

음성 신호와 심층 잔류 순환 신경망을 이용한 파킨슨병 진단

Parkinson's disease diagnosis using speech signal and deep residual gated recurrent neural network

신승수,¹ 김지연,¹ 구본미,² 김형국[†]

(Seung-Su Shin,¹ Gee Yeun Kim,¹ Bon Mi Koo,² and Hyoung-Gook Kim^{1†})

¹광운대학교 전자융합공학과, ²광운대학교 SSK 정신건강과 지역사회 연구센터

(Received February 25, 2019; accepted April 19, 2019)

초 록: 노년기 3대 질환 중 하나인 파킨슨병은 환자의 70% 이상이 음성 장애를 앓고 있으며 최근 음성 신호를 통한 파킨슨병의 진단 방법들이 고안되고 있다. 본 논문에서는 음성 특징을 이용한 심층 잔류 순환 신경망 기반의 파킨슨병 진단 방식을 제안한다. 제안하는 방식에서는 파킨슨병 진단을 위한 음성 특징을 선택하고 이를 심층 잔류 순환 신경망에 적용하여 파킨슨병 환자를 식별한다. 제안하는 심층 잔류 순환 신경망은 심층 순환 신경망에 잔류 학습 방식을 결합한 알고리즘으로 파킨슨병 진단에서 기존의 식별 알고리즘보다 더 높은 인식률을 보인다.

핵심어: 파킨슨병, 음성 신호, 음성 특징, 심층 잔류 순환 신경망

ABSTRACT: Parkinson's disease, one of the three major diseases in old age, has more than 70% of patients with speech disorders, and recently, diagnostic methods of Parkinson's disease through speech signals have been devised. In this paper, we propose a method of diagnosis of Parkinson's disease based on deep residual gated recurrent neural network using speech features. In the proposed method, the speech features for diagnosing Parkinson's disease are selected and applied to the deep residual gated recurrent neural network to classify Parkinson's disease patients. The proposed deep residual gated recurrent neural network, an algorithm combining residual learning with deep gated recurrent neural network, has a higher recognition rate than the traditional method in Parkinson's disease diagnosis.

Keywords: Parkinson's disease, Speech signal, Speech feature, Deep residual gated recurrent neural network

PACS numbers: 43.72.Ar, 43.72.Bs

1. 서 론

파킨슨병은 뇌의 신경세포 손상으로 인해 발생하는 신경퇴행성질환으로, 뇌졸중, 치매와 함께 노년기 3대 질환 중 하나이다.^[1] 전 세계적으로 2016년 기준 약 600만 명 이상의 환자가 이 질환을 앓고 있으며, 인구고령화에 따라 유병률이 증가하는 경향을 보이고 있다.^[2] 파킨슨병의 주요 증상은 운동 증상과 비운동 증상으로 구분되며, 주요 운동 증상은 떨림,

운동의 느려짐, 경축, 자세의 불안정, 보행이상 등이며, 비운동 증상으로는 인지 기능 장애, 정신 증상, 수면 장애 등이 나타난다.^[3] 또한 파킨슨병 환자들의 약 70% 이상은 운동기능저하로 인한 음성장애를 경험한다고 보고되고 있다.^[4,5] 대표적으로 단조로운 음도, 강세감소, 단조로운 강도, 거친 소리, 기식음, 부정확한 자음, 부적절한 쉬, 속도변이성 등의 특징을 보인다.^[6] 이러한 음성 특징을 감지하는 음성 신호 분석이 파킨슨병의 진단을 위해 활용되고 있으며^[7] 기계학습 방식의 발달로 음성분석의 진단 정확도가 점차 향상되고 있다.

음성 신호 분석을 통한 파킨슨병 진단의 기존 연

[†]Corresponding author: Hyoung-Gook Kim (hkim@kw.ac.kr)
Department of Electronics Convergence Engineering, KwangWoon University, 20 Gwangun-ro, Nowon-gu, Seoul 01897, Republic of Korea
(Tel: 82-2-940-5574, Fax: 82-2-913-5006)

구들^[7-12]을 살펴보면, 파킨슨병 진단과 파라미터로 제안된 F0, Shimmer, Jitter, MPT, HNR^[8]뿐만 아니라, 다양한 음성 파라미터들을 알고리즘에 적용하여 파킨슨병을 진단하였다. 파킨슨병 진단을 위한 음성 분석 연구^[7]에서는 음성 특징 선택 과정에서 GA(Genetic Algorithm)를 통해 Fhi(Hz), Fho(Hz), RAP, APQ5 등의 특징 값들을 선택하여 SVM(Support Vector Machine) 식별 방식에 적용하였고, 음성 특징을 통한 파킨슨병 식별 알고리즘의 성능 비교 연구^[11]에서는 F0(Hz), Jitter(%), RAP, PPQ, Shimmer 등의 선택된 특징 값들을 J48 및 REFTree 식별 알고리즘에 적용하여 성능을 비교 분석하였다.

기존 연구에서 적용한 식별 알고리즘은 데이터와 레이블을 이용하여 모델을 학습시키고 식별하는 기계학습 방식이다. 이러한 기계학습 방식 중의 하나인 심층 신경망은 데이터의 특징과 패턴을 학습하는 방식으로 기존의 식별 알고리즘보다 높은 인식률을 제공한다. 하지만, 인식률 개선을 위해 신경망의 깊이가 깊어지게 되어, 계산량의 증가와 함께 gradient exploding/vanishing 문제가 발생하였다.^[13] 이와 같은 제한점을 해결하기 위해 잔류 학습 방식이 개발되었고 심층 신경망에 잔류 학습 방식을 결합한 모델이 깊은 신경망 구조에서도 뛰어난 성능을 보여주고 있다.^[14]

본 논문에서는 음성 신호로부터 추출한 음성 특징과 심층 순환 신경망에 잔류 학습 방식을 결합한 심층 잔류 순환 신경망(Deep Residual Gated Recurrent Neural Network, DRGRNN)을 이용하여 파킨슨병 진단의 정확도를 높이는 방식을 제안한다.

본 논문의 구성은 다음과 같다. II장에서는 제안하는 파킨슨병 진단 방식에 관해 설명한다. III장에서는 DRGRNN 기반의 파킨슨병 진단 결과를 제시한다. 마지막으로 IV장에서 본 논문의 결론을 맺는다.

II. 파킨슨병 진단 방식의 구조

Fig. 1은 본 논문에서 제안한 음성 신호 분석을 통한 파킨슨병 진단 방식의 전체 구조도이다. 전체 구조도는 모델 학습부와 식별부로 구성된다. 모델 학습부에서는 데이터 희소성 문제를 해결하기 위해 데이터 증강 방식을 사용하여 학습 데이터를 확장한다.

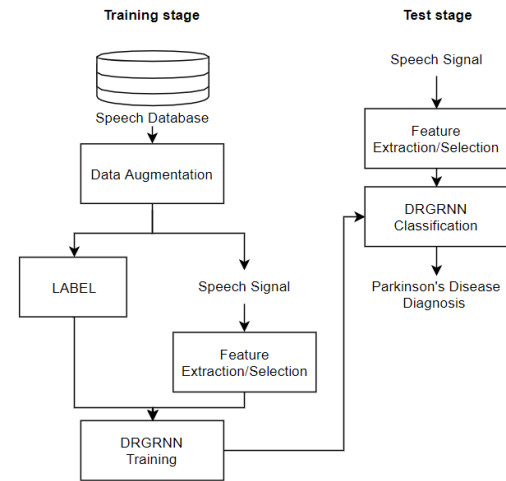


Fig. 1. Framework of training and testing procedure for proposed system.

그리고 음성 신호로부터 음성 특징들을 추출한 후, 파킨슨병 진단에 유용한 특징들을 선택한다. 선택된 음성 특징들은 DRGRNN에 적용되어 학습을 통해 파킨슨병 식별 모델을 생성한다. 식별부에서는 환자 혹은 일반인의 음성 신호로부터 음성 특징을 추출하고 식별 모델이 적용된 DRGRNN에 입력되어 파킨슨병의 유무를 판단한다.

2.1 음성 특징 선택 방식

Table 1은 본 연구에 투입된 음성 특징에 대한 설명이다. 본 논문에서는 22개의 음성 특징 중 일부를 조합하여 사용하였다. 파킨슨병 진단에 유용한 음성 특징을 선택하기 위해 심층 신경망 구조를 가진 GRU(Gated Recurrent Unit)^[13], LSTM(Long-Short Term Memory)^[15], DRGRNN 등의 각 방식에 단일 음성 특징을 적용하여 파킨슨병 식별 정확도를 비교하였다. 세 가지 식별 알고리즘에서 spread1, F0(Hz), PPE, Jitter(%), RAP, DDP 순으로 6개의 음성 특징들이 높은 파킨슨병 식별 정확도를 제공하였다.

따라서 본 논문에서는 파킨슨병 진단을 위한 음성 특징으로 F0(Hz), Jitter(%), RAP, DDP, spread1, PPE를 선택하였고 자세한 설명은 다음과 같다.

F0(Hz)는 기본 주파수의 평균값으로 목소리의 높낮이를 측정한다. 주파수는 $F_0 = \frac{1}{T_0}$ 이며 T_0 는 신호의 주기를 의미한다. spread1은 기본 주파수의 비선형

측정값으로 음성 신호의 잡음을 측정하기 위해 사용된다.^[16]

Jitter(%)는 단위 결과 값 동안 발음에서 성대의 진동으로 인한 음도의 변화를 %로 표현한 값으로 음도의 불규칙한 변이를 측정하며 계산식은 다음과 같다.^[17]

$$Jitter(\%) = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} |T_i - T_{i-1}|}{\frac{1}{N} \sum_{i=1}^N T_i}, \quad (1)$$

여기서 T_i 는 i 번째 윈도우의 기본 주파수 주기를 말하며 N 은 전체 윈도우의 개수이다.

RAP는 인접한 두 개의 구간에서 음도의 변동 값의 상대적인 변화 값으로 Eq. (2)와 같고 DDP는 인접한 Jitter의 2차 미분 값으로 Eq. (3)과 같이 정의된다.

Table 1. List of extracted speech features from speech signal.

| Features | Description |
|--------------|--|
| F0 (Hz) | Average vocal fundamental frequency |
| Fhi (Hz) | Maximum vocal fundamental frequency |
| Flo (Hz) | Minimum vocal fundamental frequency |
| Jitter (%) | MDVP jitter in percentage |
| Jitter (Abs) | MDVP absolute jitter in ms |
| RAP | MDVP relative amplitude perturbation |
| PPQ | MDVP five-point period perturbation quotient |
| DDP | Average absolute difference of differences between jitter cycles |
| Shimmer | MDVP local shimmer |
| Shimmer (dB) | MDVP local shimmer in dB |
| APQ3 | Three-point amplitude perturbation quotient |
| APQ5 | Five-point amplitude perturbation quotient |
| APQ | MDVP 11-point amplitude perturbation quotient |
| DDA | Average absolute differences between the amplitudes of consecutive periods |
| NHR | Noise-to-harmonics ratio |
| HNR | Harmonics-to-noise ratio |
| RPDE | Recurrence period density entropy measure |
| D2 | Correlation dimension |
| DFA | Signal fractal scaling exponent of detrended fluctuation analysis |
| spread1 | Nonlinear measures of fundamental frequency |
| spread2 | Nonlinear measures of fundamental frequency |
| PPE | Pitch Period Entropy |

RAP, DDP는 음도 변동에 의한 음질의 변화를 측정하는 파라미터로 음성 측정의 정확도를 높이기 위해 음도 변동 량을 다양하게 조합하여 준주기성 관련 음질을 정량적으로 나타내는 역할을 한다.

$$RAP = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} \left| T_i - \left(\frac{1}{3} \sum_{n=i-1}^{i+1} T_n \right) \right|}{\frac{1}{N} \sum_{i=1}^N T_i}, \quad (2)$$

$$DDP = \frac{\frac{1}{N-2} \sum_{i=2}^{N-1} |(T_{i+1} - T_i) - (T_i - T_{i-1})|}{\frac{1}{N} \sum_{i=1}^N T_i}. \quad (3)$$

PPE는 음도 변이를 측정할 때 주변 소음이나 정상인에게 발생하는 변이를 제거한 엔트로피로서 Eq. (4)와 같으며 음도의 비정상적인 변이 측정의 정확도를 높여준다.

$$PPE = - \sum_r P(r) \log(P(r)), \quad (4)$$

여기서 $P(r)$ 은 상대 반음 변이 발생에 대한 이산 확률 분포이다.^[1]

2.2 심층 잔류 순환 신경망

본 논문에서는 선택된 음성 특징들을 적용하여 파킨슨병을 진단하기 위해 파킨슨병 식별 알고리즘으로 DRGRNN을 사용한다.

Fig. 2는 본 논문에서 사용한 DRGRNN의 전체 구조이다. 제안한 DRGRNN은 3개의 Residual GRNN (Gated Recurrent Neural Network) Layer와 파킨슨병 증상 식별을 위해 적용된 ReLU(Rectified Linear Units)로 구성되며 Residual GRNN Layer는 3개의 GRNN과 2개의 잔류연결로 구성된다.

GRNN은 LSTM에 비해 상대적으로 간단한 구조의 GRU를 사용하여 RNN(Recurrent Neural Network)의 gradient exploding/vanishing 문제를 방지한다.^[14] LSTM은 3개의 게이트를 사용하여 내부 셀 유닛의 정보 흐름을 제어하지만 GRU는 업데이트 게이트 z 와 리셋

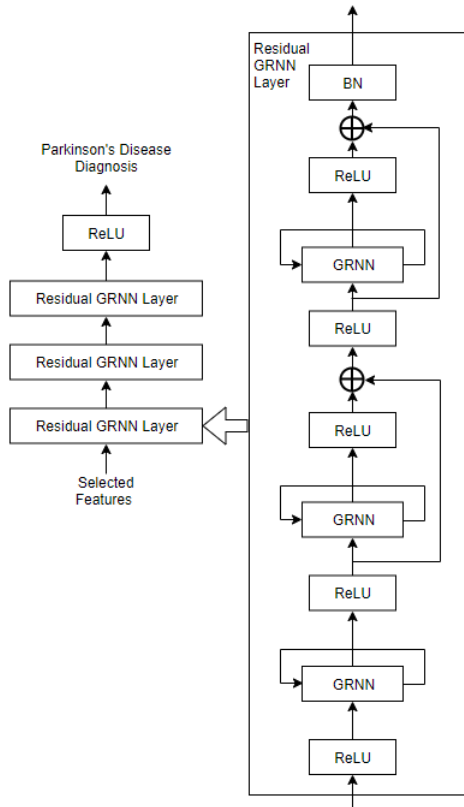


Fig. 2. Block diagram of DRGRNN.

게이트 r 로 구성된 2개의 게이트만을 사용한다. 업데이트 게이트는 메모리로 유입되는 정보를 제어하고 리셋 게이트는 메모리에서 유출되는 정보를 제어한다. 또한 GRU는 별도의 메모리 유닛 없이 게이트 유닛을 통해 유닛 내부의 정보 흐름을 변조한다. 이는 다음과 같이 나타낼 수 있다.

$$z_t = \sigma(R_z[h_{t-1}, x_t] + b_z), \quad (5)$$

$$r_t = \sigma(R_r[h_{t-1}, x_t] + b_r), \quad (6)$$

$$\tilde{h}_t = \tanh(R_h[r_t \circ h_{t-1}, x_t] + b_h), \quad (7)$$

$$h_t^{GRNN} = (1 - z_t) \circ h_{t-1} + z_t \circ \tilde{h}_t, \quad (8)$$

여기서 x_t 와 h_t, \tilde{h}_t, b 는 각각 현재의 입력, 활성화 함수, 후속 활성화 함수, 편향 벡터를 나타낸다. 또한 R 과 $\sigma(\cdot)$, \circ 는 가중치 행렬과 요소별 논리 시그모이

드 함수, 요소별 곱셈 연산을 나타낸다.

GRNN의 잔류연결은 특정 레이어에 입력된 데이터를 다음 레이어에 직접 전달하기 때문에 **gradient vanishing** 문제를 효과적으로 해결한다. 또한 잔류 학습 방식은 덧셈 연산자를 통해 **gradient**가 제한적인 레이어를 우회하여 통과하기 때문에 학습 연산이 간단해진다는 장점이 있다. 잔류연결의 연산은 다음과 같이 나타낼 수 있다.

$$h_{k+1}^{DRGRNN} = h_k^{GRNN} + \Lambda(h_k^{GRNN}, x_k; R_k), \quad (9)$$

여기서 h_k 와 h_{k+1} 은 k 번째 유닛의 입력과 출력을 나타내며 Λ 은 가중치 파라미터인 R_k 를 이용한 잔류 함수이다.

잔류 학습 방식은 **gradient**를 방해하지 않고 잔류 연결로 구성된 레이어의 출력을 정제하는데 기여하기 때문에 학습 성능을 높이면서 깊은 신경망 구성을 가능하게 한다.^[14]

III. 실험

3.1 실험 데이터

본 논문에서는 파킨슨병 환자와 일반인을 식별하기 위한 음성 특징 데이터로 UCI Machine Learning Repository^[11]에서 제공된 데이터베이스를 사용하였다.

데이터는 24명의 파킨슨병 환자를 포함한 총 32명의 피 실험자들로부터 수집된 195개 데이터로 구성되어 있다. 음성 특징은 음성 분석 기기 MDVP(Multi Dimensional Voice Program)를 이용하여 음성 신호로부터 추출되었으며^[11] 피 실험자의 번호와 파킨슨병 유무를 나타내는 이름과 상태 변수를 제외한 22개의 연속/비연속적인 음성 특징들로 구성되어 있다. 본 논문에서는 식별 알고리즘의 성능을 측정하기 위해 3-fold 교차 유효성 검증 방법을 사용하였다. 이 검증 방법에서는 데이터의 개수를 동일하게 3개 그룹으로 분할한 후, 각 그룹을 테스트 세트로 선택하고 나머지 2개의 그룹을 학습 세트로 사용한다. 본 논문에서는 파킨슨병 환자 대 일반인의 비율을 모든 그룹에 동일하게 적용하였으며 총 3번의 실험을 통해 각

Table 2. Results of Parkinson's patient classification through speech features.

| Methods | Recognition accuracy |
|---------|----------------------|
| REPTree | 84.3 % |
| kNN | 89.2 % |
| GA+SVM | 93.6 % |
| LSTM | 96.2 % |
| GRU | 96.5 % |
| DRGRNN | 98.4 % |

그룹에 대한 정확도를 측정하고, 평균을 내어 식별 알고리즘의 성능을 측정하였다.

3.2 실험 결과

실험은 음성 특징 선택부에서 선택한 6개의 음성 특징 F0(Hz), Jitter(%), RAP, DDP, spread1, PPE를 DRGRNN 식별 방식에 적용하여 성능을 테스트하고 기존의 식별 알고리즘과 비교하였다.

Table 2는 실험 결과를 나타낸다. 기존의 식별 알고리즘 중에서는 GA와 SVM을 결합한 모델의 정확도가 93.6%로 REPTree, kNN(k-Nearest Neighbor)¹²⁾보다 높게 나타났으며, 심층 신경망 구조의 LSTM과 GRU의 정확도는 각각 96.2%와 96.5%로 GA와 SVM을 결합한 기존의 알고리즘보다 약 2.6%, 2.9% 향상되었다. 또한 LSTM보다 비교적 간단한 구조를 사용하는 GRU의 정확도가 약 0.3% 정도 높게 나타났다. 최종적으로, 본 논문에서 제안한 DRGRNN은 98.4%로 비교한 방식들 중 가장 뛰어난 정확도를 나타내었다.

IV. 결 론

본 논문에서는 음성 특징을 통한 심층 잔류 순환 신경망 기반의 파킨슨병의 진단 방식을 제안하였다. 실험은 파킨슨병 진단을 위해 음성 특징으로 F0(Hz), Jitter(%), RAP, DDP, spread1, PPE를 선택하였으며 파킨슨병 식별 알고리즘으로 GRNN에 잔류 학습 방식을 결합한 DRGRNN을 적용하여 분석하였다. 그 결과 식별 정확도가 98.4%로 기존 방식보다 더 높은 정확도로 파킨슨병을 진단하였다. 이를 바탕으로 향후 파킨슨병뿐만 아니라 다양한 생체 신호를 통한 질병 식별 방법에 관해 연구할 것이다.

감사의 글

본 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(NRF-2018RID1A1B07041783). 그리고 본 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2018S1A3A2074955).

References

1. M. A. Little, P. E. McSharry, E. J. Hunter, J. Spielman, and L. O. Ramig, "Suitability of dysphonia measurements for telemonitoring of Parkinson's disease," *IEEE Trans. on Biomedical Engineering*, **56**, 1015-1022 (2009).
2. T. K. Karikari, A. Charway-Felli, K. Höglund, K. Blennow, and H. Zetterberg, "Commentary: global, regional, and national burden of neurological disorders during 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015," *Front. Neurol.* **9**, 1-2 (2018).
3. J. Jankovic, "Parkinson's disease: clinical features and diagnosis," *J. Neurol. Neurosurg. Psychiatry*, **79**, 368-376 (2008).
4. M. R. McNeil, *Clinical Management of Sensorimotor Speech Disorders* (Thieme, New York, 2009), pp. 166-186.
5. L. Hartelius and P. Svensson, "Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: a survey," *Folia Phoniatr Logop.* **46**, 9-17 (1994).
6. H. H. Kim, M. S. Lee, S. W. Kim, S. H. Choi, and W. Y. Lee, "An auditory-perceptual rating scale of dysarthric speech of patients with parkinsonism" (in Korean), *Phonetics and Speech Sciences*, **11**, 39-49 (2004).
7. M. Shahbakhti, D. T. Far, and E. Tahami, "Speech analysis for diagnosis of Parkinson's disease using genetic algorithm and support vector machine," *J. Biomedical Science and Engineering*, **7**, 147-156 (2014).
8. R. J. Holmes, J. M. Oates, D. J. Phyland, and A. J. Hughes, "Voice characteristics in the progression of Parkinson's disease," *International Journal of Language & Communication Disorders*, **35**, 407-418 (2000).
9. Y. A. Kang, K. C. Yoon, C. J. Seong, and H. Y. Park, "A preliminary study of the automated assessment of prosody in patients with Parkinson's disease" (in Korean), *The Korean Academy of Speech-Language Pathology and Audiology*, **17**, 234-248 (2012).
10. Y. M. Go, D. Y. Kim, Y. L. Choi, and H. H. Choi, "Speech rate and pause characteristics in patients with

- Parkinson's disease" (in Korean), *Phonetics and Speech Sciences*, **2**, 173-184 (2010).
11. J. W. Chung, "Performance comparison of algorithm through classification of Parkinson's Disease according to the speech feature" (in Korean), *J. Korea Multimedia Society*, **19**, 209-214 (2016).
 12. D. Gupta, A. Julka, S. Jain, T. Aggarwal, A. Khanna, N. Arunkumar, and V. H. C. de Albuquerque, "Optimized cuttlefish algorithm for diagnosis of Parkinson's disease," *Cogn. Syst. Res.* **52**, 36-48 (2018).
 13. B. Yue, J. Fu, and J. Liang, "Residual recurrent neural networks for learning sequential representations," *Information*, **9**, 56;doi:10.3390 (2018).
 14. K. Zhang, W. Zuo, S. Member, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: residual learning of deep CNN for image denoising," *IEEE Trans. on Image Processing*, **26**, 3142-3155 (2017).
 15. O. Levy, K. Lee, N. FitzGerald, and L. Zettlemoyer, "Long short-term memory as a dynamically computed element-wise weighted sum," *ACL*, **2**, 732-739 (2018).
 16. Y. Wu, P. Chen, Y. Yao, X. Ye, Y. Xiao, L. Liao, M. Wu, and J. Chen, "Dysphonic voice pattern analysis of patients in Parkinson's Disease using minimum interclass probability risk feature selection and bagging ensemble learning methods," *J. Comput. Math Methods Med.* 4201984, 11 (2017).
 17. R. A. Shirvan and E. Tahami, "Voice analysis for detecting Parkinson' disease using genetic algorithm and KNN classification method," *18th ICBME*, 278-283 (2011).

▶ 구 본 미 (Bon Mi Koo)



2015년 ~ 2016년: 미국 University of North Carolina at Chapel Hill, 의과대학, 연구인턴
 2016년 ~ 2018년: 미국 University of North Carolina at Chapel Hill, 의과대학, 연구원
 2018년 11월 ~ 현재: 광운대학교 SSK 정신건강과 지역사회 연구센터 연구교수

▶ 김 형 국 (Hyoung-Gook Kim)



1999년 ~ 2002년: 독일 SIEMENS/ Cortologic AG 책임연구원
 2002년 ~ 2005년: 독일 베를린 공과대학교 Assistant Professor
 2005년 ~ 2007년: 삼성종합기술원 수석연구원
 2007년 3월 ~ 현재: 광운대학교 전자융합공학과 교수

저자 약력

▶ 신 승 수 (Sueng-Su Shin)



2019년 2월: 광운대학교 전자융합공학과 학사
 2019년 3월 ~ 현재: 광운대학교 전자융합공학과 석박통합과정

▶ 김 지 연 (Gee Yeun Kim)



2018년 2월: 광운대학교 전자융합공학과 학사
 2018년 3월 ~ 현재: 광운대학교 전자융합공학과 석사과정