

동질적 특징을 이용한 견고한 멀티모달 분류 모델

Robust Multimodal Classification Model Using Homogeneous Features

배기웅⁰¹, 이수안², 이우기¹

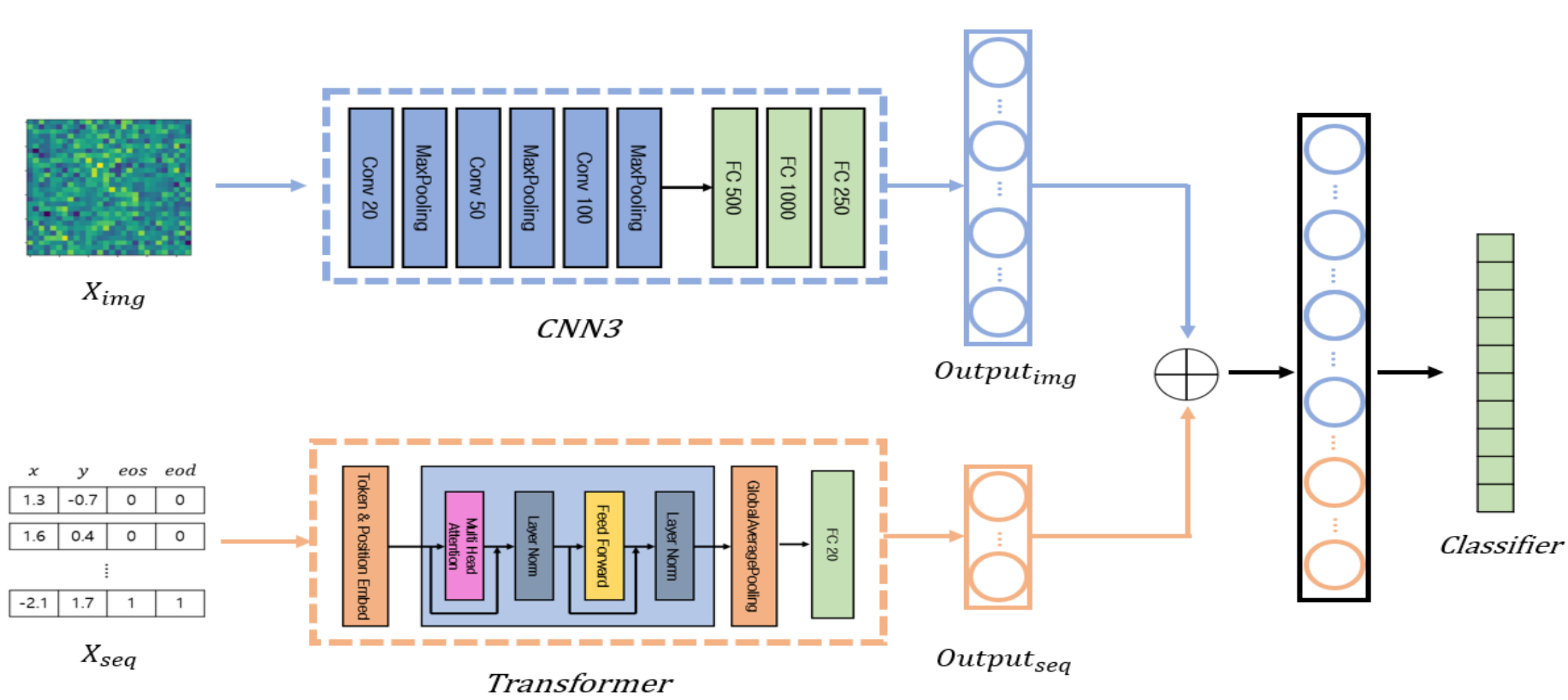
¹인하대학교 산업경영공학과, ²세명대학교 컴퓨터공학부



1. 서론

- ✓ 특징 유형이 다른 데이터를 동시에 학습시켜 결과를 제시하는 멀티모달은 모델의 일반화 과정에서 견고한 능력을 보인다고 알려져 있다
- ✓ 이러한 멀티모달의 능력은 데이터 획득비용이 높거나 정제가 어려운 경우일수록 더욱 유용하다. 이미지 MNIST 데이터셋과 시퀀스 MNIST 데이터셋은 학습하기가 쉬워 단일 CNN모델과 단일 RNN 모델로도 각각 99%, 95% 이상의 성능을 보이지만, 데이터셋에 노이즈를 추가하게 되면 성능은 크게 저하된다.
- ✓ 본 연구에서는 정제되지 않은 데이터셋을 학습시킬 경우, 단일모달보다 멀티모달을 사용하는 것이 노이즈에 견고하며 좋은 성능을 보여줄 것을 제안한다.

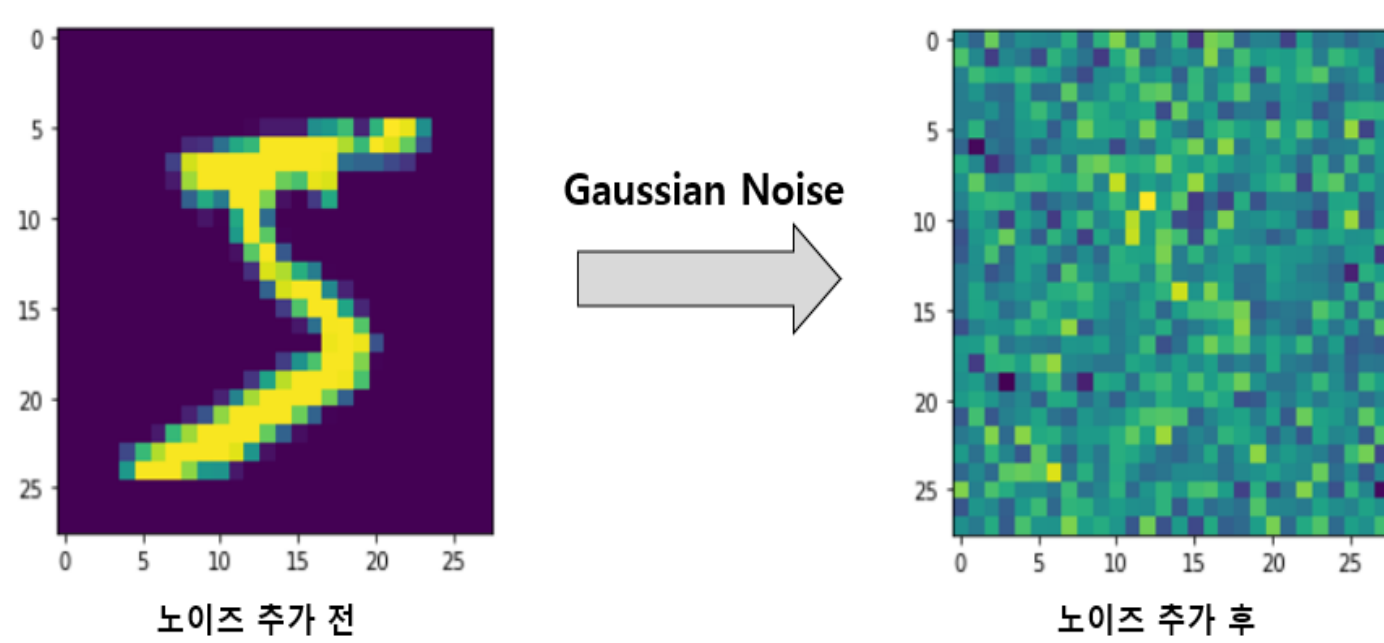
2. 동질적 특징을 이용하는 멀티모달



- ✓ 멀티모달 모델은 학습과정에서 서로 다른 유형의 데이터를 동시에 활용하고, 상호 보완하여 보다 폭넓은 특징의 포착이 가능하다는 장점이 있다.
- ✓ 따라서 멀티모달 모델은 분류 문제에 있어 노이즈에 견고하게 작동한다.
- ✓ 본 연구에서는 멀티모달 모델은 동질적 특징을 가지는 이미지 데이터셋과 시퀀스 데이터셋을 모델의 입력으로 사용하여 각 모달별 모델의 결과를 결합하였다.
- ✓ 동질적 특징을 가지는 멀티모달 학습의 성능을 평가하기 위해 CNN 계열의 이미지 분류 모델에 LSTM, BiLSTM, 그리고 트랜스포머를 결합시켜 총 3개의 멀티모달 모델을 제작하였다.

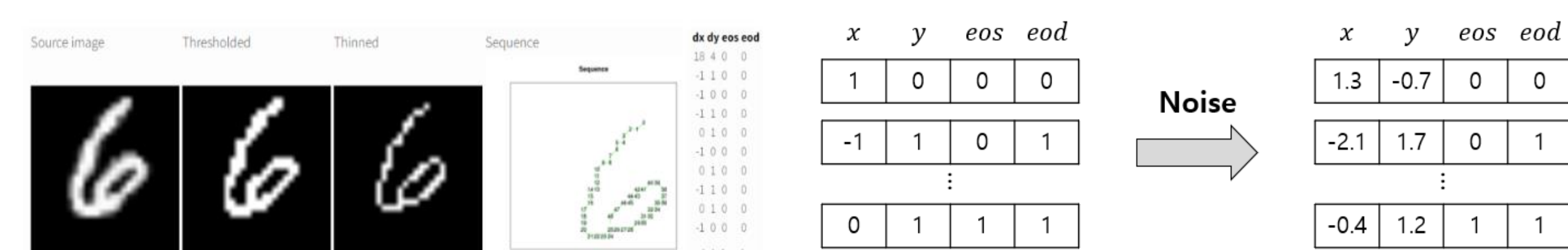
3. 노이즈를 추가한 MNIST 데이터셋

이미지 MNIST 노이즈 데이터셋



- ✓ 이미지 MNIST 데이터셋의 노이즈는 가우시안 노이즈 함수를 사용하였으며, 노이즈는 정규분포 $N(0,0.9^2)$ 을 따른다.

시퀀스 MNIST 노이즈 데이터셋

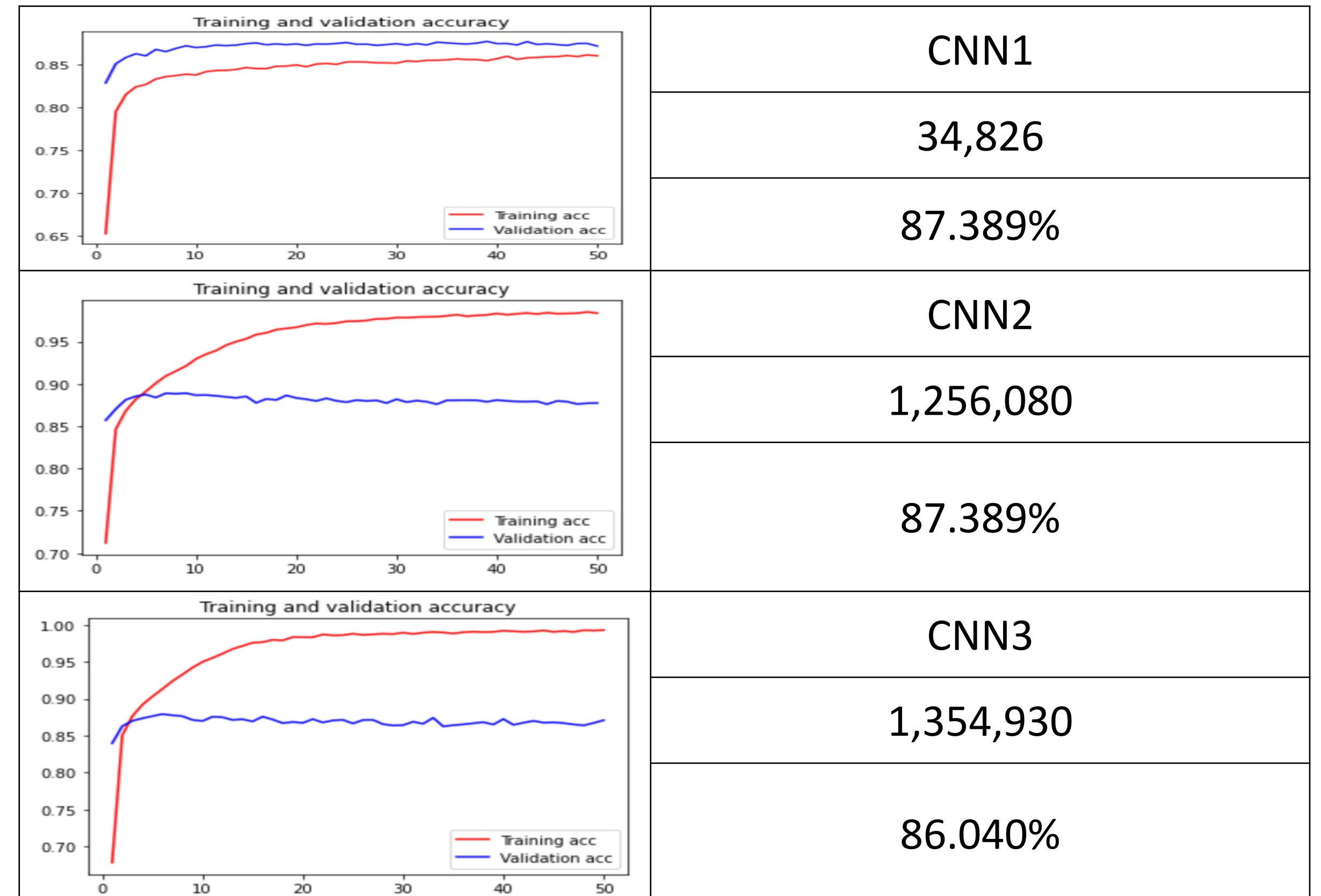


- ✓ 첫 번째와 두 번째 차원은 이미지 내 숫자의 초기 점의 좌표를 시작으로 x값과 y값의 변화 정도를 단위 벡터로 표현한 값이다. 세 번째 차원은 eos(end of stroke)로, 펜이 종이에 떨어지는 순간을 1, 나머지 경우는 전부 0으로 표현하였다. 네 번째 차원은 eod(end of digits)로, 시퀀스의 마지막 지점을 1로 표현하였다.
- ✓ 시퀀스 MNIST 데이터셋 내 1차원과 2차원 값(x값과 y값 변화 정도)에 균일분포 $U(-1.5,1.5)$ 를 따르는 노이즈를 더하였다.

4. 실험 및 결과

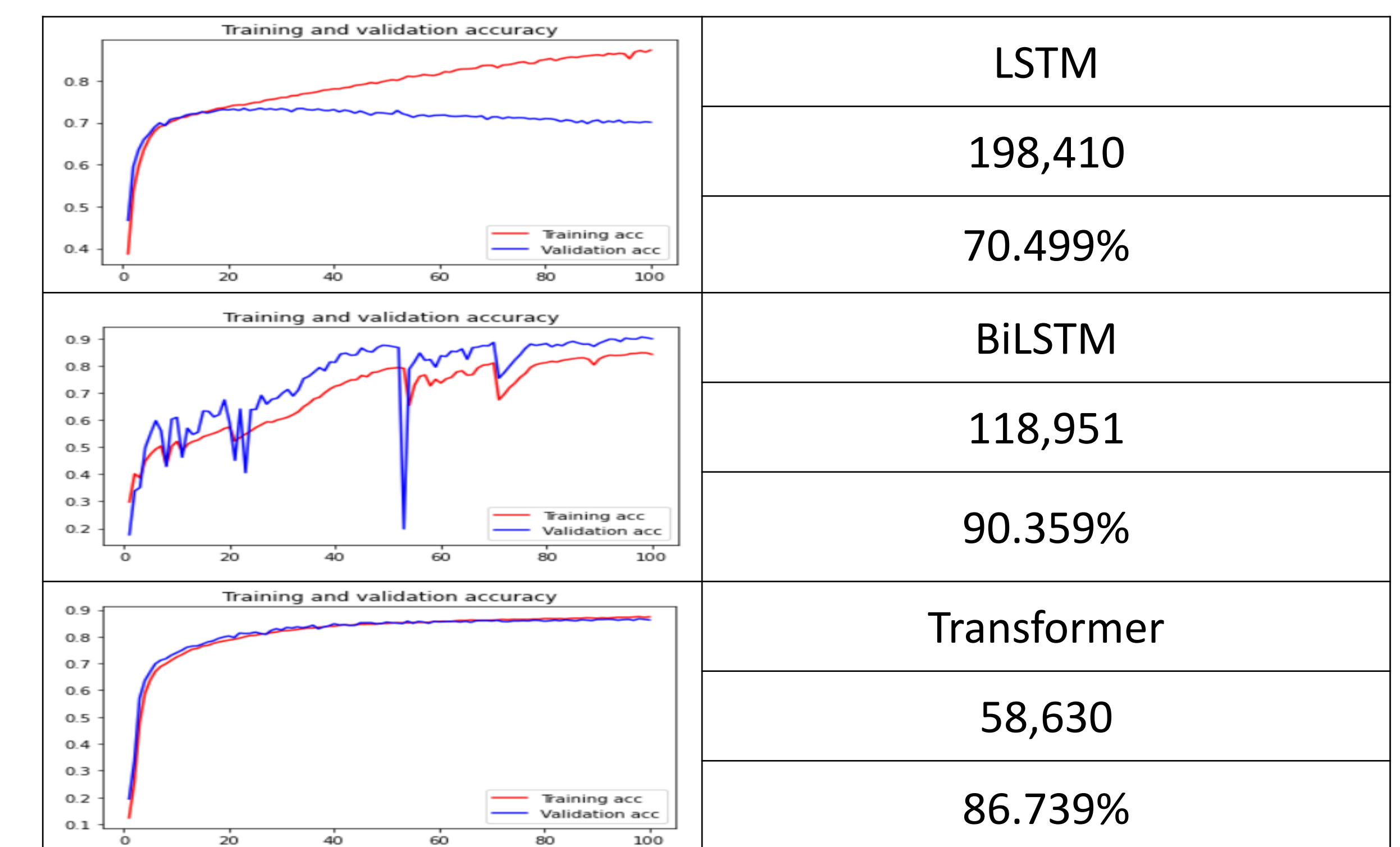
1) CNN 계열 모델 실험

50 epoch



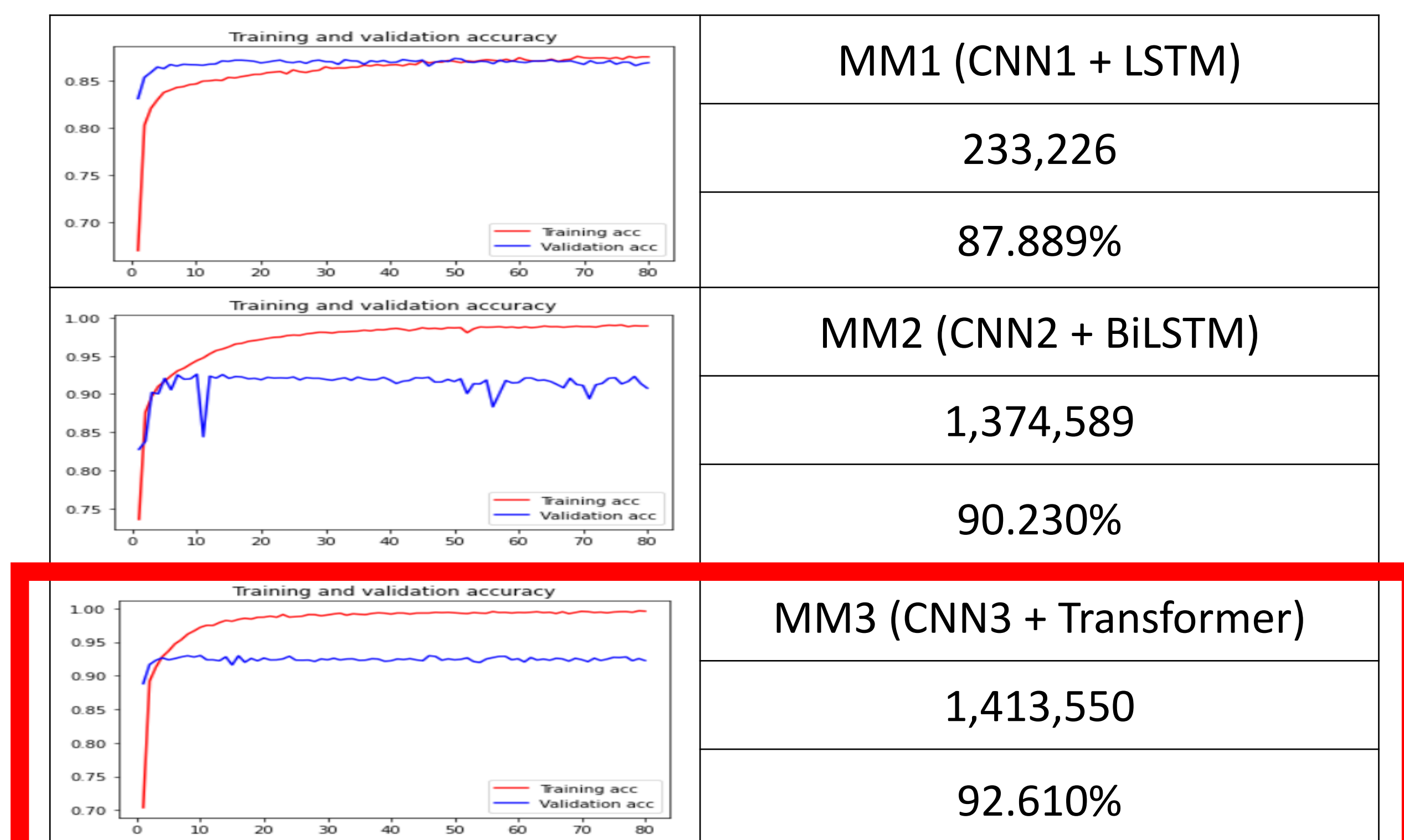
2) RNN 계열 모델 실험

100 epoch



3) 멀티모달 계열 모델 실험

80 epoch



4. 결론 및 기대효과

- ✓ 노이즈 데이터 셋을 학습시키기 위해서는 기존의 단일모달로는 좋은 성능을 내기 힘들며, 그나마 좋은 성과를 낸 BiLSTM도 학습 과정에서 불안정한 모습을 보였다.
- ✓ 반면, 멀티모달은 단일모달과 비교했을 때 전반적으로 좋은 성능을 보이며 특히 CNN3 모델과 트랜스포머 모델을 결합한 MM3 모델이 가장 높은 정확도를 보였다.
- ✓ 이를 통해 멀티모달이 견고한 능력을 가져 모델의 일반화에 우세함을 보였으며, 이러한 특징은 정제되지 않은 데이터로 학습을 진행할 때 더욱 효과를 보임을 알 수 있다.