

뇌졸중 마비 평가 자동화를 위한 CNN-GRU 시공간 특징 분석 기반의 안면마비 검출 연구

반소희*, 남효석**, 박은정**,
이화여자대학교 엘텍공과대학*, 연세대학교 의과대학**

bsh1004664@ewhain.net, hsnam@yuhs.ac, eunjeong-park@yuhs.ac

Facial Palsy Detection using CNN-GRU for Spatiotemporal Analysis in Automatic Stroke Diagnosis

Sohee Ban*, Hyosuk Nam**, Eunjeong Park**

Ewha Womans University Eltec College of Engineering*, Yonsei University College of
Medicine**

요약

본 논문에서는 딥러닝 모델을 통해 신경 손상으로 인한 뇌졸중 안면마비 평가를 자동화하는 알고리즘을 구현하고, 그 성능을 비교, 분석하였다. MTCNN 을 사용하여 이미지 내 동일한 형태의 얼굴 영역을 추출, 정렬하였으며, CNN 과 GRU 를 기반으로 안면 이미지의 시공간적 특징을 추출하였다. 제안한 모델 중 MobileNetV2-GRU 가 92%의 정확도로 동영상의 뇌졸중 안면마비 환자를 감지하였으며, Recall: 0.886, F1 Score: 0.921, AUC: 0.924 의 성능을 나타내었다. 제안된 기법인 한국인 얼굴 빅데이터 시스템과 전이학습 모델을 활용한 동영상 기반 안면마비 자동 평가 시스템은 뇌졸중 증상 이외에도 다양한 마비 증상을 일상생활에서 검출할 수 있는 모니터링 서비스에 활용될 수 있다.

I. 서론

뇌졸중은 신경 손상으로 인한 안면마비를 유발할 수 있으며, 조기에 진단하여 환자들에게 신속한 치료를 제공하는 것이 매우 중요하다[1].

안면마비 평가는 특정 표정에서의 안면 대칭에 대한 측정이 필요하며, 얼굴 주요 영역에 대한 시공간적 특징을 분석해야 한다. 안면마비 평가 자동화에 관한 연구는 얼굴 랜드마크 추출 알고리즘 기반의 머신러닝 모델과 CNN(Convolution Neural Network) 기반의 딥러닝 모델 등 여러 방법으로 시도되어 왔다[1-5]. 기존의 랜드마크 추출 기반 머신러닝 모델은 특징을 활용한 설명적 기능은 있으나, 증상을 표현할 정확한 랜드마크의 추출과 제한된 특징에 따른 한계가 있다[2]. 이에 따라 최근 의료 분야에서도 안면마비 질환의 객관적 평가를 위해 CNN 을 활용한 이미지 분석 방식이 다수 고안되었다[2-4]. 하지만 CNN 은 특정 시간 동안 발생하는 안면마비 동적 변화의 상관관계를 추출할 수는 없기 때문에 시간적 상관관계를 효과적으로 분석할 수 있는 방법이 필요하다. 따라서 본 논문에서는 정적 안면 이미지에서의 공간적 특징뿐 아니라, 안면마비의 동적 특징을 반영하기 위하여, CNN 과 GRU (Gated Recurrent Unit)를 활용한 뇌졸중 안면마비 분류를 구현하고, 그 성능을 비교, 분석한다.

II. 본론

2.1 데이터 세트

본 연구는 연세의료원 뇌졸중 집중치료실에서 수집한 환자의 안면마비 동영상 데이터 세트(Yonsei Patient Dataset)와 AI Hub 의 한국인 안면 이미지 데이터 세트(K-FACE Dataset)[6]를 이용하여 실험을 진행하였다. Yonsei Patient Dataset 의 경우, 성별, 나이, 마비 정도가 다양한 15 명의 비디오 데이터 세트로 구성되어 있다. 그중 Lux 가 낮은 비디오 1 개를 제외한

총 14 개의 비디오 데이터 세트를 실험에 이용하였으며, OpenCV 를 사용하여 등간격으로 프레임을 샘플링하였다. K-FACE Dataset 의 경우, 본 연구의 포함/배제 척도를 적용(액세서리: 착용 안 함, 일반 안경, 뿔테 안경, Light Lux: 400, 1000, 카메라 수평 각도: -15, 0, 15, 표정: 무표정, 활짝 웃음, 찡그림)하여 선별하였다.

최종적으로 Yonsei Patient Dataset 1,398 장, K-FACE Dataset 1,400 장, 총 2,798 장의 데이터를 생성하여, 7:2:1 의 비율에 따라 훈련, 검증, 테스트 데이터 세트로 나누어 모델의 훈련 및 평가에 사용하였다.

2.2 전처리

일반적으로 얼굴 검출 및 정렬 네트워크는 배경 제거, 수평 조절을 위해 모델 훈련과 추론의 초기 단계에 적용되며, Haar Cascade, HOG(Histogram of Oriented Gradients), MTCNN(Multi-Task cascaded CNN) 등 다양한 접근법이 존재한다[4]. 이때, 안면마비 증상이 없는 피실험자뿐 아니라 환자의 이미지에서도 일정 수준 이상의 정확도를 보장해야 한다. 최근 연구에서 전통적인 얼굴 정렬 기법에 비해 딥러닝 기반 얼굴 정렬은 높은 검출률을 나타내며[4], 특히 MTCNN 은 정면, 측면 모두 높은 정확도와 재현율을 보이는 것으로 연구되었다[7]. 본 실험에서는 MTCNN 을 기반으로 모든 데이터 세트에 대해 동일한 형태의 얼굴 영역을 추출, 정렬하는 전처리 과정을 수행하였다 (Fig. 1). 이후 수평 방향의 flip augmentation, min-max normalization, image resizing (222,224,3)의 전처리 작업을 수행하였다.





Fig. 1 모델 학습 및 검증에 사용된 데이터 세트 샘플

2.3 모델 구조

구현한 알고리즘은 대표적인 CNN 기반 알고리즘인 VGG16, MobileNetV2 를 백본으로 하며, 안면 이미지의 연속적인 프레임을 처리할 수 있도록 시계열 데이터 분석에 적합한 순환신경망 중 단순한 구조의 GRU 를 결합하였다. 본 연구에서 구현, 평가한 MobileNetV2 전이학습, VGG16-GRU, MobileNetV2-GRU 의 세 가지 모델은 다음과 같다.

2.3.1 MobileNetV2 전이학습

본 연구에서 전이학습을 기반으로 응급질환인 뇌졸중 안면마비의 데이터 부족과 이에 따른 모델 과적합 문제를 해결하고자 하였다. 대규모 데이터 세트에서 사전 학습된 모델의 가중치를 뇌졸중 마비 검출에 적합하도록 조정하여 MobileNetV2 전이학습 모델을 구축하였다. MobileNetV2 는 stride 가 1 인 residual block 과 downsizing 을 위한 stride 가 2 인 block, 총 두 가지의 block 으로 구성된다. FC layer 일부를 제외한 MobileNetV2 layer 를 안면 이미지의 특징 추출기로 사용한 후, 최상위층에 GlobalAveragePooling2D, 0.2 Dropout, FC layer 를 추가하여 이진 분류를 수행하였다.

2.3.2 VGG16-GRU 모델

VGG16-GRU 는 5 개의 Convolution block 과 FC layer 로 구성된 VGG16 의 FC layer 를 일부 삭제한 후 0.5 Dropout 과 GRU, FC layer 를 연결한 모델이다. VGG16 은 사전 학습된 가중치를 기반으로 인풋 프레임 내에서 안면 특징을 자동으로 추출하며, GRU 와 함께 시공간적 특징을 고려한 안면마비 유무의 정량적 평가를 수행한다.

2.3.3 MobileNetV2-GRU 모델

사전 학습된 MobileNetV2 를 백본으로 한 MobileNetV2-GRU 모델 구조는 Fig 2 와 같다. 사전 학습된 MobileNetV2 는 인풋된 프레임 내에서 얼굴의 특징을 추출하는 특징 추출기로 사용되며, GRU 와 함께 안면 이미지의 동적 특징을 분석한다.

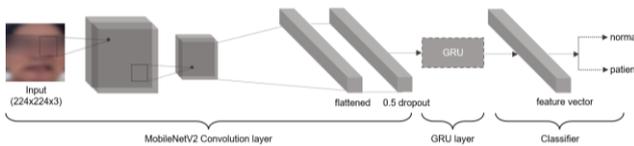


Fig. 2 MobileNetV2-GRU 모델 구조

2.4 실험 및 실험 결과

제안된 기법의 모델 구현과 성능평가를 위하여 분류 손실(cross entropy loss)을 계산하고, SGD optimizer 를 사용하여 손실 함수를 최적화하고 가중치와 편향을 갱신하였다. Cosine annealing 을 사용하여 learning rate 를 조절하고, patience 2 early stopping 을 적용하여 과적합 문제를 방지하고자 하였다. 각 모델의 epoch 에 따른 정확도는 Fig. 3 과 같다. 테스트 데이터 세트에 대한 뇌졸중 안면마비 분류 모델의 성능평가는 분류 정확도, 재현율, F1 score, AUC 로 측정하였으며, 결과는 Table 1 과 같다.

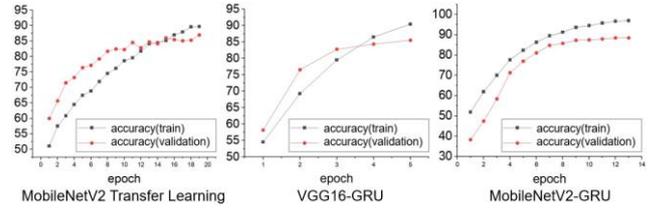


Fig. 3 학습, 검증 데이터 세트에 대한 각 모델의 정확도

Model	Accuracy	Recall	F1	AUC
MobileNetV2 TL	0.840	0.835	0.882	0.889
VGG16-GRU	0.865	0.867	0.864	0.865
MobileNetV2-GRU	0.925	0.886	0.921	0.924

Table. 1 테스트 데이터 세트에 대한 뇌졸중 안면마비 분류 모델의 성능 평가

III. 결론

본 연구는 응급증상 발현 당시의 데이터가 부족한 뇌졸중 안면마비 증상 추출을 위하여 한국인 안면 빅데이터를 활용하여 뇌졸중 안면마비 분류 모델을 구현하였다. 이미지의 공간적 특징만을 분석하는 CNN 기반의 기존 방식 대비 CNN 전이학습과 GRU 를 기반으로 안면 이미지의 시공간적 특징을 추출하여 뇌졸중 안면마비 분류 모델의 성능을 향상시켰고, 전이학습과 시계열 모델을 결합하여 데이터가 부족한 다양한 증상 검출의 기반 모델로 활용될 수 있다.

ACKNOWLEDGMENT

이 연구는 과학기술정보통신부 (NRF-2020R1A2C1013152, IITP 2021-0-00146)의 지원으로 수행되었음.

참고 문헌

- [1] Gupta, Ankit. "StrokeSave: A Novel, High-Performance Mobile Application for Stroke Diagnosis using Deep Learning and Computer Vision." arXiv preprint arXiv:1907.05358 (2019).
- [2] Wang, Ting, et al. "Automatic facial paralysis evaluation augmented by a cascaded encoder network structure." IEEE Access 7 (2019): 135621-135631.
- [3] Sajid, Muhammad, et al. "Automatic grading of palsy using asymmetrical facial features: a study complemented by new solutions." Symmetry 10.7 (2018): 242.
- [4] Bandini, Andrea, et al. "A new dataset for facial motion analysis in individuals with neurological disorders." IEEE Journal of Biomedical and Health Informatics 25.4 (2020): 1111-1119.
- [5] Mohana, Mohamed, Salem Alelyani, and Mohammed Saleh Alsaqer. "Fused Deep Neural Network based Transfer Learning in Occluded Face Classification and Person re-Identification." arXiv preprint arXiv:2205.07203 (2022).
- [6] Choi, Yeji, et al. "K-face: A large-scale kist face database in consideration with unconstrained environments." arXiv preprint arXiv:2103.02211 (2021).
- [7] 김현곤, et al. "드라마 동영상에서 강건한 등장 인물 검출을 위한 얼굴 검출 라이브러리의 조합 방법." 한국정보과학회 학술발표논문집 (2018): 905-907.