

# 머신러닝 기반 음악 장르 분류에 대한 연구

서정환<sup>1</sup>, 박재표<sup>2\*</sup>

<sup>1</sup>송실대학교 대학원 금융기술융합학과, <sup>2</sup>송실대학교 정보과학대학원

## A Study on Classification of Music Genre Based on Machine Learning

Jung-Hwan Seo<sup>1</sup>, Jae-Pyo Park<sup>2\*</sup>

<sup>1</sup>Department of Financial Technology Convergence, Soongsil University

<sup>2</sup>Graduate School of Information Science, Soongsil University

**요약** 음악 장르는 음악을 분류하기 위한 수단으로 박자, 템포, 멜로디의 높낮이 등의 수많은 특징을 사용하여 분류한다. 본 논문에서는 전처리 과정을 통하여 음악의 특징을 추출하고 머신러닝 기법으로 학습시켜 장르를 구분하는 방법을 제안하였다. CNN(Convolutional Neural Network), LSTM (Long Short-Term Memory)의 두 가지 머신러닝 알고리즘을 사용하여 알고리즘별 성능을 비교 분석하였다. 분석 결과, 경음악(light Music), 발라드(Ballad), 록(Rock), 블루스(Blues), 알앤비(R&B), 클래식(Classic) 등 크게 6가지 장르로 분류 실험을 수행한 결과로 CNN을 이용한 머신러닝 기법의 정확도가 98.7%로 가장 높았으며, 이는 특징을 자동으로 추출해주는 CNN의 특성 때문이었다. 높은 정확도를 보여준 CNN을 통해 자동 음악 추천 시스템과 비슷한 응용 분야에서의 활용 가능성을 확인하였으며, 제한된 데이터와 비 최적화된 매개변수 설정에도 불구하고 유의미한 결과를 제시하였다.

**Abstract** Music genres are classified using numerous features such as beat, tempo, and melodic pitch as a means to categorise music. This paper propose a method to classify genres by extracting music features through preprocessing and training them with machine learning techniques. Two machine learning algorithms, CNN (Convolutional Neural Network) and LSTM (Long Short-Term Memory), were used to compare the performance of each algorithm. The results of the analysis showed that the machine learning method using CNN had the highest accuracy of 98.7%, which is due to the characteristics of CNN that automatically extracts features. The high accuracy of CNN confirms the possibility of using it in applications similar to automatic music recommendation systems, and shows significant results despite limited data and non-optimal parameter settings.

**Keywords** : Music Genre, Classification of Music, Machine Learning, CNN(Convolutional Neural Network), LSTM (Long Short-Term Memory)

### 1. 서론

음악은 감정과 사상을 표현하는 예술로, 사람의 목소리나 악기를 통해 소리의 높낮이, 장단, 강약 등의 요소로 사상과 감정을 표현한다. 이 예술은 사상과 감정의 아름다움, 즐거움, 흥분, 감성, 행복, 슬픔 등을 포함하며,

동서양을 막론하고 인류 문화의 발전에 있어 오랫동안 중요한 요소로 자리 잡고 있다. 이러한 배경에서 음악 시장은 다양한 문화와 기술의 발전과 함께 형성되고 성장해 왔다. 초기에는 주로 공연을 통해 음악이 유통되었지만, 20세기에 들어서면서 레코드, 라디오, 텔레비전 등의 매체가 음악의 보급과 소비에 큰 변화를 가져왔다. 디지

\*Corresponding Author : Jae-Pyo Park(Soongsil Univ.)

email: pjerry@ssu.ac.kr

Received December 18, 2023

Accepted June 7, 2024

Revised April 29, 2024

Published June 30, 2024

텔 시대의 도래와 함께 음악 산업은 더욱 빠른 속도로 진화했다. 인터넷과 스트리밍 서비스의 등장은 어디서나 접근 가능한 음악 소비의 새로운 패러다임을 창출했다.

이러한 기술적 진보와 문화적 변화 속에서 음악의 형태와 장르도 끊임없이 진화하고 다양해지고 있다. 전통적인 클래식, 재즈, 블루스에서부터 현대의 팝, 힙합, EDM까지, 새로운 장르가 계속해서 생겨나고 있다. 이처럼 장르가 다양해짐에 따라 음악을 체계적으로 분류하고 이해하는 것이 더욱 중요해지고 있다. 각 장르는 독특한 스타일과 특성을 가지며, 이를 명확히 구분하고 분석하는 것은 음악을 더 깊이 이해하고, 소비자에게 적합한 음악을 추천하는 데 필수적이다. 따라서 음악 장르의 분류와 분석은 음악 산업에서 매우 중요한 역할을 하며, 이를 위한 기술적 접근이 필요하다.

현재 음악을 자동으로 찾는 연구들[1,2]은 활발히 진행되고 있지만, 장르를 명확히 구분하는 연구는 상대적으로 부족하여 이 분야에서 심도 있는 연구가 필요하다. 본 논문에서는 음악 장르를 자동으로 구분하기 위해 머신러닝 기법을 비교 분석하고, 이를 통한 장르 구분 방안을 제안할 것이다. 연구에 사용된 데이터는 경음악, 발라드, 록, 블루스, R&B, 클래식으로 구분된 각 장르별로 100개의 음향 데이터를 포함하며, 이는 머신러닝으로 학습되어 분류된다. 머신러닝 알고리즘으로는 CNN과 LSTM을 비교 분석하여 그 결과를 도출하고자 한다.

## 2. 장르 구분에 관한 연구

다양한 알고리즘과 정보를 바탕으로 음악 장르 분류에 관한 연구가 진행되어왔다. 음원의 MFCC (오디오 신호에서 추출할 수 있는 feature로 소리의 고유한 특징을 나타내는 수치)를 활용한 머신 학습 알고리즘을 이용하거나, 음원의 MIDI 정보에서 화성, 리듬 등의 특징을 추출하여 장르를 분류한 연구들이 있다[1-3]. 이외에도 음원에서 스펙트로그램을 순차적으로 추출하고 시그널 데이터를 활용한 분류 연구, 음원의 MFCC, 앨범, 가사를 포함한 커버까지 모든 학습 데이터를 사용한 연구, 그리고 CNN(Convolutional Neural Network)을 이용해 음악을 스펙트로그램으로 변환하여 분류하는 연구들도 진행되었다[4-6]. 또한, 가사보다는 음성 신호에 중점을 둔 내용 기반 장르 분류 연구도 있었다[7,8]. 추가적으로, 음악 관련 연구에서 자주 사용되는 데이터셋의 적절성을 평가하는 연구도 수행되었다[9,10].

다양한 연구에서 멀티모달(Multi Modal) 학습이 유니모달(Uni Modal) 모델에 비해 높은 성능을 보임을 입증했다는 사실은, 멀티모달(Multi Modal) 접근 방식이 머신러닝 및 인공지능 분야에서 매우 유망하다는 것을 시사한다. 유니모달(Uni Modal) 모델은 단일 유형의 데이터만을 활용하는 반면, 멀티모달(Multi Modal) 모델은 두 개 이상의 다른 유형의 데이터를 통합하여 사용한다. 예를 들어, 음성과 텍스트를 결합하여 감정 분석을 수행하거나, 이미지와 텍스트를 함께 사용하여 더 정밀한 이미지 분석을 실행하는 것은 멀티모달 모델의 대표적인 예이다. 이러한 모델은 각기 다른 데이터 유형에서 얻은 정보를 통합함으로써 더 풍부하고 정확한 분석 결과를 도출할 수 있는 장점을 가진다.

그러나 기존의 연구들은 주로 앨범커버와 같은 단편적인 이미지만을 활용한 멀티모달(Multi Modal) 학습에 집중해왔고, 영상을 포함한 더 다양한 형태의 데이터를 통합하는 연구는 상대적으로 부족했다. 이는 멀티모달(Multi Modal) 학습의 잠재력을 충분히 활용하지 못하고 있다는 한계를 나타낸다. 본 연구에서는 이러한 한계를 극복하고자 음성과 영상 데이터를 동시에 활용하는 새로운 멀티모달(Multi Modal) 모델을 제안한다. 이 모델은 음성의 특성과 영상 데이터에서 얻은 시각적 정보를 결합하여 음악 장르를 보다 정확하게 분류 할수있는 방법을 개발하는 것을 목표로 한다. 이는 기존 유니모달(Uni Modal) 모델이 제공할 수 없는, 향상된 분석 성능을 가능하게 할 것이다. 이와 같은 멀티모달 접근은 각 데이터 유형의 독특한 정보를 최대한 활용하여 더욱 정교하고 심층적인 분석을 제공함으로써 음악 산업의 요구에 보다효과적으로 부응할 수 있을 것이다[11,12].

## 3. 분류를 위한 머신러닝 기법 비교

머신러닝은 인공지능(AI)의 한 분야로, 데이터 기반으로 패턴을 학습하고 모델 생성을 자동화하여 데이터 분석을 수행한다. 이 과정을 통해 인간의 개입을 최소화하고 신속한 의사 결정을 가능하게 하는 것이 머신러닝의 주된 목표다. 전체 데이터를 효과적으로 분석하고 중요한 정보를 추출하는 능력이 필수적이다.

머신러닝은 컴퓨터가 특정 작업을 위해 사전에 프로그래밍 된 규칙 없이 스스로 학습할 수 있도록 설계되었다. 이는 인간의 추론 방식을 모방한 강력한 통계적 접근법을 바탕으로 한다. 시간이 지남에 따라 다양한 머신러닝

알고리즘이 개발되었으며, 새로운 기술의 발전으로 이제 빅 데이터를 빠르고 반복적으로 분석할 수 있는 자동화된 방법이 등장했다. 최근 많은 시스템들, 예를 들어 자율 주행 자동차, 맞춤형 광고, 영업 관리 등이 머신러닝 방법론을 사용하여 빠르고 효과적인 결과를 제공하고 있다.

최근 인공지능 기법을 이용한 음성 및 음악 분류, 음악 장르 분류, 음성 감정 인식에 관한 연구들이 활발히 진행되고 있다. 한상일(2021)[13]은 국내 음악방송에서 송출된 소리를 5초 단위로 구분하여 음성 및 음악 파일을 구성하고, CNN 분석을 사용하여 약 89.6%의 정확도를 보이는 딥러닝 모델을 제안하였다. 김선훈 외(2016)[14]는 멀티모달 딥러닝을 이용하여 음악 장르를 분류하는 연구를 수행하였으며, CNN과 BRNN을 동시에 이용한 모델이 75%의 정확도로 다른 구조들에 비해 우수한 성능을 보였다. 윤상혁 외(2021)[15]는 음성 데이터를 멜스펙트로그램으로 변환하고 CNN-LSTM을 이용하여 음성 감정을 인식하는 모델을 제안하였으며, 해당 모델은 88.89%의 정확도를 기록하였다. 이와 같은 연구들은 딥러닝 기법이 다양한 음성 및 음악 데이터 분석에서 높은 성능을 발휘할 수 있음을 보여준다.

본 연구에서는 두 가지 머신러닝 모델을 실험적으로 비교한다. 첫 번째 모델은 CNN(Convolutional Neural Network) 이고, 두 번째 모델은 LSTM (Long Short-Term Memory)이다.

### 3.1 CNN

CNN, 즉 컨볼루션 신경망은 합성곱 연산을 사용하는 인공 신경망(ANN)의 한 종류다. 이는 딥 러닝에서 가장 널리 사용되는 알고리즘 중 하나로, 사진, 이미지, 텍스트, 사운드 등 다양한 유형의 데이터를 효과적으로 분류할 수 있는 머신러닝 기법이다. CNN은 특징을 자동으로 학습하며, 이로 인해 수동으로 특징을 추출할 필요가 없어 학습 과정이 더 간소화된다. 더 적은 매개변수를 사용함에도 불구하고 높은 정확도로 분류 작업을 수행할 수 있다. 이러한 특성 때문에 컴퓨터 비전, 자율 주행, 얼굴 인식 등 객체 인식이 중요한 분야에서 주로 사용되고 있다.

#### 3.1.1 실험 CNN 모델 설계 구조

실험에 사용된 CNN 모델은 총 8개의 계층으로 구성된다. 첫 번째 계층에서는 합성곱 연산을 수행하고, 두 번째 계층에서는 풀링(Pooling) 연산을 진행한다. 세 번째와 네 번째 계층에서는 합성곱과 풀링을 반복한다. 다섯 번째 계층에서는 모든 커널의 결과를 1차원 형태로

변환하고 2688개의 노드로 표현하는데, 이를 플랫텐(Flatten) 계층이라 부른다. 여섯 번째 계층에서는 2688개의 노드를 30개의 노드와 완전 연결한다. 일곱 번째 계층에서는 드롭아웃(Dropout)을 적용하며, 마지막인 여덟 번째 계층에서는 다시 완전 연결 계층을 통해 출력 노드 6개와 연결되고, 출력 결과는 소프트맥스(Softmax) 함수를 통해 반환 된다.

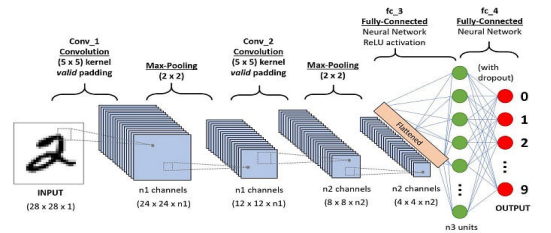


Fig. 1. CNN flatten layer

CNN은 필터의 크기, Stride, 패딩과 풀링을 조절하여 출력되는 데이터의 크기를 정하고, 필터의 개수로 출력 데이터의 차원을 결정한다. 이러한 구조는 같은 계층 크기를 가진 Fully Connected Neural Network와 비교할 때 학습 매개변수의 수가 약 20% 적다. 은닉층이 많아질수록 이 매개변수 수의 차이는 더욱 커진다. 결과적으로 CNN은 Fully Connected Neural Network에 비해 적은 매개변수로 더 높은 정확도를 달성할 수 있다고 할 수 있다.

### 3.2 LSTM

LSTM은 RNN(Recurrent Neural Network)에서 발생하는 소실 기울기 문제와 장기 의존성 문제를 해결하기 위해 설계된 알고리즘이다. 이 문제는 정보와 정보를 사용하는 지점 사이의 거리가 멀어질수록 기울기가 점차 감소함으로써 학습 능력이 크게 감소하는 현상을 말한다. LSTM은 RNN의 은닉 상태에 셀 상태(Cell State)를 추가하여, 오랜 시간이 지나도 기울기가 비교적 잘 전파될 수 있도록 한다.

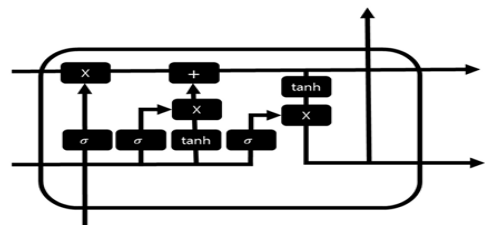


Fig. 2. LSTM Structure[15]

### 3.2.1 실험 LSTM 모델 설계 구조

실험에 사용된 LSTM 모델은 총 3개의 계층으로 구성된다. 첫 번째 계층인 LSTM 계층에서 데이터 연산을 수행한 후, 드롭아웃(Dropout)을 적용하여 과적합을 방지한다. 마지막으로, 출력 계층에서 데이터를 6개의 음악 장르로 분류한다.

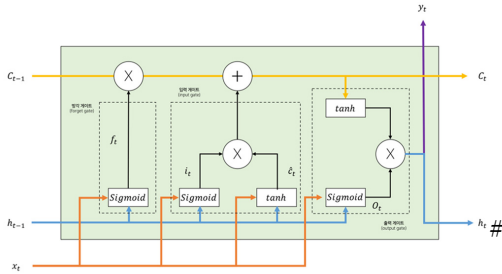


Fig. 3. LSTM Model

## 4. 실험 및 결과

### 4.1 학습 데이터 구성

학습 데이터는 경음악, 발라드, 록, 블루스, R&B, 클래식 등 6개의 장르별로 각각 100개씩, 총 600개의 약 5초 길이 음악 데이터를 사용한다. 이 데이터들은 모델의 학습과 성능 평가를 위해 학습 데이터와 테스트 데이터로 분리된다. 학습 데이터는 모델을 학습하는 데 사용되며, 테스트 데이터는 학습에 참여하지 않은 상태에서 머신러닝 모델의 성능을 평가하는 데 사용된다. 데이터 분할 비율은 학습 데이터와 테스트 데이터가 8:2로 설정되어 있다. 또한, 학습 데이터 내에서는 트레이닝 데이터와 검증 데이터로 더 세분화되어 있으며, 이 비율은 6:4로 설정되어 있다.

Fig. 4는 실험에 사용된 머신러닝 모델의 구조이다.

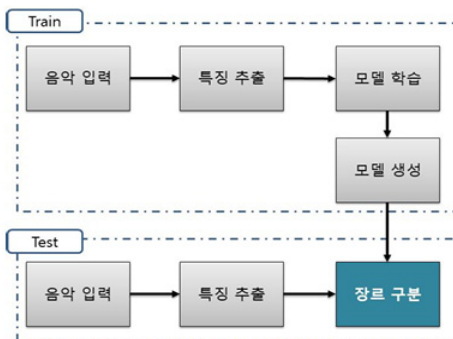


Fig. 4. Machine learning structure used in the experiment

두 가지 모델 모두 같은 전처리 과정을 거친다. 학습 단계에서는 음악을 입력받고, 전처리 과정에서 음악의 특징을 추출한다. 이 추출된 특징들은 머신러닝 모델에 입력되어 학습을 통해 모델이 구축된다. 테스트 단계에서는 학습된 모델을 사용하여 새로운 음악 입력의 특징을 추출하고, 이를 통해 음악의 장르를 분류하는 테스트를 진행한다.

### 4.2 전처리 과정

그림 5는 음악 데이터의 특징 추출 과정을 보여준다. 먼저, 음악 신호를 입력받아 고주파 증폭을 수행한다. 이 고주파 증폭은 저주파 성분에 비해 상대적으로 작은 크기의 고주파 성분을 강화하여 주파수 스펙트럼의 균형을 맞추는 데 도움을 준다. 이는 FFT(고속 푸리에 변환) 중 발생할 수 있는 수치 문제를 완화하고, 신호 잡음비(SNR)를 개선하는 데 유리하다.

그다음에는 해밍윈도우(hamming window)를 적용하고 FFT를 수행한다. 이후 Mel 스케일의 삼각 필터(일반적으로 40개 필터 사용)를 적용하고, 로그 함수를 적용한 뒤 DCT(이산 코사인 변환)를 진행한다.

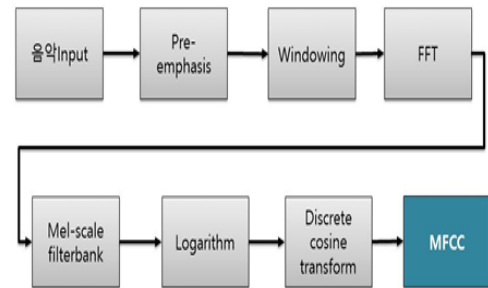


Fig. 5. Music data feature extraction process

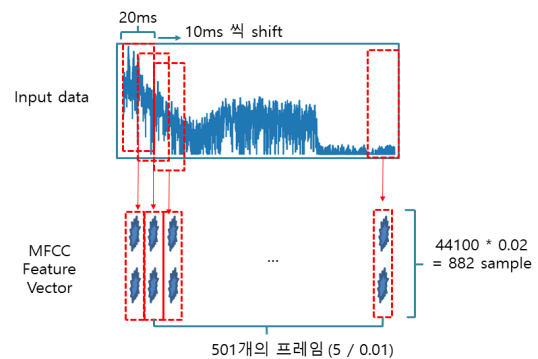


Fig. 6. Reconfiguring Data for MFCC

먼저, 특징 추출을 위해 입력된 음악 데이터를 50% 오버랩하며 20ms씩 재구성한다. 이때 각 데이터마다 501개 프레임이 생성 된다. 이에 대해 512-point FFT를 적용하여 파워 스펙트럼을 추출한다.

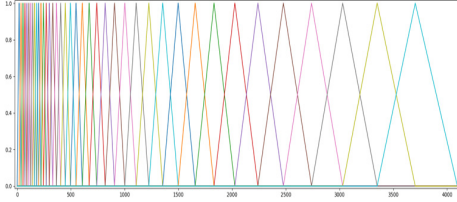


Fig. 7. MFCC Filterbanks

이후 멜 스케일의 Triangular filter를 power Spectrum에 적용하고 필터뱅크의 상관도를 줄이기 위해 DCT를 적용한다. 마지막으로 스펙트럼의 밸런스를 유지하고 SNR을 향상 시키기 위한 것으로 모든 프레임에서 각 계수의 평균을 뺀다. 위 그림은 멜스케일 필터뱅크이며, 아래 식은 필터뱅크의 식이다.

$$H_m(k) = \begin{cases} 0 & k < f(m-1) \\ \frac{k - f(m-1)}{f(m) - f(m-1)} & f(m-1) \leq k \leq f(m) \\ \frac{f(m+1) - k}{f(m+1) - f(m)} & f(m) \leq k \leq f(m+1) \\ 0 & k > f(m+1) \end{cases}$$

Fig. 8. Filterbank Equation

#### 4.3 실험 환경

학습은 Windows 10 환경에서 Python 3.7 언어로 작성된 TensorFlow v2.0 기반의 Keras v2.3.1 라이브러리를 사용했으며 CNN 모델의 학습 시에는 GPU의 병렬 처리 연산 기능을 활용하기 위하여 NVIDIA의 RTX 2080 Super 와 CUDA v10.0, cuDNN v10.0을 사용 하였다.

#### 4.4 실험 결과

데이터 세트에 대한 테스트를 위하여 본 논문에서는 두 가지의 머신러닝 모델을 비교하였다. 두 가지 모델 모두 batch size 32, epoch 500으로 설정하였다.

Table 1. CNN and LSTM accuracy and error

	CNN	LSTM
TEST_ACC	98.7	96.5
TEST_LOSS	0.09	0.26

위의 Table 1은 CNN과 LSTM의 정확도와 오차를 나타낸다. CNN모델은 98.7%의 정확도를 나타냈고, LSTM 모델은 96.5%의 정확도를 나타냈다.

아래 Fig. 9, 10은 CNN과 LSTM의 학습 그래프이다.

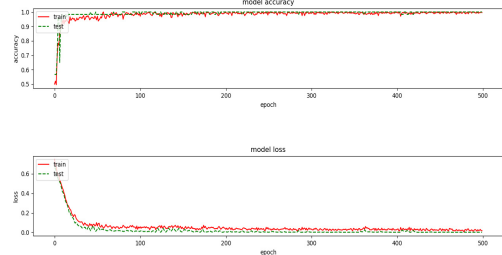


Fig. 9. CNN's learning accuracy and error

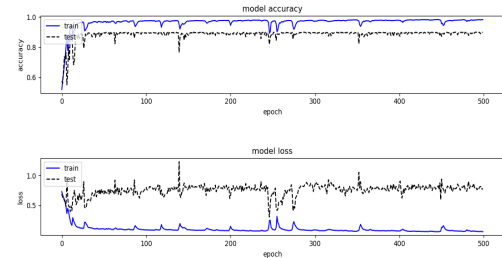


Fig. 10. LSTM's learning accuracy and error

실험 결과 CNN의 성능이 LSTM보다 높게 측정되었다. 이는 CNN의 특징을 자동으로 추출하여 학습하는 특성 때문이라고 판단된다.

Table 2. CNN의 Confusion Matrix

실제 \ 예측	0	1	2	3	4	5
0	476 (95.0%)	3	17	5	0	17
1	3	448 (89.4%)	36	13	9	16
2	9	6	417 (83.2%)	15	0	3
3	7	6	0	425 (84.8%)	0	11
4	0	13	18	15	469 (93.6%)	68
5	6	25	13	28	23	386 (77.0%)
계	501	501	501	501	501	501

Table 2는 CNN의 Confusion Matrix이다. 5초가량의 음악 데이터를 0.01 단위로 쪼개어 분류한 결과 정확도의 차이는 있지만 모두 장르를 정확히 분류했다.

## 5. 결론

본 논문에서는 MFCC 전처리를 통해 음악 장르를 구분하기 위한 머신러닝 기반의 두 가지 모델, CNN과 LSTM을 학습시키고 비교 분석하였다. 즉, 전처리 과정을 통해 음악의 특징을 추출하고, 이를 이용해 머신러닝 기법으로 학습하여 장르를 구분하는 방법을 제안하였다. 기존 연구들은 주로 자동 음악 추천 시스템에 초점을 맞추었으며, 음악 데이터에서 직접 특징을 추출하여 장르 구분에 관한 연구는 상대적으로 부족했다.

경음악, 발라드, 록, 블루스, R&B, 클래식 등 6가지 장르로 분류한 실험 결과, CNN을 사용한 경우가 LSTM을 사용했을 때보다 높은 정확도를 보였으며, 테스트 결과 CNN의 정확도는 98.7%로 매우 높았다. 이는 특징을 자동으로 추출해주는 CNN의 특성 때문으로 판단된다. CNN은 데이터 증강을 통해 학습 데이터를 인위적으로 확장하고, 모델의 일반화 성능을 향상시킨다. 또한, 노이즈를 최소화함으로써 모델이 패턴을 더 정확하게 학습할 수 있도록 지원하는 장점을 가지고 있다. 본 연구 결과를 통해 단편적으로 음악 장르를 분류하는 것뿐만 아니라, 비슷한 응용 분야에서의 가능성을 제시한다.

본 연구는 CNN과 LSTM을 활용하여 음악 장르 분류를 수행함으로써 멀티모달 학습의 효과를 검증하고, 데이터 증강 및 MFCC 전처리 기법을 적용하여 모델의 일반화 성능을 향상시켰다. 높은 정확도를 달성한 CNN을 통해 자동 음악 추천 시스템과 비슷한 응용 분야에서의 활용 가능성을 확인하였으며, 제한된 데이터와 비 최적화된 매개변수 설정에도 불구하고 유의미한 결과를 제시함으로써, 향후 더 많은 데이터와 새로운 기술을 활용한 추가 연구의 필요성을 제시하였다. 이는 음악 분류 기법에 대한 새로운 접근을 제공하고, 해당 분야의 연구 기반을 확장하는데 유의미한 결과를 도출하였다.

연구에서 사용된 데이터가 적고, 매개변수 최적화가 충분히 이루어지지 않았음에도 불구하고 우수한 결과가 나타난 것을 감안할 때, 신경망을 활용하는 머신러닝 및 인공지능 기술은 음악 장르 분류의 정확도를 크게 향상시킬 수 있을 것으로 예상된다. 본 논문에서는 각 장르별로 100개의 데이터를 사용하여 학습과 테스트를 진행하였으나, 추후 정확도를 더 높이기 위해서는 더 많은 데이터, 새로운 알고리즘 및 최적화된 매개변수 설정을 통한 추가 연구가 필요하다.

## References

- [1] S.E. Kim, E.J. Song, "A Study of New Possibility on Computer Music", Journal of Digital Contents Society, Vol.10, No.4, pp.561-568, 2009.  
<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE02270757>
- [2] J. H. Lee, "Music genre classification using Deep learning", Korean Institute of Information Scientists and Engineers, Pusan, pp.2007-2009, Dec, 2017.  
<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE07322780>
- [3] Pierce & J.Robinson, "The Science of Musical Sound", W.H. Freeman & Company Inc., USA, PP.25-48, 1992.
- [4] S.H. Kim, D.S. Kim and B.W. Suh, "Music Genre Classification using Multimodal Deep Learning", Proc. of HCI KOREA 2016, pp.389-395, Jan, 2016.  
[https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06592384&language=ko\\_KR&hasTopBanner=true](https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06592384&language=ko_KR&hasTopBanner=true)
- [5] S. Oramas, O. Nieto, F. Barbieri and X. Serra, "Multi-label music genre classification from audio, text, and images using deep features", Proc. of the 18th International Society for Music Information Retrieval Conference(ISMIR), pp.23-30, 2017.  
DOI: <https://doi.org/10.48550/arXiv.1707.04916>
- [6] S.H. Jang, H. Kim, S.W. Lee, H.M. Kim, Y.H. Kim, "Genre Classifier Using Machine Learning and Representative Segments for MIDI files", Korean Institute of Information Scientists and Engineers, Vol.36, No.2C, pp.375-378, Nov.2009.  
<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE01304542>
- [7] A. Tsaptsinos, "Lyrics-based Music Genre Classification using a Hierarchical Attentions Net work", Proc. of the 18th International Society for Music Information Retrieval Conference(ISMIR), pp. 694-701, 2017.  
DOI: <https://doi.org/10.48550/arXiv.1707.04678>
- [8] M. Fell and C. Sporleder, "Lyrics-based Analysis and Classification of Music", Proc. of the 25th International Conference on Computational Linguistics (COLING), pp. 620-631, 2014.  
<https://aclanthology.org/C14-1059/>
- [9] B. L. Sturm, "The GTZAN dataset: Its contents, its faults, their effects on evaluation, and its future use", arXivpreprint arXiv:1306.1461, 2013.  
<https://www.tandfonline.com/doi/abs/10.1080/09298215.2014.894533>
- [10] K.W. Choi, G. Fazekas, K.H. Cho and M. Sandler, "The effects of noisy labels on deep convolutional neural networks for music classification", arXiv preprint arXiv:1706.02361, 2017.  
DOI: <https://doi.org/10.48550/arXiv.1706.02361>
- [11] G. Tzanetakis, G. Essl and P. Cook, "Automatic Musical Genre Classification Of Audio Signals", Proc.

of the 2nd International Society for Music Information Retrieval Conference(ISMIR), pp.205-210. 2001.

<https://www.semanticscholar.org/paper/Automatic-Musical-Genre-Classification-of-Audio-Tzanetakis/16109ffe9d0a10e40703a3888271b3af770eaf16>

- [12] K.W. Choi, G. Fazekas and M. Sandler, "Automatic tagging using deep convolutional neural networks", Proc. of the 17th International Society for Music Information Retrieval-22 Conference(ISMIR), pp.805-811, 2016.  
DOI: <https://doi.org/10.48550/arXiv.1606.00298>
- [13] S.I. Han, "Speech-Music Discrimination Using Deep Learning", Journal of the Korea Academia-Industrial cooperation Society, Vol.22, No.10, pp.552-557, 2021.  
<https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE10619496>
- [14] S.H. Kim, D.S. Kim and B.W. Suh, "Music Genre Classification using Multimodal Deep Learning", Proc. of HCI KOREA 2016, pp.389-395, Jan, 2016.  
[https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06592384&language=ko\\_KR&hasTopBanner=true](https://www.dbpia.co.kr/journal/articleDetail?nodeId=NODE06592384&language=ko_KR&hasTopBanner=true)
- [15] S.H. Yoon, "CNN - LSTM based Speech Emotion Recognition using Variable Max Pooling", Master's Thesis, Konkuk University of Computer Science, Korea, 2022.  
<https://www.dbpia.co.kr/journal/detail?nodeId=T16082199>

## 박 재 표(Jae-Pyo Park)

[종신회원]



- 1998년 8월 : 숭실대학교 대학원 컴퓨터학과 (공학석사)
- 2004년 8월 : 숭실대학교 대학원 컴퓨터학과 (공학박사)
- 2010년 3월 ~ 현재 : 숭실대학교 정보과학대학원 교수

<관심분야>

정보보안, 보안평가 및 인증, 디지털포렌식, FinTec, 인공지능

## 서 정 환(Jung-Hwan Seo)

[정회원]



- 2020년 8월 : 숭실대학교 정보과학대학원 정보보안학과 (공학석사)
- 2024년 2월 : 숭실대학교 대학원 금융기술융합학과 박사과정 수료
- 2021년 6월 ~ 현재 : 주식회사 시스엔 SI사업부장

<관심분야>

정보경영, 정보통신