Downloaded from orbit.dtu.dk on: Apr 19, 2024

DTU Library

Automatic scoring of drug-induced sleep endoscopy for obstructive sleep apnea using deep learning

Hanif, Umaer; Kiaer, Eva Kirkegaard; Capasso, Robson; Liu, Stanley Y.; Mignot, Emmanuel J.M.; Sorensen, Helge B.D.; Jennum, Poul

Published in:
Sleep Medicine

Link to article, DOI:
10.1016/j.sleep.2022.12.015

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Hanif, U., Kiaer, E. K., Capasso, R., Liu, S. Y., Mignot, E. J. M., Sorensen, H. B. D., & Jennum, P. (2023). Automatic scoring of drug-induced sleep endoscopy for obstructive sleep apnea using deep learning. *Sleep Medicine*, *102*, 19-29. https://doi.org/10.1016/j.sleep.2022.12.015

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.

Sleep Medicine 102 (2023) 19–29

Contents lists available at ScienceDirect

Sleep Medicine

journal homepage: www.elsevier.com/locate/sleep

Automatic scoring of drug-induced sleep endoscopy for obstructive sleep apnea using deep learning

Umaer Hanif [a, d, e], Eva Kirkegaard Kiaer [b], Robson Capasso [c], Stanley Y. Liu [c], Emmanuel J.M. Mignot [d, 1], Helge B.D. Sorensen [a, 1], Poul Jennum [e, *, 1]

[a] Biomedical Signal Processing & AI Research Group, Department of Health Technology, Technical University of Denmark, Oersteds Plads 345B, 2800, Kongens Lyngby, Denmark
[b] Danish Center for Sleep Surgery, Department of Otorhinolaryngology, Head and Neck Surgery and Audiology, Copenhagen University Hospital (Rigshospitalet), Inge Lehmanns Vej 8, 2100, Copenhagen, Denmark
[c] Department of Otolaryngology/Head & Neck Surgery, Stanford University School of Medicine, 801 Welch Road, Palo Alto, CA, 94304, USA
[d] Stanford University Center for Sleep and Circadian Sciences, Stanford University, 3165 Porter Dr., CA, 94304, Palo Alto, USA
[e] Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Rigshospitalet, University of Copenhagen, Nordre Ringvej 57, 2600, Glostrup, Denmark

ARTICLE INFO

Article history:
Received 22 September 2022
Received in revised form
9 December 2022
Accepted 19 December 2022
Available online 20 December 2022

Keywords:
Obstructive sleep apnea
Drug-induced sleep endoscopy
vote classification
Deep learning

ABSTRACT

Background: Treatment of obstructive sleep apnea is crucial for long term health and reduced economic burden. For those considered for surgery, drug-induced sleep endoscopy (DISE) is a method to characterize location and pattern of sleep-related upper airway collapse. According to the VOTE classification system, four upper airway sites of collapse are characterized: velum (V), oropharynx (O), tongue (T), and epiglottis (E). The degree of obstruction per site is classified as 0 (no obstruction), 1 (partial obstruction), or 2 (complete obstruction). Here we propose a deep learning approach for automatic scoring of VOTE obstruction degrees from DISE videos.

Methods: We included 281 DISE videos with varying durations (6 s–16 min) from two sleep clinics: Copenhagen University Hospital and Stanford University Hospital. Examinations were split into 5-s clips, each receiving annotations of 0, 1, 2, or X (site not visible) for each site (V, O, T, and E), which was used to train a deep learning model. Predicted VOTE obstruction degrees per examination was obtained by taking the highest predicted degree per site across 5-s clips, which was evaluated against VOTE degrees annotated by surgeons.

Results: Mean F1 score of 70% was obtained across all DISE examinations (V: 85%, O: 72%, T: 57%, E: 65%). For each site, sensitivity was highest for degree 2 and lowest for degree 0. No bias in performance was observed between videos from different clinicians/hospitals.

Conclusions: This study demonstrates that automating scoring of DISE examinations show high validity and feasibility in degree of upper airway collapse.

© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Sources of support

This study 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 Ottilia 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 Højgaards 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

* Corresponding author. Danish Center for Sleep Medicine, Department of Clinical Neurophysiology, Nordre Ringvej 57, 2600, Glostrup, Denmark.
E-mail addresses: umaerhanif@hotmail.com (U. Hanif), eva.kirkegaard.kiaer.01@regionh.dk (E.K. Kiaer), rcapasso@stanford.edu (R. Capasso), ycliu@stanford.edu (S.Y. Liu), mignot@stanford.edu (E.J.M. Mignot), hbds@dtu.dk (H.B.D. Sorensen), poul.joergen.jennum@regionh.dk (P. Jennum).
[1] Shared last authors.

https://doi.org/10.1016/j.sleep.2022.12.015
1389-9457/© 2022 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

manuscript.

1. Introduction

Obstructive sleep apnea (OSA) is characterized by partial or complete obstruction of the upper airway during sleep, causing events with reduced airflow (hypopneas) or cessation of breathing (apneas) [1]. Presence of repeated apneas and hypopneas cause disturbances in sleep leading to daytime sleepiness [2–4], increased risk of cardiovascular diseases [5–7], motor vehicle accidents [8], and elevated mortality rates [9]. Prevalence of OSA is high; almost half a billion adults worldwide aged 30–69 years suffer from moderate to severe OSA [10]. The economic burden of undiagnosed or untreated OSA is $150 billion in the United States alone [11].

Obesity is the most frequent cause of upper airway narrowing due to presence of excess fat tissue in the tongue and around the neck area [12]. Other factors include enlarged tonsils, an anatomically narrow upper airway, the tongue falling backwards, and an underdeveloped or protracted jaw [13,14]. Continuous positive airway pressure (CPAP) [15] is the gold-standard treatment for OSA and although it is extremely effective, studies show that up to 50% of users give up on the device within a year of therapy due to various reasons [16].

For some patients, surgical procedures can be viable options to increase upper airway space, with the most common surgery being a modified uvulopalatopharyngoplasty, where excess tissue is removed from the soft palate and lateral walls of the pharynx, often combined with tonsillectomy (removal of the palatine tonsils) [17,18]. Other procedures include TORS (transoral robotic surgery) on the tongue base and epiglottis [19] and maxillomandibular advancement (advancement of the upper and lower jaw) [20]. Prior to surgery, drug-induced sleep endoscopy (DISE) is often performed to examine the location and pattern of sleep-related upper airway collapse using a fiberoptic endoscope under sedation, which is designed to simulate natural sleep [21]. The endoscope is introduced through the nasal cavity and examines the upper airway from the nares to the level of the glottis. After a DISE examination, the surgeon evaluates the sites of collapse in the upper airway according to the VOTE (velum, oropharynx lateral walls, tongue base, epiglottis) classification system, which is the most used scoring system for DISE [22]. Collapse can occur at these four sites, either individually or in combination, causing obstruction in the upper airway. The VOTE classification system assigns a degree of obstruction and pattern of collapse to each site where a collapse occurs as outlined in Table 1. VOTE obstruction degrees are classified either as 0 (no obstruction, <50%), 1 (partial obstruction, 50–75%), or 2 (complete obstruction, >75%) [22]. Additionally, there are three patterns of collapse: antero-posterior (A-P) collapse, lateral collapse, and concentric collapse [22]. Fig. 1 shows examples of V and Fig. 2 shows examples of O, T, and E with respect to obstruction degrees and collapse patterns.

Table 1
Upper airway sites where collapse can occur during sleep according to the VOTE (velum, oropharynx, tongue base, epiglottis) classification system. The degree of obstruction caused by collapse is either 0 (no collapse, <50%), 1 (partial obstruction, 50–75%), or 2 (complete obstruction, >75%). Checkmarks indicate the possible pattern of collapse at each site.

Site	Degree of obstruction	Pattern of collapse		
		Antero-posterior	Lateral	Concentric
Velum	0, 1, or 2	✓	✓	✓
Oropharynx			✓	
Tongue base		✓		
Epiglottis		✓	✓	

DISE suggests location and indication for surgical intervention and its use has been shown to improve OSA surgical outcomes [23]. The analysis however depends on the procedure and the evaluation hereof and as such presents interrater variability. First, there is an anatomical variation across subjects with respect to the upper airway, and the pattern of collapse may also be affected by the depth of sedation [24–28]. Secondly, DISE videos can appear chaotic due to several sites collapsing simultaneously in patients with severe OSA, essentially pushing the endoscope around and making it difficult to determine the sites where collapses are occurring. Mucus or saliva may also cover the endoscope, which can reduce or distort the video quality significantly and at times make it almost impossible to visually inspect the upper airway. Studies examining interscorer reliability between surgeons show poor to moderate agreement [23,29–33], demonstrating that despite a well-established classification system, interpretation remains subjective.

In this study, we hypothesize that a deep learning-based model predicting VOTE obstruction degrees from DISE videos automatically could be trained and would generalize across subjects and centers when validated on a large amount of DISE videos (+10 h of footage). Such a model could aid surgeons in the scoring of DISE videos and consequentially in the planning of surgical treatment. Deep learning techniques are chosen over other machine learning models because they allow for automatic and data-driven feature extraction from video frames through convolutional neural networks and context-based predictions through long short-term memory networks.

In a former smaller study, we evaluated the use of deep learning on DISE examinations for estimation of upper airway regions with promising results [34]. In this current study, we collect a much larger number of videos and design a model capable of predicting obstruction degree at each of the four different sites (V, O, T, and E), which is evaluated against gold-standard annotations provided by surgeons. To simplify the problem, we leave out pattern of collapse, which can affect treatment strategy at the level of the velum and lateral pharyngeal wall [35–38]. The reasoning is to ensure that the model can distinguish between no collapse/collapse and the level of obstruction at each upper airway site before attempting to differentiate collapse patterns. By demonstrating that sites of upper airway collapse and obstruction degrees can be detected accurately, which has not been attempted before, we take the first important step towards fully automatic estimation of VOTE scores from DISE and leave the inclusion of collapse pattern to a future study.

2. Methods

2.1. Data description

281 DISE videos were obtained in total from three different surgeons at two different locations: one surgeon from Copenhagen University Hospital (CUH) in Denmark (51 videos) and two surgeons at Stanford University Hospital (SUH) in California, USA (58 and 172 videos, respectively). The DISE examinations were performed for subjects with confirmed OSA in accordance with DISE procedure guidelines described by Kiaer et al. [39] and Lan et al. [40].

Each video was anonymized by 1) removing any part where the endoscope was outside of the patient's airway and 2) renaming the video file. Median duration of videos after anonymization was 2.1 min with an interquartile range of 3.33 min (min – max: 6 s–16.4 min) and the total amount of video footage was 13.7 h. Fig. 3 shows the distribution of DISE examination durations, showing that most videos in the dataset were less than 2 min long.

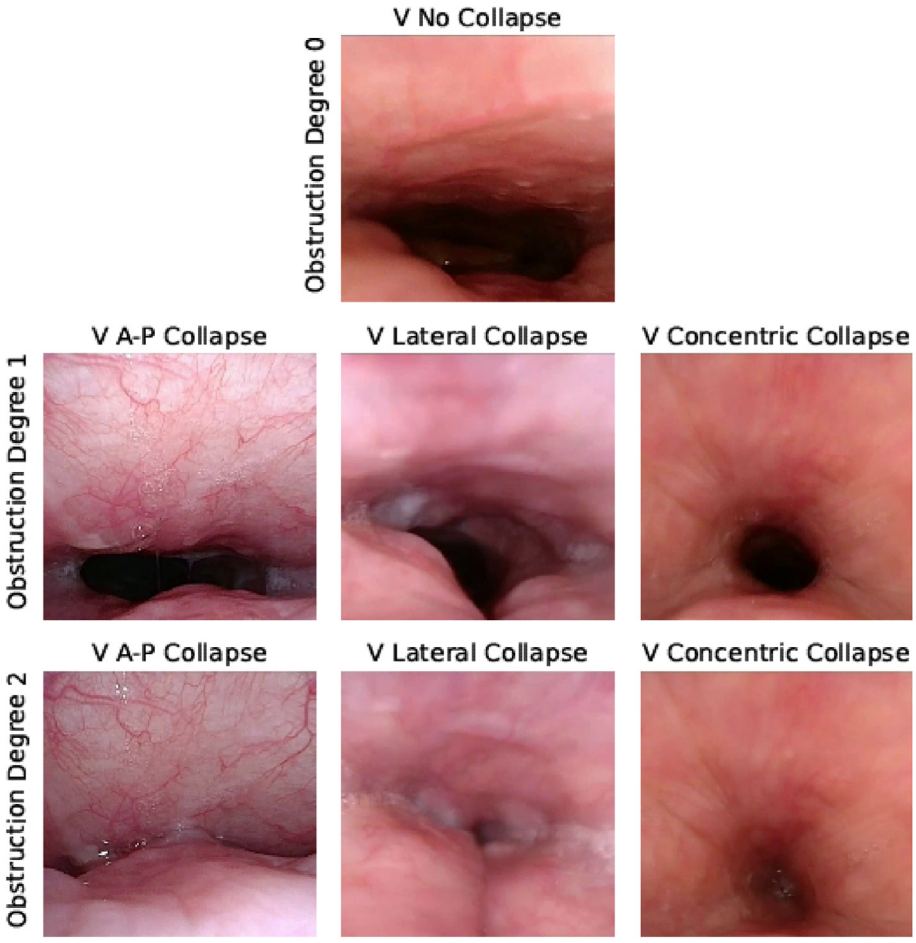

Fig. 1. Possible ways the velum (V) can collapse in the upper airway, either partially (obstruction degree 1) or completely (obstruction degree 2). The patterns of collapse are antero-posterior (A–P), lateral, or concentric.

Videos obtained from CUH had sampling rates of 25 frames per second, while videos from SUH had sampling rates of 30 frames per second.

For each examination, an annotation was obtained containing the VOTE score, i.e., obstruction degree and collapse pattern at each site as shown in Table 2. Note that several sites can collapse in the same subject (sometimes even in combination) and that a site like V can collapse in more than one way in the same subject. The distribution of obstruction degrees for each site is shown in Fig. 4. The institutions IRB approved the study under IRB-64418.

2.2. Pre-processing

Since DISE examinations varied greatly with respect to duration (Fig. 3) and there was only a one-line annotation per video (Table 2), we decided that using data as it was would be unsuitable for deep learning purposes. Consequently, all video examinations were split into 5-s clips, as shown in the top block of Fig. 5, and each clip received an annotation with respect to each site. These annotations were created in consultation with a chief surgeon in otorhinolaryngology at CUH (EKK).

5-s clips were selected due to two reasons: 1) Due to a limitation in computational resources, training the model using batches of 5-s clips was found to be feasible, while increasing the clip duration to e.g., 30 s would result in too high of a computational cost during model training; 2) The idea was for the model to be able to recognize individual collapses and not entire apneas, since this information can then subsequently be aggregated or summarized to provide the complete VOTE score using the highest observed degree of obstruction at each upper airway site.

Using 5-s clips, there are many scenarios where one or more sites are not visible. Thus, another class was introduced for such situations, denoted X, such that there were four classes (0, 1, 2, and X) for each site (V, O, T, and E) per 5-s clip, essentially amounting to a 16-class classification problem. The idea was to train and validate the proposed deep learning model on the 5-s clips and then summarize all predictions to form a single VOTE obstruction degree prediction for each DISE examination, which could then be evaluated against the surgeons' annotations.

Although sampling rates for videos from CUH and SUH were 25 and 30 frames per second, respectively, we used only every 5th and 6th frame, respectively, to reduce computational cost. Consequently, the sampling rates for 5-s clips used in the study were 5 frames per second, yielding a total of 25 frames for a 5-s clip. Instead of using all three color-channels for each clip (R,G,B), the videos were converted to grayscale. Both approaches were investigated (with and without color channels) and preserving colors did not make any noticeable difference, most likely because anatomical composition is much more important than small differences in color, so grayscale frames were chosen to reduce computational cost.

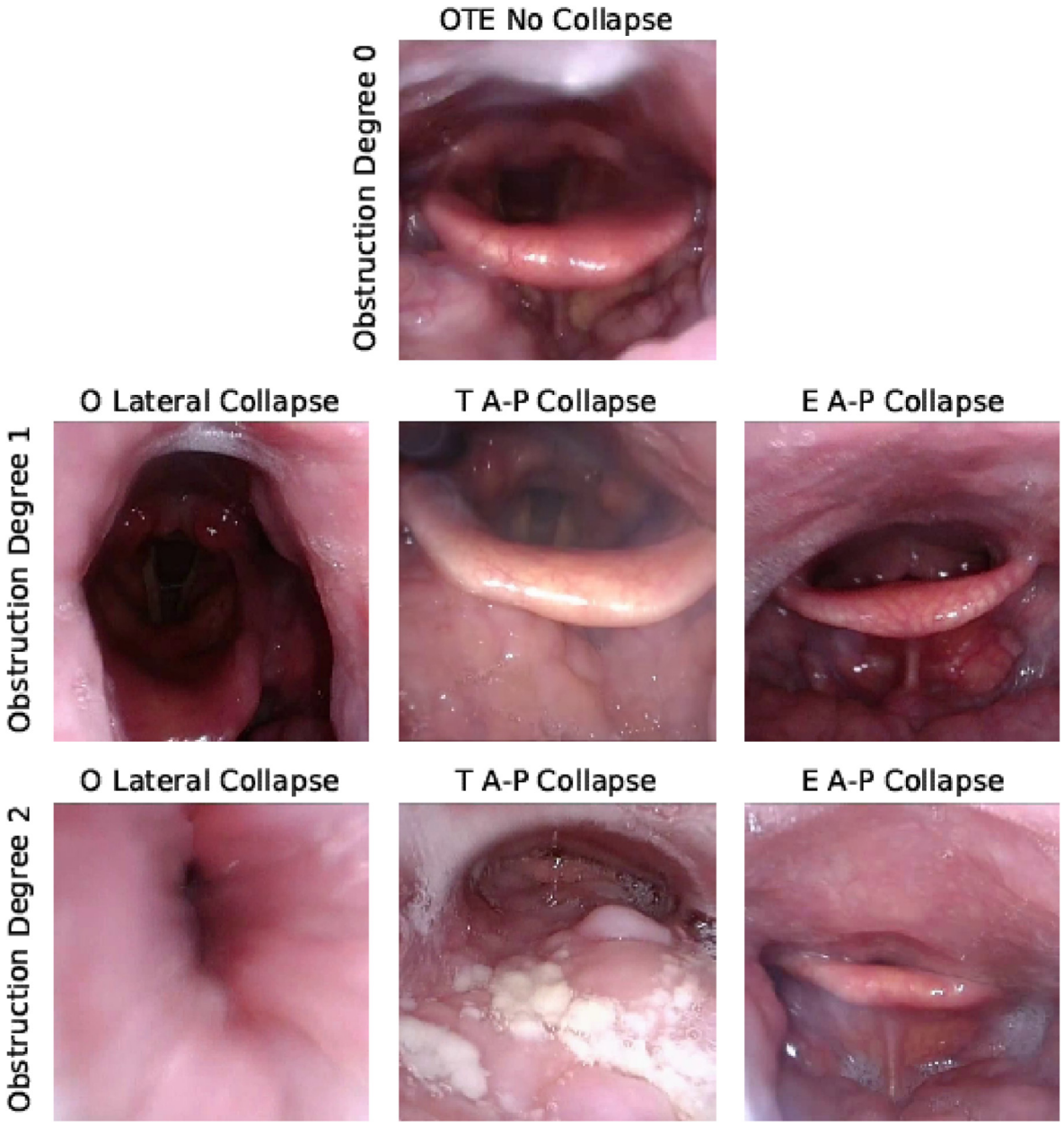

Fig. 2. Possible ways the oropharynx lateral walls (O), tongue base (T), and epiglottis (E) can collapse in the upper airway, either partially (obstruction degree 1) or completely (obstruction degree 2). The pattern of collapse is lateral for O, antero-posterior (A–P) for T, and A-P and lateral for E. Lateral collapse for E has been left out, since it is extremely rare.

2.3. Deep learning architecture

The proposed architecture is a combination of a convolutional neural network (CNN) with a Resnet18 architecture [41] and a bidirectional long short-term memory (Bi-LSTM) network [42] as illustrated in the middle block of Fig. 5. The CNN is implemented for automatic feature extraction from each video frame, while Bi-LSTM layers are included to include temporal context in both forward and backward directions for each frame. The Resnet18 network is implemented such that a 5-s clip (consisting of 25 frames) can be input one frame at a time. The output is then a feature map of size 1x512 for each frame. The feature maps for all frames are concatenated to form a 25x512 matrix, where each row is considered a time-step in the original 5-s clip and each column is a feature vector for a particular frame. This matrix is processed by a Bi-LSTM layer, followed by a dense layer, which reduces the number of features from 512 to 128 while time steps are intact, i.e., resulting matrix dimensions are 25x128. A second Bi-LSTM and dense layer reduces the number of features further from 128 to 4, yielding a matrix of dimension 25x4. From this matrix, the output at the middle time step, i.e., 13 is taken as it represents the time step where the model has most context in both directions. Finally, a softmax activation function is applied to the resulting 1x4 vector to yield a probability for each class, i.e., 0, 1, 2, and X. This architecture is repeated four times, one for each site with their own loss function. This architecture is identical to the one we presented in earlier work in greater detail [34] except for two key differences: 1) here we take the middle time step of the second Bi-LSTM and dense layer output instead of using all time steps, since there is only a single output per 5-s clip in this study instead of outputs for each time step, and 2) the number of output probabilities are 4 instead for 3.

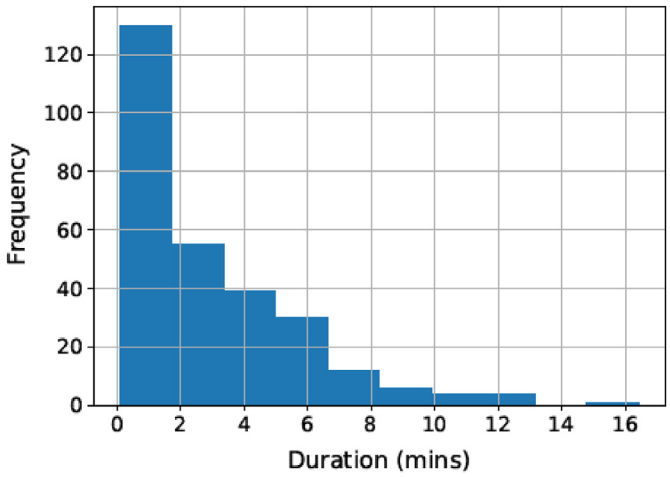

Fig. 3. Distribution of durations of the 281 drug-induced sleep endoscopy videos in the dataset.

Table 2
Examples of annotations provided by surgeons for drug-induced sleep endoscopy examinations. The number (0, 1, or 2) for each upper airway site (V, O, T, and E) indicates the obstruction degree followed by the collapse pattern (A-P, lateral or concentric). V – velum, O – oropharynx lateral wall, T – tongue base, E – epiglottis, A-P – antero-posterior.

Video	V	O	T	E
Video 1	2 A-P - Concentric	2 Lateral	1 A-P	2 Lateral
Video 2	2 A-P	2 Lateral	0	0
Video 3	1 A-P	0	2 A-P	2 A-P

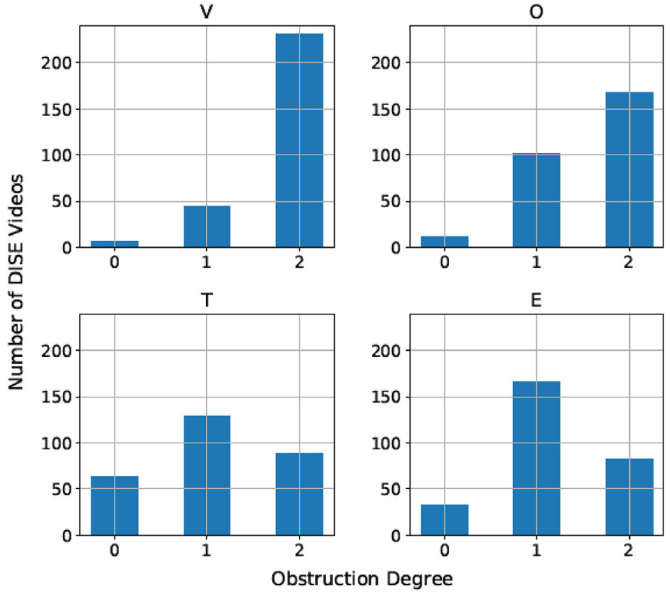

Fig. 4. Distribution of obstruction degrees (0, 1, and 2) for each of the four upper airway sites (V, O, T, and E) across 281 DISE videos. DISE – drug-induced sleep endoscopy, V – velum, O – oropharynx lateral walls, T – tongue base, E – epiglottis.

2.4. Training, validation, and testing

The proposed model was trained, validated, and tested using 10-fold cross-validation to get predictions for all DISE videos in the dataset. The loss function for each site was the cross-entropy loss, which is the loss function of choice in multi-classification settings. The loss functions for all four sites were added together and the combined loss function was used for optimizing the weights of the model using the Adam optimizer [43]. Since the degrees for each site were imbalanced as seen in Fig. 4, penalty weights were introduced during training. The weight for a particular class was calculated by dividing the number of samples for the most represented class with the number of samples for a particular class in the training set. The model was trained using batch sizes of 8 and the learning rate was set to $1 \bullet 10^{-5}$ with a weight decay of $5 \bullet 10^{-4}$. Early stopping was applied when the validation error stopped decreasing for 3 consecutive epochs, i.e., a patience of 3, to avoid overfitting to the training data. Python 3.6.10 and Pytorch 1.10.0 were used for implementation of the proposed model. Training of the model was performed using a GeForce RTX 3070 and the entire training/validation/test setup took approximately 16 h to run.

2.5. Post-processing

Post-processing steps are illustrated in the bottom blocks of Fig. 5. For each site (V, O, T, and E) in each 5-s clip, the model predicts probabilities for each of the four different classes (0, 1, 2, and X). The predicted degree for each site is the one which the model predicted the highest probability for. After a prediction is made for each site for each 5-s clip, the overall degree for each site for a particular DISE examination is calculated as the maximum predicted degree across all 5-s clips constituting a single DISE examination. The maximum degree is selected only if this degree is predicted in at least 5% of the clips which make up a full examination. This is to avoid any coincidences where a degree of e.g., 2 occurs one time by chance or because of other sites and does not reflect the true behavior of that site in a subject. In case the maximum degree does not satisfy this condition, the next greatest degree is selected if it satisfies the same condition. If this is not satisfied either, the degree is 0 by default. This is illustrated in the bottom right box of Fig. 5. A voting approach is not applied here, because surgeons annotate DISE examinations according to the highest degree observed.

2.6. Performance evaluation

The predicted obstruction degree for each site for each DISE examination was compared to the surgeons' annotations, considered as ground truth. Performance was evaluated using weighted F1 score [44]. Weighted F1 score was used instead of accuracy due to a large imbalance between the obstruction degrees. The F1 score is calculated as the harmonic mean of precision and recall or alternatively in terms of true positives (TP), false positives (FP), and false negatives (FN) expressed as:

$$F1\ score = \frac{TP}{TP + \frac{1}{2}\ (FP + FN)}$$

Using a degree of 0 as example, TP represents the number of 0's that are correctly predicted as 0's, FP represents the number of predicted 0's that are not actually 0's, and FN represents the number of predicted 1's and 2's that are actually 0's. The F1 scores for degrees 1 and 2 are calculated similarly. The weighted F1 score is calculated by averaging the F1 score of the individual degrees multiplied by their proportion in the dataset.

Cohen's kappa [45] was also used to compare model performance with interrater agreement reported in the literature with respect to degrees, either for each site or overall.

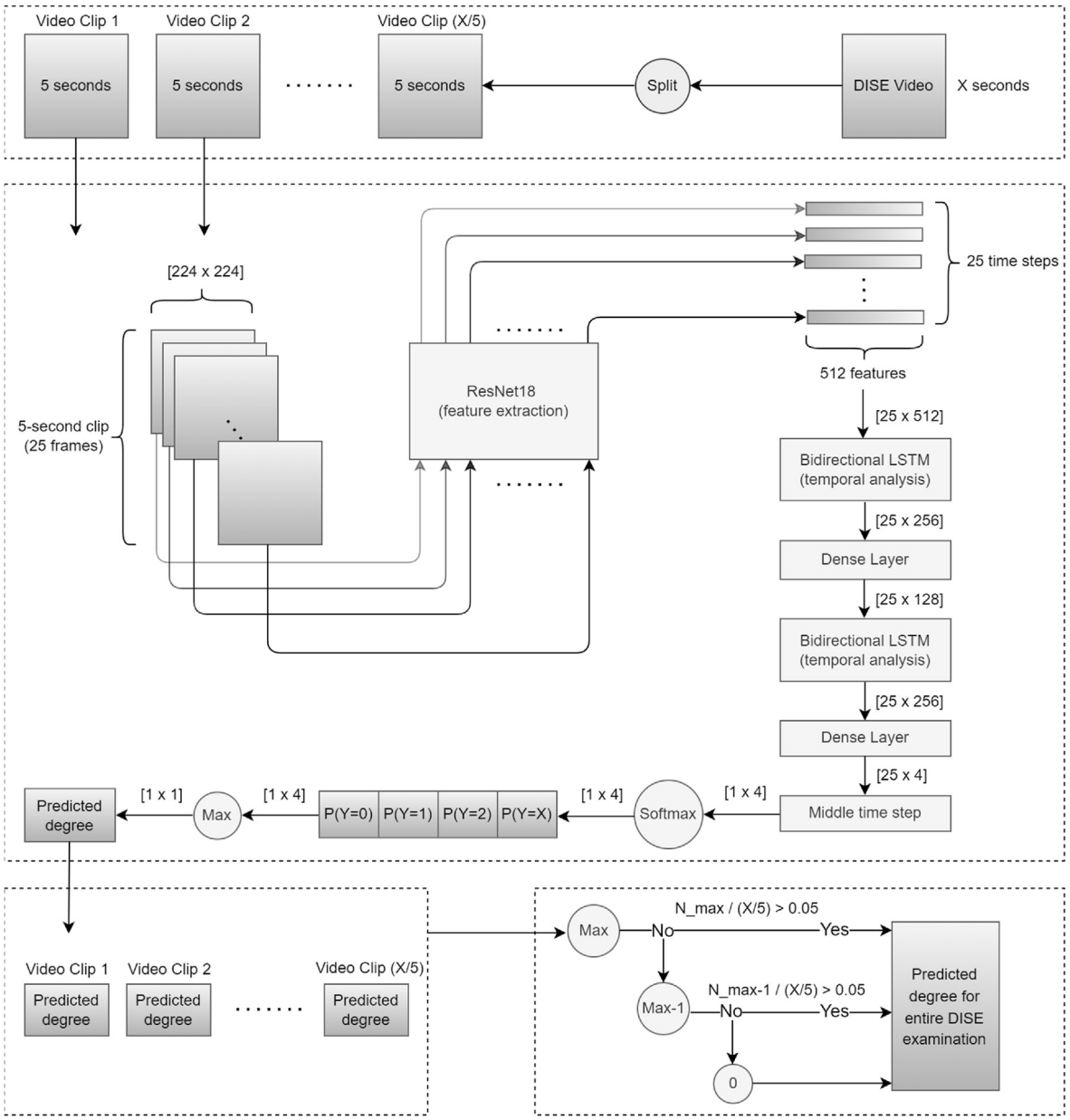

Fig. 5. Architecture of proposed model for predicting obstruction degrees based on 5-s clips from drug-induced sleep endoscopy (DISE) examinations. This architecture is repeated for each of the four upper airway sites (velum, oropharynx lateral walls, tongue base, epiglottis). Top block: A DISE examination is split into 5-s clips. Middle block: Each individual frame (grayscale) of a 5-s clip is used as input one by one for a convolutional neural network (CNN) with a ResNet18 architecture for feature extraction. All resulting feature vectors (1x512) are concatenated (25x512) and input to a bidirectional long short-term memory network (Bi-LSTM) for temporal analysis, followed by a dense layer to reduce the number of features (25x128). Another Bi-LSTM and dense layer reduce the feature vector (1x4), which is run through a softmax activation function. This yields four probabilities, one for each obstruction degree ($P(Y = 0)$, $P(Y = 1)$, $P(Y = 2)$, and $P(Y = X)$, where X means that the site is not visible). The obstruction degree with highest probability is the predicted obstruction degree. Bottom left block: Predictions for all clips within a DISE examination are collected. Bottom right block: The maximum predicted obstruction degree across all 5-s clips that make up a full examination is chosen as the overall degree if the model predicts this degree for at least 5% of all clips. Otherwise, same criterion is checked for the next highest degree and if not fulfilled either, the predicted obstruction degree is 0 by default.

3. Results

Overall F1 score for the 12-class problem, i.e., predicting the obstruction degree (0, 1, or 2) for each of the four upper airway sites across 281 DISE videos was 70% (V: 85%, O: 72%, T: 57%, E: 65%). Fig. 6 shows confusion matrices for the model's predicted degree for each site evaluated against the surgeons' annotations across all DISE videos in the dataset. Performance was also evaluated with respect to videos obtained from the three different surgeons (one from CUH and two from SUH) to investigate any biases in the model towards videos from a particular surgeon. The results are summarized in Table 3.

Overall F1 score for the 16-class problem, i.e., predicting obstruction degrees for each individual 5-s clip (including the class X for when the site is not visible) was 65 ± 14% (V: 68 ± 21%, O: 64 ± 22%, T: 64 ± 23%, E: 65 ± 23%) and was calculated by averaging all clips that make up a full DISE examination and then averaging the performance over all 281 DISE videos. Fig. 7 shows the F1 score distribution for the individual DISE examinations with respect to the 5-s clips. Fig. 8 provides an example of predicted probabilities

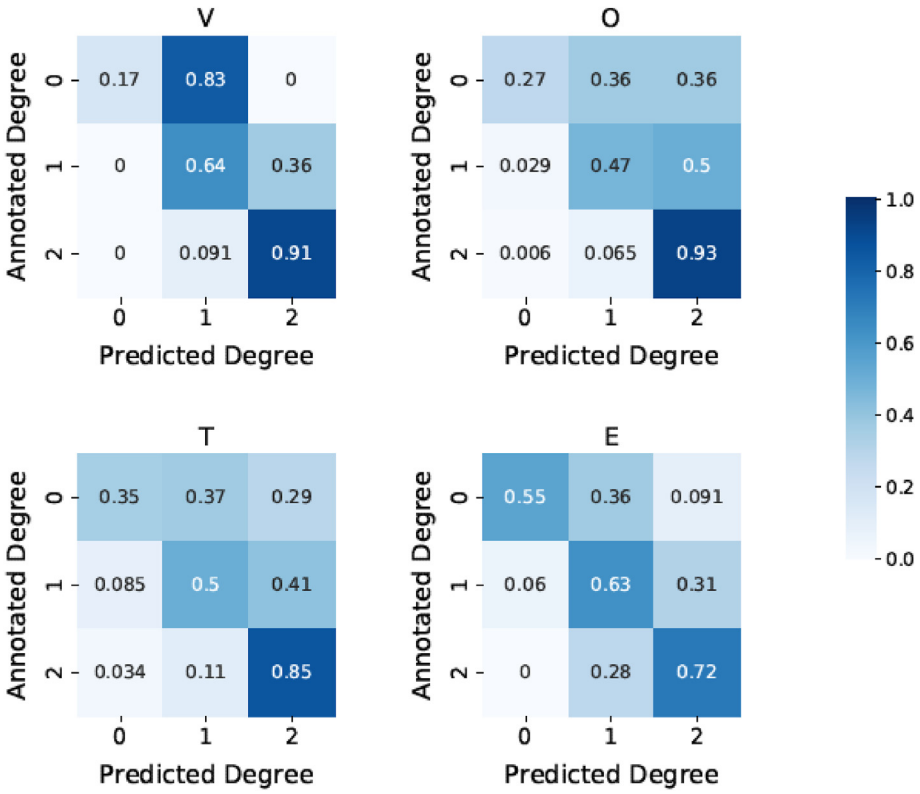

Fig. 6. Confusion matrices for the predicted obstruction degrees evaluated for 281 drug-induced sleep endoscopy videos with respect to the four different upper airway sites. V – velum, O – oropharynx lateral walls, T – tongue base, E – epiglottis.

Table 3
F1 score for drug-induced sleep endoscopy examinations with respect to each of the three surgeons who provided the videos. Performance is evaluated against the surgeons' annotations with respect to obstruction degree (0, 1, or 2) for four different upper airway sites. V – velum, O – oropharynx lateral walls, T – tongue base, E – epiglottis.

Surgeon	N Videos	V (F1)	O (F1)	T (F1)	E (F1)	Overall (F1)
S1 (CUH)	51	91%	70%	53%	58%	68%
S2 (SUH)	58	89%	64%	63%	63%	70%
S3 (SUH)	172	82%	74%	56%	67%	70%

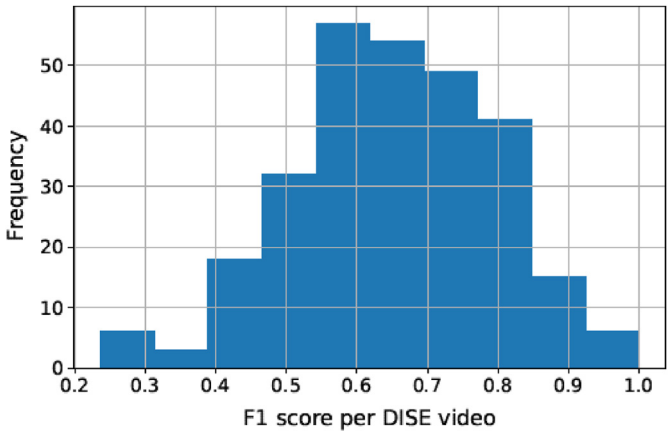

Fig. 7. Distribution of F1 scores for 281 drug-induced sleep endoscopy (DISE) videos calculated as the average F1 score of all 5-s clips making up each DISE examination with respect to four classes (0, 1, 2, and X) for the upper airway sites (velum, oropharynx lateral walls, tongue base, and epiglottis).

for 5-s clips for each site and how they change over the course of time (30 s of video).

When distinguishing between no obstruction/obstruction, i.e., combining obstruction degrees of 1 and 2 as one class, the F1 score increased to 90% (V: 98%, O: 95%, T: 78%, E: 91%). Similarly, for 5-s clips averaged over DISE examinations, where the classes are no obstruction, obstruction, and X, the F1 score increased to 74 ± 13% (V: 79 ± 19%, O: 76 ± 22%, T: 70 ± 22%, E: 71 ± 22%). Finally, Table 4 compares performance of the model (in terms of Cohen's kappa) to interrater reliabilities reported in the literature between surgeons. The comparisons are made on three levels: for each site (V, O, T, and E), region-based (palate and hypopharynx), and overall. For region-based comparison, V was compared to the palate, and OTE were combined for hypopharynx comparisons.

4. Discussion

Here we describe for the first time that deep learning can be used to reliably evaluate DISE videos with the goal of identifying site and extent of obstruction.

4.1. Performance compared to random guessing

Average F1 score for the model's VOTE obstruction degree predictions was 70% (V: 85%, O: 72%, T: 57%, E: 65%) for all DISE examinations in the dataset. If the model had instead predicted all examinations to have the most represented obstruction degree for each site (i.e., 2 for V and O, and 1 for T and E), the average F1 score would be only 48% (V: 74%, O: 45%, T: 29%, 44%). The average F1 score with respect to 5-s clips averaged over each DISE examination was 65 ± 14% (V: 68 ± 21%, O: 64 ± 22%, T: 64 ± 23%, E: 65 ± 23%). In contrast, if the model predicted the obstruction degrees for each 5-s

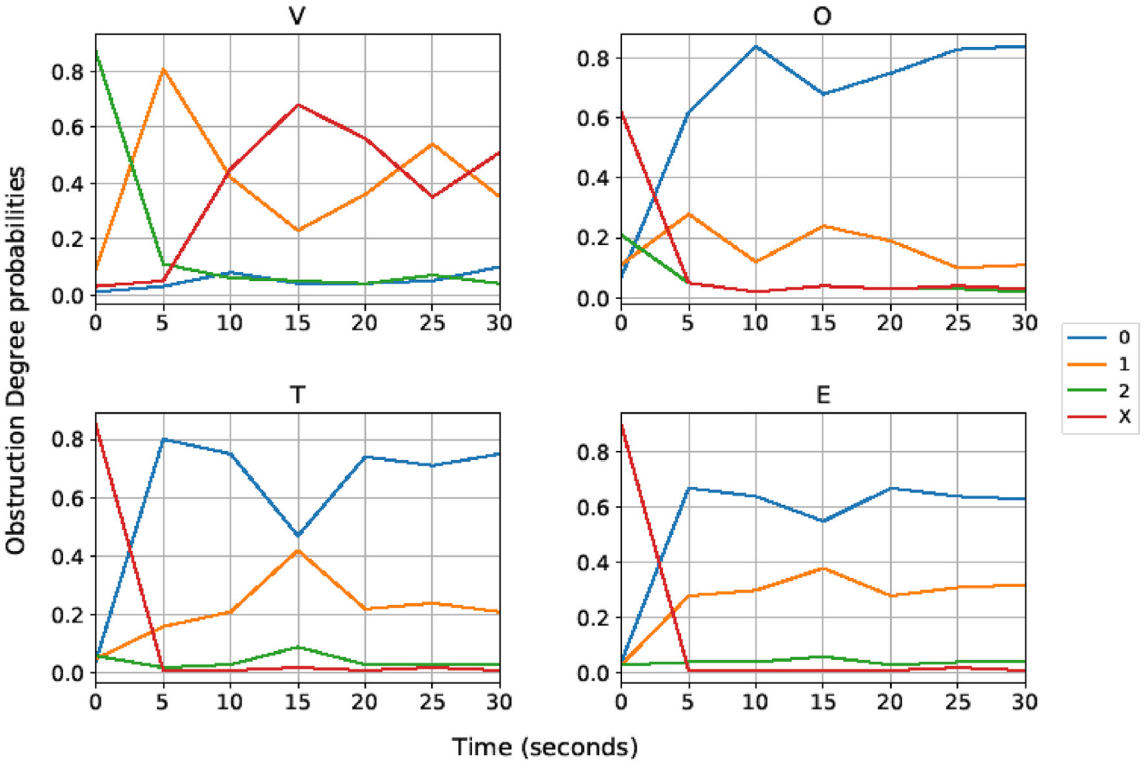

Fig. 8. Probabilities for obstruction degrees (0, 1, 2) or X (site not visible) predicted by the proposed model for 5-s clips of drug-induced sleep endoscopy for four different upper airway sites. V – velum, O – oropharynx lateral walls, T – tongue base, E – epiglottis.

clip to be the most represented obstruction degree for each site, respectively, the average F1 score would only be 20 ± 13% (V: 18 ± 21%, O: 8 ± 13%, T: 24 ± 28%, E: 31 ± 29%). This quantitative analysis supports the fact that the model performs much better than random guessing and emphasizes the large gap that otherwise would not be as apparent if a performance metric like accuracy would be used, which does not consider how well the model predicts individual classes.

4.2. Sensitivity for each obstruction degree

The model predicts especially well whether or not there is obstruction as shown by an F1 score reaching 90% (V: 98%, O: 95%, T: 78%, E: 91%) when combining degrees 1 and 2, showing that in general, the model confuses degrees 1 and 2 more often than 0 and 1 or 0 and 2. Fig. 6 shows that the highest sensitivity for degree 0 is obtained for E (55%), and that most misclassifications occur because the model predicts degree 1. In very few cases (<10%), the model predicts 2 and after inspecting the three examinations in question, it occurs for two reasons: 1) E is reflected in saliva causing a mirror image where it looks like E is collapsing, which in fact it is not, and 2) the model confuses an A-P V collapse for an E collapse, particularly when the endoscope is close to the collapse. In these examinations, however, the predicted probability for degree 2 is never higher than 40% and it only occurs in 1–2 clips per examination.

The second highest sensitivity for degree 0 is obtained for T (35%), but it is confused with both degrees 1 and 2. When degree 2 is predicted, it occurs for two reasons: the uvula or lower part of the soft palate resembles the tongue and when it collapses, the model confuses it for the tongue collapsing, and 2) when V or O are collapsing and the endoscope is extremely close to the tissue, it resembles the tissue of the tongue.

The lowest sensitivity for degree 0 is obtained for V (17%), although there are only 6 out of 281 DISE examinations where V has a degree of 0. Presence of collapse at the level of V is extremely common among OSA patients [46] which is also evident in the dataset by the lack of videos where V has a degree of 0. However, it is encouraging that the model only confuses degree 0 with degree 1 and never predicts degree 2 in those cases.

The next lowest sensitivity for degree 0 is obtained for O (27%) and the confusion is equally split between degrees 1 and 2. Again, there are only a small number of DISE examinations where O has degree 0 (11 examinations), but the cases in which they are predicted as 2 are due to three reasons: 1) A lateral collapse at the level of E which is generally not considered part of O, (2) T collapsing completely and the endoscope being very close such that the model mistakes it for a collapse at O, and 3) V collapsing and the model mistakenly predicting a contribution of O as well, which can be difficult to assess even for surgeons [47].

The highest sensitivity for degree 1 is obtained for both V and E (64% and 63%, respectively), while T and O are lower (50% and 47%, respectively). For all four sites, the model primarily confuses it with degree 2 and very rarely with degree 0 (none for V, <5% for O, <10% for T and E). For degree 2, the sensitivity is very high for V, O, and T (91%, 93%, and 85%, respectively), while the sensitivity for E is lower (72%). For all four sites, the model almost exclusively confuses degree 2 with degree 1 and almost never with degree 0 (none for V and E, <5% for O and T). Although the model confuses degree 1 with degree 2 to some extent for all sites, the model confuses degree 2 with degree 1 only for E, showing that for this site, the model appears to have most difficulty distinguishing between degrees 1 and 2.

4.3. Performance for videos from each surgeon

Table 3 shows that there is no noticeable difference in overall F1 score between videos obtained from each of the three surgeons, demonstrating that there is no meaningful bias towards any of them and suggesting that the procedures are comparable. For videos from surgeon 1 (S1) from CUH, the model yields the highest F1 score for V out of all three but also the lowest F1 score for T and E. For videos obtained from surgeon 2 (S2) from SUH, the model has the highest F1 score for T out of all three and the lowest F1 score for O. However, the gap between the highest and lowest F1 score for S2 is much smaller than for S1 and in general the discrepancy in performance between sites is lower for videos from S2. For videos obtained from surgeon 3 (S3) from SUH, the model has the highest F1 scores for both O and E compared to S1 and S2 and the lowest F1 score for V. Again, the gap between highest and lowest F1 score is much smaller than for S1.

4.4. Performance for 5-s clips

Fig. 7 shows the distribution of F1 scores per DISE examination as the average F1 score with respect to 5-s clips that make up an entire examination. It is noted that most examinations have an F1 score above 50% (86% of all videos), but the lowest F1 score is at 24%. There are three important factors which explain very low performance for some examinations: 1) The duration of an examination, since a very short examination consists of only few 5-s clips and even a few misclassified clips reduce performance by a lot, 2) the video quality, since some examinations have extremely low quality, which makes it difficult to assess the degree for each site, and 3) several sites collapsing simultaneously, causing the endoscope to be pushed back and forth and making the video chaotic to interpret.

4.5. Probability interpretation

Fig. 8, which depicts an example of predicted probabilities for each obstruction degree (and X) for each site over time, shows that the model regularly produces predictions with high confidence that are easy to interpret due to the way they covary within and across sites. Such a plot provides insight into model behavior and situations where the predictions are made with high or low confidence, which can then be directly compared to the DISE examinations. For example, for time (t) = 0 s (s), the model predicts with high probability that V has degree 2 and that O, T, and E are not visible due to the obstruction at V. The probability of O not being visible is lower than for T and E because V reopens in that clip and reveals O for a split second. This observation emphasizes a limitation of the model: predictions are made on 5-s clips, during which several events can occur, causing the model to be less confident in a single obstruction degree. For t = 5 s, V transitions from degree 2 to 1, which enables the model to see O, T, and E and predict degree 0 for all three. However, probabilities for O and E are lower because they are both on the border between having degrees 0 or 1. At t = 10 s, the model is confused whether V is not visible or has degree 1 because the endoscope is being moved down the airway and is at the border between V and O. For O, the model becomes more confident that the degree is 0 as the lateral walls separate further. At t = 15 s, probabilities for degree 0 decrease and degree 1 increase for O, T, and E because these sites move slightly, but not enough to cause any obstruction. At the same time, the model becomes more confident that V is not visible as the camera is moved further down. For the remainder of the video (t = 20–30 s), the model becomes more confident that O, T, and E have degrees 0 again, but for V, it switches between degree of 1 and X because, although the endoscope is further down in the airway, the uvula occasionally vibrates and becomes visible, causing the model to confuse whether V is visible or not.

4.6. Comparison to interrater reliability

Table 4 compares performance of the proposed model to interrater reliabilities between surgeons annotating obstruction degrees reported in the literature. The comparisons are performed at three levels: 1) for each individual site, 2) for palate (V) and hypopharynx (O, T, and E), and 3) overall (all sites combined). For V, our model achieves a higher kappa than the interrater scores. For O and E, our model has lower kappa than two studies and slightly higher than one. For T, our model has higher kappa than two studies and lower than one, demonstrating that despite our model having low performance for T, surgeons can struggle with it as well which is evident from a huge difference in kappa scores between studies (0.03–0.60). For both the palate and hypopharynx, our model produces higher kappa than one study and lower than another. For the overall evaluation of degrees, our model achieves a much higher kappa than the one study where they conduct such an analysis. This comparison adds context to the model performance and demonstrates that analysis of DISE examinations is not a trivial task, not even for experiences surgeons.

4.7. Limitations

There are two main limitations of this study: 1) the model is not able to predict the pattern of collapse, which would need to be added for the model to produce complete VOTE annotations as the surgeons do, and 2) no healthy controls were used in the study, but the model could benefit from seeing more examples of absence of collapse, particularly for V and O. For the first limitation, the absence of collapse pattern only really affects predictions for V, since O and T only have one possible collapse pattern and lateral obstructions for E are extremely rare [48]. However, the difference in collapse patterns for V (particularly concentric vs A-P or lateral) can lead to different treatment strategies at the level of the velum and lateral pharyngeal wall and is therefore important in clinical practice [49–51]. The second limitation is difficult to compensate for because DISE examinations are only performed for people with confirmed OSA as the whole point of the procedure is to identify sites contributing to upper airway collapse prior to surgery. However, such data could be gathered by utilizing DISE under sedation associated with other medical procedures.

Table 4
Comparison of the proposed model's performance in terms of Cohen's kappa (κ) to interscorer reliabilities reported in the literature between otolaryngology surgeons scoring obstruction degrees in drug-induced sleep endoscopy. Some studies use palate vs hypopharynx, where palate corresponds to V and hypopharynx corresponds to O, T, and E combined. One study reports κ overall across all sites.

Study	V (κ)	O (κ)	T (κ)	E (κ)	N scorers	N DISE
Our model	***0.55***	***0.45***	***0.38***	***0.44***	***N/A***	***281***
Vroegop et al. [29]	0.30	0.66	0.03	0.61	7	6
Llatas et al. [30]	0.17	0.67	0.35	0.43	2	31
Green et al. [23]	0.40	0.42	0.60	0.55	4	275

Study	Palate (κ)	Hypopharynx (κ)	N scorers	N DISE
Our model	***0.55***	***0.43***	***N/A***	***281***
Kezirian et al. 10 [31]	0.60	0.44	2	108
Koo et al. [32]	0.52	0.35	6	100

Study	Overall (κ)	N scorers	N DISE
Our model	***0.46***	***N/A***	***281***
Gillespie et al. [33]	0.27	3	38

4.8. Potential of artificial intelligence in sleep medicine

This study demonstrates the potential of artificial intelligence (AI) systems applied in sleep medicine, which can be explored in all aspects of the field. In recent research, there are examples of AI systems being utilized for OSA screening or diagnosis [52–55], predicting adherence to CPAP [56], OSA phenotyping [57,58], and predicting comorbidities associated with OSA [57,59]. The role of AI in sleep medicine is becoming increasingly important and has the potential to not only improve diagnosis and treatment of sleep disorders such as OSA, but also to improve understanding of the disease and provide insights into the pathology.

4.9. Future work

In future work, collapse pattern for V should be added to the output of the model. This could be done by annotating the pattern for V for all 5-s clips, i.e. A-P, lateral, concentric, and X for when there is no collapse. A network identical to the ones predicting obstruction degrees for V, O, T, and E could then be implemented and trained simultaneously along with the other networks. Future work would also benefit from using a larger, multi-scored dataset for testing purposes, such that the interrater reliability can be compared directly to the model's performance on the same data. Finally, future work should explore potential of using self-supervised learning on DISE examinations, which would require much fewer annotations and save time and resources compared to the approach followed in this study.

5. Conclusion

This study presents the first ever model for predicting sites and obstruction degrees of upper airway collapse from DISE examinations using a dedicated deep learning model. The model produces solid performance with an overall F1 score of 70% and predicts obstruction degrees for the velum and oropharynx lateral walls well but displays moderate performance for the tongue base and epiglottis. The main limitation is that it does not predict the pattern of collapse, which can affect treatment strategy at the level of the velum and lateral pharyngeal wall. The proposed model has potential to aid surgeons in interpreting DISE examinations in an automated manner but needs further validation on a multi-scored dataset and the added ability to predict collapse pattern.

CRediT authorship contribution statement

Umaer Hanif: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Eva Kirkegaard Kiaer:** Conceptualization, Data curation, Resources, Supervision, Writing – review & editing. **Robson Capasso:** Data curation, Resources, Writing – review & editing. **Stanley Y. Liu:** Data curation, Resources, Writing – review & editing. **Emmanuel J.M. Mignot:** Conceptualization, Data curation, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Helge B.D. Sorensen:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing. **Poul Jennum:** Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review & editing.

References

[1] Lévy P, et al. Obstructive sleep apnoea syndrome. Nat Rev Dis Prim 2015;1(1):1–21. https://doi.org/10.1038/nrdp.2015.43.

[2] Johns MW. A new method for measuring daytime sleepiness: the epworth sleepiness scale. Sleep 1991;14(6):540–5. https://doi.org/10.1093/SLEEP/14.6.540.

[3] Gabryelska A, Białasiewicz P. Association between excessive daytime sleepiness, REM phenotype and severity of obstructive sleep apnea. Sci Rep 2020;10(1):1–6. https://doi.org/10.1038/s41598-019-56478-9.

[4] Léger D, Stepnowsky C. The economic and societal burden of excessive daytime sleepiness in patients with obstructive sleep apnea. Sleep Med Rev 2020;51:101275. https://doi.org/10.1016/j.smrv.2020.101275.

[5] Baguet JP, Barone-Rochette G, Tamisier R, Levy P, Pépin JL. Mechanisms of cardiac dysfunction in obstructive sleep apnea. Nat Rev Cardiol 2012;9(12):679–88. https://doi.org/10.1038/nrcardio.2012.141.

[6] Gonzaga C, Bertolami A, Bertolami M, Amodeo C, Calhoun D. Obstructive sleep apnea, hypertension and cardiovascular diseases. J Hum Hypertens 2015;29(12):705–12. https://doi.org/10.1038/jhh.2015.15.

[7] McEvoy RD, et al. CPAP for prevention of cardiovascular events in obstructive sleep apnea. N Engl J Med 2016;375(10):919–31. https://doi.org/10.1056/NEJMoa1606599.

[8] Tregear S, Reston J, Schoelles K, Phillips B. Obstructive sleep apnea and risk of motor vehicle crash: systematic review and meta-analysis. J Clin Sleep Med 2009;5(6):573–81. https://doi.org/10.5664/JCSM.27662.

[9] Xie C, Zhu R, Tian Y, Wang K. Association of obstructive sleep apnoea with the risk of vascular outcomes and all-cause mortality: a meta-analysis. BMJ Open 2017;7(12):e013983. https://doi.org/10.1136/BMJOPEN-2016-013983.

[10] Benjafield AV, et al. Estimation of the global prevalence and burden of obstructive sleep apnoea: a literature-based analysis. Lancet Respir Med 2019;7(8):687–98. https://doi.org/10.1016/S2213-2600(19)30198-5.

[11] Watson NF. Health care savings: the economic value of diagnostic and therapeutic care for obstructive sleep apnea. J Clin Sleep Med 2016;12(8):1075–7. https://doi.org/10.5664/JCSM.6034.

[12] Schwartz AR, Patil SP, Laffan AM, Polotsky V, Schneider H, Smith PL. Obesity and obstructive sleep apnea: pathogenic mechanisms and therapeutic approaches. Proc Am Thorac Soc 2008;5(2):185–92. https://doi.org/10.1513/pats.200708-137MG.

[13] Dempsey JA, Veasey SC, Morgan BJ, O'donnell CP. Pathophysiology of sleep apnea. Physiol Rev 2010;90(1):47–112. https://doi.org/10.1152/physrev.00043.2008.

[14] Lee RWW, Sutherland K, Cistulli PA. Craniofacial morphology in obstructive sleep apnea: a review. Clin Pulm Med 2010;17(4):189–95. https://doi.org/10.1097/CPM.0b013e3181e4bea7.

[15] Giles T, Lasserson T, Smith B, White J, Wright J, Cates C. Continuous positive airways pressure for obstructive sleep apnoea in adults. Cochrane Database Syst Rev 2006;1. https://doi.org/10.1002/14651858.CD001106.pub2.

[16] J K, et al. Interventions for the treatment of obstructive sleep apnea in adults: a health technology assessment. Canadian Agency for Drugs and Technologies in Health; 2019 [Online]. Available: http://europepmc.org/books/NBK535532.

[17] Stuck B, Eschenhagen T, Sommer U, de Vet H, Sommer U. Uvulopalatopharyngoplasty with or without tonsillectomy in the treatment of adult obstructive sleep apnea–A systematic review. Elsevier 2018;50:152–65. https://doi.org/10.1016/j.sleep.2018.05.004.

[18] Holmlund T, et al. Tonsillectomy in adults with obstructive sleep apnea. Laryngoscope 2016;126(12):2859–62. https://doi.org/10.1002/LARY.26038.

[19] Vauterin T, Garas G, Arora A. Transoral robotic surgery for obstructive sleep apnoea-hypopnoea syndrome. Oto-Rhino-Laryngol 2018;80(3–4):134–47. https://doi.org/10.1159/000489465.

[20] Zaghi S, et al. Maxillomandibular advancement for treatment of obstructive sleep apnea: a meta-analysis. JAMA Otolaryngol Head Neck Surg 2016;142(1):58–66. https://doi.org/10.1001/JAMAOTO.2015.2678.

[21] Hohenhorst W, Ravesloot MJL, Kezirian EJ, de Vries N. Drug-induced sleep endoscopy in adults with sleep-disordered breathing: technique and the VOTE Classification system. Oper Tech Otolayngol Head Neck Surg 2012;23(1):11–8. https://doi.org/10.1016/J.OTOT.2011.06.001.

[22] Kezirian EJ, Hohenhorst W, de Vries N. Drug-induced sleep endoscopy: the VOTE classification. Eur Arch Oto-Rhino-Laryngol 2011;268(8):1233–6. https://doi.org/10.1007/S00405-011-1633-8.

[23] Green KK, et al. Drug-induced sleep endoscopy and surgical outcomes: a multicenter cohort study. Laryngoscope 2019;129(3):761–70. https://doi.org/10.1002/LARY.27655.

[24] Kotlarek KJ, Haenssler AE, Hildebrand KE, Perry JL. Morphological variation of the velum in children and adults using magnetic resonance imaging. Imaging Sci Dent 2019;49(2):153. https://doi.org/10.5624/ISD.2019.49.2.153.

[25] Gao F, et al. Upper airway morphological changes in obstructive sleep apnoea: effect of age on pharyngeal anatomy. J Laryngol Otol 2020;134(4):354–61. https://doi.org/10.1017/S0022215120000766.

[26] Diwakar R, et al. Effect of craniofacial morphology on pharyngeal airway volume measured using cone-beam computed tomography (CBCT)—a

retrospective pilot study. mdpi.com 2021;18(9):5040. https://doi.org/10.3390/ijerph18095040.
[27] Ma MA, Kumar R, Macey PM, Yan-Go FL, Harper RM. Epiglottis cross-sectional area and oropharyngeal airway length in male and female obstructive sleep apnea patients. Nat Sci Sleep 2016;8:297. https://doi.org/10.2147/NSS.S113709.
[28] Zhou N, et al. Intra-individual variation of upper airway measurements based on computed tomography. PLoS One 2021;16(11):e0259739. https://doi.org/10.1371/JOURNAL.PONE.0259739.
[29] Vroegop AVMT, et al. Observer variation in drug-induced sleep endoscopy: experienced versus nonexperienced ear, nose, and throat surgeons. Sleep 2013;36(6):947–53. https://doi.org/10.5665/SLEEP.2732.
[30] Carrasco-Llatas M, Zerpa-Zerpa V, Dalmau-Galofre J. Reliability of drug-induced sedation endoscopy: interobserver agreement. Sleep Breath 2017;21(1):173–9. https://doi.org/10.1007/S11325-016-1426-9.
[31] Kezirian EJ, White DP, Malhotra A, Ma W, McCulloch CE, Goldberg AN. Interrater reliability of drug-induced sleep endoscopy. Arch Otolaryngol Head Neck Surg 2010;136(4):393–7. https://doi.org/10.1001/ARCHOTO.2010.26.
[32] Koo SK, et al. Inter-rater reliability between experienced and inexperienced otolaryngologists using Koo's drug-induced sleep endoscopy classification system. European Archives of Oto-Rhino-Laryngology; 2019. https://doi.org/10.1007/S00405-019-05386-9.
[33] Gillespie MB, Reddy RP, White DR, Discolo CM, Overdyk FJ, Nguyen SA. A trial of drug-induced sleep endoscopy in the surgical management of sleep-disordered breathing. Laryngoscope 2013;123(1):277–82. https://doi.org/10.1002/LARY.23506.
[34] Hanif U, Kezirian E, Kiar EK, Mignot E, Sorensen HBD, Jennum P. Upper airway classification in sleep endoscopy examinations using convolutional recurrent neural networks. In: Proceedings of the annual international conference of the IEEE engineering in medicine and biology society. EMBS; 2021. p. 3957–60. https://doi.org/10.1109/EMBC46164.2021.9630098.
[35] Liu SYC, et al. Efficacy of maxillomandibular advancement examined with drug-induced sleep endoscopy and computational fluid dynamics airflow modeling. Otolaryngol Head Neck Surg 2016;154(1):189–95. https://doi.org/10.1177/0194599815611603.
[36] Liu S, et al. Lateral pharyngeal wall tension after maxillomandibular advancement for obstructive sleep apnea is a marker for surgical success: observations from drug-induced sleep endoscopy. J Oral Maxillofac Surg 2015;73(8):1575–82. https://doi.org/10.1016/j.joms.2015.01.028.
[37] Liu S, Awad M, Riley R, Capasso R. The role of the revised stanford protocol in today's precision medicine. Sleep Medicine Clin 2019;;14(1):99–107. https://doi.org/10.1016/j.jsmc.2018.10.013.
[38] Liu S, Riley R, Pogrel A, Guilleminault C. Sleep surgery in the era of precision medicine. Atlas Oral Maxillofac Surg Clin North Am 2019;27(1):1–5. https://doi.org/10.1016/j.cxom.2018.11.012.
[39] Kiær EK, et al. Propofol sedation in Drug Induced Sedation Endoscopy without an anaesthesiologist - a study of safety and feasibility. Rhinology 2019;57(2):125–31. https://doi.org/10.4193/RHIN18.066.
[40] Lan MC, Liu SYC, Lan MY, Modi R, Capasso R. Lateral pharyngeal wall collapse associated with hypoxemia in obstructive sleep apnea. Laryngoscope 2015;125(10):2408–12. https://doi.org/10.1002/LARY.25126.
[41] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. CVPR 2016:770–8. https://doi.org/10.1109/CVPR.2016.90. 2016.
[42] Schuster M, Paliwal KK. Bidirectional recurrent neural networks. IEEE Trans Signal Process 1997;45(11):2673–81. https://doi.org/10.1109/78.650093.
[43] Kingma DP, Ba J. Adam: a method for stochastic optimization. 2014. *arXiv preprint arXiv:1412.6980 [cs.LG]*.
[44] Tharwat A. Classification assessment methods. Appl Comput Inf 2020. https://doi.org/10.1016/j.aci.2018.08.003.
[45] McHugh ML. Interrater reliability: the kappa statistic. Biochem Med 2012;22(3):276. https://doi.org/10.11613/bm.2012.031.
[46] Vroegop Av, et al. Drug-induced sleep endoscopy in sleep-disordered breathing: report on 1,249 cases. Laryngoscope 2014;124(3):797–802. https://doi.org/10.1002/LARY.24479.
[47] Soares D, et al. Lateral oropharyngeal wall and supraglottic airway collapse associated with failure in sleep apnea surgery. Laryngoscope 2012;122(2):473–9. https://doi.org/10.1002/LARY.22474.
[48] Torre C, Camacho M, Liu SYC, Huon LK, Capasso R. Epiglottis collapse in adult obstructive sleep apnea: a systematic review. Laryngoscope 2016;126(2):515–23. https://doi.org/10.1002/LARY.25589.
[49] Vroegop Av, Vanderveken OM, Verbraecken JA. Drug-induced sleep endoscopy: evaluation of a selection tool for treatment modalities for obstructive sleep apnea. Respiration 2020;99(5):451–7. https://doi.org/10.1159/000505584.
[50] Susan K S, Ankur S, Omprakash C, Payal G. Management concentric collapse of velopharynx in obstructive sleep apnoea using a modified barbed palatopharyngoplasty technique. J Sleep Disord Manag 2020;6(1). https://doi.org/10.23937/2572-4053.1510028.
[51] Liu SYC, Hutz MJ, Poomkonsarn S, Chang CP, Awad M, Capasso R. Palatopharyngoplasty resolves concentric collapse in patients ineligible for upper airway stimulation. Laryngoscope 2020;130(12):E958–62. https://doi.org/10.1002/LARY.28595.
[52] Hanif U, et al. Estimation of apnea-hypopnea index using deep learning on 3-D craniofacial scans. IEEE J Biomed Health Inform 2021;25(11):4185–94. https://doi.org/10.1109/JBHI.2021.3078127.
[53] Álvarez D, et al. A machine learning-based test for adult sleep apnoea screening at home using oximetry and airflow. Sci Rep 2020;10(1):1–12. https://doi.org/10.1038/s41598-020-62223-4.
[54] Pang B, et al. Machine learning approach for obstructive sleep apnea screening using brain diffusion tensor imaging. J Sleep Res 2022:e13729. https://doi.org/10.1111/jsr.13729.
[55] Kelly JL, et al. Diagnosis of sleep apnoea using a mandibular monitor and machine learning analysis: one-night agreement compared to in-home polysomnography. Front Neurosci 2022;16. https://doi.org/10.3389/FNINS.2022.726880/FULL.
[56] Scioscia G, et al. Machine learning-based prediction of adherence to continuous positive airway pressure (CPAP) in obstructive sleep apnea (OSA). Inf Health Soc Care 2021;47(3):274–82. https://doi.org/10.1080/17538157.2021.1990300.
[57] Ma EY, Kim JW, Lee Y, Cho SW, Kim H, Kim JK. Combined unsupervised-supervised machine learning for phenotyping complex diseases with its application to obstructive sleep apnea. Sci Rep 2021;11(1):1–15. https://doi.org/10.1038/s41598-021-84003-4.
[58] Abdelwahab MA, Liu SYC. Sonographic phenotyping of the upper airway in OSA with backscattered imaging analysis by machine learning. Otolaryngol Head Neck Surg 2022;167(1):159. https://doi.org/10.1177/01945998221107672.
[59] Silva CAO, et al. Machine learning for atrial fibrillation risk prediction in patients with sleep apnea and coronary artery disease. Front Cardiovasc Med 2022;9. https://doi.org/10.3389/fcvm.2022.1050409.