



Estimation of Apnea-Hypopnea Index using Deep Learning on 3D Craniofacial Scans

Hanif, Umaer ; Leary, Eileen B.; Schneider, Logan D.; Paulsen, Rasmus R.; Morse, Anne Marie; Blackman, Adam; Schweitzer, Paula K.; Kushida, Cleto A.; Liu, Stanley Y.; Jennum, Poul

Total number of authors:
12

Published in:
IEEE Journal of Biomedical and Health Informatics

Link to article, DOI:
[10.1109/JBHI.2021.3078127](https://doi.org/10.1109/JBHI.2021.3078127)

Publication date:
2021

Document Version
Peer reviewed version

[Link back to DTU Orbit](#)

Citation (APA):
Hanif, U., Leary, E. B., Schneider, L. D., Paulsen, R. R., Morse, A. M., Blackman, A., Schweitzer, P. K., Kushida, C. A., Liu, S. Y., Jennum, P., Sorensen, H. B. D., & Mignot, E. (2021). Estimation of Apnea-Hypopnea Index using Deep Learning on 3D Craniofacial Scans. *IEEE Journal of Biomedical and Health Informatics*, 25(11), 4185-4194. <https://doi.org/10.1109/JBHI.2021.3078127>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Estimation of Apnea-Hypopnea Index using Deep Learning on 3D Craniofacial Scans

Umaer Hanif, *Member, IEEE*, Eileen B. Leary, Logan D. Schneider, Rasmus R. Paulsen, Anne Marie Morse, Adam Blackman, Paula K. Schweitzer, Clete A. Kushida, Stanley Y. Liu, Poul Jennum, Helge B. D. Sorensen, *Senior Member, IEEE*, and Emmanuel J. M. Mignot

Abstract— Obstructive sleep apnea (OSA) is characterized by decreased breathing events that occur through the night, with severity reported as the apnea-hypopnea index (AHI), which is associated with certain craniofacial features. In this study, we used data from 1366 patients collected as part of Stanford Technology Analytics and Genomics in Sleep (STAGES) across 11 US and Canadian sleep clinics and analyzed 3D craniofacial scans with the goal of predicting AHI, as measured using gold standard nocturnal polysomnography (PSG). First, the algorithm detects pre-specified landmarks on mesh objects and aligns scans in 3D space. Subsequently, 2D images and depth maps are generated by rendering and rotating scans by 45-degree increments. Resulting images were stacked as channels and used as input to multi-view convolutional neural networks, which were trained and validated in a supervised manner to predict AHI values derived from PSGs. The proposed model achieved a mean absolute error of 11.38 events/hour, a Pearson correlation coefficient of 0.4, and accuracy for predicting OSA of 67% using 10-fold cross-validation. The model improved further by adding

patient demographics and variables from questionnaires. We also show that the model performed at the level of three sleep medicine specialists, who used clinical experience to predict AHI based on 3D scan displays. Finally, we created topographic displays of the most important facial features used by the model to predict AHI, showing importance of the neck and chin area. The proposed algorithm has potential to serve as an inexpensive and efficient screening tool for individuals with suspected OSA.

Index Terms— Apnea, craniofacial scans, deep learning, mesh, multi-view

I. INTRODUCTION

Obstructive sleep apnea (OSA) is a sleep disorder characterized by recurrent collapses of the upper airway (UA) during sleep, resulting in decreased airflow (hypopnea) or total cessation of breathing (apnea), lasting until the UA reopens [1] and causing daytime sleepiness and increased cardiovascular risk [2], [3]. OSA severity is measured by the apnea-hypopnea index (AHI), representing the number of apneas and hypopneas per hour of sleep. The presence of sleep disordered breathing (SDB) is extremely common; in a recent study, the prevalence of moderate-to-severe SDB (≥ 15 events/hour) was 23.4% in women and 49.7% in men older than 40 [4]. Why and how frequently the UA collapses is due to multiple factors, such as narrow UA anatomy, poor recruitment of dilator muscles during inspiration, central control of breathing (loop gain), and inability/ability to arouse (arousal threshold) [5]. A strong contributing factor to passive anatomy is obesity, which causes fat deposits around the UA that narrow the airway during sleep [6]. Other reasons include loss of muscle tone in the UA or the tongue falling backwards into the throat [7]. Finally, research shows that several craniofacial features, mainly related to the midface, jaw, and neck are indicative of the presence of OSA [8].

Nocturnal polysomnography (PSG) is the accepted gold-standard for diagnosing OSA [9]. PSGs are performed in a sleep clinic or laboratory where the patient can sleep for a full night. While the individual is sleeping, sensors measure airflow, respiratory effort, snoring sounds, blood oxygen levels, eye movements, leg movements, and electrical activity of the heart (ECG) and brain (EEG). These signals are manually annotated by sleep technicians, who follow standard definitions for apneas, hypopneas, periodic leg movements, and sleep stages [10]. A PSG, though gold-standard for OSA

Paper submitted for review November 30, 2020. This work was supported by a grant from the Klarman Family Foundation. Additional funding was from the Technical University of Denmark, and the Danish Center for Sleep Medicine. Mr. Hanif's stay at Stanford University was funded by Danmark-Amerika Fondet, Vera og Carl Johan Michaelsens Legat, Reinholdt W. Jorck og Hustrus Fond, Torben og Alice Frimodts Fond, Christian og Otilia Brorsons Rejselegat, Marie og M.B. Richters Fond, Oberstløjtnant Max Nørgaard og hustru Magda Nørgaards Legat, William Demant Fonden, Augustinus Fonden, Rudolph Als Fondet, Knud Højgaard's Fond, Otto Mønsteds Fond, Julie von Müllens Fond, and Direktør Einar Hansen og hustru fru Vera Hansens Fond. (*Corresponding author: Emmanuel J. M. Mignot.*)

Umaer Hanif, Eileen B. Leary, Logan D. Schneider, Clete A. Kushida, and Emmanuel J. M. Mignot are with Stanford University Center for Sleep Sciences and Medicine, Stanford University, 3165 Porter Dr., CA 94304, Palo Alto, USA (e-mail: umaerhanif@hotmail.com, mignot@stanford.edu).

Umaer Hanif and Helge B. D. Sorensen are with Biomedical Signal Processing & AI Research Group, Department of Health Technology, Technical University of Denmark, Building 349, 2800 Kongens Lyngby, Denmark (e-mail: hbds@dtu.dk).

Umaer Hanif and Poul Jennum are with Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Rigshospitalet, Nordre Ringvej 57, 2600 Glostrup, Denmark (e-mail: poul.joergen.jennum@regionh.dk).

Rasmus R. Paulsen is with Department for Applied Mathematics and Computer Science, Technical University of Denmark, Richard Petersens Plads, Building 324, 2800 Kongens Lyngby, Denmark.

Anne Marie Morse is with Geisinger Commonwealth School of Medicine, Geisinger Medical Center, 525 Pine St, PA 18510, Scranton, USA.

Adam Blackman is with MedSleep, 586 Eglinton Ave E #507, Toronto, ON M4P 1P2, Canada.

Paula K. Schweitzer is with Sleep Medicine & Research Center, St. Luke's Hospital, 232 S. Woods Mill Rd., MO 63017, Chesterfield, USA.

Stanley Y. Liu is with Otolaryngology/Head & Neck Surgery, Stanford University School of Medicine, 450 Broadway St, CA 94133, Redwood City, USA.

diagnosis, is expensive and impractical; it requires a great number of modalities and a trained technician to perform, then manually view and evaluate data. Furthermore, scoring of PSG data is prone to errors, day-to-day variance, and high interscorer variability [11], [12]. Finally, a PSG is supposed to accurately portray sleep behavior, but this is often difficult as only one night is recorded, the individual is wearing equipment, and is connected to several wires, thereby reducing their comfort. Due to these limitations, research efforts have gone into finding alternative, data-driven approaches for the diagnosis of sleep disorders, including use of deep neural networks to score events [13]–[17].

Obesity and craniofacial features strongly contribute to OSA risk; thus, technology and clinical examination tools have been developed to assess these features. In clinical exams, it is frequent for the clinician to examine the size of the jaw, top of the mouth, position, and size of the tongue [18]. Imaging techniques have been proposed, including cephalometry [19]–[21], computed tomography [21], [22], and magnetic resonance imaging [23]–[25]. Furthermore, dynamic collapsibility can be identified with drug-induced sleep endoscopy (DISE) [26]. These methods are, however, rarely used in routine practice, except in case of surgery for OSA.

Since imaging modalities are cumbersome and expensive, recent research has investigated the predictive value of facial imaging [27]. Lee et al. [28] analyzed frontal and profile images of 180 patients, manually deriving measurements on the face and neck to classify subjects with or without OSA using logistic regression. Others [29]–[31] used Support Vector Regression (SVR) on similar landmarks to predict the AHI, a procedure subsequently improved by Balaei et al. [32] who used automatic instead of manual placement of landmarks. Islam et al. [33] used 3D scans from 69 subjects, which they converted to 2D depth maps of the frontal face, applied transfer learning on a VGG-16 deep convolutional neural network (CNN), and modified the network to classify subjects into OSA and non-OSA. Although these studies had some success, all used small sample sizes, and, with the exception of Islam et al. [33], all first derive possibly discriminative facial features from annotated landmarks as detected on 2D frontal and profile images, a process followed by statistical feature selection. As attempted by Islam et al. [33] using a small sample size, we believe that feature selection should be unbiased by avoiding manual extraction of features and applying CNNs instead.

In this study, taking advantage of the rapid development of depth imaging in most hand held devices, we aimed to explore how state of the art deep learning techniques as applied to a dataset of more than 1300 3D images could be used to develop a fully automatic system for the prediction of OSA severity and AHI values. This study aims to (1) investigate how accurately such a system could predict the AHI, (2) determine how accurately we can classify OSA/non-OSA based on an AHI cutoff of 15 events/hour, (3) estimate the predictive value of adding clinically relevant information like demographics and questionnaire variables to the model, and (4) use a data-driven approach to identify which regions of the face and neck

the CNN found most useful in predicting OSA severity. From a clinical standpoint, the study further aims to (5) compare our model performance to that of three sleep medicine specialists asked to predict OSA severity after inspecting the scans.

The novelty of this study is four-fold: the dataset consists of 3D surface scans instead of frontal and profile images; we compare our model predictions to those of three sleep medicine specialists imitating the task of the model; we use a purely data-driven approach to identify and reveal craniofacial features related to OSA; and our dataset is four times greater than any other study predicting OSA from craniofacial images. Facial imaging for OSA diagnosis can be performed in one minute with the current setup and presents a clear advantage; it does not require an overnight stay at a sleep clinic with several sensors connected to the body and a subsequent manual analysis by sleep technicians, thereby saving both time and resources, while being more comfortable for the patient.

II. MATERIALS

A. Data Collection

Data was collected at 11 different sleep clinic sites as part of the Stanford Technology Analytics and Genomics in Sleep (STAGES) study, which was initiated in 2018 and prematurely terminated in 2020. STAGES was designed to better understand and characterize sleep disorder phenotypes on a large scale. For each subject participating in the study, a detailed sleep questionnaire, actigraphy, psychometric testing, a PSG, a 3D craniofacial scan, and blood samples were collected. The 3D craniofacial scans were collected according to a procedure described in Hanif et al. [34]. The scans were performed using a Structure Sensor from Occipital Inc. [35], [36] attached to an iPad Pro from Apple, which was used around the subject to get a complete surface scan of face and neck. Each scan took approximately one minute to complete and was captured either at night before the PSG or in the morning after the PSG. uGo3D Inc. developed an app for STAGES which was responsible for transferring the scans to the server after they were captured. The sleep questionnaire was a modified STOP-Bang questionnaire [37] without neck circumference of the subject. The modified STOP-Bang questionnaire is used as a screening tool for OSA without the need for a specialist by asking the subject about snoring, tiredness, observed apnea, high blood pressure, sex (is the person male), age (is the person older than 65), and BMI (is the BMI greater than 35 kg/m²). Each variable gets a value of 1 if the person answers yes, so the total score from the questionnaire ranges from 0 to 7, where a score of 3 or above indicates presence of OSA. Each institution's Ethical Review Board approved all procedures involving human subjects. All participants provided written informed consent to participate in the study. The 3D scans from STAGES are not publicly available, since they count as personal identifiable data, which the Institutional Review Board would not allow to be made public. All other data from STAGES, apart from blood samples, will be made available as part of the National Sleep Research Resource (NSRR).

B. Data Description

During data collection for STAGES, a total of 1756 scans were captured from enrolled participants. However, subjects were discarded for this study if they had missing PSGs, demographics, or questionnaires, so 1366 subjects were used; 724 females and 642 males. Mean age \pm standard deviation was 45.9 ± 14.8 years, body mass index (BMI) was 30.9 ± 8.7 kg/m², and AHI was 15.5 ± 19.3 events/hour (median: 9.3, IQR: 17.5). Fig. A.1. (appendix) shows the distribution of AHI values within the dataset. AHI was derived from each PSG by summing the number of annotated obstructive apneas and hypopneas and dividing this number by the total sleep time in hours. Central apneas were excluded from the analysis because they have no known relation to craniofacial anatomy. Each scan was stored as a mesh object, defined by its vertices $V = \{v_1, v_2, \dots, v_n\}$, $V \in \mathbb{R}^3$, its triangles F , given as three interconnected vertices, and its texture coordinates $T = \{t_1, t_2, \dots, t_m\}$, $T \in \mathbb{R}^3$. A vertex is a point in 3D space described by its (x,y,z) coordinates. Three interconnected vertices form a triangle and several of these triangles are combined to approximate a surface in three dimensions. Furthermore, each scan also contains associated textures. Fig. 1 shows a typical example of a craniofacial scan where the triangles that make up the scan are also depicted. Fig. A.2. (appendix) provides a simple example of the basic components of a mesh, i.e. vertices, edges, and triangles.

III. METHODS

A. Preprocessing

Since 3D mesh scans are non-Euclidean, they cannot be used directly as inputs to CNNs. Thus, to make craniofacial scans suitable for CNNs, the multi-view consensus CNN for 3D facial landmark placement (Deep-MVLM) algorithm [38] was applied to transform each scan into a set of 2D images and depth maps captured from angles around the scan. Deep-MVLM was chosen as it outperforms state of the art algorithms [39], [40] and does not rely on pre-alignment of scans, such that they have the same orientation in 3D space. The choice of metric in Deep-MVLM also makes it more suitable for 3D surfaces than similar methods. Deep-MVLM was only used for alignment of scans and not for the subsequent prediction of AHI.

Deep-MVLM works by first applying a pre-trained neural network to automatically detect and place 73 pre-specified landmarks on each mesh object. The alignment of scans is obtained by using a least squares solution based on the detected landmarks and therefore the individual landmark errors are less important for the overall performance. Subsequently, these landmarks are utilized to align all scans in 3D space, thereby removing the influence of translation and rotation. Finally, each scan is displayed multiple times as flat 2-D images taken from different angles; this is done by rotating the scans 45 degrees around the y-axis consecutively and capturing a 2D image and a depth map at each angle. This results in eight pairs of 2D images and depth maps for each subject. Since some scans were not complete with respect to the back of the head and since this information does not

contribute to AHI prediction, only five pairs of 2D images and depth maps were used, emphasizing the frontal and profile characteristics of face and neck from several angles as shown in Fig. 2. For each subject, these images were stacked into a matrix. The 2D images had three channels each (RGB) and the depth maps had one channel each. Since there were five 2D images and five depth maps, the final input matrix per subject contained $5 \times 3 + 5 \times 1 = 20$ -channels in total and 224×224 pixels.

Each image was normalized to the range [0,1] by dividing each pixel value by 255. This was done to ensure faster convergence during training. Similarly, the patient demographics (age and bmi) were normalized using min-max normalization given by

$$\mathbf{x}_{\text{norm}} = \frac{\mathbf{x} - \min(\mathbf{x}_{\text{train}})}{\max(\mathbf{x}_{\text{train}}) - \min(\mathbf{x}_{\text{train}})}$$

where \mathbf{x} and $\mathbf{x}_{\text{train}}$ are vectors containing one of the demographics for all patients in the entire dataset and training set, respectively. No normalization was necessary for sex or the seven different variables from the modified STOP-Bang questionnaires because these values were binary.

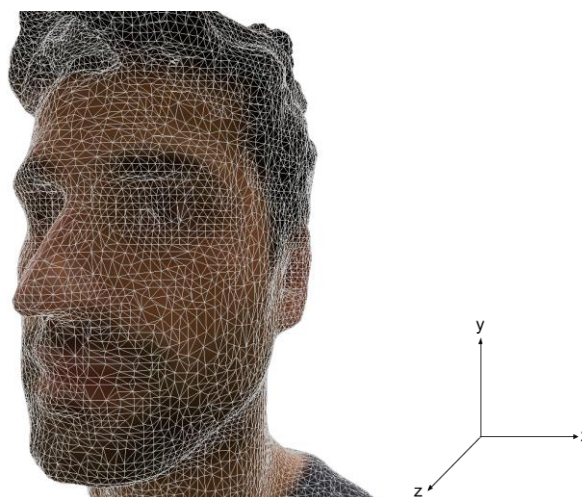


Fig. 1. An example of a 3D craniofacial scan (first author). Each triangle is formed by connecting three vertices, which are points in 3D space. Crude details are approximated using large triangles, whereas finer details require smaller triangles. The axes to the right specify the orientation of the coordinates.

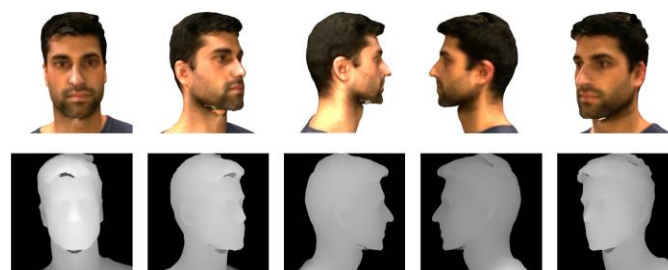


Fig. 2. Example of the five pairs of 2D images (top row) and depth maps (bottom row) captured at different angles and used as input for each subject in the multi-view convolutional neural network for predicting AHI.

B. Multi-view CNN

Our purpose for applying machine learning was to reveal data-driven mapping differences within the multi-view inputs across AHI values. For this purpose, we utilized a CNN with a ResNet18 architecture [41], which was modified to take 20 channels as input per subject. Two additional fully connected layers were added at the end, reducing 512 features to 128, followed by a ReLU activation and dropout (probability of 0.3 and 0.5, respectively), and finally a single neuron for prediction of the continuous AHI value. This architecture is illustrated in Table 1. The optimal number of neurons in the fully connected layers were found by hyperparameter tuning. Additional networks were developed for the demographics and questionnaires, respectively. Both networks were multilayer perceptrons (MLPs) with three layers and varying number of inputs (three for demographics and seven for questionnaires). This architecture is shown in Table 2. The AHI predictions of each of the three networks, using scans, demographics, and questionnaires, were averaged using an ensemble approach to form the final prediction.

Training, validation, and testing was carried out using 10-fold cross-validation. Mean squared error (MSE) was used as loss function to train the networks. This was done to penalize greater errors, since most of the AHI values in the dataset were in the range 0-15 events/hour (Fig. A.1.). Learning rate was set to $1 \cdot 10^{-5}$ with a weight decay of $5 \cdot 10^{-4}$ for the multi-view CNN, and $1 \cdot 10^{-2}$ (with the same weight decay) for the MLPs, while batch size was set to 8 for all three networks. The Adam optimizer [42] was used for training, and early stopping was applied when the validation error did not decrease for 3 consecutive epochs (patience of 3). Python 3.7.4 and Pytorch 1.3.1 were used for preprocessing and deep learning purposes. Training was carried out on a GeForce RTX 2080 and each model took one hour to train.

The measures used to evaluate model performance on the test set were mean absolute error (MAE) and Pearson Correlation Coefficient (PCC). Furthermore, predicted AHI was used to classify subjects into having OSA or not, where an AHI of 15 events/hour or greater was the clinical criterion used for defining the presence of OSA (moderate-severe versus mild or no sleep apnea). Thus, accuracy of classification was another measure used to evaluate the model performance. Bland-Altman plots [43] were used to illustrate the patterns of disagreement between true and predicted AHI values.

C. Sleep Specialists' Ability to Guess AHI

Three experienced, board certified sleep medicine physicians with in-depth knowledge of OSA were recruited to imitate the task of the proposed algorithm, estimating AHI based on inspection of the 3D scan of each subject (face and neck). The three physicians scored one third of the dataset each, while also annotating 150 of the same scans to estimate percentage agreement. When annotating a scan, each physician was shown the scan from all desired angles, having the ability to rotate the 3D image for any desired amount of time. Physicians took approximately 30 seconds to score each scan. Their first thought was to size up the person based on their estimated age, with the awareness that older individuals

often have higher AHIs (due to lax musculature, redundant tissue, and an atrophic skeletal scaffolding that result in higher risk of airway collapse). Then they would ascertain if the

TABLE 1. The applied CNN architecture for predicting AHI values based on the 20-dimensional input craniofacial images. The input dimensions are given by number of channels×height×width. The convolution layers are specified by filter size (e.g. 3×3), number of channels (e.g. 64), and a stride (e.g. /2). The same applies for the max pooling (MP) layer. The output dimensions of the feature maps are given by number of channels×height×width. The convolution layers are always followed by batch normalization. The dropout layers have keep probabilities of 0.3 and 0.5, respectively. If the skip connections (arrows) are applied, the feature maps are down sampled instead by applying 1x1 filters with a stride of 2x2. AP – Average pooling, FC – Fully connected.

| Layer | Type | Dimension | Activation | Out dim |
|----------------|---------|--------------|------------|------------|
| 0 | Input | 20×224×224 | - | - |
| Layer | Type | Convolution | Activation | Out dim |
| 1 | Conv | 7×7, 64, /2 | ReLU | 64×112×112 |
| 2 | MP | 3×3, -, /2 | - | 64×56×56 |
| Block 1 | | | | |
| 3 | Conv | 3×3, 64 | ReLU | 64×56×56 |
| 4 | Conv | 3×3, 64 | - | 64×56×56 |
| 5 | Conv | 3×3, 64 | ReLU | 64×56×56 |
| 6 | Conv | 3×3, 64 | - | 64×56×56 |
| Block 2 | | | | |
| 7 | Conv | 3×3, 128, /2 | ReLU | 128×28×28 |
| 8 | Conv | 3×3, 128 | - | 128×28×28 |
| 9 | Conv | 3×3, 128 | ReLU | 128×28×28 |
| 10 | Conv | 3×3, 128 | - | 128×28×28 |
| Block 3 | | | | |
| 11 | Conv | 3×3, 256, /2 | ReLU | 256×14×14 |
| 12 | Conv | 3×3, 256 | - | 256×14×14 |
| 13 | Conv | 3×3, 256 | ReLU | 256×14×14 |
| 14 | Conv | 3×3, 256 | - | 256×14×14 |
| Block 4 | | | | |
| 15 | Conv | 3×3, 512, /2 | ReLU | 512×7×7 |
| 16 | Conv | 3×3, 512 | - | 512×7×7 |
| 17 | Conv | 3×3, 512 | ReLU | 512×7×7 |
| 18 | Conv | 3×3, 512 | - | 512×7×7 |
| Layer | Type | Neurons | Activation | Out dim |
| 19 | AP | 512 | - | 512×1×1 |
| 20 | FC | 512 | ReLU | 512×1 |
| 21 | Dropout | - | - | 512×1 |
| 22 | FC | 128 | ReLU | 128×1 |
| 23 | Dropout | - | - | 128×1 |
| 24 | FC | 1 | - | 1×1 |

TABLE 2. The applied MLP architecture for predicting AHI values based on the 3-dimensional input demographics. The MLP using questionnaires as input is identical except for the input dimensions, which are 7×1. FC – Fully connected.

| Layer | Type | Dimension | Activation | Out dim |
|-------|-------|-----------|------------|---------|
| 0 | Input | 3×1 | - | - |
| Layer | Type | Neurons | Activation | Out dim |
| 1 | FC | 32 | ReLU | 32×1 |
| 2 | FC | 64 | ReLU | 64×1 |
| 3 | FC | 1 | - | 1×1 |

individual looked tired - droopy eyelids (ptosis), drawn face, pallor, circles/bags under the eyes, etc. - that might suggest an underlying sleep disorder. Finally, they would look for some of the high-yield characteristics: looking at the overall head and neck adiposity (fat), the characteristics of the thyromental space, the over/under-bite (to suggest a retrognathic jaw that pushes the tongue into the airway), the craniofacial complex (looking for maxillary or mandibular hypoplasia) and noting whether there was a long/thin face suggestive of life-long nasal congestion ("adenoid facies"), and the cervical lordosis (to see if subjects have their heads thrust forward, suggestive of position modification in order to ease breathing). They would then roughly estimate the AHI based on all these factors and their general knowledge of the known prevalence/proportions of varying severities of OSA. It is important to note that clinicians do not traditionally estimate AHI, but for this study their estimates served as expert level performance as a comparison for the proposed model.

D. Topographic Display of Important Features

A topographic display was created by generating saliency maps for the model using craniofacial scans. A saliency map is a visualization technique based on the gradient of the network output with respect to an input image [44]. Consequently, the pixels which contribute most to the prediction of the network can be highlighted. For the topographic displays, we averaged saliency maps for 10 subjects with the highest predicted AHI values per fold and 10 subjects with the lowest predicted AHI values per fold, yielding an average of 100 saliency maps for the highest and lowest predicted AHI values, respectively.

IV. RESULTS

A. Performance of Multi-view CNNs

The model using craniofacial scans during cross-validation converged after 6.6 ± 1.6 epochs, the model using demographics converged after 23.1 ± 6.9 epochs, and the model using questionnaires converged after 12.5 ± 6.9 epochs.

Without demographic and questionnaire information available, our model achieved a MAE of 11.38 ± 1.36 events/hour and a PCC of 0.40 ± 0.04 using 10-fold cross-validation. Fig. 3 shows the Bland-Altman plot of true and predicted AHI values. When dividing subjects from the test set into non-OSA/OSA using $AHI \geq 15$ events/hour as a criterion for OSA, an overall accuracy of $67 \pm 4\%$ was obtained. Sensitivity was $59 \pm 8\%$, specificity was $72 \pm 5\%$, and area under the receiver operating characteristic (AUC ROC) was $65 \pm 4\%$. Fig. 4 (a) shows the resulting confusion matrix of the classification task. Cohen's kappa coefficient was 0.29 for the classification.

Adding clinically relevant demographics and questionnaire scores to the model yielded a MAE of 11.05 ± 1.40 and a PCC of 0.45 ± 0.04 . Fig. 3 shows the Bland-Altman plot of the true and predicted AHI values. An accuracy of $67 \pm 4\%$ was achieved, with sensitivity of $74 \pm 7\%$, specificity of $63 \pm 7\%$, and AUC ROC of $69 \pm 3\%$. Fig. 4 (b) shows the resulting confusion matrix of the classification task. Cohen's kappa coefficient was 0.34 for the classification.

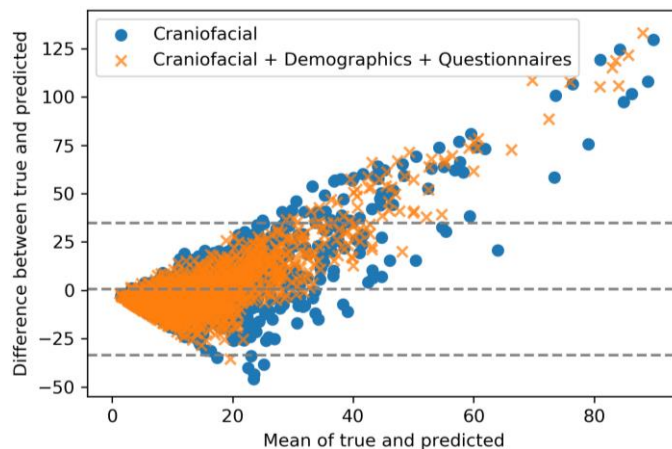


Fig. 3. Bland-Altman plot for the model using only craniofacial images and for the model using a combination of craniofacial images, demographics, and questionnaire scores. Means of true and predicted AHI values are displayed on the abscissa axis and difference between true and predicted AHI on the ordinate axis. The dashed horizontal lines above and below 0 indicate the borders of the 95% confidence interval.

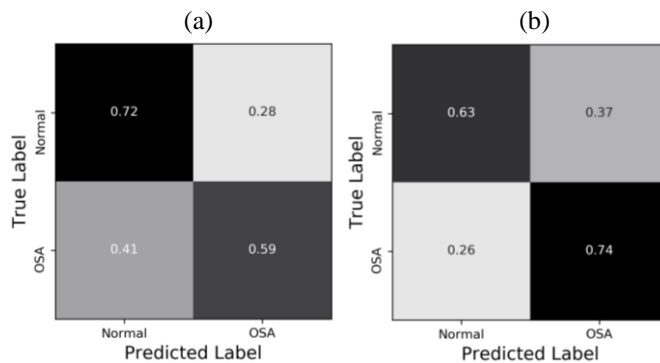


Fig. 4. (a) Confusion matrix for the model using only craniofacial images. (b) Confusion matrix for the model using a combination of craniofacial images, demographics, and questionnaire scores. The confusion matrix shows the results from classifying subjects into non-OSA/OSA using $AHI \geq 15$ as a criterion for OSA based on the predicted AHI values by the proposed models.

Table 3 compares model performance to similar work in the literature. Performance measures for comparisons are MAE, PCC, accuracy, AUC ROC, and number of subjects used in each study. As an additional comparison, Table 4 compares accuracies obtained using different methods on the same dataset, i.e. model with only demographics, or deriving a diagnosis from the modified STOP-Bang questionnaire, or a combination of both.

B. Performance of Sleep Medicine Specialists

Table 5 summarizes and compares results from the CNN to those of three sleep medicine specialists predicting AHI values based on 3D craniofacial scan displays from the test set. The percentage agreement between the three specialists was 67%. Fig. 5 compares the Bland-Altman plot of true and predicted AHIs from the model using craniofacial scans with a Bland-Altman plot of true and sleep specialists' AHIs.

TABLE 3. Comparison between the two model performances and performance achieved in the literature. MAE – Mean absolute error, PCC – Pearson correlation coefficient, AUC ROC – Area under the receiver operating characteristics curve, N – Total number of subjects in the study, N Test – Number of subjects used for testing, AMM - Active appearance model, SVR – Support vector regression, SVM – Support vector machine.

| Predictor | MAE | PCC | Accuracy | AUC ROC | N | N Test | Method | Validation scheme |
|--|-------|------|----------|---------|------|--------|---|--------------------------------|
| Craniofacial | 11.38 | 0.40 | 67% | 65% | 1366 | 1366 | Automatic landmarks with CNN. AHI prediction and OSA classification using 2D images and depth maps from five angles and multi-view CNN. | 10-fold cross-validation |
| Craniofacial + demographics + questionnaires | 11.05 | 0.45 | 67% | 69% | 1366 | 1366 | Same approach as above but with an ensemble of models using scans, demographics, and questionnaires, respectively. | 10-fold cross-validation |
| Espinoza-Cuadros et al. [29] | 12.56 | 0.37 | 71% | 67% | 285 | 285 | Automatic landmarks with AMM. AHI prediction using measurements and SVR. | Leave-one-out cross-validation |
| Nosrati et al. [30] | 13.4 | 0.52 | 68% | 75% | 180 | 180 | Manual landmarks. AHI prediction using measurements and SVR. | Leave-one-out cross-validation |
| Balaei et al. [45] | - | - | 69% | - | 376 | 204 | Automatic landmarks with SVM and cascade regression. OSA classification using measurements and logistic regression. | Training-test-set |
| Islam et al. [33] | - | - | 67% | - | 69 | 14 | OSA classification using frontal depth maps and pre-trained VGGFace. | Training-validation-test-set |
| Islam et al. [33] on our dataset | - | - | 60% | 64% | 1366 | 1366 | Same approach as above. | 10-fold cross-validation |

TABLE 4. Comparison of performance measures obtained using different variables to train and test neural networks using an ensemble approach in case of two or more modalities. The first row is obtained by using the modified STOP-Bang questionnaire as it is utilized clinically to screen for OSA, i.e. without machine learning MAE – Mean absolute error, PCC – Pearson Correlation Coefficient, AUC ROC – Area under the receiver operating characteristics curve.

| Predictor | MAE | PCC | Accuracy | AUC ROC |
|--|------------|-----------|----------|---------|
| Questionnaires | - | - | 62±4% | 65±4% |
| Questionnaire variables | 11.42±1.27 | 0.38±0.07 | 64±4% | 66±4% |
| Demographics | 11.35±1.26 | 0.40±0.06 | 64±4% | 67±3% |
| Scans | 11.38±1.36 | 0.40±0.04 | 67±4% | 65±4% |
| Demographics + Questionnaire variables | 11.24±1.28 | 0.41±0.06 | 65±4% | 67±4% |
| Scans + Demographics | 11.12±1.36 | 0.44±0.03 | 67±3% | 67±4% |
| Scans + Questionnaire variables | 11.03±1.40 | 0.45±0.04 | 67±4% | 68±4% |
| All Combined | 11.05±1.36 | 0.45±0.04 | 67±4% | 69±3% |

TABLE 5. Comparison of the main performance measures between the proposed models and three sleep medicine specialists. MAE – Mean absolute error, PCC – Pearson Correlation Coefficient, AUC ROC – Area under the receiver operating characteristics curve.

| Predictor | MAE | PCC | Accuracy | AUC ROC |
|---------------------------|------------|-----------|----------|---------|
| Craniofacial | 11.38±1.36 | 0.40±0.04 | 67±4% | 65±4% |
| Specialists data combined | 13.34±1.39 | 0.35±0.12 | 66±4% | 66±4% |
| Specialist 1 | 13.39±1.67 | 0.43±0.11 | 68±5% | 66±4% |
| Specialist 2 | 14.33±0.69 | 0.42±0.04 | 61±2% | 65±1% |
| Specialist 3 | 12.08±1.81 | 0.53±0.17 | 69±8% | 72±8% |

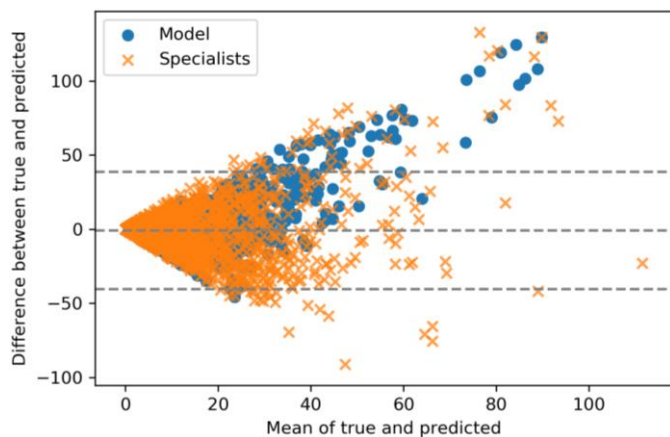


Fig. 5. Bland-Altman plot of the craniofacial model predictions compared to the sleep medicine specialists' predictions.

C. Topographic Display of Important Features

Fig. 6 (a) shows saliency maps averaged over the 100 subjects with the highest predicted AHI values, whereas Fig. 6 (b) shows saliency maps averaged over the 100 subjects with the lowest predicted AHI values.

V. DISCUSSION

Our ML model based on craniofacial images alone achieved a MAE of 11.38 events/hour, a PCC of 0.40, and an overall accuracy of 67%. In comparison, if AHI for all subjects was predicted as the mean AHI value of the dataset (i.e. 15.5 events/hour), the MAE would be 13.0 events/hour, the PCC would be -0.02, accuracy would be 34%, and AUC ROC would be 50%. This means that the average absolute deviation from true AHI per subject would be almost 2 events/hour more than the proposed model, though both correlation and accuracy would be significantly worse.

Adding the clinically relevant demographics and questionnaire scores improved the model further, yielding a MAE of 11.05, a PCC of 0.45, and an accuracy of 67%. Importantly, however, sensitivity increased (from 59% to 74%), while specificity decreased (from 72% to 63%), suggesting that the model became better at predicting subjects with OSA with the added information. Although the overall accuracy did not increase, the AUC ROC improved after introducing demographics and questionnaire scores into the model. Although this performance may appear modest, both models achieved a higher accuracy than the 62% obtained in the same dataset, respectively, using the modified STOP-Bang questionnaire. Given that questionnaires are regularly used as an early screening tool for OSA, it is encouraging to observe that accuracy yielded using craniofacial scans exceeds that of the questionnaires, with added ability to provide an estimate of disease severity (i.e., AHI), which is not possible with a simple screening questionnaire.

The Bland-Altman plots shown in Fig. 3 show significant underpredictions for subjects with very high AHI (>30 events/hour). In general, subjects with high AHI are hard to estimate, which is evident when looking at the sleep

specialists' scorings in Fig. 4, as they consistently underpredicted the higher AHI values as well. In OSA context, AHI values of 30, 60, or 90 events/hour are not significantly different, since they are all considered abnormal values. There were more than 200 subjects with AHI above 30 events/hour and 12 subjects with AHI more than 100 events/hour. Additionally, 50 subjects had an AHI of 0 events/hour, which were overpredicted on average by 10 events/hour. The subjects with very high AHI and AHI of 0 events/hour contribute to the high standard deviation (19.3) of AHI values in the dataset and increase the MAE as well.

Compared to similar work, the proposed models are at a similar level to all other studies in terms of overall accuracy (Table 3). However, our models were trained and validated on a much larger cohort collected at 11 different sleep clinics and used a very different approach than that of others that predicted AHI using landmark-based measured features [29], [30], [45]. A lot of manual work is needed to derive landmark-based, hand-selected features as opposed to using an entirely data-driven approach as we propose. As such, our study is reassuring in that the empirically identified features emphasized by the model, recapitulated clinical expertise without the manual labor or years of clinical training and experience. Only Islam et al. [33] used images directly in a data-driven manner, but these authors only used depth information and only had craniofacial scans for 69 subjects as opposed to our 1366 subjects, which is equivalent to the patient volume seen over the entire course of a clinical sleep medicine training fellowship. We implemented the algorithm proposed by Islam et al. [33] on our dataset as seen in Table 3, which decreased the overall accuracy from 67% to 60%. This makes sense because our dataset is much larger and much more diverse, since it was collected at many different sites. Furthermore, it shows that using images from several angles holds an advantage over using only frontal depth maps when predicting OSA, even when a pretrained network which has been trained on more than two million general facial images is utilized.

When comparing model performance to that of three sleep specialists guessing the AHI (Table 5), it was observed that the model was almost at the same level as two specialists and better than one in terms of overall accuracy. Fig. 5 also shows that even though both our model and sleep specialists make large underpredictions for highest AHI values, the model does not make large overpredictions in the same manner as sleep specialists. All overpredicted values by the model are within the confidence interval which most likely stems from a bias in the model towards people with low to moderate AHI values, i.e. 5-30 events/hour. Of note, however, the most underpredicted values are underpredicted by sleep specialists by a similar magnitude. The percentage agreement between the specialists was 67%, again highlighting the presence of significant interscorer variability within the field of sleep medicine.

The topographic display of Fig. 6 (a) shows that the network focuses mainly on the neck, jaw, and midface area when predicting high AHI values. These exact same regions have been reported in the literature as being the most important facial features related to OSA [8] and also the same

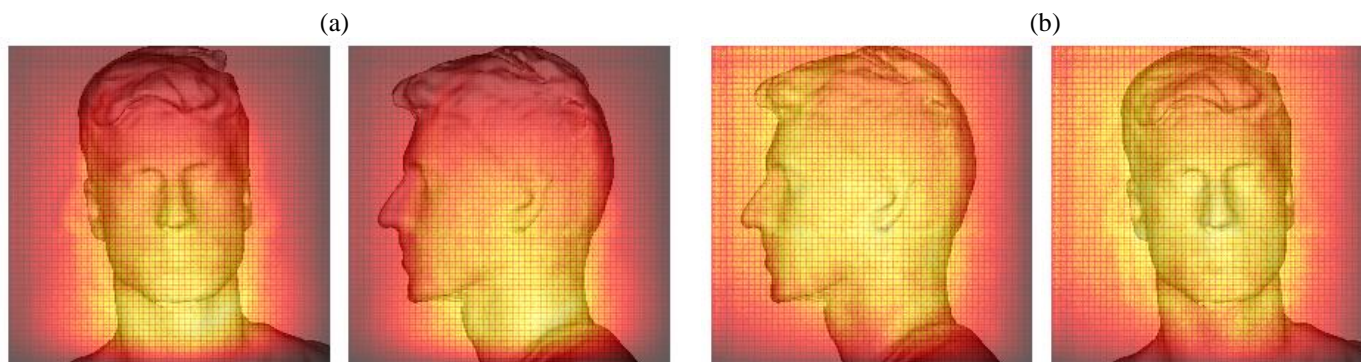


Fig. 6. (a) Saliency maps averaged over 100 subjects with the highest predicted AHI values. (b) Saliency maps averaged over 100 subjects with the lowest predicted AHI values.

regions that sleep specialists focused on when predicting AHI values. Interestingly however, when the network predicts low AHI values, as shown in Fig. 6 (b), it seems to focus more selectively on regions of the craniofacial complex that reflect skeletal anatomy (e.g. the maxilla and mandible) for predictions. This may reflect the fact that subjects with milder AHI values have a different pathophysiology where skeletal abnormalities more than body fat may be causing airflow limitations. Even though our proposed model obtains similar performance compared to sleep specialists and similar work in the literature, an average absolute error of more than 11 events/hour is still quite high. First and foremost, it is important to keep in mind that the pathology of OSA is not exclusively attributed to obesity and craniofacial factors. That is also why performance increases only moderately when adding extra information such as demographics and questionnaires. Other variables that could not be assessed from our dataset are internal upper airway anatomy factors, such as size and positioning of the tongue and palate. Other reasons for uncertainty may be physiological changes that occur with age, independent of facial anatomy or obesity such as recruitment of upper airway dilator muscles. Nonetheless, even if the features identified by the model are primarily anatomic in nature, these findings may prove useful to determine phenotypic risk for certain types of OSA. Finally, it should be noted that because our 3D scans were of structural facial features, we intentionally excluded central/mixed apneas from the AHI, in order to focus our model on the anatomical contributions to OSA. However, there are various models of this complex disorder that also account for the physiologic aspects of OSA (e.g., loop gain [46]), which is something that future modeling efforts should take into account.

Another limitation of this study was the quality and quantity of the captured 3D scans. The quality of the scans varied significantly and reflected the fact that they were captured in many different sleep clinics. Some scans had missing parts of the neck, whereas others were affected by poor lighting conditions. Furthermore, we believe that the size of the dataset was too small to truly capture the variation in craniofacial features across humans in relation to OSA in a data-driven manner. Evidently, we observe that similar performance is obtainable in smaller datasets if the features are hand-crafted like landmark-based measurements. The fact that scans were captured either at night before the PSG or in the morning after

did not have any effect on diagnostic performance, which was evident when we obtained accuracies of 68% and 66% for scans captured in the morning and at night, respectively, showing that scans can be obtained in both conditions.

Although the focus in this study is on providing a fast, efficient, and cheap screening tool for OSA, efforts to explore alternative screening methods include sleep tests at home [47], usually with very few sensors, such as sound [48], [49] and blood oxygen saturation [50], [51]. Potential of depth and thermal cameras has also been explored in breathing monitoring at an early stage [52], [53]. Even contactless bed sensors have been proposed, although only with moderate success so far [54]. The benefit of using the mentioned approaches is that the person is more comfortable sleeping in their own home wearing few or no wires. However, most studies use a small number of patients to validate their techniques and still requires a full night's sleep to reach a diagnosis.

In future work, it would be interesting to explore if, beside AHI, other clinically important variables captured by sleep studies could be better predicted. These could prove to not only make up a more accurate model but could also improve our knowledge of OSA phenotypes and their relation to facial anatomy. For example, Azarbarzin et al. [55] recently suggested that the hypoxic burden is a better measure to use compared to AHI when evaluating sleep apnea severity and resulting cardiovascular risk. Predicting hypoxic burden, oxygen desaturation index (ODI), or the duration of events instead of the AHI, or newer derivatives that better describe sleep disorder breathing heterogeneity could prove more useful and insightful.

VI. CONCLUSION

The main purpose of this study was to develop an automatic algorithm predicting AHI based on 3D craniofacial images that can be captured in a minute by a non-specialist. This was achieved by converting the 3D images into a series of 2D images and depth maps and utilizing a multi-view CNN for learning. Two models were implemented, one based exclusively on craniofacial images and one using a combination of craniofacial images, demographics, and questionnaire variables. Using these models, a MAE of 11.38 events/hour was obtained with an accuracy of 67%, which was at a similar level to performance achieved in similar work and

higher than the current screening method, i.e. questionnaires. A topographic display was created, highlighting the most important regions of the face when predicting OSA in a data-driven manner. These regions corresponded well with what is reported in the medical literature. The obtained results were at a similar level to two sleep specialists imitating the task and better than one. These sleep specialists had an overall agreement of 67% when scoring the scans. With this work, we have shown that it is possible to derive AHI values based on 3D scans and deep learning techniques for some OSA phenotypes and on a level similar to or higher than vastly experienced physicians. The proposed model has the potential to serve as a clinical screening tool for suspected OSA patients before they undergo a PSG.

ACKNOWLEDGEMENT

We would like to thank all clinic research coordinators who helped collect data as part of STAGES from the following sites: Geisinger Medical Center, Mayo Clinic, MedSleep, SleepMed, St. Luke's Hospital, and Stanford University. We would also like to thank Peter Yoo (peter@ugo3d.com), Sungwook Su (su@ugo3d.com), and Terry Izumi (terry@ugo3d.com) from uGo3D for developing software for the collection of 3D scans.

This work was supported by a grant from the Klarman Family Foundation. Additional funding was from the Technical University of Denmark, and the Danish Center for Sleep Medicine. Mr. Hanif's stay at Stanford University was funded by Danmark-Amerika Fondet, Vera og Carl Johan Michaelsens Legat, Reinholdt W. Jorck og Hustrus Fond, Torben og Alice Frimodts Fond, Christian og Otilia Brorsons Rejselegat, Marie og M.B. Richters Fond, Oberstløjtnant Max Nørgaard og hustru Magda Nørgaards Legat, William Demant Fonden, Augustinus Fonden, Rudolph Als Fondet, Knud Højgaard's Fond, Otto Mønsteds Fond, Julie von Müllens Fond, and Direktør Einar Hansen og hustru fru Vera Hansens Fond. The funding institutes played no role in the design and conduct of the study; no role in the collection, management, analysis, or interpretation of the data; and no role in the preparation, review, or approval of the manuscript.

Dr. Leary is a full-time employee of Jazz Pharmaceuticals, who, during this employment, has received stock options exercisable for, and other stock awards of, ordinary shares of Jazz Pharmaceuticals, plc.

Dr. Schneider reports other support from Sierra Pacific (VISN21) VA Mental Illness Research Education Clinical, Centers of Excellence, during the conduct of the study; personal fees from Alphabet, Inc, personal fees from Jazz Pharmaceuticals, outside the submitted work.

Dr. Morse reports personal fees from Jazz Pharmaceuticals and Rhythm Pharmaceuticals unrelated to this work. She has received research or clinical trial funding from NIH/NIMH and Jazz Pharmaceuticals unrelated to this work.

Dr. Mignot has received research or clinical trial funding unrelated to this work from Jazz Pharmaceuticals, Merck, Takeda, Sunovion and Apple. He owns stocks from Alerion, Dreem, Innoxia and Orexia, whose work is unrelated to this study, and is on an advisory board for Phillips.

The remaining authors have no conflicts of interest to declare.

APPENDIX

Figure A.1. shows the distribution of AHI values within the dataset, whereas Fig. A.2. example of the basic components of a mesh, i.e. vertices, edges, and triangles.

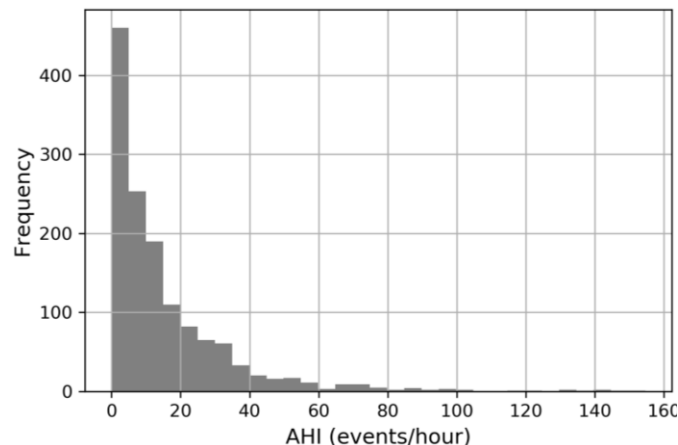


Fig. A.1. The distribution of apnea-hypopnea index (AHI) values in the dataset.

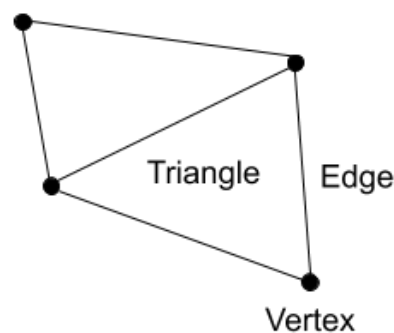


Fig. A.2. Example of four vertices that have been connected using edges, forming two triangles, which are the basic blocks of a 3D mesh surface.

REFERENCES

- [1] P. Lévy *et al.*, "Obstructive Sleep Apnoea Syndrome," *Nat. Rev. Dis. Prim.*, vol. 1, no. 1, pp. 1–21, 2015, doi: <https://doi.org/10.1038/nrdp.2015.43>.
- [2] L. F. Drager, S. M. Togeiro, V. Y. Polotsky, and G. Lorenzi-Filho, "Obstructive Sleep Apnea," *J. Am. Coll. Cardiol.*, vol. 62, no. 7, pp. 569–576, 2013, doi: <https://doi.org/10.1016/j.jacc.2013.05.045>.
- [3] A. S. M. Shamsuzzaman, B. J. Gersh, and V. K. Somers, "Obstructive Sleep Apnea Implications for Cardiac and Vascular Disease," *Jama*, vol. 290, no. 14, pp. 1906–1914, 2003, doi: <https://doi.org/10.1001/jama.290.14.1906>.
- [4] R. Heinzer *et al.*, "Prevalence of Sleep-Disordered Breathing in the General Population: the HypnoLaus Study," *Lancet Respir. Med.*, vol. 3, no. 4, pp. 310–318, 2015, doi: [https://doi.org/10.1016/S2213-2600\(15\)00043-0](https://doi.org/10.1016/S2213-2600(15)00043-0).
- [5] A. M. Osman, S. G. Carter, J. C. Carberry, and D. J. Eckert, "Obstructive sleep apnea: current perspectives," *Nat. Sci. Sleep*, vol. 10, pp. 21–34, 2018, doi: <https://doi.org/10.2147/NSS.S124657>.
- [6] A. R. Schwartz, S. P. Patil, A. M. Laffan, V. Polotsky, H. Schneider, and P. L. Smith, "Obesity and Obstructive Sleep Apnea: Pathogenic Mechanisms and Therapeutic Approaches," *Proc. Am. Thorac. Soc.*, vol. 5, no. 2, pp. 185–192, Feb. 2008, doi: <https://doi.org/10.1513/pats.200708-137MG>.
- [7] J. A. Dempsey, S. C. Veasey, B. J. Morgan, and C. P. O'donnell, "Pathophysiology of Sleep Apnea," *Physiol. Rev.*, vol. 90, no. 1, pp. 47–112, 2010, doi: <https://doi.org/10.1152/physrev.00043.2008>.
- [8] R. W. W. Lee, K. Sutherland, and P. A. Cistulli, "Craniofacial

Morphology in Obstructive Sleep Apnea: A Review,” *Clin. Pulm. Med.*, vol. 17, no. 4, pp. 189–195, 2010, doi: <https://doi.org/10.1097/CPM.0b013e3181e4bea7>.

[9] C. A. Kushida *et al.*, “Practice Parameters for the Indications for Polysomnography and Related Procedures: An Update for 2005,” *Sleep*, vol. 28, no. 4, pp. 499–523, 2005, doi: [10.1093/sleep/28.4.499](https://doi.org/10.1093/sleep/28.4.499).

[10] R. B. Berry *et al.*, “The AASM Manual for the Scoring of Sleep and Associated Events: Rules, Terminology, and Technical Specifications Version 2.2,” *Am. Acad. Sleep Med.*, 2015, [Online]. Available: www.aasmnet.org.

[11] N. Collop, “Scoring Variability between Polysomnography Technologists in Different Sleep Laboratories,” *Sleep Med.*, vol. 3, no. 1, pp. 43–47, 2002, doi: [https://doi.org/10.1016/S1389-9457\(01\)00115-0](https://doi.org/10.1016/S1389-9457(01)00115-0).

[12] M. Younes *et al.*, “Reliability of the American Academy of Sleep Medicine Rules for Assessing Sleep Depth in Clinical Practice,” *J. Clin. Sleep Med.*, vol. 14, no. 2, pp. 205–213, 2018, doi: <https://doi.org/10.5664/jcs.m.6934>.

[13] J. B. Stephansen *et al.*, “Neural Network Analysis of Sleep Stages Enables Efficient Diagnosis of Narcolepsy,” *Nat. Commun.*, vol. 9, no. 1, pp. 1–15, Dec. 2018, doi: <https://doi.org/10.1038/s41467-018-07229-3>.

[14] U. Hanif *et al.*, “Non-invasive Machine Learning Estimation of Effort Differentiates Sleep-disordered Breathing Pathology,” *Physiol. Meas.*, vol. 40, no. 2, p. 025008, Feb. 2019, doi: [10.1088/1361-6579/ab0559](https://doi.org/10.1088/1361-6579/ab0559).

[15] A. Brink-Kjaer *et al.*, “Automatic Detection of Cortical Arousals in Sleep and Their Contribution to Daytime Sleepiness,” *Clin. Neurophysiol.*, vol. 131, no. 6, pp. 1187–1203, Jun. 2020, doi: <https://doi.org/10.1016/j.clinph.2020.02.027>.

[16] M. Piriyaajitakonkij *et al.*, “SleepPoseNet: Multi-View Learning for Sleep Postural Transition Recognition Using UWB,” *IEEE J. Biomed. Heal. Informatics*, pp. 1–1, 2020, doi: [10.1109/JBHI.2020.3025900](https://doi.org/10.1109/JBHI.2020.3025900).

[17] N. Banluesombatkul *et al.*, “MetaSleepLearner: A Pilot Study on Fast Adaptation of Bio-signals-Based Sleep Stage Classifier to New Individual Subject Using Meta-Learning,” *IEEE J. Biomed. Heal. Informatics*, 2020, doi: [10.1109/JBHI.2020.3037693](https://doi.org/10.1109/JBHI.2020.3037693).

[18] R. J. Schwab *et al.*, “Digital Morphometrics: a New Upper Airway Phenotyping Paradigm in OSA,” *Chest*, vol. 152, no. 2, pp. 330–342, 2017, doi: <https://doi.org/10.1016/j.chest.2017.05.005>.

[19] C. Guilleminault, R. Riley, and N. Powell, “Obstructive Sleep Apnea and Abnormal Cephalometric Measurements: Implications for Treatment,” *Chest*, vol. 86, no. 5, pp. 793–794, 1984, doi: <https://doi.org/10.1378/chest.86.5.793>.

[20] A. Jamieson, C. Guilleminault, M. Partinen, and M. A. Quera-Salva, “Obstructive Sleep Apneic Patients Have Craniomandibular Abnormalities,” *Sleep*, vol. 9, no. 4, pp. 469–477, 1986, doi: <https://doi.org/10.1093/sleep/9.4.469>.

[21] A. A. Lowe, J. A. Fleetham, S. Adachi, and C. F. Ryan, “Cephalometric and Computed Tomographic Predictors of Obstructive Sleep Apnea Severity,” *Am. J. Orthod. Dentofac. Orthop.*, vol. 107, no. 6, pp. 589–595, 1995, doi: [https://doi.org/10.1016/s0889-5406\(95\)70101-x](https://doi.org/10.1016/s0889-5406(95)70101-x).

[22] T. Ogawa, R. Enciso, W. H. Shintaku, and G. T. Clark, “Evaluation of Cross-Section Airway Configuration of Obstructive Sleep Apnea,” *Oral Surgery, Oral Med. Oral Pathol. Oral Radiol. Endodontology*, vol. 103, no. 1, pp. 102–108, 2007, doi: <https://doi.org/10.1016/j.tripleo.2006.06.008>.

[23] R. J. Schwab *et al.*, “Identification of Upper Airway Anatomic Risk Factors for Obstructive Sleep Apnea with Volumetric Magnetic Resonance Imaging,” *Am. J. Respir. Crit. Care Med.*, vol. 168, no. 5, pp. 522–530, 2003, doi: <https://doi.org/10.1164/rccm.200208-866OC>.

[24] M. Okubo *et al.*, “Morphologic Analyses of Mandible and Upper Airway Soft Tissue by MRI of Patients with Obstructive Sleep Apnea Hypopnea Syndrome,” *Sleep*, vol. 29, no. 7, pp. 909–915, 2006, doi: <https://doi.org/10.1093/sleep/29.7.909>.

[25] R. W. W. Lee *et al.*, “Relationship between Surface Facial Dimensions and Upper Airway Structures in Obstructive Sleep Apnea,” *Sleep*, vol. 33, no. 9, pp. 1249–1254, 2010, doi: <https://doi.org/10.1093/sleep/33.9.1249>.

[26] E. J. Kezirian, W. Hohenhorst, and N. de Vries, “Drug-Induced Sleep Endoscopy: the VOTE Classification,” *Eur. Arch. Oto-Rhino-Laryngology*, vol. 268, no. 8, pp. 1233–1236, 2011.

[27] R. W. W. Lee, A. S. L. Chan, R. R. Grunstein, and P. A. Cistulli, “Craniofacial Phenotyping in Obstructive Sleep Apnea—A Novel Quantitative Photographic Approach,” *Sleep*, vol. 32, no. 1, pp. 37–45, 2009, doi: <https://doi.org/10.5665/sleep/32.1.37>.

[28] R. W. W. Lee, P. Petocz, T. Prvan, A. S. L. 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, doi: <https://doi.org/10.5665/sleep/32.1.46>.

[29] F. Espinoza-Cuadros, R. Fernández-Pozo, D. T. Toledano, J. D. Alcázar-Ramírez, E. López-Gonzalo, and L. A. Hernández-Gómez, “Speech Signal and Facial Image Processing for Obstructive Sleep Apnea Assessment,” *Comput. Math. Methods Med.*, vol. 2015, 2015, doi: <https://doi.org/10.1155/2015/489761>.

[30] H. Nosrati, N. Sadr, and P. de Chazal, “Apnoea-Hypopnoea Index Estimation using Craniofacial Photographic Measurements,” *CinC 2016, IEEE*, pp. 1033–1036, 2016.

[31] P. De Chazal, A. T. Balaei, and H. Nosrati, “Screening Patients for Risk of Sleep Apnea using Facial Photographs,” *EMBC 2017, IEEE*, pp. 2006–2009, 2017, doi: <https://doi.org/10.1109/EMBC.2017.8037245>.

[32] A. T. Balaei, K. Sutherland, P. A. Cistulli, and P. de Chazal, “Automatic Detection of Obstructive Sleep Apnea using Facial Images,” *ISBI 2017, IEEE*, pp. 215–218, 2017, doi: <https://doi.org/10.1109/ISBI.2017.7950504>.

[33] S. M. S. Islam, H. Mahmood, A. A. Al-Jumaily, and S. Claxton, “Deep Learning of Facial Depth Maps for Obstructive Sleep Apnea Prediction,” *iCMLDE 2018, IEEE*, pp. 154–157, 2018, doi: <https://doi.org/10.1109/iCMLDE.2018.00036>.

[34] U. Hanif, R. R. Paulsen, E. B. Leary, E. Mignot, P. Jennum, and H. B. D. Sorensen, “Prediction of Patient Demographics using 3D Craniofacial Scans and Multi-view CNNs,” in *EMBC 2020, IEEE*, 2020, pp. 1950–1953, doi: <https://doi.org/10.1109/EMBC44109.2020.9176333>.

[35] M. Kalantari and M. Nechifor, “Accuracy and utility of the Structure Sensor for collecting 3D indoor information,” *Geo-spatial Inf. Sci.*, vol. 19, no. 3, pp. 202–209, Jul. 2016, doi: [10.1080/10095020.2016.1235817](https://doi.org/10.1080/10095020.2016.1235817).

[36] P. G. M. Knoops *et al.*, “Comparison of three-dimensional scanner systems for craniomaxillofacial imaging,” *J. Plast. Reconstr. Aesthetic Surg.*, vol. 70, no. 4, pp. 441–449, Apr. 2017, doi: [10.1016/j.bjps.2016.12.015](https://doi.org/10.1016/j.bjps.2016.12.015).

[37] F. Chung, H. R. Abdullah, and P. Liao, “STOP-Bang Questionnaire: A Practical Approach to Screen for Obstructive Sleep Apnea,” *Chest*, vol. 149, no. 3, pp. 631–638, 2016, doi: <https://doi.org/10.1378/chest.15-0903>.

[38] R. R. Paulsen, K. A. Juhl, T. M. Haspang, T. Hansen, M. Ganz, and G. Einarsson, “Multi-View Consensus CNN for 3D Facial Landmark Placement,” *Asian Conf. Comput. Vis.*, pp. 706–719, 2018, doi: https://doi.org/10.1007/978-3-030-20887-5_44.

[39] S. Zulqarnain Gilani, F. Shafait, and A. Mian, “Shape-Based Automatic Detection of a Large Number of 3D Facial Landmarks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4639–4648.

[40] C. M. Grewe and S. Zachow, “Fully Automated and Highly Accurate Dense Correspondence for Facial Surfaces,” in *European Conference on Computer Vision*, 2016, pp. 552–568.

[41] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” *CVPR 2016*, pp. 770–778, 2016, doi: <https://doi.org/10.1109/CVPR.2016.90>.

[42] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *arXiv Prepr. arXiv1412.6980 [cs.LG]*, 2014.

[43] D. G. Altman and J. M. Bland, “Measurement in Medicine: the Analysis of Method Comparison Studies,” *J. R. Stat. Soc. Ser. D (The Stat.)*, vol. 32, pp. 307–317, 1983, doi: <https://doi.org/10.2307/2987937>.

[44] K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps,” *arXiv Prepr. arXiv1312.6034 [cs.CV]*, 2013.

[45] A. T. Balaei, K. Sutherland, P. Cistulli, and P. de Chazal, “Prediction of Obstructive Sleep Apnea using Facial Landmarks,” *Physiol. Meas.*, vol. 39, p. 094004, 2018, doi: <https://doi.org/10.1088%2F1361-6579%2Faaadb35>.

[46] D. J. Eckert, D. P. White, A. S. Jordan, A. Malhotra, and A. Wellman, “Defining phenotypic causes of obstructive sleep apnea:

- Identification of novel therapeutic targets,” *Am. J. Respir. Crit. Care Med.*, vol. 188, no. 8, pp. 996–1004, Oct. 2013, doi: 10.1164/rccm.201303-0448OC.
- [47] F. Mendonça, S. S. Mostafa, A. G. Ravelo-García, F. Morgado-Dias, and T. Penzel, “Devices for home detection of obstructive sleep apnea: A review,” *Sleep Medicine Reviews*, vol. 41. W.B. Saunders Ltd, pp. 149–160, Oct. 01, 2018, doi: 10.1016/j.smr.2018.02.004.
- [48] H. Nakano, T. Furukawa, and T. Tanigawa, “Tracheal sound analysis using a deep neural network to detect sleep apnea,” *Journal of Clinical Sleep Medicine*, vol. 15, no. 8. American Academy of Sleep Medicine, pp. 1125–1133, Aug. 15, 2019, doi: 10.5664/jcsm.7804.
- [49] S. Hayashi, M. Tamaoka, T. Tateishi, Y. Murota, I. Handa, and Y. Miyazaki, “A New Feature with the Potential to Detect the Severity of Obstructive Sleep Apnoea via Snoring Sound Analysis,” *mdpi.com*, doi: 10.3390/ijerph17082951.
- [50] S. Nikkonen, I. O. Afara, T. Leppänen, and J. Töyräs, “Artificial neural network analysis of the oxygen saturation signal enables accurate diagnostics of sleep apnea,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Dec. 2019, doi: 10.1038/s41598-019-49330-7.
- [51] F. Mendonça, S. Mostafa, F. Morgado-Dias, and A. G. Ravelo-García, “An Oximetry Based Wireless Device for Sleep Apnea Detection,” *mdpi.com*, doi: 10.3390/s20030888.
- [52] A. Al-Naji, A. J. Al-Askery, S. K. Gharghan, and J. Chahl, “A System for Monitoring Breathing Activity Using an Ultrasonic Radar Detection with Low Power Consumption,” *J. Sens. Actuator Networks*, vol. 8, no. 2, p. 32, May 2019, doi: 10.3390/jsan8020032.
- [53] A. Procházka, M. Schätz, ... O. T. ... on I. P., and undefined 2014, “The MS Kinect image and depth sensors use for gait features detection,” *ieeexplore.ieee.org*, Accessed: Feb. 05, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7025460/?casa_token=UE_j-wXrA4UAAAAA:PMTN_gm81suiUV9yJGekUC5i1hGjwqKYe77-8bEVa9S11EI6JKZnoJzOWlte5Oj1dd5cFVAvohU.
- [54] I. Sadek, T. Tan Soon Heng, E. Seet, and B. Abdulrazak, “A new approach for detecting sleep apnea using a contactless bed sensor: Comparison study,” *J. Med. Internet Res.*, vol. 22, no. 9, p. e18297, Sep. 2020, doi: 10.2196/18297.
- [55] A. Azarbarzin *et al.*, “The Hypoxic Burden of Sleep Apnoea Predicts Cardiovascular Disease-Related Mortality: the Osteoporotic Fractures in Men Study and the Sleep Heart Health Study,” *Eur. Heart J.*, vol. 40, pp. 1149–1157, 2019, doi: <https://doi.org/10.1093/eurheartj/ehy624>.