

약물 유도 수면내시경에서 VOTE 영역 폐쇄양상 분류 시스템

양현지^{1*}, 오현서^{2*}, 이명은³, 양형정^{1,3**}, 양형채^{2,4}

¹전남대학교 인공지능융합학과

²전남대학교 의과대학 의학과

³전남대학교 초광역 연합 의료 AI 연구센터

⁴전남대학교병원 이비인후과

yhj22@jnu.ac.kr, hyunseo.oh03@gmail.com, myungeun07@gmail.com, hhyang@jnu.ac.kr,
blessed@chonnam.ac.kr

VOTE Region Closure Pattern Classification System in Drug-Induced Sleep Endoscopy

Hyeon-Ji Yang^{1*}, Hyunseo Oh^{2*}, 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

요약

폐쇄성 수면무호흡증(Obstructive Sleep Apnea, OSA)은 수면 중 상기도 폐쇄로 인해 호흡이 저하되거나 중단되는 질환이다. 약물 유도 수면내시경(Drug-induced sleep endoscopy, DISE)은 이러한 상기도의 폐쇄 위치를 확인하는 데 유용하지만, 기존의 VOTE(Velum, Oropharynx, Tongue base, Epiglottis) scoring 방식은 판독자 간 평가 불일치가 크다는 한계를 가진다. 따라서 본 연구는 DISE 영상에서 VOTE 영역 폐쇄양상 분류 및 시계열 분석 기반 진단 보조 시스템을 개발하였다. 데이터는 전남대학교 병원에서 진단받은 532명의 환자로부터 수집된 DISE 영상을 사용하였으며, EfficientNet 모델을 기반으로 한 폐쇄양상 분류 모델은 8가지 클래스를 대상으로 80%의 정확도와 79%의 F1-Score를 나타냈다. 폐쇄양상의 시간적 변화를 분석하기 위해 추가로 Bi-LSTM(Bi-directional LSTM) 모델을 추가하였다. 적용한 결과 12가지 클래스에 대해 정확도와 F1-Score가 모두 89%로 우수한 성능을 나타냈다. 본 연구에서 개발한 시스템은 DISE 영상 진단의 객관성 및 재현성을 높여 의료진 간의 판독 편차를 감소시키고, 환자 맞춤형 치료 계획 수립의 정확성과 효율성을 높일 수 있다.

1. 서론

폐쇄성 수면무호흡증(Obstructive Sleep Apnea, OSA)은 수면 중 상기도가 반복적으로 붕괴되어 호흡이 부분적 또는 완전히 중단되는 질환이다[1]. 생리적 요인(각성 역치, 기도 확장 근육의 기능 저하)[2] 혹은 해부학적 요인(두개 안면 기형, 비만, 좁은 기도)[3]에 의해 발생할 수 있으나, 결과적으로 상기도의 해부학적 폐쇄가 핵심 병태생리이다. 효과적인 치료를 위해서는 상기도의 폐쇄 부위를 정확하게 파악하는 것이 필수적이며[4], 폐쇄 주요 부위는 연구개(Velum), 구인두(Oropharynx), 혀(Tongue),

후두개(Epiglottis) 등이 있다. 이러한 폐쇄 부위를 평가하기 위한 대표적인 방법으로 약물 유도 수면내시경(Drug-Induced Sleep Endoscopy, DISE)이 사용되며[6], VOTE 분류 시스템[5]을 기반으로 폐쇄의 위치, 형태, 정도를 평가한다.

그러나 DISE 영상을 통한 진단은 판독자 간의 평가 일관성이 낮고 렌즈 흔들림이나 점액, 타액 등으로 인한 영상 품질이 저하되어 정확성과 효율성 측면에서 한계가 있다[7]. 따라서 본 연구에서는 이러한 한계를 극복하기 위해 DISE 영상에서 VOTE 영역 폐쇄양상 분류 및 시계열 분석 기반 진단 보조 시스템을 개발하고자 한다.

이를 위해 본 연구진은 전문 의료진이 라벨링한 DISE 영상 데이터를 활용하여 EfficientNet 기반의

* 공동 제 1저자
** 교신저자

합성곱 신경망(CNN)을 통해 폐쇄양상을 분류하고, 시계열 분석이 가능한 양방향 장단기 기억(Bi-LSTM) 신경망을 통해 시간에 따른 폐쇄양상의 변화를 예측하는 시스템을 구축하였다.

2. 연구 방법

2.1 DISE 영상 폐쇄양상 분류 모델

약물 유도 수면내시경 판독 모델 구조는 EfficientNet[8] 모델을 기반으로 하였다. EfficientNet 모델은 CNN 기반 모델로, 깊이를 늘리기보다는 효율적인 모델 크기 조정을 통해 더 나은 성능을 발휘하는 네트워크다. 본 연구에서는 EfficientNet 모델이 DISE 영상에서 VOTE 영역을 효과적으로 분리하고, 기도 폐쇄의 정도와 양상을 정확히 분석할 수 있도록 설계되었다. EfficientNet 모델은 VOTE 영역에서 중요한 특징을 추출하는 데 강점을 가지며, 적은 파라미터로 높은 성능을 달성할 수 있다.

모델 학습은 CrossEntropyLoss를 손실 함수로 사용하고, 최적화 기법으로 Adam Optimizer를 사용하였다.

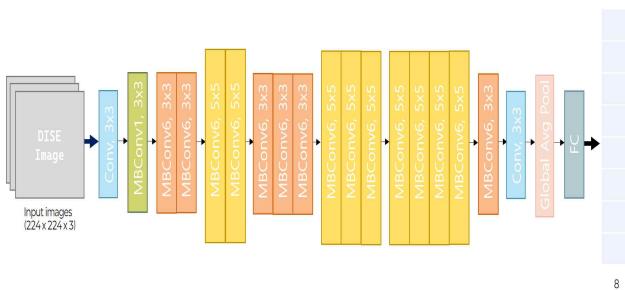


그림 1. EfficientNet 모델 구조

2.2 시계열 데이터 분석 모델

시간에 따른 기도 폐쇄양상의 시계열적 변화를 분석하기 위해, Bi-LSTM(Bi-directional LSTM)[9] 모델을 사용하였다. Bi-LSTM 모델은 양방향 시퀀스 학습을 통해, 데이터의 시간적 전후 관계를 효과적으로 반영할 수 있다. 이를 통해 폐쇄양상의 시간적 변화를 정확히 추적하고 예측할 수 있다.

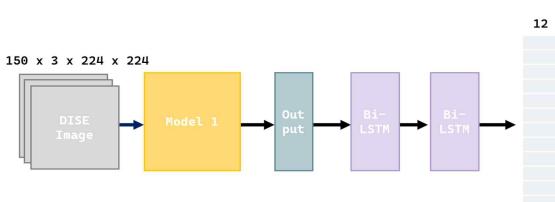


그림 2. Bi-LSTM 모델 구조

3. 실험 결과

모델 성능 평가를 위해 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1-Score를 성능지표로 사용하였다.

3.1 데이터셋

본 연구는 전남대학교 병원 이비인후과에서 폐쇄성 수면무호흡증 진단을 받은 환자 532명의 DISE 영상 데이터를 사용하였다. 사용된 데이터는 IRB 심사를 거쳐 제공받았다.

데이터셋 라벨링은 숙련된 이비인후과 의료진이 수행하였다. 연구개(V) 및 후두개(E)의 폐쇄 여부와 폐쇄양상에 따라 구분하였으며, 영상 품질 저하로 인해 정확한 분류가 어려운 경우는 ‘진단 불가(X)’ 클래스로 분류하였다.

데이터는 환자 수 기준으로 학습 238명, 검증 147명, 테스트 147명으로 나누어 구성하였다. DISE 영상 폐쇄양상 분류 모델 데이터셋은 총 26,341개의 이미지로 구성되었으며, 클래스별 분포는 V10(C) (5,862개), V10(AP) (3,080개), V10(L) (2,108개), V2 (4,330개), E10(AP) (4,219개), E10(L) (2,186개), E2 (2,340개), X (2,216개)로 총 8개의 클래스다. 시계열 데이터 분석 모델 데이터셋은 10초 길이의 영상 3,767개로 구성하였으며, 클래스별 분포는 V0(AP) (286개), V0(C) (261개), V0(L) (372개), V1(AP) (213개), V1(C) (421개), V1(L) (174개), V2 (386개), E0(AP) (221개), E0(L) (644개), E1(AP) (358개), E1(L) (118개), E2 (313개)로 총 12개의 클래스다.

3.2 DISE 영상 폐쇄양상 분류 모델 결과

EfficientNet 모델을 활용하여 DISE 영상 데이터에 대한 폐쇄양상 분류 성능을 평가한 결과는 표 1과 같다. 전체 클래스에 대한 모델의 정확도 및 F1-score는 평균 79%로 나타났다. 제시된 클래스별 성능 분석 결과, E2 클래스가 정확도 85%, F1-score 0.89로 가장 우수한 성능을 보였다. 반면, V10(L) 클래스는 정밀도가 56%로 다른 클래스 대비 상대적으로 낮은 성능을 나타냈다. 이는 해당 클래스가 다른 클래스와의 혼동 가능성성이 더 높다는 것을 나타낸다.

Confusion Matrix 결과인 그림 3을 살펴보면, 대부분의 클래스에서 높은 정확도를 보였으나, 영상 왜곡 클래스(X)의 경우 다른 클래스 대비 오분류율이 높게 나타났다. 특히 V10(AP)과 V10(C) 클래스

간에 혼동이 일부 발생하는 것을 확인할 수 있다.

<표 1> EfficientNet 모델의 클래스별 성능

성능지표 클래스 \	정확도(%)	정밀도(%)	재현율(%)	F1-score
V10(AP)	79	70	79	0.74
V10(C)	67	70	67	0.69
V10(L)	78	56	78	0.65
V2	86	76	86	0.80
E10(AP)	88	83	88	0.86
E10(L)	78	99	78	0.87
E2	85	93	85	0.89
X	69	96	60	0.80

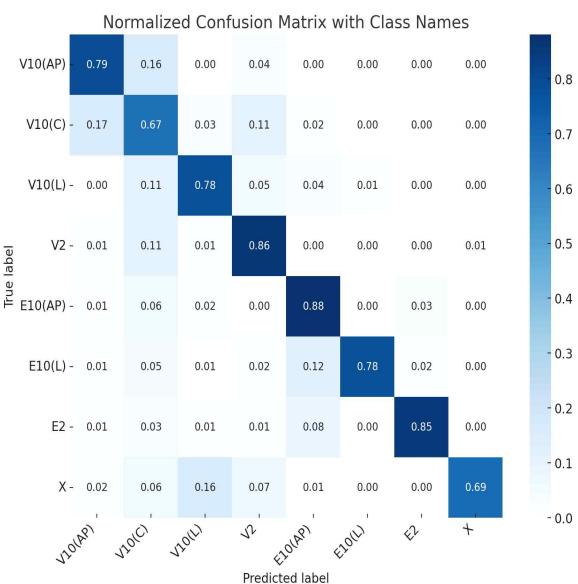


그림 3. EfficientNet 모델의 Confusion Matrix 결과

3.3 시계열 데이터 분석 모델 결과

외부 검증 데이터를 기반으로 한 Bi-LSTM 모델의 클래스별 성능 평가 결과는 표 2와 같다. 전체 클래스 평균 정확도와 F1-score는 약 89%로 나타났다. 세부적으로는 E2와 E1(L) 클래스가 정확도, 정밀도, 재현율 모두 100%(F1-score=1.0)로 가장 우수했다. 반면 V1(AP) 클래스는 정확도 77%, 정밀도 61%, 재현율 77%, F1-score 0.68로 상대적으로 성능이 낮았다.

Confusion Matrix 결과인 그림 4에 따르면 대부분의 클래스는 높은 예측 성능을 보였으나, 일부 클래스(V1(AP), V1(C)) 간의 혼동이 확인되었다.

<표 2> Bi-LSTM 모델의 클래스별 성능

성능지표 클래스 \	정확도(%)	정밀도(%)	재현율(%)	F1-score
V0(AP)	92	96	92	0.92
V0(C)	81	90	81	0.85
V0(L)	96	96	96	0.96
V0(AP)	77	61	77	0.68
V1(C)	61	70	61	0.65
V1(L)	69	92	69	0.79
V2	96	63	96	0.76
E0(AP)	100	89	100	0.94
E0(L)	99	99	99	0.99
E1(AP)	95	98	95	0.96
E1(L)	100	100	100	1.0
E2	100	100	100	1.0

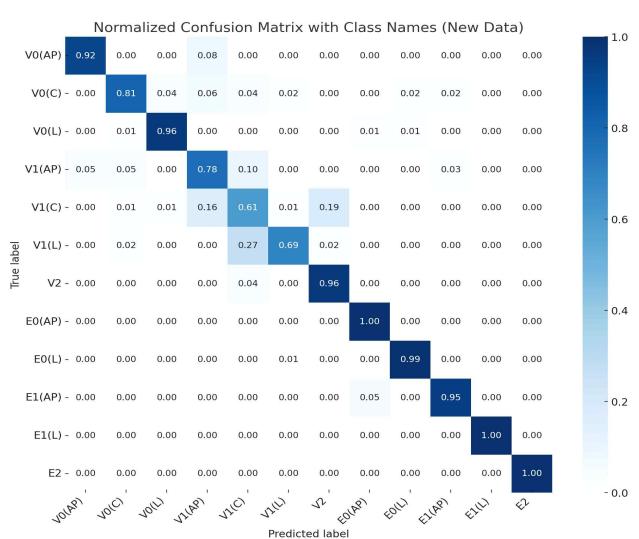


그림 4. Bi-LSTM 모델의 Confusion Matrix 결과

4. 폐쇄양상 분류 시스템 구현

본 연구에서 개발한 VOTE 영역 폐쇄양상 분류 시스템은 DISE 영상의 폐쇄양상 분류와 시계열 분석 결과를 웹 기반으로 제공한다. Flask 프레임워크 기반의 백엔드와 HTML 템플릿을 활용한 프론트엔드로 구현하였다.

개발된 시스템은 사용자가 DISE 영상을 웹사이트에 업로드하면 EfficientNet 기반 모델을 통해 영상의 상기도 폐쇄양상을 자동으로 분석하고, 그 결과를 웹페이지에서 시작화하여 제공한다. 분석 결과는 각 클래스별 분류 결과가 포함된 엑셀 파일로 다운로드 받을 수 있다. 동시에 웹 인터페이스를 통해

폐쇄양상 분류 결과를 그래프 형태로 직관적으로 확인할 수 있다. 또한, EfficientNet 모델을 통해 분류된 결과를 기반으로 Bi-LSTM 모델을 활용한 시계열 분석 결과를 추가로 제공한다.

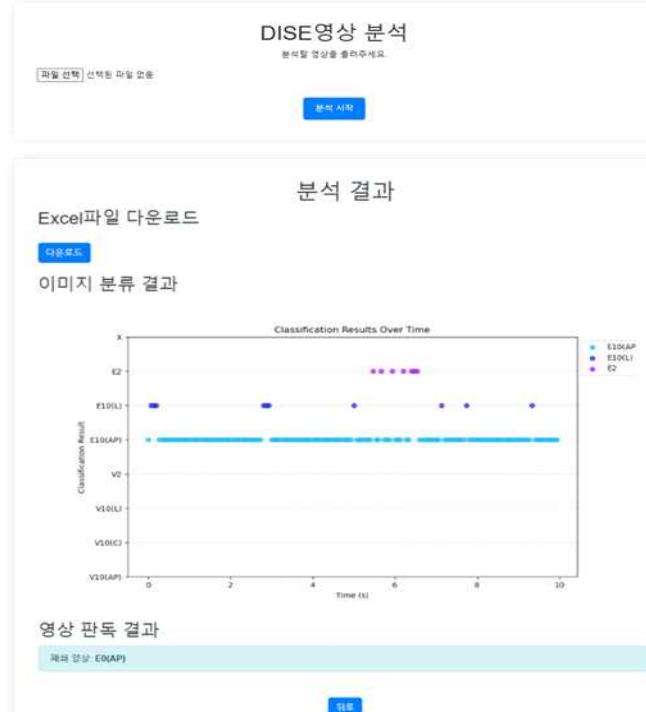


그림 5. 시스템 웹 인터페이스

5. 결론

본 연구에서는 DISE 영상에서 VOTE 영역 폐쇄 양상 분류 및 시계열 분석 기반 진단 보조 시스템을 개발하였다. 제안된 시스템은 기존 VOTE scoring 방식의 주관성과 판독자 간 일관성 부족 문제를 해결할 수 있는 객관적이고 신뢰할 만한 대안을 제시한다.

그러나 본 연구는 단일 기관에서 수집한 데이터 셋을 기반으로 수행되었기 때문에 데이터의 다양성 및 일반화 측면에서 한계가 있다. 향후 다기관의 다양한 데이터를 포함한 추가적인 외부 검증을 통해 시스템의 신뢰성과 일반화 성능을 향상시키고, 실제 임상 환경에서도 안정적으로 활용될 수 있는 시스템으로 발전시키고자 한다.

사사

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

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 인공지능융합혁신재양성사업 연구 결과로 수

행되었음(IITP-2023-RS-2023-00256629)

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터사업의 연구결과로 수행되었음 (IITP-2025-RS-2024-00437718)

참고문헌

- [1] P. Lévy, M. Kohler, W. T. McNicholas, F. Barbé, R. D. McEvoy, V. K. Somers, L. Lavie, J. Pépin, Obstructive sleep apnoea syndrome, *Nature Reviews Disease Primers*, vol. 1, no. 1, pp. 1 - 21, 2015.
- [2] A. M. Osman, S. G. Carter, J. C. Carberry, D. J. Eckert, Obstructive sleep apnea: current perspectives, *Nature and Science of Sleep*, vol. 10, pp. 21 - 34, 2018.
- [3] R. W. W. Lee, K. Sutherland, P. A. Cistulli, Craniofacial Morphology in Obstructive Sleep Apnea: A Review, *Clinical Pulmonary Medicine*, vol. 17, no. 4, pp. 189 - 195, 2010.
- [4] K. K. Green et al., Drug-Induced Sleep Endoscopy and Surgical Outcomes: A Multicenter Cohort Study, *Laryngoscope*, vol. 129, pp. 761 - 770, 2019.
- [5] E. J. Kezirian, W. Hohenhorst, N. de Vries, Drug-induced sleep endoscopy: the VOTE classification, *European Archives of Oto-Rhino-Laryngology*, vol. 268, pp. 1233 - 1236, 2011.
- [6] W. Hohenhorst, M. J. L. Ravesloot, E. J. Kezirian, N. de Vries, Drug-induced sleep endoscopy in adults with sleep-disordered breathing, *Operative Techniques in Otolaryngology*, vol. 23, no. 1, pp. 3 - 10, 2012.
- [7] E. J. Kezirian, D. P. White, A. Malhotra, W. Ma, C. E. McCulloch, A. N. Goldberg, Interrater Reliability of Drug-Induced Sleep Endoscopy, *Archives of Otolaryngology - Head & Neck Surgery*, vol. 136, no. 4, pp. 393 - 397, 2010.
- [8] 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.
- [9] A. Graves, J. Schmidhuber, "Framewise phone me classification with bidirectional LSTM and other neural network architectures," *Neural Networks*, vol. 18, no. 5-6, pp. 602 - 610, 2005.