

# Vision Transformer를 이용한 자동변조인식 기술

이민주\*, 채명호\*, 임완수<sup>o</sup>

## Research on Automatic Modulation Recognition Using Vision Transformer

Minju Lee\*, Myoungcho Chae\*, Wansu Lim<sup>o</sup>

요약

자동변조인식 (AMR, Automatic Modulation Recognition)은 무선 통신 시스템에서 핵심적인 역할을 하는 기술로, 데이터 통신의 효율성 향상 및 무선 통신 시스템의 신뢰성과 보안 강화에 기여한다. 최근 딥러닝 기술 발전으로 AMR 분야도 딥러닝을 활용하여 변조 인식 성능을 향상하는 연구가 매우 활발히 수행되고 있다. 이에 본 논문은 시계열 이미지 데이터 처리 능력이 뛰어난 ViT (Vision Transformer) 모델 기반 AMR 기술을 제안한다. ViT 모델은 입력 이미지를 작은 이미지 단위로 나눈 후, 각 패치에 순서를 할당하여 Transformer Encoder의 입력으로 사용한다. ViT 기반 AMR 모델은 각 변조 방식의 성상도를 학습하여 변조 방식을 인식한다. 제안한 변조 인식 기법은 낮은 SNR에서도 변조 인식 정확도가 평균 약 2% 향상되었다.

**키워드** : 자동변조인식, Vision transformer(ViT), Transformer encoder, 무선통신시스템, 성상도

**Key Words** : Automatic modulation recognition, Vision transformer(ViT), Transformer encoder, Wireless communication systems, Constellation

### ABSTRACT

Automatic Modulation Recognition (AMR) is a technology that plays a key role in wireless communication systems, contributing to improving the efficiency of data communication and enhancing the reliability and security of wireless communication systems. Recently, due to the development of deep learning technology, research using deep learning has been actively conducted in the field of AMR. In this paper, we propose an AMR technique based on the ViT (Vision Transformer) model, which has excellent time series data processing capabilities. The ViT model divides the input image into patches, which are small image units, and assigns an order to each patch, which is used as an input to the transformer encoder. By doing so, the ViT-based AMR model learns the characteristics of each modulation scheme and automatically recognizes the modulation scheme. By using the ViT-based AMR model, we were able to achieve an average classification accuracy improvement of about 2% even at low SNR.

※ 본 연구는 방위사업청의 재원으로 국방과학연구소의 지원을 받아 수행된 연구임 (311JJ5-912967201).

◆ First Author : Kumoh National Institute of Technology, Department of Electronic Engineering, apfhs0927@naver.com, 학생회원

◦ Corresponding Author : Sungkyunkwan University, School of Electronic and Electrical Engineering, wansu.lim@skku.edu, 정회원

\* Agency for Defense Development, mhchae4940@naver.com

논문번호 : 202401-015-B-RN, Received January 17, 2024; Revised March 26, 2024; Accepted April 11, 2024

## I. 서 론

자동변조인식 (AMR, automatic modulation recognition)은 무선 통신 시스템에서 신호의 변조 방식을 자동으로 감지하고 구분하는 기술로 통신의 효율성과 신호 간 간섭 관리에 큰 영향을 미친다<sup>[1]</sup>. AMR은 주로 무선 통신 및 레이더 시스템에서 활용하며, 신호의 변조 방식을 정확히 식별하여 신호의 내용을 이해하거나 분석하는데 사용된다. AMR을 사용한 신호 분류 및 식별은 무선 통신의 다양한 측면에서 이점을 제공한다<sup>[2]</sup>. 다양한 변조 방식을 구별하여 신호의 종류를 파악하므로 특정 신호를 구분하여 처리하는 데 도움을 준다. 이는 신호의 안정성, 오류 감지 및 수정 능력을 높이며, 특정 환경에서 최적의 성능을 제공하도록 돕는다. 따라서 한정된 대역폭 내에서 최대한 많은 정보를 전송할 수 있게 된다. AMR 기술은 다양한 통신 환경에서 보안 강화와 운영 비용 절감 등에 기여할 수 있으므로 기존 신호처리 응용부터 딥러닝 활용까지 다양한 기법의 AMR 연구가 활발히 진행되고 있다.

기존 신호처리 기술 응용에서는 푸리에 변환, 웨이블릿 변환, 그리고 모멘트 등이 대표적으로 사용된다. [7]는 자동변조인식을 위하여 FFT 기반 정상도 PDF(Probability Density Function) 분석 기법을 제안했다. 정상도 PDF가 변조 방식에 따라 주파수 특성이 다름을 이용하여 변조 방식을 인식했다. 제안한 기법은 수신 심벌 수가 적으면 타 변조 방식 검출 기법에 비해 변조 인식 성능이 크게 저하되는 단점이 있다. [8]는 변조 신호의 웨이블릿 변환에서 추출한 특징을 이용하여 신호를 분류했다. 특히, 시간 영역에서 신호를 작게 나눈 후 각각에 주파수 정보를 부여하여 시간과 주파수를 동시에 분석했다. 이는 신호의 특징 추출에는 유용하지만, 입력 데이터의 길이에 따라 계산복잡도가 선형적으로 증가하므로 대규모 데이터에서는 확장성이 제한되는 한계가 있다. [9]는 고차 큐물런트를 이용한 위상 매핑 식별 및 변조 분류 알고리즘을 제안했다. 변조 방식마다 고차 큐물런트 특징을 추출하여 분류 알고리즘을 설계했다. 계산복잡도를 최소화하기 위해 신호의 실수부만 사용했으나 복소수의 위상 정보를 잃게 되어 변조 방식 간 세부적인 차이를 구분하는 성능은 저하되었다.

딥러닝 기반 AMR 연구에서는 완전 연결 네트워크, CNN(convolutional neural network) 및 LSTM(long short-term memory) 등의 방식을 주로 활용한다. [1]은 완전 연결 네트워크 기반 자동 변조 인식 기법을 제안했으며, 시간, 주파수, 그리고 시간-주파수 혼합 영역에서

변조 인식 성능을 분석했다. 완전 연결 모델은 공간적 구조를 배제하고 모든 입력 데이터를 같은 차원의 뉴런에서 처리하기 때문에 신호의 형상에 담긴 중요 정보를 활용할 수 없는 한계가 있다.

[2]는 자동변조분류를 위해 CNN을 사용했고, CNN의 그래디언트 소실 문제를 완화하기 위해 ResNet의 스킵 커넥션 구조를 적용했다. 또한, 네트워크의 계산복잡도를 줄이기 위해 비대칭 컨볼루션 필터 구조를 적용했으며, 정규화된 프레임(N)은 송신된 복소 신호인 실수부와 허수부로 나누어  $2 \times N$  매트릭스로 구성했다. 프레임을 확장하기 위해  $2 \times N$  매트릭스를 복사 후 순서를 역으로 나열한 다음 기존  $2 \times N$  매트릭스와 연결했다. 이러한 방식으로 프레임을  $2 \times N$ 에서  $4 \times N$ 로 확장하여 모델에 적용했다. 그러나 제안한 방법은 다양한 크기의 데이터( $2 \times N$ ,  $4 \times N$ )를 동일한 구조의 모델에 적용했을 때, 데이터의 크기가 증가할수록 계산복잡도가 증가하는 문제점이 있다.

[3]은 자동 변조 분류를 위해 CNN-LSTM 기반 이중 스트림 구조를 사용했으며, 제안한 모델은 CNN과 LSTM을 결합하여 신호의 공간적 및 시간적 특징을 효과적으로 추출했다. CNN은 공간적 특징 추출에 우수한 성능을 보이지만 시계열 데이터에서 시간적 정보를 학습하는 데에는 효과적이지 않기 때문에 시계열 데이터에서 지속적인 특징을 학습하는 데 뛰어난 LSTM과 결합하여 사용했다. 또한, 이중 스트림 구조를 통해 I/Q 형식과 A/P 표현의 신호를 동시에 고려하여 성능을 향상시켰다.

본 논문은 CNN과 LSTM의 장점을 효율적으로 결합하여 시계열 데이터 처리에 있어 우수한 성능을 보이는 ViT를 이용한 자동변조인식 알고리즘을 제안한다. ViT는 self-attention 메커니즘으로 위치 임베딩을 실시하여 이미지의 공간적 특징을 파악한다<sup>[10],[11]</sup>. 또한, 이미지를 작은 패치로 나눈 후 각 패치 간 관계 학습을 통해 전체 이미지 구조를 파악하므로 CNN보다 더 효율적으로 공간적 특징을 추출한다. 패치기반 접근법은 입력 이미지의 크기와 상관없이 고정된 크기의 패치를 이용함으로써 이미지의 크기가 증가하면 패치의 수는 증가하지만, self-attention의 계산복잡도는 패치의 크기에 따라 결정되므로 패치 수는 계산복잡도에 큰 영향을 주지 않는다. 따라서 ViT는 다양한 크기의 데이터에 대해 일관된 모델을 적용할 수 있다.

## II. Vision Transformer를 이용한 AMR

그림 1은 제안한 시스템 구조이고, ‘Data Generation’,

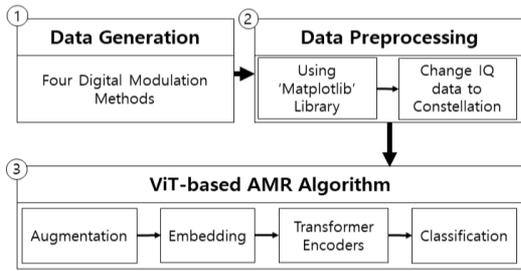


그림 1. 제안하는 시스템 구조  
Fig. 1. Proposed system structure

‘Data Preprocessing’, ‘ViT-based AMR Algorithm’로 구성한다. ‘Data Generation’ 과정에서 통신 신호 자동 변조 인식을 위해 다양한 디지털 통신 신호 변조 형태와 SNR을 사용하여 신호를 생성했다. ‘Data Preprocessing’ 과정은 신호를 성상도 형태로 변환하여 데이터로 사용할 수 있도록 하였다. 데이터를 이미지 형태로 변환함으로써 통신 신호의 변조 패턴은 이미지의 공간적 특징으로 나타난다. 마지막으로, ‘ViT based AMR algorithm’ 과정은 ViT 아키텍처를 기반으로 Transformer Encoder를 사용하여 동작한다. 이러한 시스템 구성을 통해 통신 신호의 변조를 효과적으로 감지하고 분류할 수 있다.

### 2.1 Data Generation

본 논문은 ViT 모델의 학습과 검증에 사용할 데이터를 생성하여 연구를 수행했다. 생성한 데이터는 무선 통신 시스템에서 활용되고 있는 다양한 디지털 변조 방식 중 BPSK, QPSK, 8PSK, 16QAM 등 총 4가지 변조 방식을 사용했다. 무선 통신 채널은 AWGN(additive white Gaussian noise) 환경을 사용했고, 잡음 파워에 따라 SNR(signal-to-noise ratio) 값을 변경했다. 최신 딥러닝 기술이 변조 인식에 미치는 영향에 집중하고자 비교적 간단한 채널 모델을 사용했다. 신호는 IQ (inphase quadrature) 1,024개로 생성하였고, SNR 값은 -20dB부터 20dB까지 2dB씩 증가하도록 설정하였다. 각 변조 인식과 SNR의 조합마다 2,000개의 데이터를 생성하여 총 168,000개의 데이터를 사용했으며 학습, 검증, 그리고 테스트 데이터는 각각 8:1:1 비율로 설정하여 사용했다.

### 2.2 Payload 제거

그림 2는 Data Preprocessing 과정을 시각적으로 보여준다. ViT 모델은 이미지 처리에 유용한 모델이므로 신호 데이터를 이미지 데이터로 변환하여 사용해야 한

```

    IQ data : [[-0.05551403+0.89513428j]
              [ 0.95873596-0.09280268j]
              ...
              [ 0.12388173-0.92579078j]
              [-0.7959748 -0.7063038j ]]
    
```

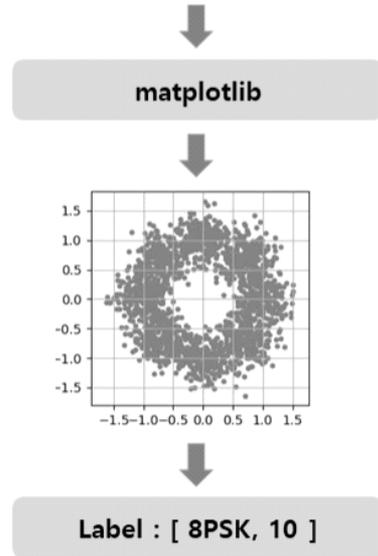


그림 2. 데이터 전처리 과정  
Fig. 2. Data preprocessing

다. 이를 위해 신호의 IQ 데이터를 성상도로 변환하여 이미지 데이터로 활용했다. 먼저 mat 형식으로 저장된 신호의 IQ 데이터를 ‘matplotlib’ 라이브러리를 활용하여 성상도로 변환한다. ‘matplotlib’은 다양한 차트, 그래프 및 플롯을 생성하는 데 사용하는 파이썬 라이브러리로, 이를 통해 신호를 시각적으로 분석하고 이해할 수 있다. 성상도는 모델의 입력 크기에 맞춰 32×32 pixel로 제작하였다.

데이터 처리 과정에서는 ‘numpy’ 라이브러리를 활용하여 다차원 배열을 효율적으로 관리하였다. ‘numpy’의 기능은 복잡한 배열 연산 및 조작을 쉽게 해주기 때문에 이미지 데이터와 변조 유형 및 SNR 등 정보를 효율적으로 처리할 수 있었다. 예를 들면 변조 유형 및 SNR을 그림 2와 같이 [8PSK, 10] 형태로 저장한다. 이렇게 2차원으로 구성된 배열 형태의 라벨은 분석 및 학습 단계에서 입력 형태로 활용된다.

### 2.3 ViT based AMR algorithm

그림 3은 제안한 ViT 기반 AMR 알고리즘 구조이고, 증강, 임베딩, Transformer Encoder, MLP Head, 분류 등 총 5단계로 구성한다.

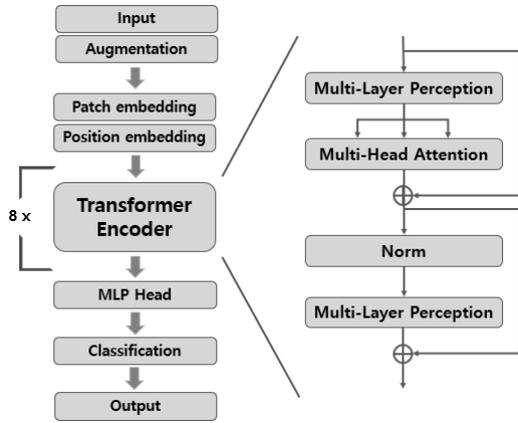


그림 3. ViT 구조  
Fig. 3. ViT architecture

2.3.1 데이터 증강 (Augmentation)

ViT의 첫 단계로 학습 데이터의 다양성 증가를 위해 정규화, 리사이징, 랜덤 플립, 랜덤 회전, 랜덤 줌 등 증강 기법을 사용했다. 데이터 정규화는 각 픽셀 값이 0과 1 사이의 범위에 들어가도록 조정하여 모델의 학습 속도를 향상시켰고, 리사이징은 이미지 크기를 일정하게 유지하였다. 랜덤 플립은 이미지를 수평 또는 수직으로 무작위로 뒤집는 것이며, 랜덤 회전은 이미지를 무작위로 회전한다. 마지막으로 랜덤 줌은 이미지를 무작위로 확대 또는 축소한다. 이러한 방식으로 원본 이미지에 다양한 변형을 가함으로써 모델이 학습하는 데이터의 다양성을 증가시켰다. 이러한 증강 기법은 모델이 신호 분석 문제에 적절하게 대응할 수 있도록 도움을 준다.

2.3.2 CSP 알고리즘

그림 4는 임베딩을 진행하는 전체적인 과정이다. 임베딩은 차원 축소를 통해 입력 데이터를 효과적으로 표현하는 방법이다. 이미지 처리 분야에서는 특정 객체, 개념, 또는 위치와 관련된 정보를 벡터로 임베딩한다. 이러한 임베딩은 고차원 시각적 데이터를 저차원 벡터로 표현하여 이미지의 중요한 특성을 간결하게 표현할 수 있다. ViT는 패치 임베딩과 위치 임베딩을 수행한다.

**패치 임베딩 (Patch Embedding) :**

ViT에서는 이미지를 패치로 나누어 처리하는데, 이는 전통적인 CNN 아키텍처와 차별화된 특징이다. 그림 4 (a)와 같이 이미지를 격자 형태의 패치로 나누고 각 패치를 벡터로 변환하여 패치 임베딩을 수행한다. 이를 통해 ViT는 이미지의 전역적인 구조와 지역적인 세부 정보를 함께 고려하여 학습할 수 있다. 각 패치의

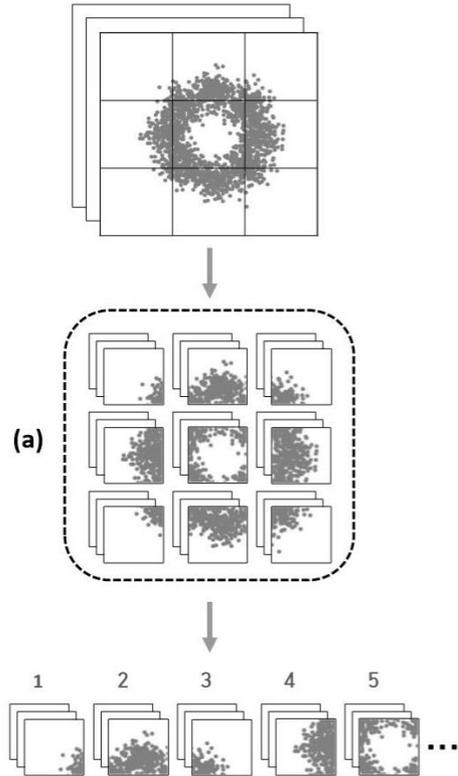


그림 4. 임베딩 진행 방법  
Fig. 4. Embedding process

벡터는 주변 패치와 관계를 고려하여 이미지 내의 중요한 특징을 효과적으로 포착할 수 있다.

또한, 패치 임베딩을 사용하면 입력 이미지 크기에 대한 계산복잡도 민감도를 줄일 수 있다. ViT의 계산복잡도는 나누어진 패치의 수가 아닌 고정된 패치의 크기에 따라 결정되므로 이미지를 고정된 크기의 패치로 처리하는 기법은 모델의 계산복잡도에 큰 영향을 끼치지 않는다. 따라서, 기존 변조 인식 기법은 입력 데이터 크기가 증가함에 따라 계산복잡도가 증가하지만, ViT는 다양한 크기의 입력 데이터에 일관된 구조의 모델을 사용할 수 있다.

**위치 임베딩 (Position Embedding) :**

다음으로 그림 4 (b)와 같이 위치 임베딩을 사용하여 패치의 상대적 위치 정보를 파악한다. 이때, 가우시안 함수를 이용하여 패치 위치에 관한 임베딩을 생성하거나, 사인 및 코사인 함수를 활용하여 위치 정보를 임베딩하는 방식을 사용한다. 이러한 위치 임베딩은 모델이 각 패치의 절대적인 위치 및 상대적인 관계를 학습할 수 있게 도와주며, 이미지 내의 패치 간 공간적인 구조

를 적절히 유지하면서 효과적으로 학습할 수 있다.

### 2.3.3 Transformer Encoder

ViT는 Transformer Encoder를 여러 번 반복하여 사용한다. Transformer Encoder는 Self-Attention, Multi-Layer Perception 그리고, Multi-Head Attention 으로 구성한다. ViT 모델의 Self-Attention 메커니즘은 입력 데이터의 모든 위치 간 상호 작용을 고려하므로 데이터의 전역적인 구조를 파악할 수 있을 뿐 아니라 데이터의 중요한 정보에 집중할 수 있다. 또한, ViT는 위치 임베딩을 사용하므로 시계열 데이터 처리에도 성능이 우수하다. 따라서, ViT는 CNN과 LSTM의 장점을 모두 제공하고 다양한 데이터 형태에 유연하게 학습하므로 데이터의 패턴 인식과 분류를 효과적으로 수행할 수 있다.

#### Multi-Layer Perception(MLP) :

MLP는 각 패치에 대한 Self-Attention 결과를 다양한 방식으로 조합하여 비선형 특징을 학습하기 위해 사용한다. 이를 통해 이미지 내의 세부적인 패턴과 특징을 더 효과적으로 학습할 수 있으며, 이미지의 전역적인 특성도 명확하게 이해할 수 있도록 돕는다. 따라서, ViT 모델이 복잡한 패턴을 학습하고, 높은 수준의 추론을 수행할 수 있도록 돕는다.

#### Multi-Head Attention(MHA) :

MHA는 여러 개 Attention Head로 구성된 메커니즘이며, 입력 정보를 여러 관점에서 파악하므로 다양한 특징을 추출할 수 있다. 각 Attention Head는 입력 정보의 서로 다른 부분에 집중하며, 이를 통해 복잡한 패턴을 효과적으로 학습할 수 있다. 또한, MHA는 병렬 학습이 가능하여 입력 데이터의 특징을 빠르게 파악할 수 있다.

### 2.3.4 분류 (Classification)

ViT의 Classification 파트는 이미지의 특징을 학습한 후, 다양한 클래스로 분류를 수행한다. 이를 위해 Transformer의 마지막 레이어에서 각 패치에 대한 정보를 종합하고 전역적인 특징을 고려하여 최종적으로 클래스에 대한 확률을 예측한다. 이때, Softmax 활성화 함수는 학습을 통해 얻은 가중치를 이용하여 클래스 예측 정확도를 높인다.

## III. 시뮬레이션 및 성능평가

### 3.1 실험 환경

이번 절은 ViT 모델 학습을 위한 파라미터 설정을 설명한다. 옵티마이저는 AdamW로 지정하였고 학습률 초기값과 가중치 감쇠율을 각각 0.001과 0.0001로 설정하여 손실을 최소화하였다. 배치사이즈는 256으로 설정하여 모델을 반복 학습하였고 트랜스포머 레이어는 총 8개 사용하였다. 트랜스포머 레이어 내의 유닛 수는 [128, 64]로 설정했고 최종 MLP 헤드에 사용된 유닛의 수는 [2048,1024]이고 드롭아웃 비율은 0.5를 사용했다. 논문에서 사용한 ViT 파라미터는 표 1에 자세히 정리했다. ViT 모델에 가장 적합한 Epoch를 찾기 위해 여러 Epoch에 대한 정확도 비교하였고 그림 5에 정리하였다. 이 과정에서 모델의 과적합 및 과소 적합을 파악했고, 특히 SNR이 0dB 이하일 때 모델의 정확도가 점진적으로 향상되는 양상을 확인했다. 이를 바탕으로 0dB 전후에서 학습이 수렴하기 시작하는 지점을 찾아내어 모델의 일반화 능력이 최대화되는 적절한 Epoch를 설정했다. 또한, 과적합을 피하려고 정확도의 상승이 미미하게 변하거나 감소하는 지점의 Epoch를 훈련에 사용했다.

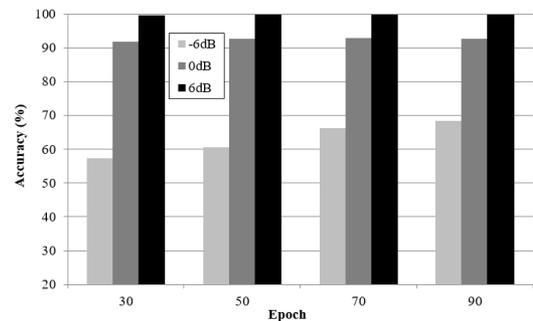


그림 5. Epoch별 정확도 비교  
Fig. 5. Accuracy comparison depending on Epoch

표 1. 제안한 ViT 파라미터  
Table 1. Proposed ViT parameters

Optimizer	AdamW	Learning rate	0.001
		Weight Decay	0.0001
Batch Size	256	Epoch	100
Transformer Encoder	8	Number of Layer units	[128, 64]
Activation Function	Softmax	Number of MLP unitx	[2048,1024]
Dropout	0.1	Final MLP Dropout	0.5

### 3.2 실험 결과

그림 6은 학습 데이터 크기와 정확도 간 관계를 나타낸 그래프이다. Large data는 변조 방식과 SNR 조합별 1,600개 데이터를 학습 과정에 활용했고, Small data는 각 800개의 데이터를 활용했다. 0dB 이상에서는 두 모델 모두 비슷한 정확도를 나타냈으나, 0dB 이하에서는 Large data가 평균적으로는 약 2.7%, 최대 8%의 더 높은 정확도를 보였다. 이러한 결과로 학습 데이터의 크기가 모델 성능을 높이는 데 중요한 역할을 한다는 것을 확인할 수 있었다.

그림 7은 SNR에 따른 변조 방식별 정확도를 혼동행렬(confusion matrix)로 표현했다. 그림 7 (d)와 (e)를 보면, QPSK를 8PSK로 많이 혼동한다는 것을 알 수 있으며, 그림 7 (f)를 보면 다른 3가지 변조 방식에 비해 낮은 SNR에서 8PSK의 분류 정확도가 매우 높은 것을 확인할 수 있다. 각 변조 방식별 전체적인 정확도 추이를 확인하기 위해 SNR에 대한 4가지 변조 방식 정확도를 그림 8에 나타냈다. 이는 그림 7보다 직관적으로 각 변조 방식의 정확도를 비교할 수 있게 한다. BPSK와 16QAM은 -10dB 이하에서는 정확도가 10% 이하로 거의 분류가 되지 않지만, -8dB 이상에서는 매우 높은 정확도를 보이며 -2dB 이상 SNR에서는 분류 정확도가 100%이다. QPSK 정확도는 전체 데이터의 정확도 그래프와 비슷한 양상을 보이며, 4dB 이상 범위에서는 정확도가 100%이다. 8PSK의 정확도를 보면, -20dB부터 약 70% 이상 분류 정확도를 보여준다. 이러한 특징이 나타나는 이유는 8PSK와 16QAM 모두 많은 위상을 사용하여 각 심볼이 나타내는 정보량이 많은 고밀도 모달리티 특징을 갖고 있으나, 16QAM은 8PSK에 비해 더 많은 위상을 사용하므로 심볼 간 거리가 더 가까워져 잡음이나 왜곡에 민감하다는 특징을 갖기 때문이다. 이러한 이유로, 8PSK가 다른 변조 방식에 비해 낮은 SNR에서도 우수한 분류 정확도를 보이는 것으로 파악할 수 있다.

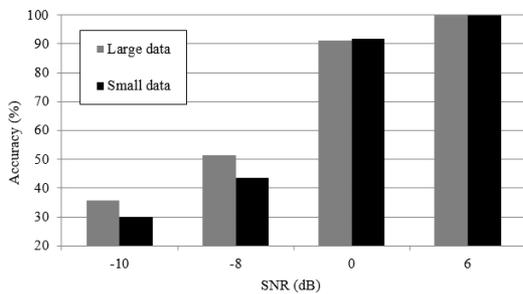


그림 6. 학습 데이터 크기 별 모델 정확도 비교  
Fig. 6. Comparison of model accuracy by training data size

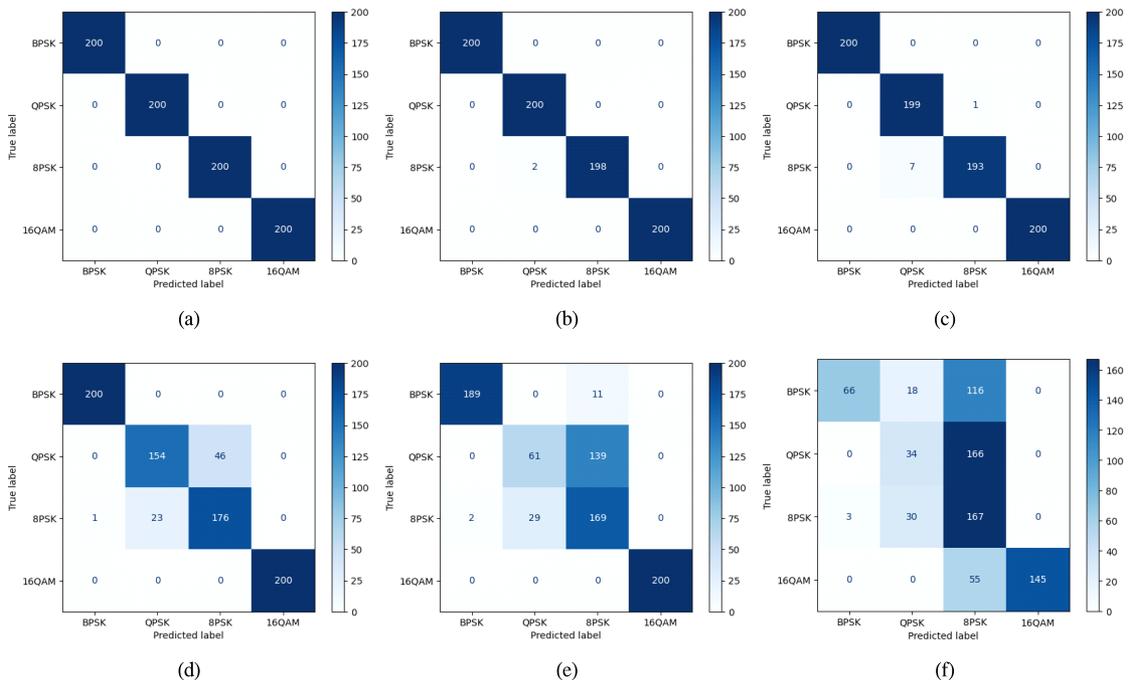


그림 7. SNR에 따른 혼동행렬 : (a) 10dB, (b) 6dB, (c) 4dB, (d) 0dB, (e) -4dB, (f) -8dB  
Fig. 7. Confusion matrix as a function of SNR : (a) 10dB, (b) 6dB, (c) 4dB, (d) 0dB, (e) -4dB, (f) -8dB

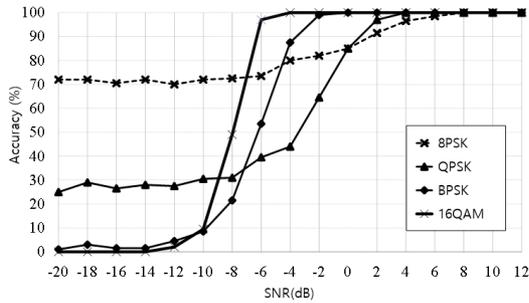


그림 8. ViT 모델 기반 변조별 정확도  
Fig. 8. Accuracy comparison by modulation type

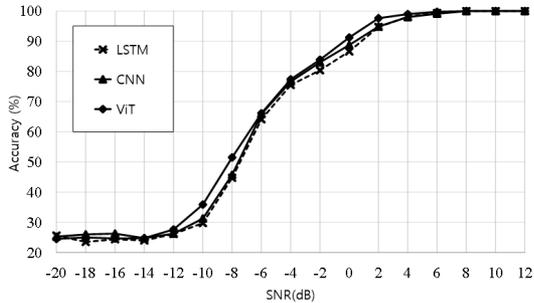


그림 9. ViT, CNN, LSTM의 정확도 비교  
Fig. 9. Accuracy comparison of ViT, CNN, and LSTM

본 논문은 ViT 모델의 AMR 성능을 평가하기 위해 CNN과 LSTM과 정확도 결과를 비교했다. 그림 9는 각 모델의 정확도를 비교한 그래프이다. 이를 보면 0dB에서 CNN, LSTM 그리고 ViT는 88.75%, 86.63% 그리고 91.25%의 예측 정확도를 보였고, 6dB에서는 99.13%, 99.38% 그리고 99.75%의 정확도를 가진다. ViT 정확도는 0dB에서 CNN과 LSTM보다 각각 2.5%와 4.62% 높았고, 6dB에서는 0.62%와 0.37% 높았다. 특히, 다른 모델에 대비하여 -8dB에서 CNN과 LSTM 각각 최대 5.75%와 6.75%의 향상된 정확도를 기록하며 ViT의 성능의 우수성을 입증했다. -10~6dB에서는 ViT 모델이 특히 뛰어난 성능을 보였으며, 12dB 이상에서는 변조 인식 정확도가 100%였다. 결론적으로 ViT 모델이 다양한 환경에서 안정적이고 강력한 성능을 보였다. 본 논문은 AWGN 채널에서 ViT의 우수성을 충분히 검증하였고, 이를 기반으로 페이딩, 다중 경로, 간섭 등 다양한 무선 환경을 고려한 추가 연구가 필요하다.

#### IV. 결론

본 논문은 ViT 모델 기반 자동변조인식 기법을 제안했다. 변조 방식은 BPSK, QPSK, 8PSK, 16QAM 등

4가지를 사용했고, 각 변조의 정상도를 학습하여 분류에 사용했다. 시뮬레이션 결과 SNR이 12dB 이상에서 제한한 기법의 분류 정확도는 100%였으며, SNR이 낮은 -6dB에서도 제한한 변조 인식 기법이 CNN과 LSTM을 사용한 기법보다 분류 정확도가 각각 5.75%와 6.75% 향상되었다. 이를 통해 본 연구에서 사용한 ViT 모델이 변조 방식을 효과적으로 인식하고 분류할 수 있음을 검증하였다.

#### References

- [1] J. J. Kang and J. H. Kim, "Performance analysis in automatic modulation classification based on deep learning," *J. KIICE*, vol. 25, no. 3, pp. 427-432, Mar. 2021.
- [2] S. H. Kim, C. B. Moon, J. W. Kim, and D. S. Kim, "Design of deep learning-based automatic modulation classification with extended frame," *J. KICS*, vol. 46, no. 8, pp. 1227-1236, Aug. 2021.
- [3] Z. Zhang, H. Luo, C. Wang, C. Gan, and Y. Xiang, "Automatic modulation classification using CNN-LSTM based dual-stream structure," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13521-13531, Nov. 2020.
- [4] H. Kim, H. Kim, J. Je, and K. Kim, "A deep learning method for the automatic modulation recognition of received radio signals," *J. KIICE*, vol. 23, no. 10, pp. 1275-1281, Oct. 2019.
- [5] S. H. Kim and D. S. Kim, "Design of convolutional neural networks based on frame extension for automatic modulation classification," *2021 Winter Conf. KICS*, pp. 579-580, Feb. 2021.
- [6] B. Jdid, K. Hassan, I. Dayoub, W. H. Lim, and M. Mokayef, "Machine learning based automatic modulation recognition for wireless communications: A comprehensive survey," *IEEE Access*, vol. 9, pp. 57851-57873, Apr. 2021.
- [7] S. W. Ha, K. H. Lee, J. H. Jang, and H. J. Choi, "An enhanced constellation PDF analysis method based on FFT for automatic modulation classification," in *Proc. Symp.*

- KICS*, pp. 1262-1263, 2012.
- [8] C. S. Park, S. P. Nah, J. W. Yang, and J. H. Choi, "Automatic recognition of digital modulation types using wavelet transformation," *The IEEE Telecommun.*, vol. 45, no. 4, pp. 22-30, Apr. 2008.
- [9] J. Lee, S. Ahn, J. Choi, and D. Yoon, "Phase mapping recognition and modulation classification algorithm using new Higher-order Cumulants," *J. KIIT*, vol. 15, no. 2, pp. 153-160, Feb. 2017.
- [10] A. Steiner, A. Kolesnikov, X. Zhai, R. Wightman, J. Uszkoreit, and L. Beyer, "How to train your vit? data, augmentation, and regularization in vision transformers," *arXiv preprint arXiv:2106.10270*, Jun. 2022. (<https://doi.org/10.48550/arXiv.2106.10270>)
- [11] Z. Li, S. Li, and X. Yan, "Time series as images: Vision transformer for irregularly sampled time series," *Advances in NIPS*, vol. 36, 2024.

이 민 주 (Minju Lee)

2019년 3월~현재 : 금오공과대학교 전자공학과  
<관심분야> 자동변조인식, 머신러닝, 최적화

채 명 호 (Myoungho Chae)

2014년 2월~현재 : 국방과학연구소 선임연구원  
<관심분야> 통신 프로토콜, 프로토콜 역공학  
[ORCID:0000-0001-7741-1818]

임 완 수 (Wansu Lim)

2024년 3월~현재 : 성균관대학교 전자전기공학부 교수  
<관심분야> 통신 프로토콜, 기계학습, 자동변조인식  
[ORCID:0000-0003-2533-3496]