# The procedure for selecting the training data is explained in Step.1:

**Team AZ_KA: Azin Al Kajbaf, Kaveh Faraji Najarkolaie**

## Step.1. Selecting Dataset for training

### 1.1. Read data from csv files

```
clc;
clear;
train =readtable("train_metadata.csv"); % Provide the location of the train_metadata.csv
trainlabels = readtable("train_labels.csv"); % Provide the location of the train_labels.csv
```

### 1.2. Finding "Non-Confident" videos

In this part, we found the "Non-Confident" videos. Non-Confident videos are the ones that have inconsistency between their crowd score and stalled labels. We checked the crowd scores and their associated train labels. We noticed that some of them might be inconsistent. We decided to remove the IDs with crowd score lower than 0.6 and train label of 1 and also the ones with crowd score over 0.3 and train label of 0.

```
jj=1;
for ii=1:size(trainlabels,1)
    if (trainlabels.stalled(ii)==1 && train.crowd_score(ii)<0.6 ) ||...
      (trainlabels.stalled(ii)==0 && train.crowd_score(ii)>0.3)
        AllNonConfidentIDs{jj,1}=ii;
        jj=jj+1;
    end
end
```

### 1.3. Deleting "Non-Confident" videos

In this section, the "Non-Confident" videos are deleted from trainlabels and train files.

```
confident_train=train;
confident_trainlabels=trainlabels;
ANC_ID=cell2mat(AllNonConfidentIDs(:,1));
confident_train(ANC_ID,:)=[];
confident_trainlabels(ANC_ID,:)=[];
```

### 1.4. Finding IDs of stalled videos

There are very few videos in the dataset that are categorized as stalled. So, we decided to include all confident stalled videos in our training set.

```
% The IDs of stalled videos are found
Count=0;
jj=1;
for ii=1:size(confident_trainlabels,1)
    if confident_trainlabels.stalled(ii)==1
        Allstalledvideosid(jj,1)=ii;
```

```
            Count=Count+1;
            jj=jj+1;
        end
    end

    % In here the stalled videos are found
    stalled_train=confident_train(Allstalledvideosid,:);
    Stalled_labels=confident_trainlabels(Allstalledvideosid,:);
    % In here the flowing videos are found
    flowing_train=confident_train;
    flowing_train(Allstalledvideosid,:)=[];
```

## 1.5. Selecting videos for training set comprised of 10% stalled and 90% flowing

```
    % In here flowing videos are randomly selected.
    stalled_fraction=0.1;
    flowing_fraction=0.9;
    rng(1)
    trainind_flowing=randperm(size(flowing_train,1),floor(size(Allstalledvideosid,1)*...
      (flowing_fraction/stalled_fraction)))';
    flowing_Rand=flowing_train(trainind_flowing,:);
```

## 1.6. Adding new dataset (milli) column to original train table

```
    % The IDs of stalled and flowing videos are found and then assigned in
    % the main table as 'True' and 'False'
    idx=ismember(train.filename(:), flowing_Rand.filename(:))+...
      ismember(train.filename(:), Stalled_labels.filename(:));
    % idx_mili=table(cell(size(idx,1),1));
    % idx_mili{idx==0,1}={'False'};
    % idx_mili{idx==1,1}={'True'};
    % train.milli=idx_mili.Var1;
    idx_mili=cell(size(idx,1),1);
    idx_mili(idx==0,1)={"False"};
    idx_mili(idx==1,1)={"True"};
    train.milli=string(idx_mili(:));
    % The new main training table is written to a csv file
    writetable(train,'train_metadata_V1.csv');
```

## 1.7. Making datastore of all video files

```
    if ~exist('Datastore',"dir")
     mkdir Datastore
    end
    Dir_save='Datastore\';
    Dir_read='E:\DeepAlzheimer\'; % Provide the location of the main dataset
    mp4location = fullfile(Dir_save,"mp4.mat");
    if exist(mp4location,'file')
    load(mp4location,'mp4datastore')
    else
    mp4datastore = fileDatastore(Dir_read,"ReadFcn",@VideoReader);
```

```
    save(mp4location,"mp4datastore");
end
```

## 1.8. Making a folder and transferring selected training files to it

```
if ~exist('Input_train',"dir")
 mkdir Input_train
end
filedest='Input_train\';
millitrain = train(train.milli(:) == 'True',:);
filenamesmili=millitrain.filename(:);
%namev=zeros(size(mp4datastore.Files,1),1);
filesmp4=mp4datastore.Files;
sz=size(filesmp4,1);
parfor ii=1:sz
    namev{ii,1}=convertCharsToStrings(erase(filesmp4{ii},Dir_read));
     if sum(strcmp(namev{ii,1}, filenamesmili(:)))
         filepath{ii}=filesmp4{ii};
         copyfile(filepath{ii},filedest)
     end
end
```