

PET-CT 영상 알츠하이머 분류에서 유전 알고리즘 이용한 심층학습 모델 최적화

이상협[†], 강도영^{**}, 송종관^{***}, 박장식^{****}

Optimization of Deep Learning Model Using Genetic Algorithm in PET-CT Image Alzheimer's Classification

Sanghyeop Lee[†], Do-Young Kang^{**}, Jongkwan Song^{***}, Jangsik Park^{****}

ABSTRACT

The performance of convolutional deep learning networks is generally determined according to parameters of target dataset, structure of network, convolution kernel, activation function, and optimization algorithm. In this paper, a genetic algorithm is used to select the appropriate deep learning model and parameters for Alzheimer's classification and to compare the learning results with preliminary experiment. We compare and analyze the Alzheimer's disease classification performance of VGG-16, GoogLeNet, and ResNet to select an effective network for detecting AD and MCI. The simulation results show that the network structure is ResNet, the activation function is ReLU, the optimization algorithm is Adam, and the convolution kernel has a 3-dilated convolution filter for the accuracy of dementia medical images.

Key words: Alzheimer's Disease Classification, Genetic Algorithm, Deep Learning, ResNet.

1. 서 론

최근 과학과 의학의 발전으로 인간의 삶의 질이 향상됨에 따라 수명도 연장되고 있다. 우리나라의 경우 고령화 현상이 매우 빠르게 진행되고 있으며 고령 인구가 매년 증가하고 있다. 고령화 사회를 준비할 시점에 이르고 있다. 고령화 사회에서 크게 대두되는 문제 중 하나가 노인성 질환이다. 급속히 고령화 사회가 진행됨에 따라 노인성 질환도 급증하고 있다. 노인성 질환인 치매는 65세 이상 노인에게 발병률이 크게 증가하는데 그 수는 매년 크게 증가한

다. 우리나라에서도 치매환자의 수가 2030년에 100만 명, 2050년에는 200만 명에 이를 것으로 예측된다. 치매는 장기적인 치료가 필요하고 사회적, 경제적 부담이 큰 질환이며, 치매로 인한 노인성 질환에 대한 진료비의 약 30%를 차지하는 수준으로 2017년 치매의 진료비는 2조 9,226억원이며 진료 건수는 501만 9,701건이다[1].

치매는 완치가 아닌 치매 증상을 경감시키고 억제하는 치료만이 진행되고 있다. 따라서 조기 발견을 통하여 치매의 진행을 늦추는 것이 필요하다. 최근 이러한 치매 문제를 해결하기 위해 딥러닝 영상분석

* Corresponding Author : Jangsik Park, Address: (48434) Suyoung-ro 309, Nam-gu, Busan, Korea, TEL : +82-51-663-4768, FAX : +82-51-663-4768, E-mail : jsipark@ks.ac.kr

Receipt date : Jan. 6, 2020, Revision date : Jul. 29, 2020
Approval date : Aug. 19, 2020

[†] Dept. of Electronic Eng., Kyungshung University
(E-mail : soundslock05@naver.com)

^{**} Dept. of Nuclear Medicine, Donga University College of Medicine (E-mail : dykang@dau.ac.kr)

^{***} Dept. of Electronic Eng., Kyungshung University
(E-mail : jsipark@ks.ac.kr)

^{****} Dept. of Electronic Eng., Kyungshung University

※ This research was supported by the National Research Foundation (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2018R1A2B2008178) & Brain Busan 21+ Program

을 이용하여 다양한 연구가 진행되고 있다.

Suk 등은 Autoencoder를 통해 치매환자(Alzheimer's disease, AD), 경도인지장애 환자(mild cognitive impairment, MCI) 그리고 정상인(normal control, NC) 3가지 클래스(Class)로 분류하는 방법을 제안하였다. 저수준의 특징을 가진 패턴을 결합하여 치매 분류에 강력한 모델을 제안하였다[2].

Zhang 등은 뇌질환 관련 다중 영상은 일반적으로 불완전한 단점이 있다. 추정을 통해 불완전한 단점을 해결하고 치매환자를 분류하는 방법을 제안하였다[3].

Carlos 등은 치매분류를 위해 각 입력에 대해 최적화 된 분류기의 출력을 결합하여 새로운 분류기를 통해 분류하는 방법을 제안하였다[4].

Chen 등은 기존의 단면만을 사용하는 치매환자 분류에서 컨볼루션 네트워크(convolution neural network)와 재귀 신경망(recursive neural network)을 결합하여 3차원 모델링을 생성하여 치매 환자를 분류하는 방법을 제안하였다[5].

Gariali 등은 뇌 영상에서 해부학적 관심영역 116 개를 지정하고 이를 21개의 영역을 분할하여, 서포트 벡터 머신(support vector machine, SVM)으로 치매 환자 분류를 제안한다[6].

본 논문에서는 효과적인 치매 분류를 위하여 의료 영상에 데이터셋에 적합한 심층학습 네트워크와 파라미터를 사전 실험을 통해 선정하고, 유전 알고리즘을 활용하여 심층학습 모델에 대한 최적의 파라미터를 찾는 방법을 제안한다. 일반적으로 심층학습 기법의 검출 또는 분류 성능은 대상 영상, 심층학습 네트워크 구조(architecture), 컨볼루션 커널(convolution kernel), 활성화함수(activation function), 최적화 알고리즘(optimization algorithm) 등에 의해 결정된다[7]. 본 논문에서는 성능비교를 위해 VGG-16, GoogLeNet 그리고 ResNet 네트워크 구조에 따른 성능을 비교분석하고, 활성화함수와 최적화 알고리즘에 따른 학습과 정확도를 비교 분석한다. 컨볼루션 커널 비교 분석은 가장 성능이 좋은 네트워크 구조, 활성화함수 그리고 최적화 알고리즘을 선정 후에 모든 컨볼루션 계층(convolution)에 팽창 컨볼루션 커널(dilated convolution kernel)을 적용에 따른 성능을 비교 분석하고, 활성화함수, 최적화 알고리즘 그리고 컨볼루션 커널을 유전 알고리즘의 유전자로 선택하여 성능을

사전실험과 비교분석하여, 의료 영상에 적합한 활성화함수, 최적화 알고리즘, 컨볼루션 커널 설정 방안을 제안한다. 학습을 위한 영상 데이터셋 구성은 AD, MCI 그리고 NC 3가지 클래스로 분류한다. 각 클래스 영상은 소뇌를 기준으로 정수리까지 환자 1명당 68장의 축상면(Axial) 영상을 사용한다.

시뮬레이션 결과 네트워크 구조는 ResNet, 활성화함수는 ReLU, 최적화 알고리즘은 Adam 그리고 컨볼루션 커널은 팽창계수(dilation factor)가 3인 컨볼루션 커널이 치매 의료 영상에 대해 정확도가 우수하다.

2. 유전 알고리즘과 심층학습 신경망

2.1 영상기반 알츠하이머 분류

Suk 등은 기존의 단면만을 사용하는 치매환자 분류가 아닌 3차원 모델링을 통해 복잡한 패턴을 발견하여 이를 통해 치매환자를 분류하는 연구를 제안하였다[8]. 치매환자를 분류할 때에 단면영상만을 사용하는 것이 아니라 3D 영상 전체에서 나타나는 패턴을 활용하여 이를 분석하였다. 정확도는 AD 95.35%, MCI 85.67%, NC 74.58%의 성능을 나타내었다.

Lui 등은 종래의 Autoencoder에 대한 병목현상을 극복하기 위해 소프트맥스(softmax) 출력 계층을 포함한 네트워크 구조를 설계하였다. 이 구조는 적은 학습 데이터셋에도 높은 성능을 보여주며, 커널의 크기와 활성화함수를 조절하는 방법을 제안하였다[9]. 정확도는 NC vs. AD는 87.76%, NC vs. MCI는 76.92%의 성능을 나타내었다.

검출 또는 분류 심층학습 네트워크의 성능은 대상 영상 데이터셋과 네트워크 파라미터에 따라서 차이가 있다. 본 논문에서는 의료 영상 데이터셋에 효과적인 심층학습 네트워크와 파라미터를 선정하기 위하여 Suk 등의 활용한 패턴 분석과 Lui 등이 성능 향상에 고려한 활성화함수에 대한 설정에 대한 실험을 통하여 분석한다. 의료 영상의 종류는 파라미터에 영향이 있을 것으로 판단한다.

2.2 심층학습 신경망

2.2.1 VGG-16

VGG-16 모델은 16개의 계층(Layer)으로 구성된다. 224×224 입력 영상에 3×3 커널을 반복적으로 사용한다. VGG-16은 기본적으로 AlexNet의 구조를

활용하며, 3×3 커널을 반복 사용하여 커널의 크기가 큰 계층을 대체 사용할 수 있음을 보였다[10].

2.2.2 GoogLeNet

GoogLeNet은 22개의 계층으로 CNN의 성능을 향상시키기 위해 직접적인 영향을 주는 계층을 증가시켰다. GoogLeNet에서는 1×1, 3×3, 5×5의 컨볼루션 연산을 각각 수행하여 특징을 추출한다. 3×3의 최대값 풀링(max pooling)을 수행하며, 1×1의 컨볼루션에 넣어 채널을 줄였다가 확장을 할 때 3×3 그리고 5×5 컨볼루션을 한다. GoogLeNet의 기본적인 구조는 인셉션 모듈이다. 인셉션 모듈을 9번을 수행을 하면 연산의 양이 줄어드는 효과가 있다. 인셉션 모듈의 풀링은 채널의 수가 이전과 동일하게 만들기 위해 1×1의 컨볼루션을 사용한다[11].

2.2.3 ResNet

심층학습의 망이 깊어지면, 기울기 소멸(vanishing gradient) 또는 포화(exploding gradient)와 같은 문제 때문에 학습의 성능이 저하되는 문제(degradation problem)가 발생한다[12].

일반적인 학습망에서 $H(x)$ 가 최적화되도록 학습을 수행한다. 잔여학습(residual learning)은 개념을 달리하여 $H(x)-x$ 를 최적화하도록 한다. 여기서, $F(x)$ 를 식 (1)로 정의한다.

$$F(x) = H(x) - x \tag{1}$$

따라서, $H(x)-x$ 의 최적화는 $F(x)$ 가 0으로 수렴하는 것이다. 이것은 사전 조정(pre-conditioning)의 역할을 하게 된다. $F(x)$ 가 거의 0이 되는 방향으로 학습하게 되고, 식 (1) 과 같이 출력과 입력의 차로 학습하기 때문에 잔여학습이라고 하고, 잔여학습을 사용하는 심층학습 모델을 ResNet이라고 한다[12].

ResNet은 단순히 출력에 입력을 더하는 단축 연결(shortcut connection)이 추가 때문에 계수가 증가하지 않으며, 덧셈 계산만 추가된다. 단축 연결을 통하여 깊은 망에서도 성능 저하 없이 최적화가 가능하고, 깊어진 망으로 정확도를 개선할 수 있다.

2.3 심층학습 파라미터

2.3.1 활성화함수

활성함수란 입력 데이터를 비선형으로 바꾸는 역

할을 수행 한다. 심층학습에서 입력 데이터가 선형일 경우 심층학습의 망이 깊어지지 않기 때문에, 이를 비선형으로 바꾸는 작업이 필요하다. 망이 깊어지면 매개변수가 줄어들고, 필요한 연산의 수가 줄어든다. 이를 통해 입력 데이터에 적절한 함수를 적용하여 출력 값을 조절하는 역할을 한다.

최근 가장 많이 사용되는 ReLU은 입력 값이 0보다 작으면 0으로 설정하는 활성화 함수이다[13].

$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases} \tag{2}$$

ReLU는 음수를 0으로 출력하여 파라미터 갱신이 되지 않는 문제점을 가지고 있다. 이 문제점을 해결하기 위해 Leaky ReLU는 입력 값이 음수이면 0이 아닌 아주 작은 값을 가지도록 한다[14].

$$f(x) = \begin{cases} x, & x \geq 0 \\ scale * x, & x < 0 \end{cases} \tag{3}$$

Clipped ReLU는 ReLU의 값이 너무 커져 기울기 포화가 발생 하는 것을 방지하기 위해 상승한계(Ceiling)를 설정하여 상승한계를 넘으면 상승한계 값을 가지도록 한다[15].

$$f(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \leq x < ceiling \\ ceiling, & x \geq ceiling \end{cases} \tag{4}$$

2.3.2 최적화 알고리즘

최적화 알고리즘이란 손실함수(loss function)의 결과값을 최소화하는 모델의 파라미터를 찾는 알고리즘이다.

경사 하강법(gradient descent)은 손실함수의 최소값을 찾기 위해 비용 함수(cost function)의 기울기의 반대방향으로 간격을 이동하면서 가중치와 편향을 갱신한다. 수식은 다음과 같다.

$$\theta_{\ell} + 1 = \theta_{\ell} - a \nabla E(\theta_{\ell}) \tag{5}$$

여기서, ℓ 은 반복 횟수이고, $a > 0$ 은 학습률이고, θ 는 파라미터 벡터이고, $E(\theta)$ 는 손실 함수이다. 손실 함수의 기울기 $\nabla E(\theta)$ 는 훈련 데이터셋을 사용하여 계산되며, 경사 하강법은 한 번에 전체 데이터셋을 사용한다.

확률적 경사 하강법(stochastic gradient descent, SGD)은 각 반복에서 훈련 데이터의 일부를 사용하여 기울기를 계산하고 파라미터를 갱신한다. 손실함

수를 계산할 때 전체 데이터셋(batch) 대신 데이터셋의 일부(mini batch)에 대해서만 계산한다. 데이터셋의 일부를 사용하여 계산되는 파라미터 갱신은 전체 데이터셋을 사용하여 계산되는 파라미터 갱신의 잡음이 있는 추정 값이기 때문에 확률적이다. 또한 최적해를 향해 가는 최속강하법 경로에서 진동이 발생할 수 있다. 이러한 진동을 줄이는 한 가지 방법으로 파라미터 갱신에 모멘텀(momentum)항을 추가한다. 모멘텀을 사용한 확률적 경사 하강법(stochastic gradient descent with momentum, SGDM) 갱신은 다음과 같이 표현된다[16].

$$\theta_{\ell}+1=\theta_{\ell}-a\nabla E(\theta_{\ell})+r(\theta_{\ell}-\theta_{\ell-1}) \quad (6)$$

여기서 r 는 직전 기울기 스텝이 현재 반복에서 차지하는 비중을 결정합니다.

SGDM은 모든 파라미터에 대해 고정된 학습률을 사용한다. 최적화 알고리즘 중에는 손실함수에 자동으로 적용할 수 있는 학습률을 사용하는 알고리즘도 있다. 그 중 하나가 제곱평균제곱근 전파(root mean square propagation, RMSProp)이다. RMSProp은 파라미터 기울기를 계산하기 위해 지수이동평균(exponential moving average)을 사용한다.

$$v_{\ell}=\beta_2 v_{\ell-1}+(1-\beta_2)[\nabla E(\theta_{\ell})]^2 \quad (7)$$

v_{ℓ} 는 누적 기울기 제곱이고, β_2 는 제곱 기울기 이동평균의 감쇠율이고, 제곱 기울기의 대응되는 평균 길이는 $1/(1-\beta_2)$, 파라미터 갱신과 같다. 이러한 이동평균을 사용하여 각 파라미터의 갱신을 개별적으로 정규화 한다.

$$\theta_{\ell}+1=\theta_{\ell}-\frac{a\nabla E(\theta_{\ell})}{\sqrt{v_{\ell}+\epsilon}} \quad (8)$$

RMSProp를 사용하면 기울기가 큰 파라미터에 대해서는 학습률이 효과적으로 감소하고, 기울기가 작은 파라미터에 대해서는 학습률이 증가한다.

적응적 모멘트 추정(Adaptive Moment Estimation, Adam)은 모멘텀 항이 추가된 것을 제외하면 RMSProp와 유사한 파라미터 갱신을 사용한다. 다음과 같이 파라미터 기울기와 그 제곱 값의 요소별 이동평균을 모두 유지한다.

$$m_{\ell}=\beta_1 m_{\ell-1}+(1-\beta_1)\nabla E(\theta_{\ell}) \quad (9)$$

$$v_{\ell}=\beta_2 v_{\ell-1}+(1-\beta_2)[\nabla E(\theta_{\ell})]^2 \quad (10)$$

m_{ℓ} 는 모멘텀의 누적 기울기 벡터이고, β_1 는 기울기 이동평균의 감쇠율이고, β_2 는 제곱 기울기 이동평균의 감쇠율이다. Adam은 두 이동평균을 사용하여 네트워크 파라미터를 다음과 같이 갱신한다.

$$\theta_{\ell}+1=\theta_{\ell}-\frac{am_{\ell}}{\sqrt{v_{\ell}+\epsilon}} \quad (11)$$

ϵ 는 0이 아닌 아주 작은 임의의 숫자이다. 반복 동안 기울기가 비슷한 경우에 기울기의 이동평균을 사용하면 파라미터 갱신이 특정 방향에서 모멘텀을 갖게 된다. 기울기의 대부분이 잡음인 경우, 기울기의 이동평균이 작아지므로 파라미터 갱신도 작아져 빠르게 손실함수의 최소값을 찾을 수 있다[17].

2.4 유전 알고리즘

유전 알고리즘은 생물의 진화를 모방한 최적화 기법 중 하나이다. 생물의 적응 정도를 계산하는 척도로서 평가함수의 적응도(fitness)를 사용하여 최적화하는 기법이다. 적응도는 생존율, 출생률 및 그것들의 조합에 기인하는 번식의 성공도(reproductive success)로 정의할 수 있다. 생물에 있어서 중요한 것은 자신과 유사한 것을 만드는 능력과, 유사하지만 미묘하게 다른 다수의 생물개체가 종이라는 집단을 구성해서 존재하는 것이다. 생물집단에 어떤 종의 제약(전 개체수가 일정하다는 것과 생물의 도태)이 부과되면 집단 내에 경합관계가 생겨나게 되고 보다 환경에 적응한 개체군이 번식에 성공하도록 된다. 즉, 적응도가 높은 개체군이 번식하여 개체를 보존하게 된다. 이와 같은 생물계의 현상을 이용한 것이 유전 알고리즘이다[18].

유전 알고리즘은 Fig. 1과 같이 유전자 초기화(initialize), 적응도(fitness) 계산 및 선택(selection), 교배(crossover), 돌연변이(mutation)와 같은 단계를 통해 하나의 세대를 생성하고 목표 값을 만족할 때까지 반복한다.

유전 알고리즘은 최적화 대상 모델이 설정되면, 그 모델의 최적해를 구하기 위해서 설정모델을 유전 알고리즘으로 사상(mapping)시켜야 한다. 이 과정에서 유전자(gene)를 개체(string)로서 부호화해야 하며 해공간을 탐색하기 위해 개체의 길이와 개체수를 얼마정도로 할 것 인가, 어느 정도의 정밀도로 해공간을 탐색(resolution)할 것인가 등을 설정해야한다.

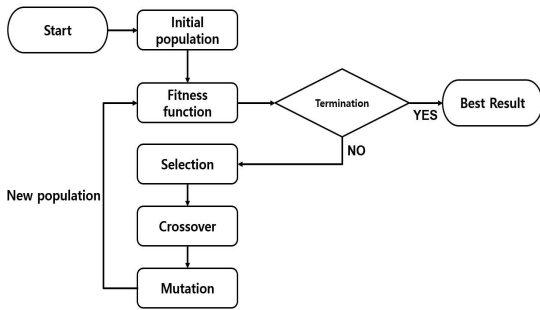


Fig. 1. Flow chart of Genetic algorithm.

초기집단을 정해진 개체길이와 개체수로 임의로 구성한다. 이를 초기화라고 한다. 생성된 초기집단이 탐색초기에 편향된 분포로 인해 국소최저점에 빠지는 것을 방지하기 위해 다양한 분포법을 이용하여 균등하게 초기화한다.

집단 내의 이진수로 표현된 개체를 십진수로 바꿔 적응도를 평가하고 집단의 평균적응도를 고려하여 확률적인 취사선택을 한다. 각 개체의 적응도는 목적함수에 의해 결정된다. 여기서 목적함수는 개체의 생성과 소멸을 결정한다. 여기서 목적함수는 개체의 생성과 소멸을 결정하는 요인이다. 정해진 세대수만큼 반복하여 탐색을 종료한 후 최상의 적응도 값을 갖는 개체를 평가하여 그 값을 최적해로 한다.

선택은 개체가 다음 세대에 자신의 유전자를 남기는 것과 관련한 기능이다. 개체에 선택압을 가할 때에 중요한 것은 어느 개체끼리 교배시키는 가이다. 즉, 다음 세대에 우수한 자손을 번식시키기 위해 적응도가 높은 개체를 선택하는 방법이다[19].

교차과정은 선택교배 전략에 의해 선택된 염색체에 대하여 교배지(mating pool)에서 수행하는 조작이다. 교배할 개체 쌍을 랜덤하게 선택한 후 부모 개체의 bit를 교차율에 따라 선택하여 자손의 개체를 만든다.

돌연변이는 유전자를 일정한 확률로 변화시키는 조작으로, 집단의 유전자의 다양성을 유지하여 가능한 넓은 해공간을 탐색하도록 하며 초기수렴 즉, 국소최저점에 빠지는 것을 방지한다.

3. 유전 알고리즘을 이용한 심층학습 모델 최적화

심층학습 모델의 성능은 입력 영상, 심층학습 네트워크 구조, 컨볼루션 커널, 활성화함수, 최적화 알고리

즘 등에 의해 결정된다. 전역 탐색을 통해 모든 파라미터에 대하여 탐색을 하는 것은 비효율적이다. 이를 유전 알고리즘을 통해 최적의 파라미터를 찾아낸다.

전역 탐색(exhaustive search)을 이용하는 경우 모든 경우의 수를 조합하여 탐색을 하는데, 파라미터가 늘어날수록 계산 복잡도가 기하급수적으로 증가한다. 유전 알고리즘은 특성 상 다양한 초기 집단을 생성하여 가장 우수한 염색체를 선택하게 되는데, 가장 우수한 염색체가 전역 최적해에 근접한 최적해가 된다. 따라서 전역 탐색을 하지 않고도 전역 최적해에 가까운 해를 찾을 수 있다.

본 논문에서는 유전 알고리즘을 이용하여 치매 분류에 최적화 된 심층학습 모델의 파라미터를 찾는 것이 목표이다. Fig. 2에 제안하는 유전 알고리즘의 전체적인 흐름도를 나타낸다.

초기화는 유전자 균일 분포법(gene uniform distribution)을 이용한다. 유전자 균일 분포법은 유전자 집단의 50%을 임의의 이진수로 초기화한 후, 집단의 나머지 50%을 앞의 이진수의 보수(complement)로 초기화한다[20]. 유전자의 길이는 2-bit이고, 염색체의 길이는 총 6-bit이다. 앞 3-bit에 대해 임의의 이진수로 할당하고, 나머지 3-bit는 앞 3-bit의 보수 처리하여 균일한 비율의 초기 개체를 생성한다. 이 방법에 의해 집단 내의 각 개체들의 유전자 0과 1의 집단 내 분포는 1:1로 균일하게 된다. 이후 2-bit씩 2진수에서 10진수로 변환한다.

10진수로 변환된 한 세대에 대해 적응도 계산을 통해 다음 세대에 전할 염색체를 선택한다. 이때에

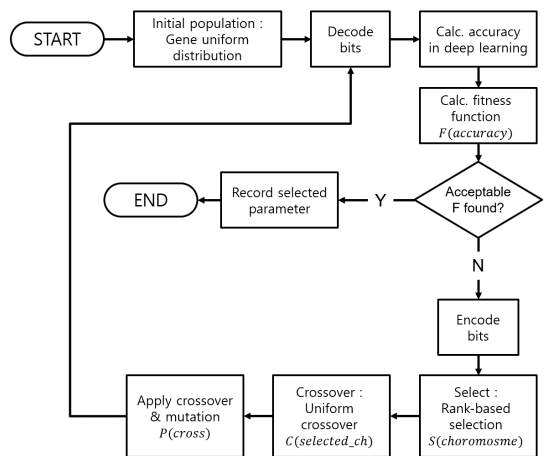


Fig. 2. Block diagram of Genetic algorithm.

적응도 f 를 심층학습 모델에 대해 정확도로 지정한다. 정확도는 활성함수, 최적화 알고리즘, 컨볼루션 커널 파라미터를 통해 계산한다. 목표한 적응도를 계산하지 못한 경우, 염색체를 다시 10진수에서 2진수로 변환한다. 적응도 계산은 순위 선택 기반에 따라 정확도가 높은 순서에 따라 정렬 후, 상위 50%의 염색체를 선택한다[21].

선택된 염색체들은 균등 교차(uniform crossover)를 통하여 다음 세대를 생성한다. 균등 교차는 부모 염색체를 임의로 선택하여 임계확률을 임의로 설정 후 각 개체에 대하여 임의의 난수를 생성한다. 난수가 임계확률보다 높으면 부모 1의 개체를 낮거나 같으면 부모 2의 개체를 선택하여 새로운 염색체를 생성한다.

변이는 5세대마다 고정적으로 나타나며 임의의 염색체를 선택하고 임의의 개체 3-bit를 보수 계산하여 새로운 염색체를 생성한다. 이를 최적의 해 또는 지정된 횟수를 찾을 때 까지 반복한다.

성능비교를 위해 사전에 3개의 네트워크 모델에 대하여 모든 경우의 활성함수와 최적화 알고리즘에 따른 성능을 비교하고 분석하여 최적의 네트워크 모델은 선택한다. 유전자는 활성함수, 최적화, 컨볼루션 커널로 Table 1과 같이 구성한다.

4. 시뮬레이션 결과 및 검토

본 논문에서는 유전 알고리즘의 유전자를 활성함수, 최적화 그리고 컨볼루션 커널로 정하고 사전 실험 결과와 비교하여 최적의 파라미터를 도출하는지 실험한다.

Table 1. Configuration of gene

	Gene	encoding	Binary
Activation	ReLU	0	00
	Leaky ReLU	1	01
	Clipped ReLU	2	10
Optimizer	SGDM	0	00
	RMSProp	1	01
	Adam	2	10
Convolution Kernel	no dilated conv.	0	00
	1 - dilated conv.	1	01
	2 - dilated conv.	2	10
	3 - dilated conv.	3	11

CNN 모델을 학습하기 위해 AD 159명에 대한 10,812장, MCI 130명에 대한 8,840장, NC 85명에 대한 5,780장의 전처리가 완료된 NII 영상으로 데이터셋을 구성한다.

공간 정규화(spatial normalization)[22]와 강도 정규화(count normalization)[23]를 적용하여 사람마다 다른 뇌의 크기를 통일하고, 사람마다 밝기 차이가 거의 없는 소뇌를 중심으로 밝기값을 조정한다.

추론을 위하여 학습에 사용되지 않고 완전히 분리된 AD 17명, MCI 14명, NC 9명에 대한 이미지 각 1,156장, 952장, 612장의 추론 데이터셋을 구성한다.

유전 알고리즘의 성능 비교를 위해 심층학습 모델 파라미터에 대해 사전 학습 및 추론 결과를 분석한다. 각 심층학습 모델의 파라미터에 대한 성능을 분석하고 이를 바탕으로 최적 심층학습 모델을 위한 유전 알고리즘의 유전자를 선정한다.

사용된 심층학습 모델 구조는 VGG-16, GoogLe Net, ResNet을 사용하였고, 심층 학습 모델의 파라미터는 최적화 알고리즘, 활성함수로 지정하여 성능을 비교 분석한다. 최적화 알고리즘에서 SGDM의 모멘텀 값은 0.9를 사용하고, RMSProp의 β_2 값은 0.999를 사용하고, Adam의 β_1 값은 0.9, β_2 값은 0.999를 사용한다. 활성함수에서 Leaky ReLU의 Scale 값은 0.01을 사용하고, Clipped ReLU의 Ceiling 값은 10으로

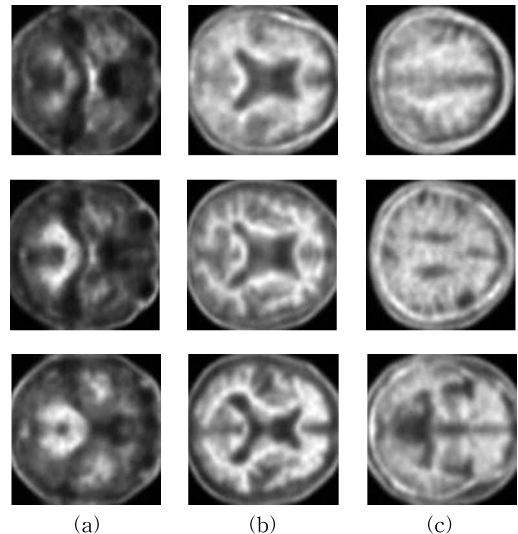


Fig. 3. Training example image of Alzheimer's disease classification, (a) normal control, (b) mild cognitive impairment, (c) Alzheimer's disease.

사용한다.

VGG-16 모델에 대한 실험 결과는 Table 2과 같다. 실험의 결과는 각 모델의 정확도로 나타낸다.

GoogLeNet 모델에 대한 실험 결과는 Table 3과 같다.

ResNet 모델에 대한 실험 결과는 **Table 2**과 같다.

사전 실험 결과 ResNet, Adam, ReLU 그리고, 팽창계수가 3인 컨볼루션 커널을 사용했을 때 성능이 86.91%로 가장 높은 것을 알 수 있다. 사전 결과 ResNet이 전반적으로 성능이 뛰어나 유전알고리즘의 네트워크 모델을 ResNet로 선택하고 실험한다.

Table 5는 유전 알고리즘의 결과를 계산했을 때의

Table 2. Experiment result of deep learning with VGG-16

		No-dilated	1-dilated	2-dilated	3-dilated
SGDM	ReLU	51.39%	51.96%	50.84%	50.36%
	Leaky ReLU	37.19%	38.45%	38.64%	34.68%
	Clipped ReLU	48.18%	44.68%	48.64%	49.86%
RMSProp	ReLU	53.59%	56.16%	56.48%	54.83%
	Leaky ReLU	53.35%	55.63%	56.48%	54.86%
	Clipped ReLU	53.06%	55.64%	55.66%	54.78%
Adam	ReLU	57.79%	55.24%	58.74%	54.89%
	Leaky ReLU	57.19%	58.49%	58.97%	58.72%
	Clipped ReLU	56.19%	54.79%	58.59%	59.66%

Table 3. Experiment result of deep learning with GoogLeNet

		No-dilated	1-dilated	2-dilated	3-dilated
SGDM	ReLU	66.87%	66.48%	68.18%	64.96%
	Leaky ReLU	50.56%	55.20%	55.48%	49.46%
	Clipped ReLU	68.41%	67.45%	66.64%	70.49%
RMSProp	ReLU	67.84%	65.82%	65.25%	64.96%
	Leaky ReLU	62.20%	64.72%	62.70%	64.47%
	Clipped ReLU	62.42%	67.72%	67.41%	64.69%
Adam	ReLU	70.45%	63.64%	70.25%	70.89%
	Leaky ReLU	68.52%	69.66%	67.89%	68.16%
	Clipped ReLU	67.97%	64.96%	68.79%	64.69%

Table 4. Experiment result of deep learning with ResNet

		No-dilated	1-dilated	2-dilated	3-dilated
SGDM	ReLU	83.98%	81.47%	82.54%	82.22%
	Leaky ReLU	69.59%	82.90%	84.15%	83.79%
	Clipped ReLU	83.45%	80.12%	81.97%	82.57%
RMSProp	ReLU	75.94%	78.91%	77.95%	74.91%
	Leaky ReLU	78.63%	76.63%	78.48%	75.42%
	Clipped ReLU	72.45%	71.09%	72.80%	74.14%
Adam	ReLU	85.20%	85.22%	83.89%	86.91%
	Leaky ReLU	84.12%	85.69%	81.78%	84.77%
	Clipped ReLU	82.43%	84.63%	80.91%	83.54%

Table 5. Configuration of chromosomes in results generation

order	number	chromosome	cost
1	6	001011	10.0
2	4	001011	8.9
3	5	001011	7.8
4	1	001011	6.7
5	10	001011	5.6
6	9	001011	4.5
7	8	001011	3.4
8	2	001011	2.3
9	3	001011	1.2
10	7	100111	0.1

염색체의 구조이다. 염색체의 번호는 초기에 생성된 염색체의 순서이다. 동일한 염색체인 경우, 이전 세대에서 적응도가 높은 염색체가 우선순위를 가지게 된다. 실험 결과 일정 세대가 지나면 변이를 제외한 모든 염색체들이 동일하게 구성됨을 확인할 수 있다. 염색체를 해석하면 최적화된 파라미터는 Adam, ReLU 그리고, 팽창계수가 3인 컨볼루션 커널로 고정되는 것을 알 수 있고, 사전 실험과 동일한 염색체가 선택되는 것을 확인할 수 있다.

본 논문에서 제안하는 유전 알고리즘과 다른 최적화 알고리즘인 타부 탐색(tabu search)과 입자 군집 최적화(particle swarm optimization)의 성능 비교를 한다. 타부 탐색은 임의로 생성된 초기의 단일해 기반으로 현재해의 이웃에서 후보해를 탐색하고 이동하여 최적화를 수행한다. 탐색 과정을 일정 기간 기억하여 금지된 이동(tabu move)을 기억하여 이전 해로 재방문하는 것을 방지하는 것이 특징이다. 입자 군집 최적화는 유전 알고리즘과 유사하게 다수의 후보해를 동시에 취급하는 특징을 가지고 있다. 또 다른 특징으로는 개별 후보해의 움직임은 후보해 전체

Table 6. Performance comparison of optimization algorithm

	running total	computation time (min)
genetic algorithm	21	1,236
tabu search	42	2,354
particle swarm optimization	32	1,885
exhaustive search	36	2,146

에 영향을 끼치며, 또한 전체 후보해의 움직임은 개별 후보해의 움직임에 영향을 끼친다. 후보해 전체가 지금까지 얻어낸 최소 목적 함수의 값을 활용하여 최적화가 되도록 움직이는 확률론적 접근법이다.

다음 Table 6는 유전 알고리즘과 타부 검색, 입자 군집화에 대해 파라미터의 계산량과 소요시간을 나타낸다.

5. 결 론

본 논문에서는 유전 알고리즘을 이용하여 PET 영상에 최적의 심층 학습 파라미터를 찾는 것을 목표로 하였다. 유전 알고리즘의 결과와 사전 실험의 결과를 비교하여 최적의 파라미터를 찾는 것을 확인한다.

사전 실험 결과 PET 영상에 최적화 된 심층학습 모델인 ResNet을 선정하고 유전자로 최적화 알고리즘, 활성화함수, 컨볼루션 커널을 선택하여 유전 알고리즘 수행한 결과, Adam, ReLU 그리고, 팽창계수가 3인 컨볼루션 커널이 선택되었을 때 86.71%으로 PET 영상에 대하여 가장 높은 성능을 나타낸다. 또한, 다른 최적화 알고리즘인 타부 검색과 입자 군집 최적화 알고리즘과 비교하였을 경우 계산하는 파라미터 수와 수행하는 시간이 다른 알고리즘에 비해 약 1.2배 빠른 것을 확인할 수 있다. 환자의 학습 데이터셋을 보완함으로써 성능을 개선할 수 있을 것으로 판단된다.

REFERENCE

[1] S.G. Hwang and H.K. Park, "An Analysis on Prescribing Patterns of Alzheimer's Dementia Treatment and Choline Alfoscerate Using HIRA Claims Data," *Korean Journal of Clinical Pharmacy*, Vol. 29, No. 1, pp. 1-8, 2019.

[2] H.I. Suk and D. Shen, "Deep Learning-based Feature Representation for AD/MCI Classification," *Medical Image Computing and Computer-Assisted Intervention*, Vol. 8150, No. 1, pp. 583-590, 2013.

[3] R. Li, W. Zhang, H.I. Suk, W. Li, J. Li, D. Shen, et al., "Deep Learning Based Imaging Data Completion for Improved Brain Disease Diagnosis," *Medical Image Computing and*

- Computer-Assisted Intervention*, Vol. 8675, No. 1, pp. 305-312, 2014.
- [4] C. Carlos and M. Silveira, "Classification of Alzheimer's Disease from FDG-PET Images Using Favourite Class Ensembles," *Proceeding of 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2477-2480, 2013.
- [5] J. Chen, L. Yang, Y. Zhang, M. Alber, and D.Z. Chen, "Combining Fully Convolutional and Recurrent Neural Networks for 3D Biomedical Image Segmentation," *Proceeding of Advances in Neural Information Processing Systems*, pp. 3036- 3044, 2016.
- [6] I. Gariali, M. Adel, S. Bourennane, and E. Guedj, "Region-based Brain Selection and Classification on Pet Images for Alzheimer's Disease Computer Aided Diagnosis," *Proceeding of IEEE International Conference on Image Processing*, pp. 1473-1477, 2015.
- [7] S.Y. Kwon, Y.J. Kim, and G.G. Kim, "An Automatic Breast Mass Segmentation Based on Deep Learning on Mammogram," *Journal of Korea Multimedia Society*, Vol. 21, No. 12, pp. 1363-1369, 2018.
- [8] H.I. Suk, S.W. Lee, and D. Shen, "Hierarchical Feature Representation and Multimodal Fusion with Deep Learning for AD/MCI Diagnosis," *Neuro Image*, Vol. 101, No. 1, pp. 569-582, 2014.
- [9] S. Liu, S. Liu, W. Cai, S. Pujol, R. Kikinis, and D. Feng, "Early Diagnosis of Alzheimer's Disease with Deep Learning," *Proceeding of IEEE International Symposium on Biomedical Imaging*, pp. 1015-1018, 2014.
- [10] X. Zhang, J. Zou, K. He, and J. Sun, "Accelerating Very Deep Convolutional Networks for Classification and Detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 38, No. 10, pp. 1943-1955, 2015.
- [11] C. Szegedy, W. Liu, Y. Ji, P. Sermanet, S. Reed, D. Anguelov, et al., "Going Deeper with Convolutions," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1-9, 2015.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.
- [13] V. Nair and G.E. Hinton, "Rectified Linear Units Improve Restricted Boltzmann Machines," *Proceedings of International Conference on Machine Learning*, pp. 807-814, 2010.
- [14] A.L. Maas, A.Y. Hannun, and A.Y. Ng, "Rectifier Nonlinearities Improve Neural Network Acoustic Models," *Proceedings of International Conference on Machine Learning*, pp. 3-9, 2013.
- [15] A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, et al., "Deep Speech: Scaling Up End-to-end Speech Recognition," *arXiv Preprint arXiv:1412.5567*, 2014.
- [16] K.P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press, Cambridge, Massa., 2012.
- [17] D.P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *arXiv Preprint arXiv: 1412.6980*, 2014.
- [18] C.F. Geoffrey and N. Mansour, "A Hybrid Genetic Algorithm for Task Allocation in Multicomputers," *Proceedings of International Conference on Genetic Algorithms*, pp. 446-473, 1991.
- [19] W.M. Spears, "Crossover or Mutation?," *Foundations of Genetic Algorithms*, Vol. 2, No. 1, pp. 221-337, 1993.
- [20] Z.W. Geem, J.H. Kim, and G.V. Loganathan, "A New Heuristic Optimization Algorithm: Harmony Search," *SIMULATION: Transactions of The Society for Modeling and Simulation International*, Vol. 76, No. 2, pp. 60-68, 2001.
- [21] L.D. Whitley, "The GENITOR Algorithm and

Selection Pressure: Why Rank-based Allocation of Reproductive Trials is Best,” *Proceedings of International Conference on Genetic Algorithms*, pp. 116-123, 1989.

- [22] H.J. Park, “Methodological Review on Functional Neuroimaging Using Positron Emission Tomography,” *Nuclear Medicine and Molecular Imaging*, Vol. 41, No. 2, pp. 71-77, 2007.
- [23] S. Misciagna, “Basic PET Data Analysis Techniques,” *Positron Emission Tomography-Recent Developments in Instrumentation, Research and Clinical Oncological Practice*, IntechOpen, London, UK, 2013.



이 상 협

2016년 동의대 멀티미디어공학과 졸업
 2018년 동의대 대학원 공학석사
 2018년~현재 경성대학교 공학박사 진행
 관심분야: 기계학습 및 영상 분석, 음향처리



강 도 영

1991년 동아대 의과대학 졸업
 1995년 동아대 의료원 핵의학과 석사
 2000년 동아대 의료원 핵의학과 박사
 2000년~현재 동아대 의료원 핵의학과 교수

관심 분야: 퇴행성 뇌 질환 이미지 분석



송 종 관

1989년 부산대 전자공학과 졸업
 1991년 한국과학기술원 대학원 공학석사
 1995년 동 대학원 공학박사
 1997년~현재 경서대학교 전자공학과 교수

관심분야: 영상 처리 및 부호화, 기계학습, 컴퓨터비전, 심층학습



박 장 식

1992년 부산대 전자공학과 졸업
 1994년 부산대 전자공학과 석사
 1999년 부산대 전자공학과 박사
 1997년~2011년 동의과학대학 전자과 교수
 2011년~현재 경성대학교 전자공학과 교수

관심분야: 신호처리, 기계학습, 컴퓨터비전, 심층학습, 임베디드시스템