Prediction Model for Obstructive Sleep Apnea from Facial Depth Maps using Transfer Learning

Ramadevi Javvaji , Ravikiran Biroju, Hanuja Korukonda

Department of Electronics and Communication Engineering

Sree Dattha Group of Institutions, Hyderabad, Telangana, India.

ABSTRACT

Obstructive Sleep Apnea (OSA) occurs when obstruction happens repeatedly in the airway during sleep due to relaxation of the tongue and airway-muscles. Usual indicators of OSA are snoring, poor night sleep due to choking or gasping for air and waking up unrefreshed. OSA diagnosis is costly both in the monetary and timely manner. That is why many patients remain undiagnosed and unaware of their condition. Previous research has shown the link between facial morphology and OSA. In this paper, investigated the application of deep learning techniques to diagnose the disease through depth map of human facial scans. Depth map will provide more information about facial morphology as compared to the plain 2-D color image. Even with very less amount of sample data, we can get around 69 validation accuracy using transfer learning. We are predicting patients with above moderate > 15 or below moderate ≤ 15 OSA. Finally, the simulations revealed that the proposed VGG 19 resulted in superior performance as compared to existing model.

Keywords: Obstructive Sleep Apnea, Transfer learning, Deep Learning, Facial Depth Map.

1. INTRODUCTION

Social and personal activities are significantly affected by poor sleep. There are different types of sleep disorders, and it is costing us at different levels. As [1] shows that only in Australia sleep disorder costs the economy around $5.1 billion per year that comprises health care, associated medical conditions, productivity, and non-medical costs. And among all sleep disorder, OSA is the most common cause [2]. Normally during sleep, our upper airway remains open due to relaxed but strong enough muscles, lining the upper throat. But in OSA, someone can have a recurring blockage in upper airway due to different reasons, for more than 10 sec for each blockage, which causes the lungs out of oxygen and person to wake, which will restore the airway [3]. If more than 15 apneas occur, then the diagnosis of OSA is made. History of the patient, physical examination, polysomnography (PSG) test, and imaging are being used to diagnose OSA. The gold standard to diagnoses is PSG test. In which a person needs to sleep in a unit in a hospital with some sensors to monitor breathing patterns, Oxygen level, heart rate, and body movements. Some devices are also helping to conduct these tests at the patient's own home, but there will be a question mark on the reliability of the test and have not been proved to be as accurate as PSG [4]. After the test Apnea-Hypopnea Index (AHI) is computed. This index points out the severity of sleep apnea. Due to cost in term of money and time, invasiveness of the PSG, non-specific nature of symptoms associated with OSA and the limited access to sleep clinics, many OSA patients remain undiagnosed until significant symptoms appear [5].

Many attempts have been made in the past to predict OSA based on questionnaires. For example, the Berlin questionnaire predicts the level of risk based on snoring, tiredness, blood pressure and body mass index information while the Epworth Sleepiness questionnaire assesses sleepiness in various situations during the day. Although they are self-administered and low-cost, they have shortcomings in accurately identifying affected individuals.

2. LITERATURE SURVEY

Hillman et al. Studied the economic impact of sleep disorders demonstrates financial costs to Australia of $5.1 billion per year. This comprises $270 million for health care costs for the conditions themselves, $540 million for care of associated medical conditions attributable to sleep disorders, and about $4.3 billion largely attributable to associated productivity losses and non-medical costs resulting from sleep loss-related accidents. Based on the high prevalence of such problems and the known impacts of sleep loss in all its forms on health, productivity, and safety, it is likely that these poor sleep habits would add substantially to the costs from sleep disorders alone.

Lam et a. estimated the prevalence rates of OSA have been in the range of 2 to 10 per cent worldwide, and the risk factors for obstructive sleep apnoea include advanced age, male sex, obesity, family history, craniofacial abnormalities, smoking and alcohol consumptionEarly recognition and treatment of obstructive sleep apnoea may prevent from adverse health consequences. some of the epidemiological aspects of obstructive sleep apnoea in adults are reviewed.

Obstructive sleep apnoea should be considered in a range of presentations. Diagnosis is based on history, examination, investigation and, occasionally, a trial of therapy. Management options should start with lifestyle management. Further options include surgery, dental splints, and continuous positive airway pressure. Continuous positive airway pressure requires long term input by both the patient and the general practitioner. Common issues with the use of machines for the management of sleep apnoea are also discussed.

Lam et al. determined whether the craniofacial profile predicts the presence of OSA, the upper airway and craniofacial structure of 239 consecutive patients (164 Asian and 75 white subjects) referred to two sleep centres (Hong Kong and Vancouver) were prospectively examined for suspected sleep disordered breathing. A crowded posterior oropharynx and a steep thyromental plane predict OSA across two different ethnic groups and varying degrees of obesity.

Barrera et al. determined the anatomic dimensions of airway structures are associated with airway obstruction in obstructive sleep apnea (OSA) patients. Twenty-eight subjects with (n = 14) and without (n = 14) OSA as determined by clinical symptoms and sleep studies; volunteer sampleThe soft palate thickness, mandibular plane-hyoid (MP-H) distance, posterior airway space (PAS) diameters and area, and tongue volume were calculated.

Lee et al. studied confirms of hypothesis that there is a relationship between surface facial dimensions and upper airway structures in subjects with OSA using MRI during wakefulness. In particular, the strongest correlations were demonstrated between the volume of the tongue and the widths of the midface and lower face. Significant relationships between some surface facial measurements and anthropometrics of obesity were also demonstrated. Surface facial dimensions in combination were strong determinants for tongue volume.

Pae et al. determined the shape difference of the face and tongue of obstructive sleep apnea (OSA) patients, in comparison to those of non-apneic patients. A set of anatomical landmarks were selected for outlines of the face and the tongue on cephalograms. X and Y coordinates of each landmark were utilized as variablesDespite many limitations, this paper demonstrated that the supine cephalometric during wakefulness can be a useful adjunctive diagnostic tool for OSA, when cephalograms are analyzed in a coordinate data form.

Lee et al. compared the craniofacial morphological phenotype of subjects with and without obstructive sleep apnea (OSA) using a quantitative photographic analysis technique. Standardized frontal-profile craniofacial photographic imaging performed prior to polysomnography. Photographs were analyzed for the computation of linear, angular, area and polyhedral volume measurements representing dimensions and relationships of the various craniofacial regions. Craniofacial phenotypic differences in OSA in Caucasian subjects can be demonstrated using a photographic analysis technique.

Cuadros et al. investigated the use of both image and speech processing to estimate the apnea-hypopnea index, AHI (which describes the severity of the condition), over a population of 285 male Spanish subjects suspected to suffer from OSA and referred to a Sleep Disorders Unit. A set of local craniofacial features related to OSA are extracted from images after detecting facial landmarks using Active Appearance Models (AAMs). Support vector regression (SVR) is applied on facial features and i-vectors to estimate the AHI.

3. DATASET AND METHODOLOGY

Sleep data and 3D scans were collected from the patients appearing to Genesis Sleep Care for different sleep issues who undergo home-based/lab-based sleep studies. A total of 39 male and 30 female adults has participated so far in the study which had been approved by ECU Human Research Ethics Committee.

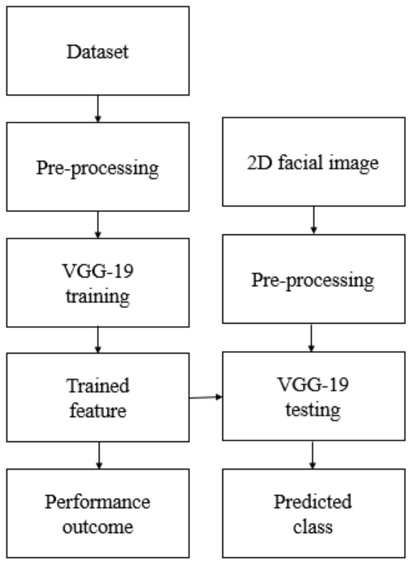


Fig. 1. Block diagram of the proposed deep learning based OSA prediction framework.

3.1 Dataset

The dataset contains the two classes, such as “Abnormal and normal”. Here, abnormal contains the 99 number of images and normal contains the 110 number of images. Sleep data and 3D scans were collected from the patients appearing to Genesis Sleep Care for different sleep issues who undergo home-based/lab-based sleep studies. A total of 39 male and 30 female adults has participated so far in the study which had been approved by ECU Human Research Ethics Committee. Overview of steps in all our methodology is shown in fig. 1. The 3D scans are captured by Artec Eva through Artec Studio [20]. These scans are recorded by different groups at different places that caused the variations in pose and produced some extra artifacts. As shown in fig. 2.

Fig. 2. Sample raw images in the collected 3D dataset.

While converting these 3D scans to frontal 2D depth maps, we want to reduce these unwanted variations. We use Artec Studio to make the corrections in all the 3D scans. As shown in fig. 2.

Fig. 3. Sample pre-processed images corresponding to raw images in Fig. 2

After making corrections in default poses and other factors, we choose the maximum and minimum scale to get higher resolution across depth values as shown in fig. 4.

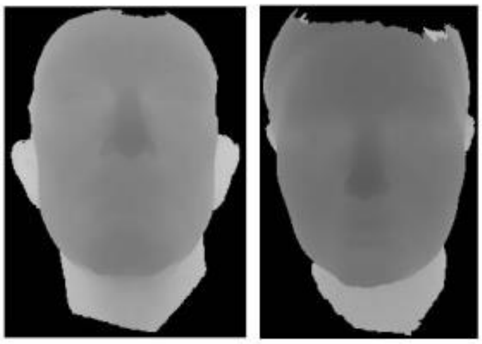


Fig. 4. 2D Facial depth maps.

3.2 Pre-processing

Digital image processing is the use of computer algorithms to perform image processing on digital images. As a subfield of digital signal processing, digital image processing has many advantages over analogue image processing. It allows a much wider range of algorithms to be applied to the input data — the aim of digital image processing is to improve the image data (features) by suppressing unwanted distortions and/or enhancement of some important image features so that our AI-Computer Vision models can benefit from this improved data to work on.

To train a network and make predictions on new data, your images must match the input size of the network. If you need to adjust the size of your images to match the network, then you can rescale or crop your data to the required size.

Resize image: In this step-in order to visualize the change, we are going to create two functions to display the images the first being a one to display one image and the second for two images. After that, we then create a function called processing that just receives the images as a parameter.

Need of resize image during the pre-processing phase, some images captured by a camera and fed to our AI algorithm vary in size, therefore, we should establish a base size for all images fed into our AI algorithms.

3.3 VGG-19

Training CNN from scratch needs a large amount of sample data, which in our case is very less. So, we choose three different networks which are pre-trained for face recognition. We choose VGGFace [22] Pose-Aware CNN Models (PAMs) for Face Recognition [23] for transfer learning with our dataset. Choosing the networks which are already trained on faces, although not on facial depth maps, provide a great jump start on learning. And in our experimentation, fine-tuning facial recognition for facial depth maps proves to be advantageous. VGG-Face is trained on 2.6 million images and performed well with 98.95% accuracy. This network is implemented on VGG-Very-Deep-16 CNN architecture. [23] provided two pretrained networks for face recognition with AlexNet [24] and 19-layer VGGNet [25]. To make these network classify for two classes, last fully connected layers are replaced by new fully connected layer and a softmax layer in each network type. After Adding the last layers below is the block diagram of all three networks.

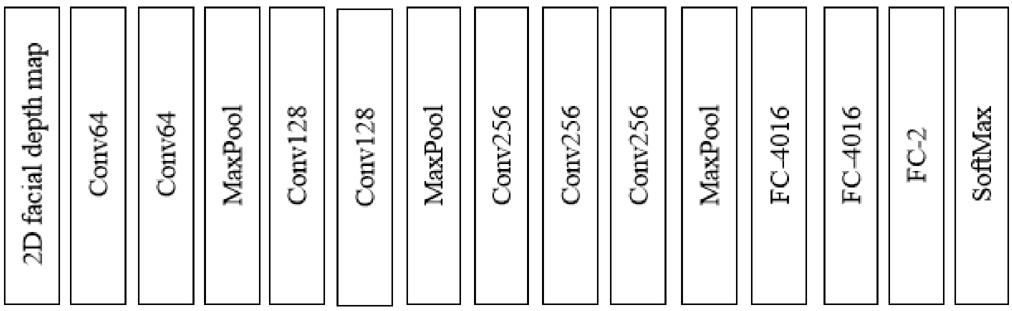


Fig. 5. Pre-Trained VGG Face with edition in last layers.

Background

AlexNet came out in 2012 and it improved on the traditional Convolutional neural networks, so we can understand VGG as a successor of the AlexNet, but it was created by a different group named as Visual Geometry Group at Oxford's and hence the name VGG, It carries and uses some ideas from its predecessors and improves on them and uses deep Convolutional neural layers to improve accuracy.

Let's explore what VGG19 is and compare it with some of other versions of the VGG architecture and see some useful and practical applications of the VGG architecture.

Before diving in and looking at what VGG19 Architecture is let's look at ImageNet and a basic knowledge of CNN.

So, in simple language VGG is a deep CNN used to classify images. The layers in VGG19 model are as follows:

Table.1: Layers description.

Conv1	3x3	64
Conv2	3x3	64
MaxPool	-	-
Conv3	3x3	128
Conv4	3x3	123
MaxPool	-	-
Conv5	3x3	256
Conv6	3x3	256
Conv7	3x3	256
MaxPool	-	-
Conv8	3x3	512
Conv9	3x3	512
Conv10	3x3	512
Conv11	3x3	512
MaxPool	-	-
Fully connected	-	4096
Fully connected	-	4096
Fully connected	-	1000
SoftMax	-	-

Architecture

- A fixed size of (224 * 224) RGB image was given as input to this network which means that the matrix was of shape (224,224,3).
- The only preprocessing that was done is that they subtracted the mean RGB value from each pixel, computed over the whole training set.
- Used kernels of (3 * 3) size with a stride size of 1 pixel, this enabled them to cover the whole notion of the image.
- Spatial padding was used to preserve the spatial resolution of the image.
- Max pooling was performed over a 2 * 2-pixel windows with sride 2.

- This was followed by Rectified linear unit (ReLu) to introduce non-linearity to make the model classify better and to improve computational time as the previous models used tanh or sigmoid functions this proved much better than those.
- Implemented three fully connected layers from which first two were of size 4096 and after that a layer with 1000 channels for 1000-way ILSVRC classification and the final layer is a softmax function.

3.4 CNN basics

Convolution layer is the primary layer to extract the features from a source image and maintains the relationship between pixels by learning the features of image by employing tiny blocks of source data. It's a mathematical function which considers two inputs like source image $I(x, y, d)$ where x and y denotes the spatial coordinates i.e., number of rows and columns. d is denoted as dimension of an image (here $d = 3$, since the source image is RGB) and a filter or kernel with similar size of input image and can be denoted as $F(k_x, k_y, d)$.

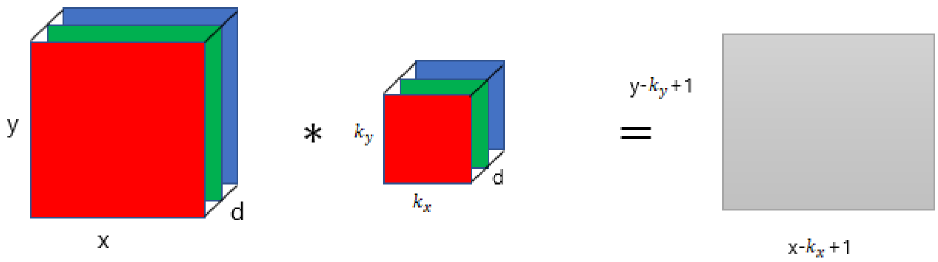
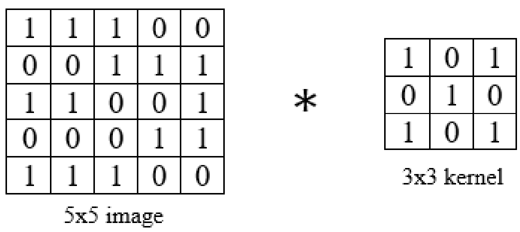
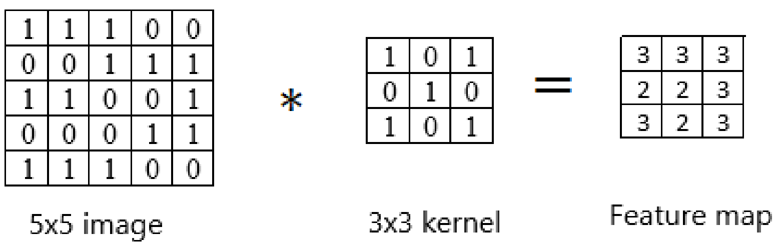


Fig. 6: Representation of convolution layer process

The output obtained from convolution process of input image and filter has a size of $C\left((x - k_x + 1), (y - k_y + 1), 1\right)$, which is referred as feature map. Let us assume an input image with a size of 5×5 and the filter having the size of 3×3.



(a)



(b)

Fig. 7: Example of convolution layer process (a) an image with size $\mathbf{5} \times \mathbf{5}$ is convolving with $\mathbf{3} \times \mathbf{3}$ kernel (b) Convolved feature map

3.4.1 ReLU layer

Networks those utilizes the rectifier operation for the hidden layers are cited as rectified linear unit (ReLU). This ReLU function $\mathcal{G}(\cdot)$ is a simple computation that returns the value given as input directly if the value of input is greater than zero else returns zero. This can be represented as mathematically using the function $max(\cdot)$ over the set of 0 and the input x as follows:

$$\mathcal{G}(x) = \max\{0, x\}$$

3.4.2 Max pooing layer

This layer mitigates the number of parameters when there are larger size images. This can be called as subsampling or down sampling that mitigates the dimensionality of every feature map by preserving the important information. Max pooling considers the maximum element form the rectified feature map.

3.4.3 Softmax classifier

Here, X is the input of all the models and the layers between X and Y are the hidden layers and the data is passed from X to all the layers and Received by Y. Suppose, we have 10 classes, and we predict for which class the given input belongs to. So, for this what we do is allot each class with a particular predicted output. Which means that we have 10 outputs corresponding to 10 different class and predict the class by the highest probability it has.

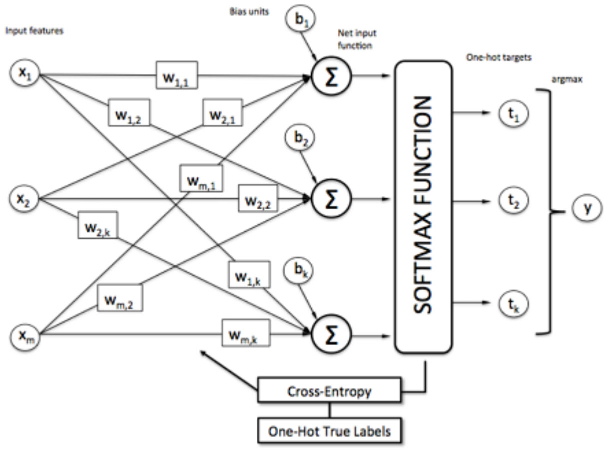


Fig. 8: Softmax classifier.

4. RESULTS AND DISCUSSION

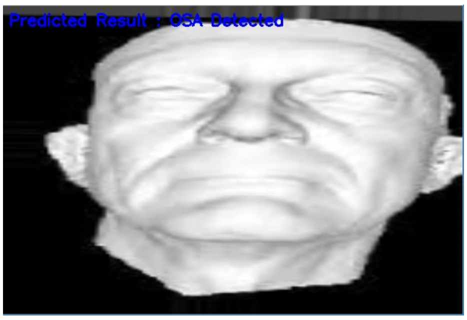
This section gives the detailed analysis of simulation results implemented using “python environment”. Further, the performance of proposed method is compared with existing methods using same dataset.

4.1 Modules

1) Upload OSA Faces Dataset: using this module we will upload dataset to application

2) Pre-process Dataset: using this module we will read all images and then resize all images to equal size and then normalize all pixel values
3) Build VGG-19 Model: processed images will be input to VGG-19 algorithm to train a model
4) Upload Test Data & Predict OSA: using this module we will upload new test image and then applied VGG19 trained model to predict test image is normal or contains OSA disease
5) Accuracy Comparison Graph: using this module we will plot VGG19 training accuracy and loss graph

(a) (b)



(c)

Fig. 9: Predicted result recognized as OSA detected.

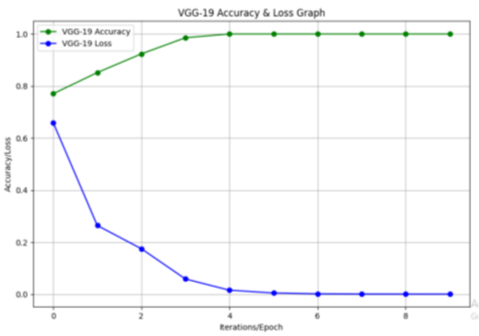


Fig. 10: Accuracy performance comparison.

5. CONCLUSION

In this paper, we propose the first facial depth map-based sleep apnea detection. Patients' dataset is small, to overcome this limitation we took advantage of transfer learning. We analyze three pre-trained models and among them, VGG face performs the best. Our method shows comparable performances to the state-of-the-art results in terms of getting prediction straight from depth facial data using end-to-end deep learning. In future, pose correction problem will be solved through a 3D morphable model. Hole filled depth maps will be created through an automatic procedure. This work gets good results with a very small dataset and with more 3D scans of OSA and non-OSA patients, we will enhance the performance for diagnoses.

REFERENCES

[1] D. R. Hillman and L. C. Lack, “Public health implications of sleep loss: the community burden,” Med J Aust, vol. 199, no. 8, pp. S7–S10, 2013.

[2] J. C. Lam, S. Sharma, B. Lam et al., “Obstructive sleep apnoea: definitions, epidemiology & natural history,” Indian Journal of Medical Research, vol. 131, no. 2, p. 165, 2010.

[3] “Obstructive sleep apnoea (osa; sleep apnea) information,” Jul 2018. [Online]. Available: https://www.myvmc.com/diseases/obstructivesleep-apnoea-osa-sleep-apnea

[4] K. Kee, M. T. Naughton et al., “Sleep apnoea-a general practice approach,” Australian family physician, vol. 38, no. 5, p. 284, 2009.

[5] B. Lam, M. Ip, E. Tench, and C. Ryan, “Craniofacial profile in asian and white subjects with obstructive sleep apnoea,” Thorax, vol. 60, no. 6, pp. 504–510, 2005.

[6] J. E. Barrera, C. Y. Pau, V.-I. Forest, A. B. Holbrook, and G. R. Popelka, “Anatomic measures of upper airway structures in obstructive sleep apnea,” World journal of otorhinolaryngology-head and neck surgery, vol. 3, no. 2, pp. 85–91, 2017.

[7] R. J. Schwab, M. Pasirstein, R. Pierson, A. Mackley, R. Hachadoorian, R. Arens, G. Maislin, and A. I. Pack, “Identification of upper airway anatomic risk factors for obstructive sleep apnea with volumetric magnetic resonance imaging,” American journal of respiratory and critical care medicine, vol. 168, no. 5, pp. 522–530, 2003.

[8] Lowe, J. A. Fleetham, S. Adachi, and C. F. Ryan, “Cephalometric and computed tomographic predictors of obstructive sleep apnea severity,” American Journal of Orthodontics and Dentofacial Orthopedics, vol. 107, no. 6, pp. 589–595, 1995.

[9] T. Vidigal, L. Oliveira, T. Moura, F. Haddad, K. Sutherland, P. Cistulli, R. Schwab, A. Pack, U. Magalang, S. Leinwand et al., “Can intra-oral and facial photos predict osa in the general and clinical population?” Sleep Medicine, vol. 40, p. e339, 2017.

[10] R. W. Lee, K. Sutherland, A. S. Chan, B. Zeng, R. R. Grunstein, M. A. Darendeliler, R. J. Schwab, and P. A. Cistulli, “Relationship between surface facial dimensions and upper airway structures in obstructive sleep apnea,” Sleep, vol. 33, no. 9, pp. 1249–1254, 2010.

[11] E.-K. Pae, A. A. Lowe, and J. A. Fleetham, “Shape of the face and tongue in obstructive sleep apnea patientsstatistical analysis of coordinate data,” Clinical orthodontics and research, vol. 2, no. 1, pp. 10–18, 1999.

[12] R. W. Lee, A. S. Chan, R. R. Grunstein, and P. A. Cistulli, “Craniofacial phenotyping in obstructive sleep apneaa novel quantitative photographic approach,” Sleep, vol. 32, no. 1, pp. 37–45, 2009.

[13] R. W. Lee, P. Petocz, T. Prvan, A. S. Chan, R. R. Grunstein, and P. A. Cistulli, “Prediction of obstructive sleep apnea with craniofacial photographic analysis,” Sleep, vol. 32, no. 1, pp. 46–52, 2009.

[14] F. Espinoza-Cuadros, R. Fernandez-Pozo, D. T. Toledano, J. D. Alc ´ azar- ´ Ram´ırez, E. Lopez-Gonzalo, and L. A. Hern ´ andez-G ´ omez, "Speech sig- ´ nal and facial image processing for obstructive sleep apnea assessment," Computational and mathematical methods in medicine, vol. 2015, 2015.

[15] H. Nosrati, N. Sadr, and P. de Chazal, "Apnoea-hypopnoea index estimation using craniofacial photographic measurements," in Computing in Cardiology Conference (CinC), 2016. IEEE, 2016, pp. 1033–1036.

[16] T. Balaei, K. Sutherland, P. A. Cistulli, and P. de Chazal, "Automatic detection of obstructive sleep apnea using facial images," in Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on. IEEE, 2017, pp. 215–218.

[17] S. M. Islam, M. S. Goonewardene, and M. Farella, "A review on three-dimensional facial averaging for the assessment of orthodontic disorders," in Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering. Springer, 2015, pp. 391–397.

[18] J. Moss, "The use of three-dimensional imaging in orthodontics," The European Journal of Orthodontics, vol. 28, no. 5, pp. 416–425, 2006.

[19] S. M. Banabilh, A. Suzina, S. Dinsuhaimi, A. Samsudin, and G. Singh, "Craniofacial obesity in patients with obstructive sleep apnea," Sleep and Breathing, vol. 13, no. 1, pp. 19–24, 2009.

[20] "Artec eva." [Online]. Available: https://www.artec3d.com/portable-3dscanners/artec-eva

[21] "Meshlab." [Online]. Available: http://www.meshlab.net

[22] O. M. Parkhi, A. Vedaldi, A. Zisserman et al., "Deep face recognition." in BMVC, vol. 1, no. 3, 2015, p. 6.

[23] Masi, S. Rawls, G. Medioni, and P. Natarajan, "Pose-Aware Face Recognition in the Wild," in CVPR, 2016.

[24] Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in Advances in neural information processing systems, 2012, pp. 1097–1105.

[25] Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman, "Return of the devil in the details: Delving deep into convolutional nets," arXiv preprint arXiv:1405.3531, 2014.