

약물 유도 수면내시경에서 딥러닝을 이용한 VOTE 분류 및 완전폐쇄 예측

양현지¹, 이승준¹, 오현서², 이명은³, 양형정^{1,3*}, 양형채^{2,4*}

¹전남대학교 인공지능융합학과

²전남대학교 의과대학

³전남대학교 초광역 연합 의료 AI 연구센터

⁴전남대학교 의과대학, 전남대학교병원 이비인후과

yhj22@jnu.ac.kr, dltmdwns3462@naver.com, hyunseo.oh03@gmail.com,

myungeun07@gmail.com, hjyang@jnu.ac.kr, blessed@chonnam.ac.kr

Deep Learning-Based VOTE Classification and Full Occlusion Prediction in Drug-Induced Sleep Endoscopy

Hyeon-Ji Yang¹, Seung-Jun Lee¹, Hyun-Seo Oh², Myungeun Lee³,

Hyung-Jeong Yang^{1,3*}, Hyung-Chae Yang^{2,4*}

¹Dept. of Artificial Intelligence Convergence, Chonnam National University

²Dept. of Medicine, Chonnam National University

³Hyper-wide Federated Medical AI Research Center, Chonnam National University

⁴Dept. of Otolaryngology, Chonnam National University Hospital

요 약

폐쇄성 수면 무호흡증(Obstructive Sleep Apnea, OSA)은 수면 중 상기도의 폐쇄로 호흡 저하 및 중단이 발생하는 질환이다. 약물 유도 수면내시경(Drug-induced sleep endoscopy, DISE)은 이러한 상기도의 폐쇄 위치를 파악하는 데 사용되며, VOTE(Velum, Oropharynx, Tongue base, Epiglottis) scoring을 통해 DISE 결과를 판독한다. 하지만 VOTE scoring은 의료진이 주관적으로 각 해부학적 위치의 폐쇄 정도를 판단하기 때문에, 의료진의 해석에 따라 진단 결과가 달라질 수 있으며, 이로 인한 판독자간 판독 불일치가 상당하다. 따라서 본 연구는 딥러닝 모델을 활용하여 DISE 영상에서 VOTE 영역을 자동으로 분류하고 상기도의 완전폐쇄 여부를 예측하고자 한다. 연구개(V), 연구개 완전폐쇄(V degree 2), 구인두-혀 기저부-후두덮개(OTE), 구인두-혀 기저부-후두덮개 완전폐쇄(OTE degree 2), 왜곡(X) 총 5가지 클래스로 분류하였다. 이러한 분류 체계는 상기도의 부위별 폐쇄를 더 정확하고 체계적으로 예측할 수 있게 해 준다. 이를 위해 13명의 환자로부터 수집된 DISE 영상 데이터셋과 임상 정보를 기반으로 ResNet, DenseNet, VGG, EfficientNet과 같은 딥러닝 모델을 구현하고 성능을 비교했다. 그 결과 EfficientNet 모델이 80%의 정확도로 가장 뛰어난 성능을 보였다. 향후 연구에서는 DISE 영상에서 상기도 폐쇄 양상을 더욱 세분화하여 분석하고자 한다.

1. 서론

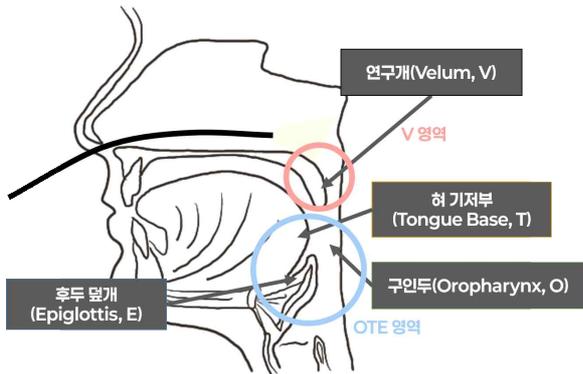
폐쇄성 수면 무호흡증(Obstructive Sleep Apnea, OSA)은 수면 중 상기도가 반복적으로 또는 완전히 막혀 기류가 감소(저호흡)하거나 호흡이 중단(무호흡)되는 현상이다[1]. 이 상태는 수면 패턴을 방해하고 만성 저산소혈증을 유발하며, 고혈압, 심혈관 질환, 인지 장애를 포함한 다양한 2차 건강 문제로 이어질 수 있다.

진단은 크게 수면다원검사(Polysomnography, PSG)와 약물 유도 수면내시경 검사(Drug-induced

sleep endoscopy, DISE)를 통해 진행된다. 먼저 PSG를 통해 무호흡-저호흡 지수(Apnea-Hypopnea Index, AHI)를 측정하여 질환의 중증도를 평가한다. 그 후, 상기도 폐쇄 부위와 폐쇄 정도를 확인하고 치료 방법을 결정하기 위해 DISE를 시행한다. DISE는 자연 수면 상태를 모방하기 위해 약물 유도 하에 수면 상태를 만들고 상기도의 해부학적 상태를 섬유광학 내시경을 사용하여 촬영하는 기술이다[2]. 상기도에서 폐쇄가 일어나는 부위를 확인할 수 있고, 환자에 대해 여러 치료 방법을 시뮬레이션하여 각 치료 방법의 효과를 본 후 치료 방법을 결정할 수 있다.

* 교신저자

DISE 검사 과정에서 의료진은 VOTE(Velum, Oropharynx, Tongue base, Epiglottis) scoring 기준을 통해 상기도 폐쇄를 평가한다. VOTE scoring 기준은 그림 1과 같이 DISE를 통해 관찰되는 해부학적 부위를 연구개(Velum, V), 구인두(Oropharynx, O), 혀 기저부(Tongue Base, T), 후두덮개(Epiglottis, E)로 구분하여 각 영역마다 폐쇄의 정도를 평가한다. 폐쇄가 없으면(폐쇄율 0~50%) degree 0, 부분 폐쇄(폐쇄율 50~75%)되었거나 진동하면 degree 1, 완전폐쇄(폐쇄율 75~100%)되었으면 degree 2로 평가한다[3]. 이러한 평가를 통해 의료진은 환자의 상기도 폐쇄를 해부학적으로 관찰하여 진단할 수 있다.



(그림 1) 상기도의 개략적 시각화

하지만 현재 VOTE 판독은 수치 측정이나 영상의 기계적 분석 없이 의료진이 주관적으로 각 해부학적 위치의 폐쇄 정도를 평가하는 방식이다. 이로 인해 의료진의 해석에 따라 진단 결과가 달라질 수 있어, 판독자간의 판독 불일치가 상당하다.

본 연구는 이러한 문제를 해결하기 위해 DISE 영상에 대해 내시경이 상기도의 어느 영역을 촬영하고 있는지 분류하고 영역별로 완전폐쇄 발생 여부를 파악하고자 한다.

2. 관련 연구

기존의 DISE 영상 분석 연구는 주로 특정 영역을 분류하거나, 폐쇄 정도를 예측하는 방식으로 진행되었다. DISE 영상을 분석하여 상기도 붕괴 부위를 자동으로 분류하기 위한 Convolutional Recurrent Neural Network(CRNN) 연구[4]는 ResNet18 아키텍처와 양방향 LSTM을 결합하여 5초 분량의 DISE 영상을 세 가지 범주(V, OTE, X)로 분류하는 작업을 수행하였다. V와 OTE 두 영역을 구분하는 분류 모델을 제안하여, DISE 영상에서 상기도 붕괴 위치,

유형 및 정도를 예측하는 가능성을 보였다.

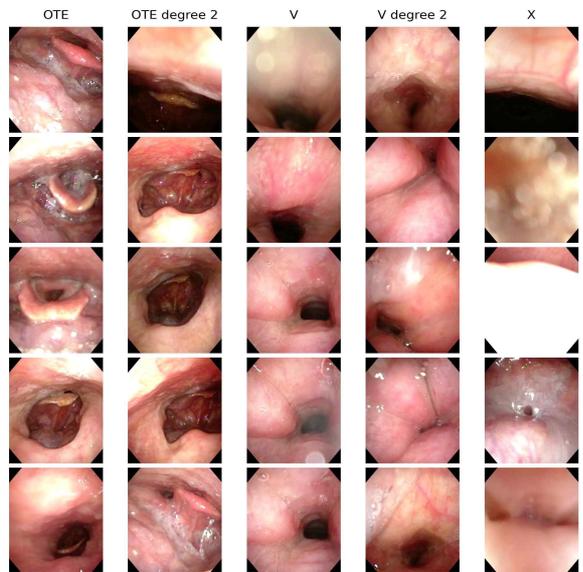
DISE 영상에서 상기도 붕괴 정도를 예측하는 연구[5]는 VOTE scoring을 사용하여 상기도의 네 가지 부위인 연구개(V), 구강 인두(O), 혀(T), 후두개(E)에 대한 붕괴 정도를 평가했으며, 각 부위에 대해 0(폐쇄 없음), 1(부분 폐쇄), 2(완전폐쇄)로 분류했다. 281개의 DISE 영상을 대상으로 각 5초 단위의 클립에 대해 VOTE 붕괴 정도를 의료진의 라벨과 비교하여 모델을 학습시켰다.

본 연구는 V와 OTE 영역을 분류하는 것뿐만 아니라, 각 영역에 대해 완전폐쇄가 발생하는 경우까지 분류하여 더 정밀한 정보를 제공하는 분류 체계를 제안한다.

3. 데이터셋

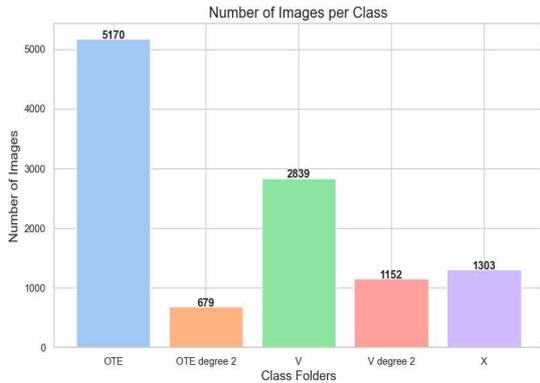
본 연구는 전남대학교 병원 이비인후과에서 폐쇄성 수면 무호흡증으로 진단받은 13명의 내시경 영상을 활용했다. 연구에 사용된 데이터는 전남대학교 병원 의료진으로부터 제공받았으며, 축적된 영상 데이터를 사용하기 위해 원내 규정에 따른 윤리적 승인을 받았다.

데이터셋은 총 11,143개의 영상으로 구성되었다. 데이터의 정확한 분류를 위해 각 영상에 대한 라벨 작업은 숙련된 의료진이 직접 수행하였다. 연구개(V), 연구개 완전폐쇄(V degree 2), 구인두-혀 기저부-후두덮개(OTE), 구인두-혀 기저부-후두덮개 완전폐쇄(OTE degree 2), 영상 왜곡으로 인해 부위를 확인할 수 없는 경우의 왜곡 클래스(X) 총 5가지 클래스로 분류하였다(그림 2).



(그림 2) 클래스별 활용한 DISE 샘플 영상

영상에 심각한 노이즈나 촬영의 결함이 있는 경우는 제외하였다. 실험에 사용된 클래스별 영상 수는 그림 3과 같다.



(그림 3) 클래스별 영상 수

데이터셋은 환자 수를 기준으로 학습, 검증, 테스트 세트, 각각 9:2:2 비율로 나누었다. 데이터 중복을 방지하고 모델의 성능을 보다 신뢰성 있게 검증하기 위해, 같은 환자의 데이터가 학습과 테스트 세트에 동시에 포함되지 않도록 구성했다.

4. 딥러닝 모델

본 연구에서는 DISE 영상 데이터셋에 대해 4가지(ResNet101, DenseNet169, VGG16, EfficientNet)의 CNN 기반 딥러닝 모델을 구현하였다.

4.1 ResNet101[6]

ResNet(Residual Network)은 심층 신경망 훈련의 어려움을 해결하기 위해 잔여(Residual) 블록을 도입한 모델이다. 잔여 블록은 신경망의 입력을 직접 출력에 더하는 방식으로, 신경망이 깊어질수록 발생할 수 있는 소실 경사 문제를 완화한다. 이 방법은 네트워크가 더 많은 레이어를 가질 수 있도록 하여, 더 복잡한 기능을 학습할 수 있게 한다.

4.2 DenseNet169[7]

DenseNet(Deeply Supervised Network)은 각 레이어가 이전 모든 레이어의 출력을 입력으로 사용하는 구조를 통해 네트워크의 정보 흐름을 향상시키고, 파라미터 효율성을 높인 모델이다. 이 모델은 조밀하게 연결된 블록을 사용하여 정보 손실을 줄이고, 학습 과정을 가속하며 더 적은 파라미터로 높은 성능을 발휘한다. 이는 영상 내의 미세한 구조적 차이와 복잡한 패턴을 학습하는 데 유리하며, 특히 다

양한 해부학적 특성을 잘 포착할 수 있다.

4.3 VGG16[8]

VGG(Visual Geometry Group Network)는 비교적 단순하지만 깊은 네트워크 구조다. 작은 필터 크기(3x3)는 상세한 영상 특징을 추출하는 데 유리하며, 여러 층을 쌓아 네트워크의 깊이를 늘리는 방식으로 복잡한 영상을 효과적으로 인식한다. 따라서 다양한 해부학적 구조를 더 세밀하게 분석할 수 있다.

4.4 EfficientNet[9]

EfficientNet은 네트워크의 깊이, 폭, 해상도를 균형있게 조정하는 복합 스케일링을 도입하여 성능을 최적화한다. 이 모델은 적은 파라미터 수로 높은 정확도를 달성하도록 설계되었다.

5. 실험 결과

모델의 성능 평가를 위해 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-Score를 성능지표로 사용했다. 표 1은 전체 데이터셋에 대해 각 딥러닝 모델의 성능을 보여준다. 실험에 적용한 4개의 모델 중, EfficientNet 모델의 정확도 및 재현율은 각각 80%, 정밀도 86%, F1-score 0.80으로 가장 우수한 성능을 보였다.

<표 1> 딥러닝 모델의 성능

성능지표 모델	정확도(%)	정밀도(%)	재현율(%)	F1-score
ResNet101	72	79	72	0.74
DenseNet169	76	82	76	0.77
VGG16	76	80	76	0.76
EfficientNet	80	86	80	0.80

EfficientNet 모델의 클래스(OTE, OTE degree 2, V, V degree 2, X)별 성능 분석 결과는 표 2와 같다. OTE와 OTE degree 2 클래스에서 매우 높은 정밀도(98%, 97%)와 재현율(84%, 90%)을 보였다. 특히 OTE degree 2 클래스에서 높은 F1-score(0.94)를 기록하여 세부적인 폐쇄 정도를 정확하게 예측하였다. 그러나 V degree 2 클래스에서는 상대적으로 낮은 재현율(48%)과 F1-score(0.62)로 완전폐쇄 상태를 구분하는 데 어려움이 있었다. 왜곡 클래스(X)에 대해서는 높은 재현율(96%)을 보였으나, 정밀도가 낮아(42%) 오 탐

지율이 높을 수 있다.

<표 2> EfficientNet 모델의 클래스별 성능

성능지표 클래스	정확도(%)	정밀도(%)	재현율(%)	F1-score
OTE	84	98	84	0.90
OTE degree 2	90	97	90	0.94
V	94	78	94	0.85
V degree 2	48	88	48	0.62
X	96	42	96	0.59

6. 결론

본 연구에서는 DISE 영상 데이터에서 VOTE 분류 및 완전폐쇄 예측에 딥러닝 모델 활용 가능성을 보여주었다. 적용한 딥러닝 모델 중 EfficientNet 모델의 성능이 가장 우수했으며, 클래스별 분류 결과 OTE와 OTE degree 2의 결과가 가장 우수했다. 하지만, 몇 가지 한계점이 여전히 남아있다. 첫째, 13명의 DISE 영상 데이터를 활용했기에 더 많은 증례를 수집할 필요가 있다. 둘째, V degree 2 및 X 클래스의 경우 다른 클래스 대비 낮은 성능 결과를 보였는데, 이는 클래스별 데이터의 불균형 문제가 원인이 될 수 있기에 향후 더 다양한 증례를 수집 후 추가 실험이 필요하다.

몇 가지 한계점에도 불구하고, 기존 연구에서 시도되지 않은 완전폐쇄 상태까지 예측할 수 있어, 임상에서의 활용도가 높을 것으로 예상된다.

향후 연구에서는 다양한 증례를 통해 모델의 견고성 및 재현성을 확립하고, VOTE 영역이 폐쇄 시 어떤 방식으로 폐쇄되는지 폐쇄 양상에 대해 분류하고자 한다.

사 사

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 인공지능융합혁신인재양성사업 연구 결과로 수행되었음(IITP-2023-RS-2023-00256629)

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행되었음 (IITP-2024-RS-2024-00437718)

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (RS-2023-00208397).

참고문헌

[1] P. Lévy, M. Kohler, W. T. McNicholas, F.

Barbé, R. D. McEvoy, V. K. Somers, ... & J. L. Pépin, "Obstructive sleep apnoea syndrome", Nature reviews Disease primers, vol. 1, pp. 1-21, 2015.

[2] E. J. Kezirian, "Drug-induced sleep endoscopy", Operative Techniques in Otolaryngology-Head and Neck Surgery, vol. 17, no. 4, pp. 230-232, 2006.

[3] E. J. Kezirian, W. Hohenhorst, and N. de Vries, "Drug-induced sleep endoscopy: the VOTE classification", European Archives of Oto-Rhino-Laryngology, vol. 268, no. 8, pp. 1233-1236, 2011.

[4] U. Hanif, E. J. Kezirian, E. K. Kiar, E. Mignot, H. B. D. Sorensen, and P. Jennum, "Upper Airway Classification in Sleep Endoscopy Examinations using Convolutional Recurrent Neural Networks", Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Mexico City, 2021, pp. 3957-3960.

[5] U. Hanif, E. K. Kiaer, R. Capasso, S. Y. Liu, E. J. Mignot, H. B. Sorensen, and P. Jennum, "Automatic scoring of drug-induced sleep endoscopy for obstructive sleep apnea using deep learning", Sleep Medicine, vol. 102, pp. 19-29, 2023.

[6] S. Jian, H. Kaiming, R. Shaoqing, and Z. Xiangyu, "Deep residual learning for image recognition", in IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016, pp. 770-778.

[7] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks", in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, 2017, pp. 4700-4708.

[8] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", arXiv preprint arXiv:1409.1556, 2014.

[9] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks", in International Conference on Machine Learning, Long Beach, 2019, pp. 6105-6114.