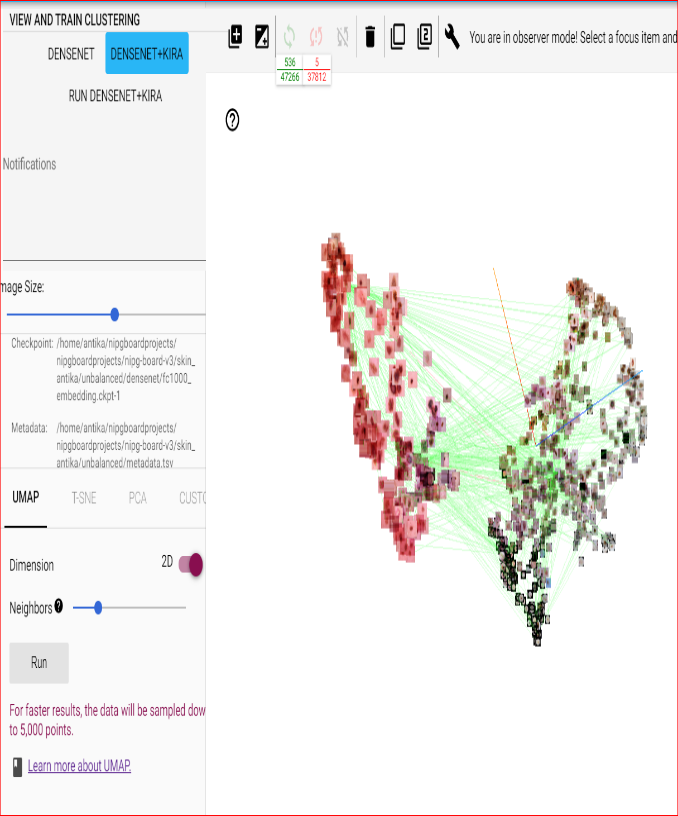
*Kira Parameters* : epoch 50,

Pairwise constraints: 15 % of TN and FN (connected graph)

**T-SNE**

****

**Umap : default :neighbor (15)**

****

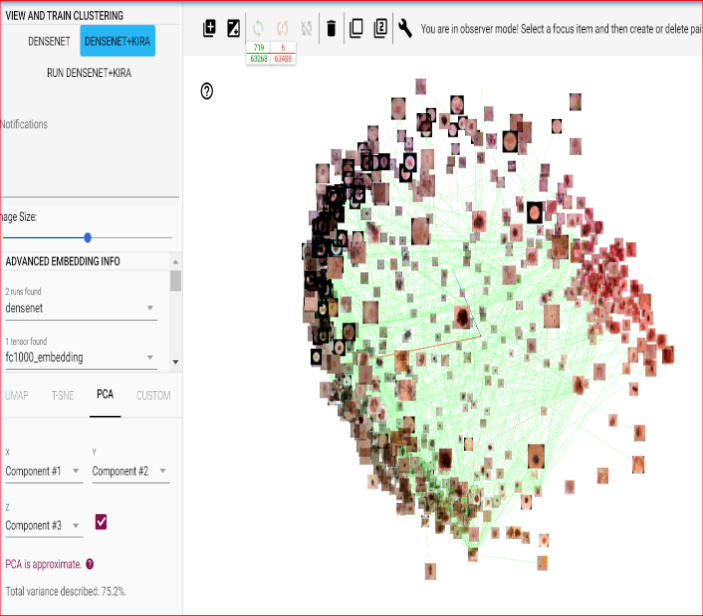
**Experiment 4:**

*Category* : balanced [412FN + 412 TN🡪 training ( 1:2)**]**

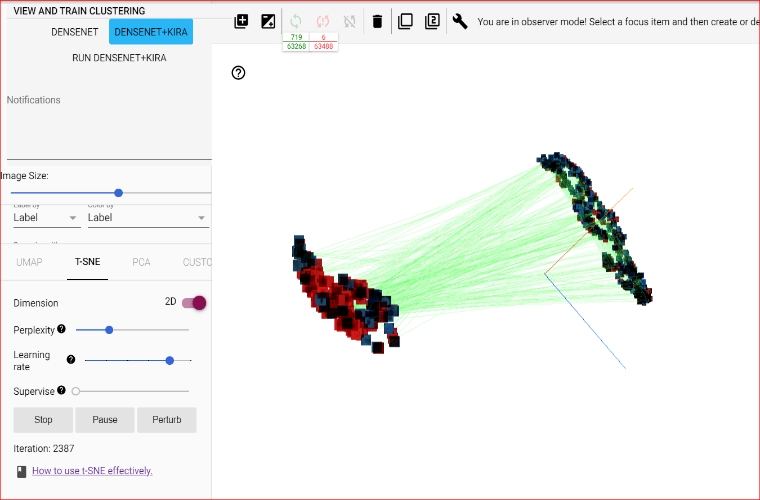
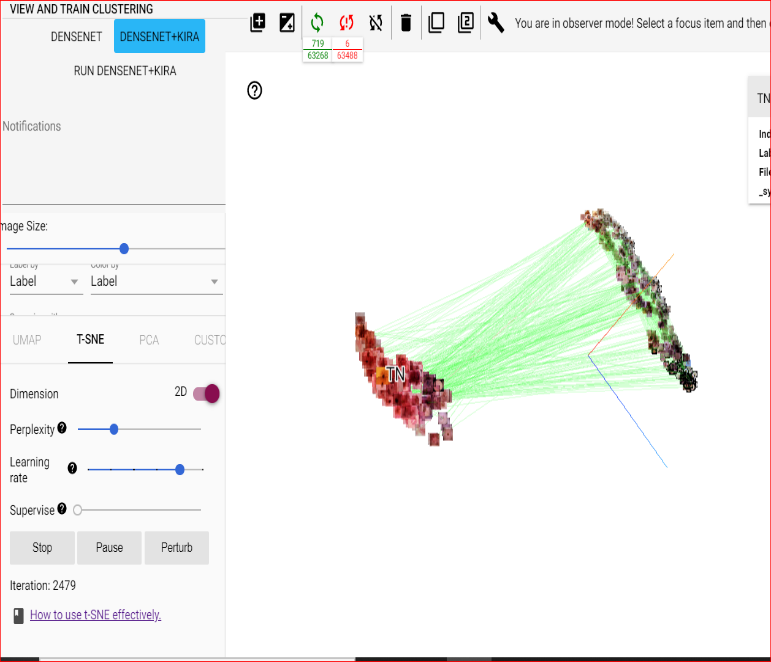
*Kira Parameters* : epoch 10,

Pairwise constraints: 15 % of TN and FN (connected graph)

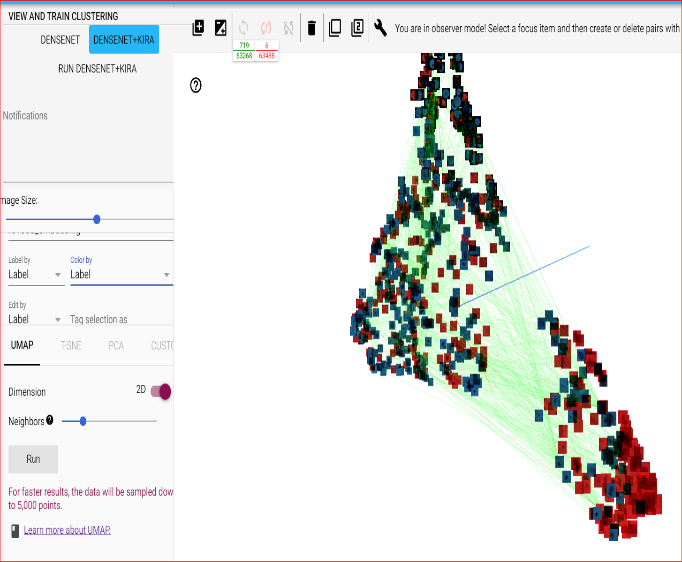
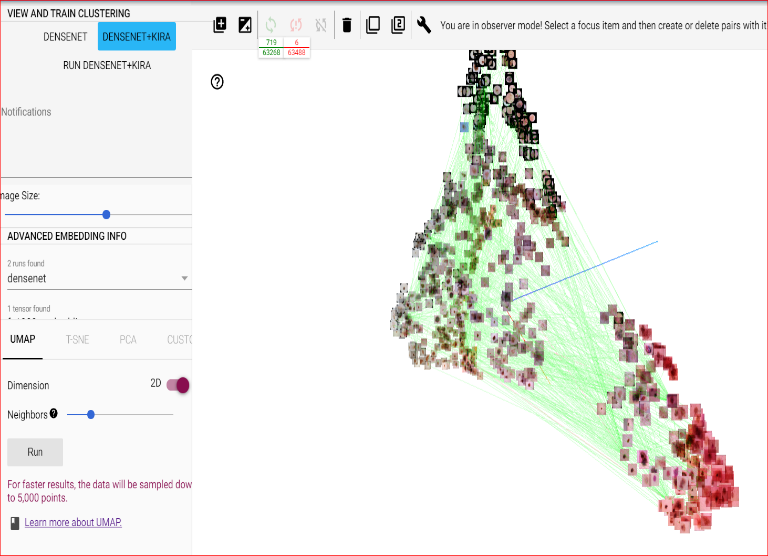
**PCA**:



**T-SNE**



**Umap:Neighbors default (15)**



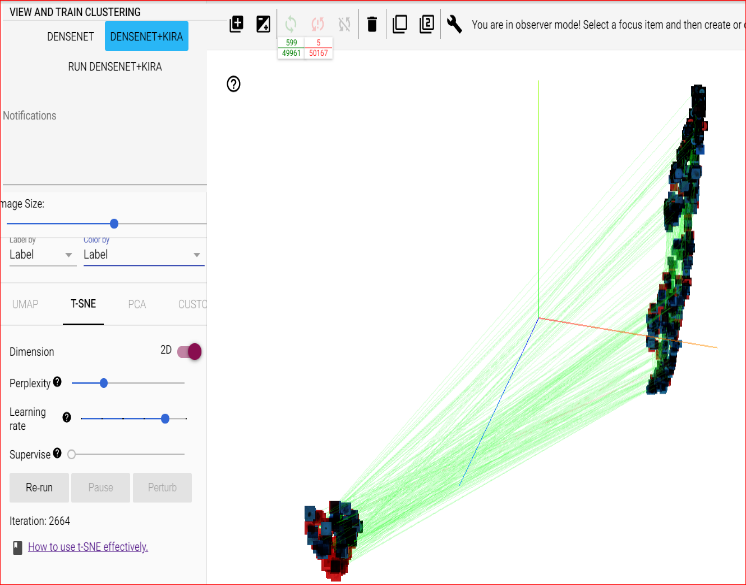
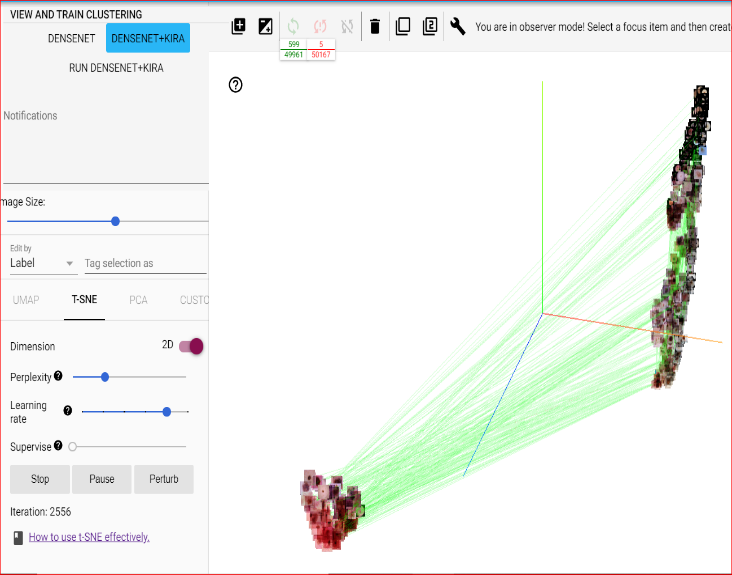
**Experiment 5:**

*Category* : balanced [412 FN + 412TN🡪 training ( 1:2)**]**

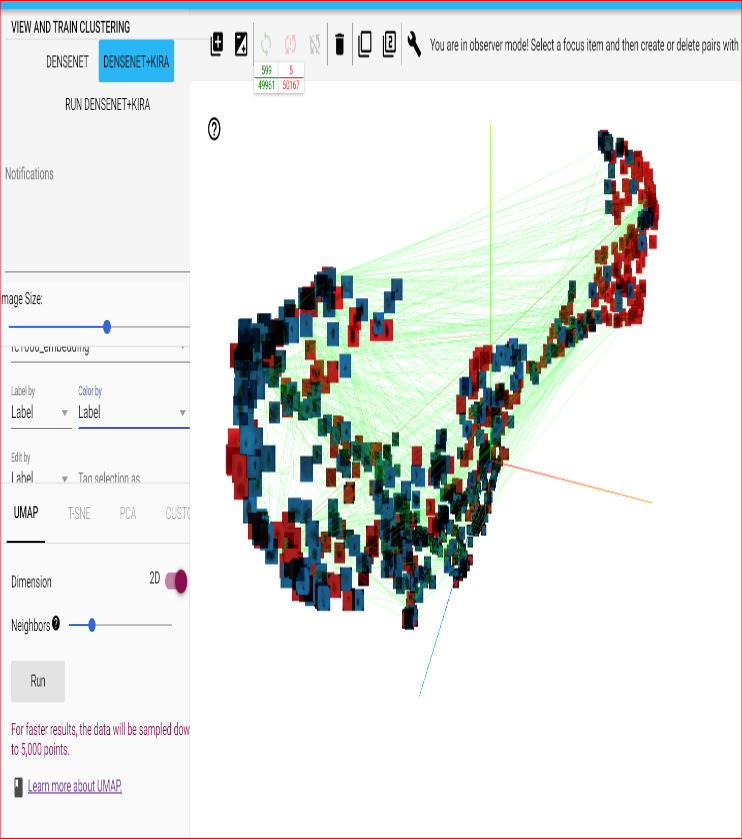
*Kira Parameters* : epoch 35,

Pairwise constraints: 15 % of TN and FN (connected graph)

**T-SNE**

****

**Umap:Neighbors default (15)**



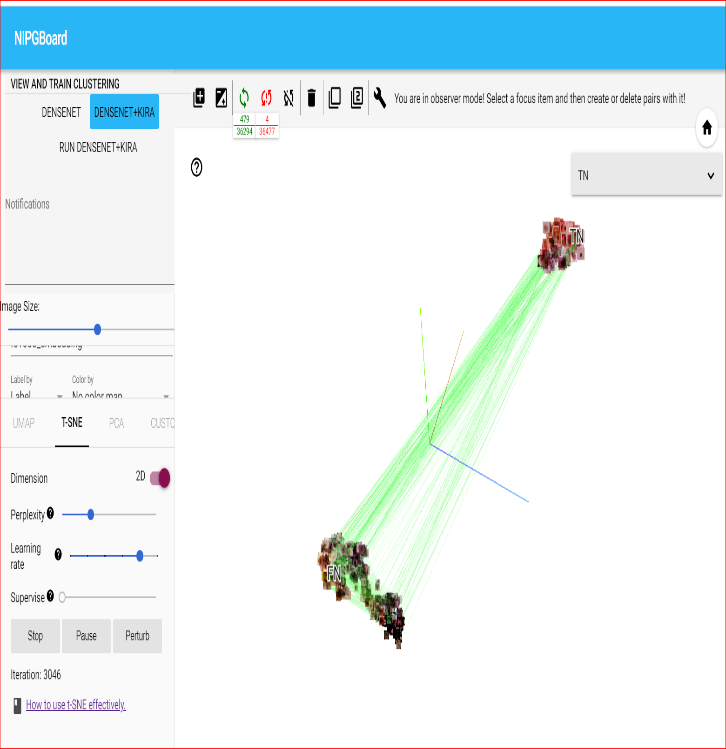
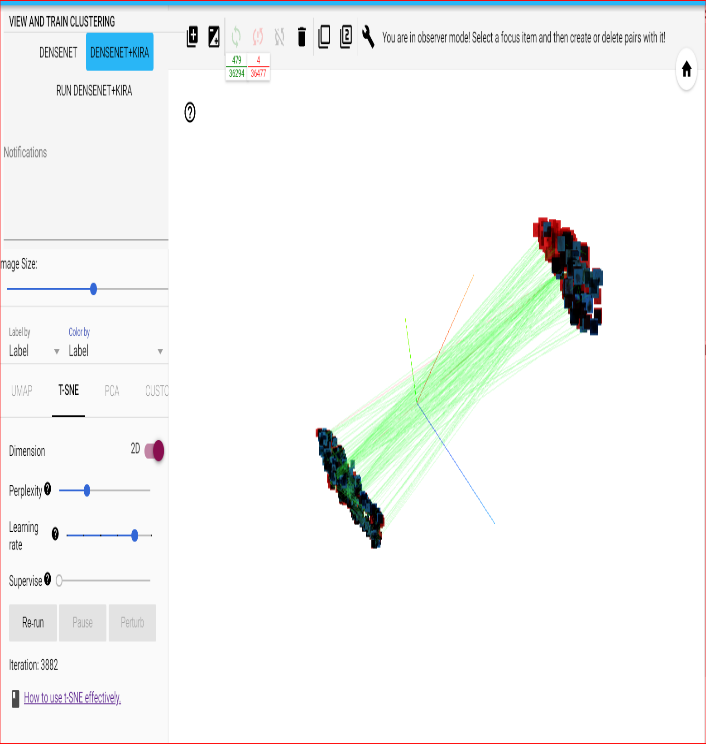
**Experiment 6.**

*Category* : balanced [412 FN + 412TN🡪 training ( 1:2)**]**

*Kira Parameters* : epoch 50,

Pairwise constraints: 15 % of TN and FN (connected graph)

**T-SNE:**



**Umap**

