## **Text S1**

## **Supplementary text: K-fold cross-validation**

Cross-validation is a statistical method for estimating how accurate a machine learning algorithm is expected to perform on new unseen data. It is especially useful when the available data sample is small as is generally the case in medical applications. The method is popular because it is simple to implement and gives an estimate that is less biased than other methods of how accurate the machine learning algorithm is expected to perform on unseen data.

The procedure of k-fold cross-validation is as follows. (1) Divide the data into k random and mutually exclusive groups. (2) Hold out one group as test data and use the remaining k-1 groups for training the machine learning algorithm from scratch. (3) Use the learned algorithm to calculate the results (accuracy) on the test data. Remember that the test data was not used for training the algorithm. (4) Repeat step (2) and (3) by holding out each of the k groups one at a time. (5) Report the average accuracy of the k experiments as well as the standard deviation in the accuracy.

Commonly used values of k are 5 and 10. We used 10-fold cross-validation in this paper. Note that with k-fold cross-validation, every data sample is guaranteed to be tested by the algorithm while trained on mutually exclusive training data. Moreover, each sample gets to be used k-1 times for training the algorithm. If we were to divide the data once only into a training and test set, it would count as a single-fold validation. In this case, not every sample gets tested and the test set may contain samples that are more (or less) challenging than we would expect to generally find in new data. Hence, our calculated estimate of how well the algorithm is expected to perform on new unseen cases may be too conservative or speculative.

Once the cross-validation is complete and the algorithm's expected performance is acceptable, it is trained on the complete data and deployed in the real world to test new samples.

## **Text S2**

The code for the final algorithm used is provided as a supplement to this manuscript. This code can be used by anyone to train an LDA classifier using their own data. Further, this code can be used to "test" their data to find the accuracy of multiple faces scanned at their site. Please note that for a 100% replication of our results in this paper, the user would require the 3D faces that were used.



% angles, etc. % GT Labels =  $n \times 1$  vector of ground truth labels. OSA is denoted as % '1' and nonOSA is denoted as '0'. % Outputs: % LDA vec = An m x 1 vector obtained by Linear Discriminant Analysis % of Training Data. % Mu $\overline{OSA}$  = A scalar representing the mean of OSA class in LDA space. % Mu\_nonOSA = A scalar representing the mean of nonOSA class in LDA space. % copyright 2019. Syed Zulqarnain Gilani and Ajmal Mian. This code can only % be used for research purposes. samples=Fea'; OSA= find(GT\_Labels==1); nonOSA= find(GT\_Labels==0); X1= samples(OSA,:); X2= samples(nonOSA,:); Mu1=mean(X1)'; Mu2=mean(X2)'; S1=cov(X1); S2=cov(X2); SW=S1+S2; SB=(Mu1-Mu2)\*(Mu1-Mu2)'; invSw=pinv(SW); invSwbySb=invSw\*SB; [V,~]=eig(invSwbySb); LDA\_vec= $V(:,1);$ newX1=LDA\_vec'\*X1'; newX2=LDA\_vec'\*X2'; Mu\_OSA=mean(newX1); Mu\_nonOSA=mean(newX2);

```
function [Label]=Test_LDA_AHI_Classifier(Fea,LDA_vec,Mu_OSA,Mu_nonOSA) 
%Test_LDA_AHI_Classifier: Projects a feature vec of a sample in LDA space 
%and predicts the label (OSA=1, nonOSA=0). 
% Inputs: 
% Fea = A m x p matrix of features of the test samples where
% 'm' are the features and 'n' are the samples. 
% Features can be distances, angles, etc. 
% 'p' can be one for a single subject. 
% LDA vec = An m x 1 vector obtained by Linear Discriminant Analysis
% of Training Data. 
% Mu\overline{OSA} = A scalar representing the mean of OSA class in LDA
% space during training. 
% Mu_nonOSA = A scalar representing the mean of nonOSA class in LDA
% space during training. 
% Outputs: 
% Label = A p x 1 vector of predicted lables obtained by
% projecting the features of test samples in LDA space and 
% finding the distance to the two class means. The label 
% of the closest class-mean is ascribed to the test 
% sample. 
% copyright 2019. Syed Zulqarnain Gilani and Ajmal Mian. This code can only 
% be used for research purposes. 
TestSamples=Fea'; 
Label = zeros(size(TestSamples,1),1); 
for ii=1:size(TestSamples,1) 
   TC1=TestSamples(ii,:); 
   newTC=LDA_vec'*TC1'; 
   dOSA=dist([Mu_OSA,newTC]); 
   dOSA=dOSA(2); 
   dnonOSA=dist([Mu_nonOSA,newTC]); 
   dnonOSA=dnonOSA(2); 
  if dOSA<dnonOSA
     Label(ii)=1; 
   else 
     Label(ii)=0; 
   end 
end
```
## **Figure S1**

Angle =  $\text{acos}\left(\frac{a.b}{|a|,|b|}\right)$ 



**Figure S1.** Angles were determined between three points *A*, *B* and *C*. Linear dimensions only were used for these analyses. Each point is represented by its x, y and z coordinates in the three dimensional space.  $\vec{a}$  and  $\vec{b}$ are vectors from points *B* to *A* and points *B* to *C* respectively. Angular measurements represent an angle made up between the three listed points with the middle landmark representing the vertex of the angle.

There is a significant difference between a 3D angle measurement (as has been calculated in the current study) and a 2D angle measurement (as has been calculated in all previous studies). In the present study angles are calculated between two vectors in 3D space (i.e. each vector has an *x*, *y* and *z* coordinate). Hence, it is a composite of a 2D frontal view and a 2D lateral view. Take for example the Sn-N-Gn angle in Figure 1. In the frontal view this would be the angle between three points lying on an almost straight line. However, in the lateral view (see Lee *et al*. Craniofacial phenotyping in obstructive sleep apnea - a novel quantitative photographic approach, *Sleep* 32:37-45, 2009) the angle is substantial. It is notable that, in the Lee et al study, the angle between Sn-N-Sl is not significantly different between the non-OSA and OSA groups, whereas the angle between Sn-N-Gn, in the current study, is highly significantly different between the non-OSA and OSA groups (Supplementary Table 3).

**Table S1.** *Geodesic distances* between landmarks in individuals without OSA (AHI<5) and with OSA (AHI≥5).



\* Comparisons were performed using t-tests with p-value adjusted using Bonferroni Correction (Alpha/25). Avg; average of distances from left and right sides of the face. Values are Mean ±SD.

**Table S2.** *Linear distances* between landmarks in individuals without OSA (AHI<5) and with OSA (AHI≥5).





\* Comparisons were performed using t-tests with p-value adjusted using Bonferroni Correction (Alpha/25). Avg; average of distances from left and right sides of the face. Values are Mean ±SD.

**Table S3.** *Anglular measurements* between landmarks in individuals without OSA (AHI<5) and with OSA (AHI≥5).



Angular measurements represent an angle made up between the three listed points with the middle landmark representing the vertex of the angle. \*Comparisons were performed using t-tests with p-value adjusted using Bonferroni Correction (Alpha/10). Avg; average of angles from left and right sides of the face.

Values are Mean ±SD.

**Table S4.** *Geodesic distances* between landmarks in individuals without OSA (AHI<5) and those with mild OSA (5 $\leq$ AHI<15), moderate OSA (15 $\leq$ AHI<30) and severe OSA (AHI  $\geq$  30).





\*Comparisons between groups were performed using ANOVA and Tukey-Kramer post hoc testing. The p-value adjusted using Bonferroni Correction (Alpha/50). Values are Mean ±SD.

**Table S5***. Linear distances* between landmarks in individuals without OSA (AHI<5) and those with mild OSA (5<AHI<15), moderate OSA ( $15\leq$ AHI<30) and severe OSA (AHI  $\geq$  30).





\*Comparisons between groups were performed using ANOVA and Tukey-Kramer post hoc testing. The p-value adjusted using Bonferroni Correction (Alpha/50). Values are Mean ±SD.

**Table S6.** *Angular mesurements* between landmarks in individuals without OSA (AHI<5) and those with mild OSA (5 $\leq$ AHI $<$ 15), moderate OSA (15 $\leq$ AHI $<$ 30) and severe OSA (AHI $\geq$ 30).



Angular measurements represent an angle made up between the three listed points with the middle landmark representing the vertex of the angle. \*Comparisons between groups were performed using ANOVA and Tukey-Kramer post hoc testing. The p-value adjusted using Bonferroni Correction (Alpha/10). Values are Mean ±SD.