

석사학위논문

다중 오토인코더를 통한 치매음성데이터  
이상신호 감지

Detection of abnormal signals of dementia voice data  
through multiple autoencoders

김 영 진

한양대학교 공학대학원

2023 년 8 월

석사학위논문

다중 오토인코더를 통한 치매음성데이터  
이상신호 감지

Detection of abnormal signals of dementia voice data  
through multiple autoencoders

지도교수 조 인 휘

이 논문을 공학 석사학위논문으로 제출합니다.

2023 년 8 월

한양대학교 공학대학원

컴퓨터 공학 전공

김 영 진

이 논문을 김 영 진의 석사학위논문으로 인준함

2023 년 8 월

심사위원장 이 병 호 (인)

심사위원 조 인 휘 (인)

심사위원 손 규 식 (인)

한양대학교 공학대학원

## 국문요지

본 연구에서는 효과적인 이상 신호 감지를 위해 다중 오토 인코더 모델을 제안한다. 이 모델을 기존의 오토 인코더와 컨볼루션 오토인코더 모델과 비교하였다. 이를 통해, 다중 오토인코더 모델이 기존 모델들에 비해 뛰어난 성능을 보임을 입증하였다.

제안된 다중 오토인코더는 이상신호 감지 오토인코더와 노이즈 제거 오토인코더를 결합하여 구성하였다. 이 두가지 기능은 복잡하고 노이즈가 많은 음성 데이터에서 이상신호를 정확하게 감지하는데 중요한 역할을 하였다.

학습이 진행된 후에는, 재구성 오류(Reconstruction Error) 분포를 분석하여 모델의 학습 상태를 확인하였다. 이를 통해 모델이 데이터를 어떻게 인식하고, 이상치와 정상치를 어떻게 구분하고 있는지를 분석하였다.

그 다음 단계에서는 여러 Threshold 값들을 실험하여, 모델의 성능을 최적화하는 Threshold 값을 찾았다. 최적의 Threshold 값은 모델의 성능 예측에 중요한 역할을 하였다. 이 값을 토대로 Confusion Maxtrix를 사용하여 모델의 성능을 시각화하고 성능평가를 진행하였다.

이런 과정들을 통해, 제안된 다중 오토인코더 모델은 기존 오토인코더에 비해 약 4% 이상의 성능향상을 보였다. 또한 컨볼루션 오토인코더와 비교하였을때도 미세하게 높은 성능을 보였다.

## 차 례

국문요지.....	i
차 례.....	ii
그림 차례.....	iv
표 차례.....	v
제 1 장 서 론.....	1
1.1. 연구 배경 및 필요성.....	1
1.2. 연구 내용.....	4
제 2 장 관련연구.....	8
2.1. 오토인코더 소개.....	8
2.1.1. 다중 오토인코더.....	9
2.1.2. 컨볼루션 오토인코더.....	10
2.2. 이상치 탐지 소개.....	12
2.3. 음향과 음성의 차이.....	15
2.4. 음성 데이터 전처리 음향 데이터 전처리 차이.....	17
2.5. 치매 음성 분석 및 이상신호 감지의 기존연구.....	20
제 3 장 설계 및 구현.....	22
3.1. 모델 설계 및 구축.....	22
3.1.1. 모델 설계.....	22
3.1.2. 모델 구축.....	24
3.2. 성능 예측 및 성능평가.....	27
3.2.1. 재구성 분포 (Reconstruction Error).....	27
3.2.2. 최적의 Threshold 범위 탐색.....	29

3.2.3. 최적의 Threshold.....	32
3.2.4. Confusion Matrix.....	33
제 4 장 결론 및 향후 연구.....	36
4.1. 결론.....	36
4.2. 향후 연구.....	38
참고문헌.....	40
ABSTRACT.....	41
감사의 글.....	43

## 그림 차례

그림 1.1 2015년에서 2050년까지 예상되는 치매 환자수 .....	1
그림 2.1 오토인코더 그림.....	8
그림 2.2 이상치탐지 그림.....	12
그림 2.4 음성데이터 스펙트로그램.....	17
그림 3.1 다중 오토인코더 구조.....	22
그림 3.2 오토인코더 구축 코드.....	25
그림 3.3 다중 오토인코더 구축 코드.....	26
그림 3.4 컨볼루션 오토인코더 구축 코드.....	26
그림 3.5 오토인코더 모델 재구성 분포.....	28
그림 3.6 다중 오토인코더 모델 재구성 분포.....	28
그림 3.7 컨볼루션 오토인코더 모델 재구성 분포.....	29
그림 3.8 오토인코더 모델 최적의 Threshold 범위탐색.....	30
그림 3.9 다중 오토인코더 모델 최적의 Threshold 범위탐색.....	30
그림 3.10 컨볼루션 오토인코더 모델 최적의 Threshold 범위탐색.....	31
그림 3.11 오토인코더 모델 최적의 Threshold 그래프.....	32
그림 3.12 다중 오토인코더 모델 최적의 Threshold 그래프.....	33
그림 3.13 컨볼루션 오토인코더 모델 최적의 Threshold 그래프.....	33
그림 3.14 오토인코더 모델 Confusion Matrix.....	34
그림 3.15 다중 오토인코더 모델 Confusion Matrix.....	34
그림 3.16 컨볼루션 오토인코더 모델 Confusion Matrix.....	34
그림 4.1 3가지 오토인코더 모델 결과 그래프.....	37

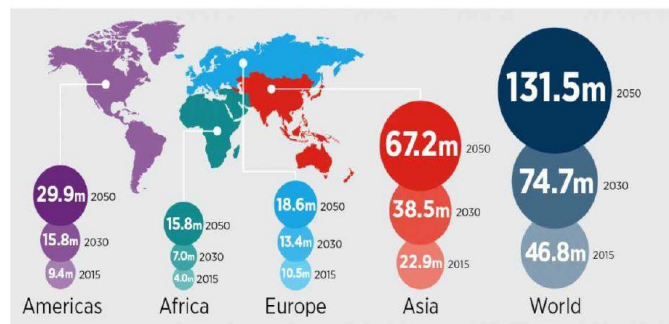
## 표 차례

표 1 3가지 오토인코더 모델 비교.....	36
--------------------------	----



# 제 1 장 서 론

## 1 . 1 . 연구 배경 및 필요성



(출처: World Alzheimer report 2015)

<2015년에서 2050년까지 예상되는 치매 환자 수>

그림 1.1 2015년에서 2050년까지 예상되는 치매 환자 수

치매는 기억력 저하, 인지 능력의 감소, 행동 및 성격 변화 등을 수반하는 신경퇴행성 질환으로, 전세계적으로 인구 고령화가 진행됨에 따라 발생률이 증가하고 있으며, 이는 개인, 가족, 사회에 큰 부담을 주고 있다.

세계 보건기구(WHO)에 따르면, 현재 전 세계에서 치매 환자의 수는 약 5000만 명이며, 이 수치는 매년 약 1000만 명씩 증가하고 있다. 또한, 인구 고령화 추세에 따라 이런 상황은 앞으로 더욱 심각해질 것으로 예상된다. 이에 따라 치매에 대한 연구 및 치료 개발은 전 세계적으로 절실한 과제로 떠오르고 있다.

그러나 치매는 복합적인 원인으로 인해 발생하며, 아직까지 정확한 치료 방

법이 개발되지 않았다. 따라서, 현재로서는 질병의 진행을 늦추거나 증상을 완화시키는 것이 최선의 대응 방안으로 여겨지고 있다. 이를 위해서는 치매의 조기 진단이 중요하게 여겨지는데, 이는 치료의 효과를 극대화하고 환자의 삶의 질을 개선하는 데 기여하기 때문이다.

그러나 현재의 치매 진단 방법은 여러가지 한계를 가지고 있다.

먼저 첫째로 뇌 스캔이나 인지능력 테스트와 같은 방법들은 고비용이 소요되며, 전문적인 의료 인력을 필요로 한다. MRI나 CT 등의 뇌 스캔은 매우 정밀한 정보를 제공할 수 있지만, 이러한 장비의 구매 및 유지 비용은 매우 높다. 또한, 이러한 검사 결과를 정확하게 해석하고 진단을 내리기 위해서는 전문적인 의료 지식과 경험이 필요하다.

둘째로 현재의 치료 방법은 치매의 초기 증상을 감지하는데 한계가 있다. 치매는 초기에는 미묘한 증상으로 나타나기 때문에, 환자나 가족들이 이를 인지하기 어렵고, 이로 인해 초기에 진단받는 것이 어렵다. 또한, 뇌 스캔이나 인지능력 테스트는 주로 중증이거나 발병 후 일정 시간이 지난 환자에게서만 명확한 결과를 보여주므로, 조기 진단에는 제한적이다.

셋째로, 현재의 진단 방법은 환자에게 상당한 스트레스를 주게 된다. 뇌 스캔은 물리적 불편함을 줄 수 있으며, 인지능력 테스트는 환자에게 정신적 압박을 줄 수 있다. 이로 인해 환자의 건강 상태가 더 악화될 수 있으며, 진단 과정 자체가 환자의 삶의 질을 저하시킬 수 있다.

이러한 배경에서, 본 연구는 다중 오토 인코더를 사용하여 치매 환자의 음성 데이터에서 이상 신호를 감지하는 새로운 접근 방식을 제안한다. 이 방법은 기존의 치매 진단 방법과 비교하여 다음과 같은 이점을 가진다.

첫째로, 이 방법은 비교적 저렴하고 쉽게 수집할 수 있는 음성 데이터를 활용한다. 이는 대규모의 데이터 수집 및 분석을 가능하게 하며, 이로 인해 더 정확하고 빠른 진단이 가능해진다. 또한, 이 방법은 전문적인 의료 지식이나 특별한 장비 없이도 실행될 수 있으므로, 의료 인프라가 미비한 지역에서도 활용 가능하다.

둘째로, 이 방법은 뇌 스캔이나 인지능력 테스트와 달리 비침습적이다. 즉, 환자에게 물리적 불편함이나 정신적 스트레스 없이 데이터를 수집할 수 있다. 이는 환자의 삶의 질을 저하시키지 않으며, 진단 과정이 환자의 건강 상태에 부정적인 영향을 미치는 것을 방지할 수 있다.

셋째로, 이 방법은 음성 데이터의 복잡한 패턴을 학습하고 이상 신호를 감지하는 데에 뛰어난 성능을 보인다. 최근의 연구들은 치매 환자들이 특정한 음성 패턴을 보인다는 것을 밝혀냈다. 이러한 패턴은 치매의 초기 증상을 포함할 수 있으므로, 이를 감지하는 것은 조기 진단에 매우 중요하다. 본 연구에서 제안하는 다중 오토 인코더는 이러한 패턴을 효과적으로 학습하고 감지하는 데 사용될 수 있다.

마지막으로, 이 방법은 기계 학습의 발전을 통해 더욱 정교해질 수 있다. 다중 오토 인코더는 비지도 학습 알고리즘의 한 종류로, 입력 데이터의 패

턴을 스스로 학습하며, 이를 바탕으로 이상 신호를 감지한다. 이러한 알고리즘은 데이터의 양이 많을수록 더 높은 성능을 보이므로, 음성 데이터의 수집과 분석이 확대됨에 따라 그 성능은 계속해서 향상될 것이다.

따라서, 본 연구는 현재의 치매 진단 방법의 한계를 극복하고, 환자와 의료 서비스 제공자 모두에게 이익을 가져다 줄 수 있는 새로운 방법을 제시한다. 또한, 이 방법은 치매 진단 뿐만 아니라 다양한 의료 분야에서도 활용될 수 있으므로, 의료 분야 전반의 진단 및 감시 방법을 혁신할 수 있는 가능성을 가지고 있다. 이러한 점에서 본 연구는 치매 진단에 대한 새로운 관점을 제시하며, 기계 학습과 의료의 교차 분야에서의 혁신적인 발전을 이끌어 낼 것이다.

## 1.2. 연구 내용

본 연구는 치매 환자의 음성 데이터를 수집하여 다중 오토인코더를 훈련하는 방식으로 진행된다. 치매는 환자의 언어능력과 음성 패턴에 변화를 초래하며, 이러한 변화는 음성 데이터의 이상신호로서 표현된다. 이러한 이상 신호를 효과적으로 감지하고 분석함으로써, 우리는 치매 환자의 조기 진단과 치료에 기여할 수 있다.

우리의 연구 접근법은 크게 세 가지 단계로 이루어진다.

첫째, 치매환자의 음성 데이터를 수집하고 전처리하는 과정이다. 이과정에서 우리는 음성 데이터의 높낮이, 강도, 음성의 일관성 등의

특징을 추출하고, 이를 다중 오토인코더의 입력으로 사용한다.

이러한 특징은 치매 환자의 언어능력 저하와 음성 변화를 반영하는 중요한 지표로 작용한다.

둘째, 다중 오토인코더를 통해 이상 신호를 학습하고 감지하는 과정이다. 다중 오토인코더는 비지도 학습 방법을 사용하여 데이터의 복잡한 패턴과 구조를 학습하고 재현한다. 이 과정에서, 다중 오토인코더는 음성 데이터의 이상 신호를 학습하고, 이를 통해 새로운 음성 데이터에서 이상 신호를 감지할 수 있게 된다. 이러한 방식을 통해, 우리는 치매 환자의 음성에서 나타나는 미묘한 변화를 감지하고 분석할 수 있다.

셋째, 성능예측과 성능평가 과정이다. 이 과정에서 우리는 다중 오토인코더의 성능을 측정하고, 이를 기반으로 최적의 모델을 선정한다. 이를 위해, 우리는 '재구성 오차(Reconstruction Error)'와 '최적의 Threshold', 'Confusion Matrix' 라는 세 가지 주요 평가 지표를 사용한다.

재구성 오차(Reconstruction Error)는 모델이 입력 데이터를 얼마나 잘 재현하는지를 측정하는 지표이다. 이는 원래의 입력 데이터와 모델이 생성한 출력 데이터 사이의 차이를 측정하여 계산된다. 이 오차가 크면 클수록 모델이 데이터를 정확하게 재현하지 못한 것이며, 이는 이상 신호가 존재할 가능성이 높다는 것을 나타낸다. 따라서, 이 오차를 최소화하는 것이 우리의 주요 목표 중 하나이다.

다음으로, 최적의 Threshold는 재구성 오차를 기반으로 이상 신호를

감지하는데 사용되는 기준 값이다. 재구성 오차가 이 Threshold 값보다 크면, 우리는 그 데이터를 이상 신호로 분류한다. 이 Threshold 값은 여러 실험을 통해 탐색하며, 이를 통해 우리는 가장 높은 이상 신호 감지 정확도를 가진 모델을 선정한다.

최적의 Threshold를 설정한 후에는 Confusion Matrix를 사용하여 성능 평가를 진행한다. Confusion Matrix는 실제 값과 모델이 예측한 값이 얼마나 일치하는지를 시각적으로 표현하는 행렬이다. 이 행렬은 True Positive, False Positive, True Negative, False Negative의 네 가지 분류로 이루어져 있다.

True Positive는 이상 신호를 정확하게 감지한 경우, True Negative는 정상 신호를 정확하게 감지한 경우를 나타낸다. 반면, False Positive는 정상 신호를 이상 신호로 잘못 분류한 경우, False Negative는 이상 신호를 정상 신호로 잘못 분류한 경우를 나타낸다.

Confusion Matrix를 통해, 우리는 모델의 정확도, 정밀도, 재현율, F1 스코어 등 다양한 성능 지표를 계산할 수 있다. 이를 통해 우리는 모델이 이상 신호를 얼마나 잘 감지하는지, 그리고 어떤 유형의 오류를 주로 범하는지에 대한 깊은 이해를 얻을 수 있다.

이 세 가지 평가 지표를 통해, 우리는 다중 오토인코더의 성능을 깊이 있게 평가하고, 이를 바탕으로 모델의 성능을 최적화한다. 또한, 이를 통해 우리는 모델이 실제 음성 데이터에서 이상 신호를 얼마나 잘 감지하는지를

정량적으로 이해할 수 있다.

이러한 과정을 통해, 우리는 다중 오토인코더를 이용한 이상 신호 감지 방법의 성능을 신뢰성 있게 평가하고, 이를 바탕으로 모델의 성능을 개선하는 데 필요한 정보를 얻을 수 있다. 이를 통해, 본 연구는 치매 환자의 음성 데이터 분석 및 조기 진단에 중요한 도구로서의 가능성을 보여준다. 이를 통해, 본 연구는 치매 환자의 조기 진단과 치료에 기여하며, 이에 따라 그들의 삶의 질을 향상시키는 데 중요한 역할을 할 것이다.

## 제 2 장 관련연구

### 2.1. 오토인코더 소개

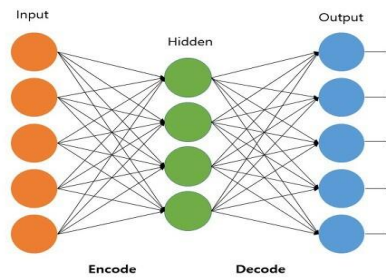


그림 2.1 오토인코더 그림

오토 인코더는 비지도 학습 방법의 하나로, 딥러닝 및 신경망의 구조 중 하나이다. 입력데이터의 유용한 특성을 학습하는 데 사용되는 이 알고리즘은 입력을 중요한 특징을 포착하는 낮은 차원의 표현으로 압축한다. 이를 인코더라 부른다. 그런 다음 이 표현을 사용하여 원래의 입력을 재구성하는 과정을 거친다. 이를 디코더라 부른다.

오토 인코더는 크게 세 부분으로 구성된다. 첫 번째 부분은 '인코더'이다. 인코더는 입력 데이터를 받아들여 더 낮은 차원의 '잠재 공간'이라는 새로운 표현으로 변환한다. 이 잠재 공간은 원래의 입력 데이터에서 가장 중요한 정보를 캡처하여 표현한다. 두 번째 부분은 '잠재 공간' 자체이며, 이 공간에서 데이터는 압축된 형태로 표현된다. 세 번째 부분은 '디코더'이다. 디코더는 잠재 공간의 표현을 다시 원래의 입력 데이터와 같은 차원으로 재구성한다.



훈련 과정에서, 오토 인코더는 원본 입력과 재구성된 입력 사이의 차이, 즉 재구성 손실을 최소화하려고 한다. 이를 통해, 오토 인코더는 입력 데이터의 중요한 특성을 학습하게 된다. 이러한 특성은 다양한 응용 분야에서 사용될 수 있다. 예를 들어, 차원 축소, 노이즈 제거, 그리고 이상치 탐지 등에 사용된다.

이상치 탐지에서, 오토 인코더는 일반적으로 '정상' 데이터만으로 훈련된다. 그런 다음 훈련된 모델을 사용하여 새로운 데이터를 인코딩하고 디코딩한다. 만약 새로운 데이터가 '정상' 데이터와 유사하다면, 오토 인코더는 이를 잘 재구성할 수 있을 것이다. 그러나 만약 새로운 데이터가 이상치라면, 오토 인코더는 이를 잘 재구성하지 못하며, 이를 통해 이상치를 탐지할 수 있게 된다.

#### 2.1.1. 다중 오토인코더

다중 오토 인코더는 기본적인 오토 인코더 구조를 확장한 모델로, 더 많은 은닉층을 통해 데이터를 인코딩하고 디코딩하는 구조를 가지고 있다. 이러한 은닉층은 복잡한 데이터 구조를 더 잘 모델링하는 데 도움이 될 수 있다.

본 연구에서는 이상 신호 감지 오토 인코더와 노이즈 제거 오토 인코더, 두 가지 형태의 다중 오토 인코더를 사용하였다. 이상 신호 감지 오토 인코더는 '정상' 음성 데이터를 잘 재구성하도록 학습되며, 이상한 패턴을 가진 데이터(즉, 이상 신호)는 잘 재구성하지 못한다.

따라서, 이 모델은 재구성 오류를 통해 이상 신호를 감지할 수 있다.

한편, 노이즈 제거 오토 인코더는 잡음이 섞인 음성 데이터를 원래의 '정상' 데이터로 재구성하도록 학습된다. 이 모델은 데이터의 핵심 특성을 학습하도록 강제되며, 이는 복잡하고 노이즈가 많은 음성 데이터에서 이상 신호를 감지하는 데 도움이 될 수 있다.

이전의 연구에서는 다중 오토 인코더를 이용한 이상치 탐지가 산업 분야에서 음향 데이터를 처리하는 데 효과적이었다. 그러나 본 연구는 이러한 접근 방식을 음성 데이터에 적용하였다는 점에서 차별화된다. 음향과 음성은 물리적 특성과 사용되는 컨텍스트가 다르기 때문에, 이 두 분야에서 이상치 탐지는 별도의 고려 사항을 필요로 한다. 본 연구는 이러한 고려 사항을 탐색하고, 이상치 탐지 기술을 음성 데이터에 적용하는 새로운 방법을 제시한다.

#### 2.1.2. 컨볼루션 오토인코더

컨볼루션 오토인코더(Convolutional Autoencoder, CAE)는 오토인코더의 한 종류로, 특히 이미지 데이터와 같은 고차원 데이터를 처리하는 데에 특화되어 있다. 본 연구에서는 이상 신호 감지 오토인코더와 노이즈 제거 오토인코더를 결합한 다중 오토인코더와 비교 대상으로 컨볼루션 오토인코더를 선택하였는데, 이는 컨볼루션 오토인코더가 고차원 데이터를 처리하는 데 강점을 가지고 있어, 복잡한 오디오 데이터를 처리하는 데 적합하다고 판단되었기 때문이다.

컨볼루션 오토인코더는 전통적인 오토인코더와 마찬가지로 데이터의 고유한 특성을 학습하고 재구성하는 데 목적을 두고 있지만, 컨볼루션 신경망 (Convolutional Neural Network, CNN)의 개념을 도입하여 이를 실현한다. CNN은 이미지 처리에서 특히 강점을 보이는 신경망 구조로, 이를 통해 컨볼루션 오토인코더는 고차원 데이터의 지역적 특성과 패턴을 더 효과적으로 학습하고 캡처할 수 있다.

오토인코더와 마찬가지로, 컨볼루션 오토인코더도 인코더와 디코더 두 부분으로 구성된다. 인코더는 입력 데이터를 저차원 특성 공간으로 압축하고, 디코더는 이 저차원 특성을 다시 원래의 고차원 공간으로 재구성한다. 하지만, 컨볼루션 오토인코더의 인코더와 디코더는 컨볼루션 계층과 역컨볼루션 (또는 전치 컨볼루션) 계층을 사용하여 이 과정을 수행한다. 이를 통해 컨볼루션 오토인코더는 고차원 데이터의 지역적인 패턴을 더 효과적으로 학습하고 재구성할 수 있다.

오디오 데이터는 시간적 차원과 주파수 차원을 가지는 고차원 데이터이다. 이러한 특성 때문에 오디오 데이터는 이미지 데이터와 비슷한 방식으로 처리될 수 있다. 특히, 스펙트로그램과 같은 시간-주파수 변환을 통해 오디오 데이터는 2D 이미지 데이터와 매우 유사한 형태로 표현될 수 있다. 이는 컨볼루션 오토인코더가 오디오 데이터를 처리하는 데에 효과적인 방법임을 시사한다.

더불어, 컨볼루션 오토인코더는 지역적인 특성을 잘 캡처할 수 있다. 이는 오디오 데이터에서 매우 중요한 점으로, 예를 들어 사람의 음성에서는 일정 시간 동안의 패턴이 중요한 의미를 가질 수 있다. 또한, 노이즈 같은 이상

신호도 일반적으로 지역적인 특성을 가진다. 따라서, 컨볼루션 오토인코더는 이러한 지역적인 특성을 효과적으로 학습하고, 이상 신호를 감지하는 데에 유용할 수 있다.

본 연구에서는 이러한 이유로 컨볼루션 오토인코더를 선택하여 이상 신호 감지와 노이즈 제거 오토인코더로 구성된 다중 오토인코더와의 비교를 진행하였다. 이를 통해, 본 연구의 다중 오토인코더가 고차원 오디오 데이터를 처리하는 데 얼마나 효과적인지를 정량적으로 평가할 수 있었다.

## 2.2. 이상치 탐지 소개

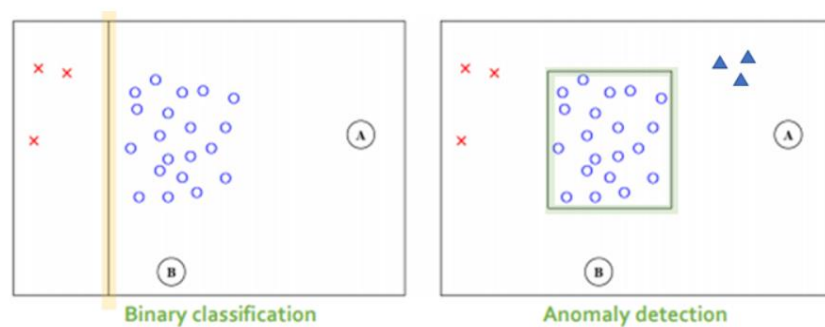


그림 2.2 이상치탐지 그림

이상치 탐지는 데이터 과학과 머신러닝의 중요한 분야로, 데이터의 패턴에서 통계적으로 또는 구조적으로 이상한 관찰 값을 찾아내는 과정이다. 이상치는 일반적으로 '정상' 데이터의 패턴에서 벗어난, 예상치 못한 데이터 포인트를 가리킨다. 이런 데이터 포인트는 종종 중요한 정보를 가지고 있으며, 예를 들어 부정 행위, 사기, 기계 고장, 또는 본 연구와 관련하여, 치매와 같은 질병의 조기 신호를 포착하는 데 사용될 수 있다.

이상치 탐지의 기본적인 전제는 '정상' 데이터와 '이상' 데이터가 서로 다른 패턴을 갖는다는 것이다. 이상치 탐지 알고리즘의 목표는 이러한 패턴을 학습하고, 새로운 데이터가 '정상'인지 '이상'인지를 판별하는 것이다. 이상치 탐지 알고리즘은 일반적으로 두 가지 방식으로 작동한다. 두가지 방식은 바로 지도 학습 방식과 비지도 학습 방식이다.

지도 학습 방식에서는 알고리즘이 '정상' 데이터와 '이상' 데이터 모두의 예시를 받는다. 알고리즘은 이러한 예시를 사용하여 '정상'과 '이상'을 구분하는 패턴을 학습한다. 그러나 이 방식의 단점은 '이상' 데이터의 예시를 얻는 것이 종종 어렵거나 불가능하다는 것이다. 이 경우에는 비지도 학습 방식이 유용하다.

비지도 학습 방식에서는 알고리즘이 '정상' 데이터만으로 훈련된다. 알고리즘은 '정상' 데이터의 패턴을 학습하고, 이 패턴과 크게 다른 새로운 데이터를 '이상'으로 간주한다. 이 방식은 '이상' 데이터의 예시를 얻을 필요가 없으므로, 많은 실제 상황에서 더 실용적이다.

오토 인코더는 이상치 탐지에서 비지도 학습 방식을 구현하는 데 특히 유용하다. 오토 인코더는 '정상' 데이터를 인코딩하고 디코딩하는 방법을 학습한다. 그런 다음, 이를 이용하여 새로운 데이터를 인코딩하고 디코딩한다. 새로운 데이터가 '정상' 패턴과 유사하다면, 오토 인코더는 이를 잘 재구성할 것이다.

그러나 새로운 데이터가 '정상' 패턴과 크게 다르다면, 오토인코더는 이를

잘 재구성하지 못하며, 이를 통해 '이상' 데이터를 감지한다. 이렇게 재구성 오류가 큰 데이터 포인트는 이상치로 간주할 수 있다.

이상치 탐지는 신경망, 클러스터링, 통계적 방법 등 다양한 방법을 사용하여 수행될 수 있다. 그러나 각 방법은 특정 유형의 데이터와 문제에 가장 적합하며, 각각의 장단점을 가지고 있다. 따라서 적절한 이상치 탐지 방법을 선택하는 것은 문제의 성격, 사용 가능한 데이터, 그리고 특정 문제에 대한 도메인 지식에 크게 의존한다.

본 연구에서는, 다중 오토인코더를 사용하여 음성 데이터에서 이상치를 탐지하는 방법을 탐구한다. 다중 오토인코더는 데이터의 복잡한 구조를 모델링하고, 이를 이용하여 이상치를 정밀하게 탐지할 수 있다. 또한, 오토인코더는 비지도 학습 방식을 사용하므로, '이상' 음성 데이터의 예시를 얻을 필요가 없다. 이는 본 연구와 같이, 이상치의 예시를 얻기 어려운 문제에서 매우 유용하다.

음성 데이터는 독특한 특성을 가지고 있으며, 이는 이상치 탐지에 중요한 역할을 한다. 예를 들어, 음성의 톤, 강도, 빠르기 등은 사람의 건강 상태를 반영할 수 있다. 이러한 특성을 캡처하고 모델링하는 것은 이상치 탐지에서 중요한 도전 과제이며, 이는 본 연구의 주요 관심사 중 하나이다.

### 2.3. 음향과 음성의 차이

본 연구의 주요 관심사는 음성 데이터이며, 이는 기존에 음향 데이터에 초점을 맞춘 다른 연구와는 차별화된 점이다. 이 두 분야는 서로 연관되어 있지만, 음향과 음성은 명확한 차이점이 있으며, 이들의 차이는 이상치 탐지에 큰 영향을 미친다. 이 섹션에서는 음향과 음성의 주요 차이점을 탐구하고, 이러한 차이가 이상치 탐지에 어떻게 영향을 미치는지를 설명할 것이다.

음향학은 소리의 생성, 전파, 수용에 관한 학문이며, 이는 소리의 물리적 측면을 중심으로 한다. 이에 반해, 음성은 사람의 언어 의사소통을 위한 소리에 관련된 현상을 의미하며, 이는 물리적, 생물학적, 심리적, 문화적 측면을 모두 포괄한다.

이런 차이점은 음향과 음성 데이터를 처리하는 방법과 이들에게서 이상치를 탐지하는 방법에 영향을 미친다. 예를 들어, 음향 데이터는 일반적으로 물리적 속성, 예를 들어 주파수, 진폭, 위상 등에 중점을 두며, 이상치 탐지는 이러한 물리적 속성에서 이탈하는 것에 주로 초점을 맞춘다. 이는 예를 들어, 기계 장비의 고장이나 구조물의 결함 같은 산업 응용 분야에서 특히 중요하다.

그러나 음성 데이터는 이보다 훨씬 복잡하다. 음성은 언어, 감정, 개인적 특징 등과 같은 다양한 정보를 포함하며, 이러한 정보는 음향 데이터의 물리적 속성만으로는 완전히 이해할 수 없다. 이상치 탐지는 이러한 복잡한

정보를 모델링하고 감지하는 데 집중해야 한다. 예를 들어, 본 연구에서는 치매를 조기에 감지하기 위해 이상한 패턴을 찾는 것에 중점을 두고 있다. 이는 음성의 톤, 강도, 빠르기 등과 같은 속성이 변화하는 것을 감지하는 것을 포함할 수 있다.

따라서, 음성 데이터에서 이상치를 탐지하려면, 기존 음향 데이터에 사용되는 방법론을 그대로 적용하는 것이 아니라, 음성 데이터의 고유한 특성과 복잡성을 고려한 새로운 접근 방식이 필요하다. 이는 데이터의 전처리, 특성 추출, 모델링, 그리고 평가 방법에 영향을 미친다. 예를 들어, 음성 데이터에서는 음향 특성뿐만 아니라 발화자의 성별, 나이, 발음 패턴 등 추가적인 정보를 고려할 수 있어야 한다.

또한, 음성 데이터는 일반적으로 시간에 따라 변화하는 동적 특성을 가지고 있다. 이러한 동적 특성은 스피치의 의미와 감정 표현에 중요한 역할을 하며, 이상치 탐지에도 중요한 정보를 제공한다. 이와 반대로, 대부분의 음향 데이터는 시간에 따른 동적 변화보다는 단일 시점에서의 속성에 더 많은 관심을 가진다.



## 2.4. 음성 데이터 전처리 음향 데이터 전처리 차이

음성과 음향 데이터 전처리는 오디오 데이터를 분석하고 처리하는 과정이지만, 그들의 목표와 접근 방식에는 중요한 차이점이 있다.

이 차이를 이해하는 것은 우리가 올바른 전처리 기법을 선택하고 적용하여 데이터를 효과적으로 분석할 수 있도록 돕는다.

음성 데이터 전처리는 사람의 음성을 분석하는 것을 목표로 한다. 이는 주로 음성 인식, 스피커 식별, 감정 분석 등의 작업을 수행하는 데 사용된다. 이런 음성 데이터는 통상적으로 웨이브(WAV) 파일 형식으로 저장되며, 이 형식은 신호의 모든 샘플을 디지털 형식으로 직접 저장한다.

전처리 과정에서는 일반적으로 먼저 음성 데이터를 스펙트로그램으로 변환한다. 스펙트로그램은 시간에 따른 주파수 분포를 시각적으로 표현하므로, 음성의 주요 특징을 명확하게 볼 수 있다. 이렇게 변환된 데이터는 머신러닝 모델에 직접 입력으로 제공되거나, 추가적인 특징 추출 단계를 거쳐 처리될 수 있다.

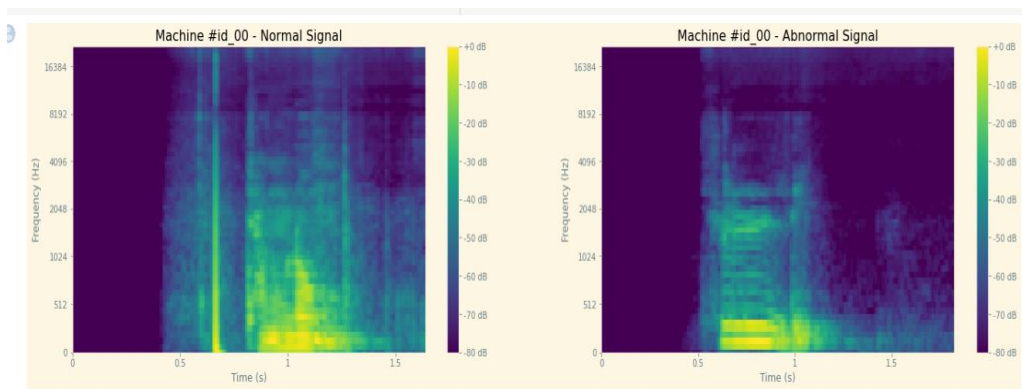


그림 2.4 음성 데이터 스펙트로그램

한편, 음향 데이터 전처리는 음향 환경을 분석하는 것을 목표로 한다. 이는 도시의 소음 분석, 야생 동물의 울음소리 분석, 음악 분류 등 다양한 작업에 적용될 수 있다. 따라서 음향 데이터 전처리는 다양한 유형의 소리를 인식하고 분류하는 능력이 필요하다.

음성 데이터와 음향 데이터는 그들을 생성하는 원천인 '사람의 음성'과 '일반적인 소리'에서부터 차이가 나타난다. 이 차이는 전처리 방법과 분석 접근법에도 반영된다. 예를 들어, 음성 데이터는 일반적으로 스펙트럼 분석과 같은 고급 시그널 처리 기법을 통해 주요 특징을 추출하며, 이러한 특징은 머신러닝 모델이 음성을 인식하고 분석하는 데 사용된다.

반면, 음향 데이터 전처리는 소리의 특성과 환경을 파악하는 데 중점을 둔다. 음향 데이터는 주로 소음 수준, 주파수 내용, 시간 도메인에서의 에너지 분포 등을 분석하며, 이 정보를 바탕으로 소리의 출처를 파악하거나 음향 환경을 분류하는 데 사용한다.

음향 데이터는 대체로 더 복잡하며 다양한 소리가 섞여 있을 수 있기 때문에, 전처리 과정도 복잡하다. 일반적으로 푸리에 변환, 멜 스펙트럼 변환 등의 방법을 사용하여 소리의 주파수 성분을 분석하고, 이를 바탕으로 머신러닝 알고리즘에 적합한 피처를 추출한다.

이처럼 음성 데이터와 음향 데이터는 목표와 원천이 다르기 때문에, 그들을 처리하는 방법과 기술도 다르다. 음성 데이터는 음성의 특징을 중점적으로 분석하며, 음향 데이터는 소리의 다양한 특성을 포괄적으로 분석한다.

음성 데이터와 음향데이터의 주요 차이점은 '목표'와 '전처리 접근 방식'에 있다. 음성 데이터는 사람의 음성을 분석하고 이해하는 것에 초점을 맞추고, 음향 데이터는 소리의 전체적인 특성과 환경을 파악하는 것이 목표다. 따라서, 전처리 방법과 사용하는 기술, 그리고 데이터 분석에 대한 접근법 역시 상이하다.

최종적으로 이러한 차이점들은 우리가 데이터를 어떻게 이해하고, 어떤 문제에 적용하며, 어떤 결과를 예측하는지에 큰 영향을 미친다. 이 둘 사이의 차이를 이해하는 것은 오디오 데이터를 처리하고 분석하는데 있어 핵심적인 요소이며, 이를 바탕으로 우리는 올바른 데이터 전처리 전략을 선택하고 효과적인 머신러닝 모델을 구축할 수 있다.

## 2.5. 치매 음성 분석 및 이상신호 감지의 기존연구

치매 진단에 대한 필요성과 중요성은 많은 연구에서 강조되어 왔다. 특히 음성 분석을 통한 치매 진단은 이 분야에서 주목받는 방법 중 하나이다. 본 섹션에서는 치매 환자의 음성 데이터에서 이상 신호를 탐지하기 위해 사용된 기존 방법들과 그 결과에 대해 논의하고, 본 연구의 접근법과 비교해 볼 것이다.

기존의 많은 연구에서는 다양한 특성 추출 방법과 머신 러닝 기법을 이용하여 치매 음성 데이터의 분석을 시도하였다. 특히, MFCCs, spectral contrast, chroma features 등과 같은 특성을 이용하여 음성에서 정보를 추출하였다. 이러한 특성들은 음성의 주파수, 피치, 강도 등을 포착할 수 있어, 치매 환자의 음성에서 나타나는 변화를 감지하는데 유용하였다.

이러한 특성들을 기반으로, 다양한 머신 러닝 기법이 치매 진단에 적용되었다. 예를 들어, 서포트 벡터 머신(SVM), 의사결정 나무(Decision Trees), 랜덤 포레스트(Random Forests), 그래디언트 부스팅(Gradient Boosting) 등의 방법들이 사용되었다. 이러한 기법들은 특성을 바탕으로 치매 환자와 비 치매 환자를 분류하거나, 음성 데이터에서 이상 신호를 탐지하는 데 사용되었다.

그러나, 이러한 기존 방법들은 몇 가지 한계점을 가지고 있다. 첫째, 이러한 방법들은 주로 전통적인 머신 러닝 기법에 의존하며, 딥러닝 방법론의 적용을 상대적으로 덜 보여주고 있다. 둘째, 특성 추출 과정이

수동적이고, 특정 특성에 의존하는 경우가 많아, 일부 중요한 정보가 누락될 가능성이 있다.

본 연구에서는 이러한 문제점을 극복하기 위해 다중 오토인코더를 이용하였다. 다중 오토인코더는 비지도 학습법을 기반으로 하며, 입력 데이터의 중요한 특성을 자동으로 학습하고 추출할 수 있는 능력을 가지고 있다. 이는 치매 환자의 음성 데이터에서 중요한 정보를 포착하고 이상 신호를 감지하는 데 매우 유용하게 사용될 수 있다.

다중 오토인코더는 입력 데이터를 압축하고 재구성하는 과정을 통해 데이터의 핵심 특성을 학습한다. 이 과정에서 모델은 음성 데이터의 복잡한 패턴과 구조를 포착하며, 이를 바탕으로 이상 신호를 감지한다. 이는 전통적인 머신 러닝 기법이 다루기 어려운 고차원 및 비선형 패턴을 포착하는 데 특히 유용하다.

본 연구의 방법론은 기존 연구의 한계를 극복하고, 치매 음성 데이터 분석의 정확성과 효율성을 향상시키는 데 기여하였다. 이를 통해 치매 진단의 정확도를 향상시키고, 조기 진단의 가능성을 높이는 데 도움이 될 것으로 기대된다. 그러나, 이 방법론의 성능과 효과를 더욱 검증하기 위한 추가적인 연구가 필요하며, 이에 대한 논의는 후속 연구에서 이루어질 예정이다.

## 제 3 장 설계 및 구현

### 3.1. 모델 설계 및 구축

#### 3.1.1. 모델 설계

본 연구에서 제안하는 다중 오토 인코더 모델은 두 가지 주요 기능을 통합한 설계로, 이상 신호 감지와 노이즈 제거라는 두 가지 기능을 가지고 있다. 이들 기능은 각각 이상 신호 감지 오토 인코더와 노이즈 제거 오토 인코더로 구현되었다.

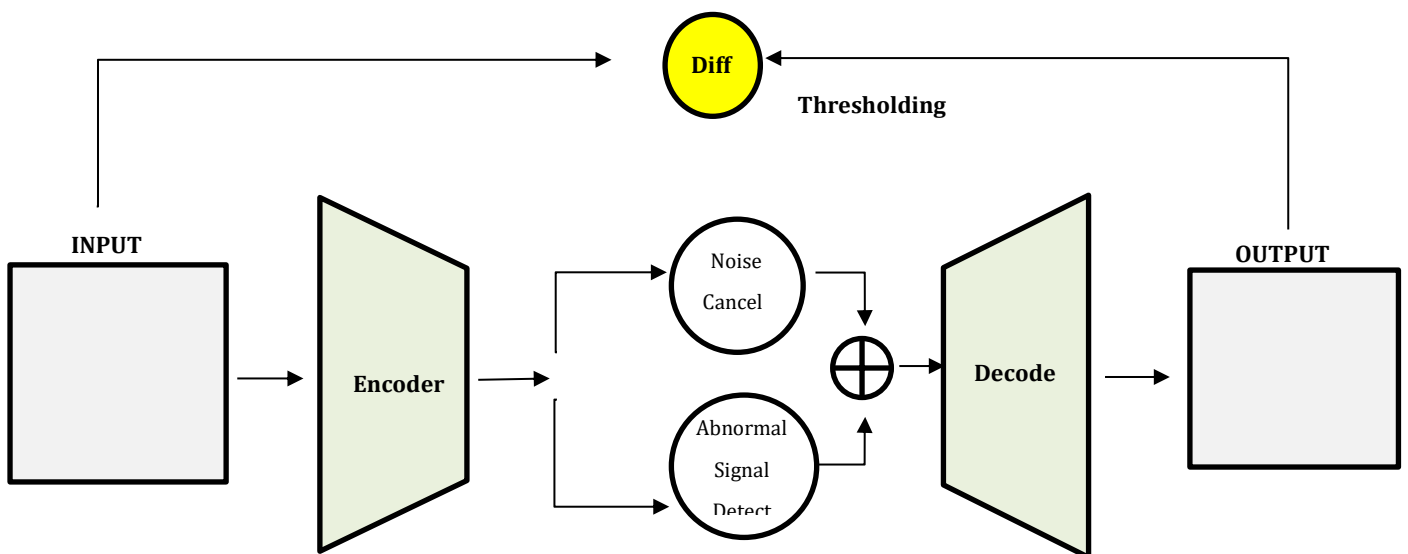


그림 3.1 다중오토인코더 구조

이상 신호 감지 오토 인코더는 '정상' 음성 데이터를 잘 재구성하는 것을 목표로 하였다. 학습 과정에서는 정상 데이터를 이용하여 오토 인코더를 학습시키고, 이를 통해 모델이 정상 패턴을 학습하게 되었다. 학습 후에는 이상한 패턴을 가진 데이터를 잘 재구성하지 못하는 특성을 이용하여 이상 신호를 감지하였다. 즉, 재구성 오류가 큰 데이터를 이상 데이터로 판단하였다.

노이즈 제거 오토 인코더는 잡음이 섞인 음성 데이터를 '정상' 데이터로 재구성하는 것을 목표로 하였다. 학습 과정에서는 잡음이 추가된 음성 데이터와 원래의 정상 데이터를 이용하여 학습하였다. 잡음이 추가된 데이터를 입력으로, 원래의 정상 데이터를 목표로 설정하여 오토 인코더를 학습시키면, 오토 인코더는 입력된 데이터에서 잡음을 제거하고 정상 데이터를 재구성하게 된다.

이 두 오토 인코더를 합치기 위해, 공유 인코더를 사용하는 아키텍처를 선택하였다. 이상 신호 감지 오토 인코더와 노이즈 제거 오토 인코더는 동일한 인코더를 공유하고, 각각 다른 디코더를 사용하여 각각의 목표에 맞게 재구성하였다. 이렇게 함으로써, 두 오토 인코더가 동일한 특성을 학습하고, 이를 다른 방식으로 활용할 수 있도록 하였다

이 구조는 두 가지 방식으로 이상 신호를 감지하는 능력을 강화하였다.

첫째, 이상 신호 감지 오토 인코더는 '정상' 음성 데이터를 잘 재구성하도록 학습되었다.

이상한 패턴을 가진 데이터는 잘 재구성하지 못하는 특징을 이용하여 재구

성 오류를 통해 이상 신호를 감지하였다.

둘째, 노이즈 제거 오토 인코더는 잡음이 섞인 음성 데이터를 원래의 '정상' 데이터로 재구성하도록 학습되었다. 이는 복잡하고 노이즈가 많은 음성 데이터에서 이상 신호를 감지하는 데 도움이 되었다.

모델의 학습은 주로 재구성 오류를 최소화하는 방향으로 진행되었다. 모델이 학습하는 동안, 우리는 재구성 오류가 어떻게 변화하는지 관찰하였고, 이를 통해 학습의 진행 상황을 모니터링하였다. 이런 방법으로, 모델이 정상적인 데이터 패턴을 학습하고 이상한 패턴을 감지하는 능력을 갖추게 되었다.

### 3.1.2. 모델 구축

본 연구의 핵심은 다중 오토인코더 모델의 구축이다. 이 모델은 기존의 오토인코더를 기반으로, 이상 신호 감지 오토인코더와 노이즈 제거 오토인코더를 결합하여 구성하였다.

다중 오토인코더의 구성은 먼저 이상 신호 감지 오토인코더를 만드는 것으로 시작되었다. 이 구성 요소는 기본적으로 오토인코더와 동일하게 작동하되, 특이치로 인해 '정상' 패턴과 다르게 나타나는 데이터를 재구성하는 데 어려움을 겪는 특성을 학습하도록 설계되었다. 이를 통해, 이상 신호 감지 오토인코더는 정상 데이터를 잘 재구성하고, 이상한 패턴을 가진 데이터는 잘 재구성하지 못하는 특성을 갖게 되었다.

또한, 노이즈 제거 오토인코더를 만들기 위해 추가적인 작업을 진행하였다.



이 오토인코더는 잡음이 섞인 음성 데이터를 '정상' 데이터로 재구성하는 학습 과정을 거쳤다. 이 과정에서 모델은 복잡하고 노이즈가 많은 데이터에 적응하여, 이러한 데이터에서도 '정상' 데이터를 잘 재구성하는 능력을 키웠다.

이 두 가지 컴포넌트를 결합하여 다중 오토인코더 모델을 완성하였다. 이를 통해 이상 신호와 노이즈를 더욱 효과적으로 감지할 수 있는 강력한 모델을 만들어냈다.

또한, 다중 오토인코더의 성능을 검증하기 위해 기존의 오토인코더 모델과 컨볼루션 오토인코더 모델을 동일한 실험 조건하에 구축하였다. 컨볼루션 오토인코더에서는 음성 데이터의 시간적인 특성을 캡처하기 위하여 1차원 컨볼루션 레이어, 즉 Conv1D를 사용하였다. (이미지 데이터의 경우 2차원 데이터 이므로 Conv2D를 사용한다)

이 세 가지 모델을 동일한 데이터셋에 적용하여 성능을 비교하였다. 비교를 위한 평가 지표로는 재구성 오류의 평균과 분산, 이상 신호 감지 정확도 등을 사용하였다.

```
[ ] # 일반 오토인코더

def autoencoder(input_dims):
    input_layer = Input(shape=(input_dims,))
    h = Dense(64, activation='relu')(input_layer)
    h = Dense(64, activation='relu')(h)
    h = Dense(8, activation='relu')(h)
    h = Dense(64, activation='relu')(h)
    h = Dense(64, activation='relu')(h)
    h = Dense(input_dims, activation=None)(h)

    return Model(inputs=input_layer, outputs=h, name='autoencoder')
```

그림 3.2 오토인코더 모델 구축 코드

```

# 다중 오토인코더 모델 (잡음 제거 오토인코더 + 이상신호 감지 오토인코더)

from keras.layers import Input, Dense, Concatenate, Lambda
from keras.models import Model

#잡음 제거 오토인코더
def noise_cancellation_autoencoder(input_dias):
    input_layer = Input(shape=(input_dias,))
    h = Dense(64, activation='relu')(input_layer)
    h = Dense(32, activation='relu')(h)
    h = Dense(16, activation='relu')(h)
    encoded = Dense(8, activation='relu')(h)

    h = Dense(16, activation='relu')(encoded)
    h = Dense(32, activation='relu')(h)
    h = Dense(64, activation='relu')(h)
    decoded = Dense(input_dias, activation=None)(h)

    return Model(inputs=input_layer, outputs=decoded, name='noise_cancellation_autoencoder')

#이상신호 감지 오토인코더
def abnormal_signal_detection_autoencoder(input_dias):
    input_layer = Input(shape=(input_dias,))
    h = Dense(64, activation='relu')(input_layer)
    h = Dense(32, activation='relu')(h)
    h = Dense(16, activation='relu')(h)
    encoded = Dense(8, activation='relu')(h)

    h = Dense(16, activation='relu')(encoded)
    h = Dense(32, activation='relu')(h)
    h = Dense(64, activation='relu')(h)
    decoded = Dense(input_dias, activation=None)(h)

    return Model(inputs=input_layer, outputs=decoded, name='abnormal_signal_detection_autoencoder')

# 다중 오토인코더
def multi_autoencoder(input_dias, noise_cancel_dias, abnormal_dias):
    input_layer = Input(shape=(input_dias,))
    noise_cancel_input = Lambda(lambda x: x[:, :noise_cancel_dias])(input_layer)
    abnormal_input = Lambda(lambda x: x[:, noise_cancel_dias:])(input_layer)

    noise_cancel_autoencoder = noise_cancellation_autoencoder(noise_cancel_dias)
    noise_cancel_output = noise_cancel_autoencoder(noise_cancel_input)

    abnormal_signal_autoencoder = abnormal_signal_detection_autoencoder(abnormal_dias)
    abnormal_output = abnormal_signal_autoencoder(abnormal_input)

    merged_output = Concatenate()([noise_cancel_output, abnormal_output])
    decoded = Dense(input_dias, activation=None)(merged_output)

    return Model(inputs=input_layer, outputs=decoded, name='multi_autoencoder')

```

그림 3.3 다중오토인코더 모델 구축 코드

```

[] from keras.layers import Input, Dense, Conv1D, MaxPooling1D, UpSampling1D, BatchNormalization, Dropout
from keras.models import Model

# Define the input shape of the audio data
input_audio = Input(shape=(None, 1))

# Define the encoder layers
x = Conv1D(64, 3, activation='relu', padding='same')(input_audio)
x = BatchNormalization()(x)
x = Dropout(0.25)(x) # Adding dropout
x = Conv1D(64, 3, activation='relu', padding='same')(x)
x = BatchNormalization()(x)
x = MaxPooling1D(2, padding='same')(x)
x = Conv1D(32, 3, activation='relu', padding='same')(x)
x = BatchNormalization()(x)
x = Dropout(0.25)(x) # Adding dropout
x = Conv1D(32, 3, activation='relu', padding='same')(x)
x = BatchNormalization()(x)
x = MaxPooling1D(2, padding='same')(x)
x = Conv1D(16, 3, activation='relu', padding='same')(x)
x = BatchNormalization()(x)
x = Dropout(0.25)(x) # Adding dropout
x = Conv1D(16, 3, activation='relu', padding='same')(x)
x = BatchNormalization()(x)
x = MaxPooling1D(2, padding='same')(x)
encoded = Conv1D(8, 3, activation='relu', padding='same')(x)

# Define the decoder layers
x = Conv1D(16, 3, activation='relu', padding='same')(encoded)
x = BatchNormalization()(x)
x = Dropout(0.25)(x) # Adding dropout
x = Conv1D(16, 3, activation='relu', padding='same')(x)
x = BatchNormalization()(x)
x = UpSampling1D(2)(x)
x = Conv1D(32, 3, activation='relu', padding='same')(x)
x = BatchNormalization()(x)
x = Dropout(0.25)(x) # Adding dropout
x = Conv1D(32, 3, activation='relu', padding='same')(x)
x = BatchNormalization()(x)
x = UpSampling1D(2)(x)
x = Conv1D(64, 3, activation='relu', padding='same')(x)
x = BatchNormalization()(x)
x = Dropout(0.25)(x) # Adding dropout
x = Conv1D(64, 3, activation='relu', padding='same')(x)
x = BatchNormalization()(x)
x = UpSampling1D(2)(x)
decoded = Conv1D(1, 3, activation='sigmoid', padding='same')(x)

# Create the autoencoder model
autoencoder = Model(input_audio, decoded)

```

그림 3.4 컨볼루션 오토인코더 모델 구축 코드

## 3.2. 성능 예측 및 성능평가

### 3.2.1. 재구성 분포 (Reconstruction Error)

본 연구에서는 다중 오토 인코더 모델의 성능을 평가하기 위해 재구성 오류 (Reconstruction Error) 분포를 활용하였다. 재구성 오류는 모델이 입력 데이터를 얼마나 잘 재구성하는지를 나타내는 척도로, 이를 통해 모델의 학습 상태와 성능을 판단할 수 있다.

재구성 오류 분포는 모델이 입력 데이터를 재구성하는 과정에서 발생하는 오류의 분포를 나타낸다. 이 분포는 모델이 데이터를 어떻게 인식하고 처리하는지, 또한 정상 데이터와 이상 데이터를 어떻게 구분하는지에 대한 중요한 정보를 제공한다. 따라서, 재구성 오류 분포를 분석함으로써, 모델의 학습 상태와 성능을 직관적으로 이해할 수 있다.

본 연구에서는 각 모델의 재구성 오류 분포를 시각화하여 비교하였다. 이를 통해, 다중 오토 인코더 모델이 기존의 오토 인코더 모델과 컨볼루션 오토 인코더 모델에 비해 어떠한 장점을 가지는지 확인하였다. 또한, 이를 통해 모델이 데이터를 어떻게 처리하고, 이상 데이터를 어떻게 감지하는지에 대한 깊은 이해를 얻을 수 있었다.

또한, 재구성 오류 분포를 통해 모델의 과적합 상태를 확인하였다. 과적합은 모델이 학습 데이터에 지나치게 적응하여, 새로운 데이터에 대한 성능이 떨어지는 현상을 말한다. 본 연구에서는 재구성 오류 분포를 통해 모델이

과적합 상태에 빠지지 않았는지를 확인하였다. 이를 통해, 모델이 안정적으로 학습되었음을 입증하였다.

결과적으로, 재구성 오류 분포 분석은 모델의 성능 예측에 중요한 도구로 작용하였다. 이를 통해, 다중 오토 인코더 모델이 기존 모델들에 비해 뛰어난 성능을 가지고 있음을 입증하였으며, 이는 본 연구의 중요한 결과 중 하나이다.

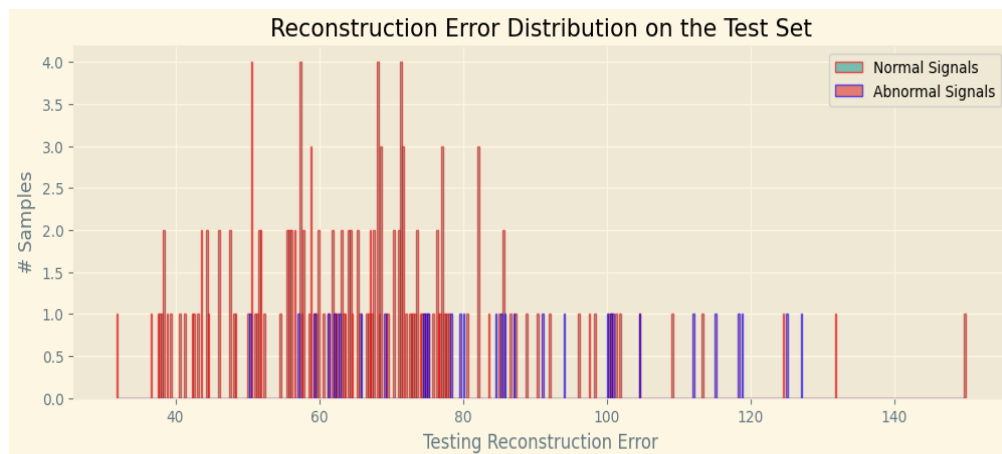


그림 3.5 오토인코더 모델 재구성 분포

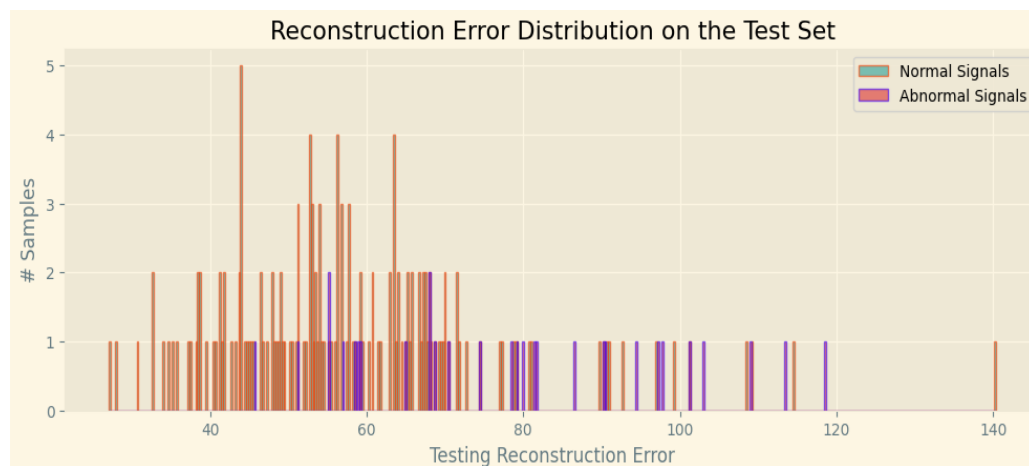


그림 3.6 다중오토인코더 모델 재구성 분포

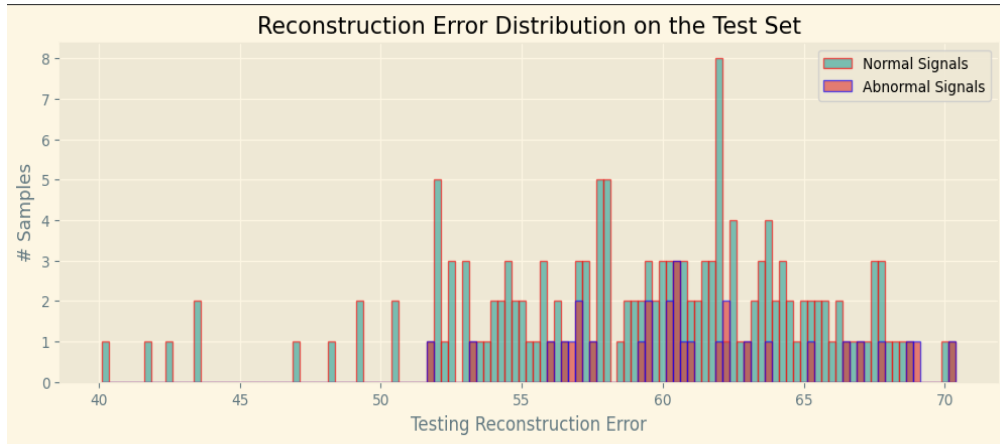


그림 3.7 컨볼루션 오토인코더 모델 재구성 분포

### 3.2.2. 최적의 Threshold 범위 탐색

재구성 오류(Reconstruction Error)를 통해 성능을 예측한 다음 단계로, 본 연구에서는 최적의 Threshold 값을 찾기 위해 Threshold 범위 탐색을 진행하였다. Threshold는 결정 경계를 의미하며, 이를 통해 모델이 이상치를 탐지하는 기준을 설정할 수 있다.

Threshold 범위 탐색은 모델의 성능을 최적화하는 중요한 과정이다. 이 과정에서는 다양한 Threshold 값들을 실험해보며, 모델의 성능을 최적화하는 Threshold 값을 찾게 된다. 이때, 성능을 평가하는 지표로는 정밀도, 재현율, F1 점수 등을 사용할 수 있다.

모델의 성능은 Threshold 값에 크게 의존하며, Threshold 값이 너무 낮으면 모델은 너무 많은 이상치를 탐지하게 된다. 이 경우, 실제로는 이상치가 아닌 데이터까지 이상치로 분류하게 되어 모델의 성능이 떨어지게 된다. 반

면, Threshold 값이 너무 높으면 이상치를 충분히 탐지하지 못하게 된다. 이 경우, 실제로는 이상치인 데이터를 정상치로 분류하게 되어 역시 성능이 떨어지게 된다.

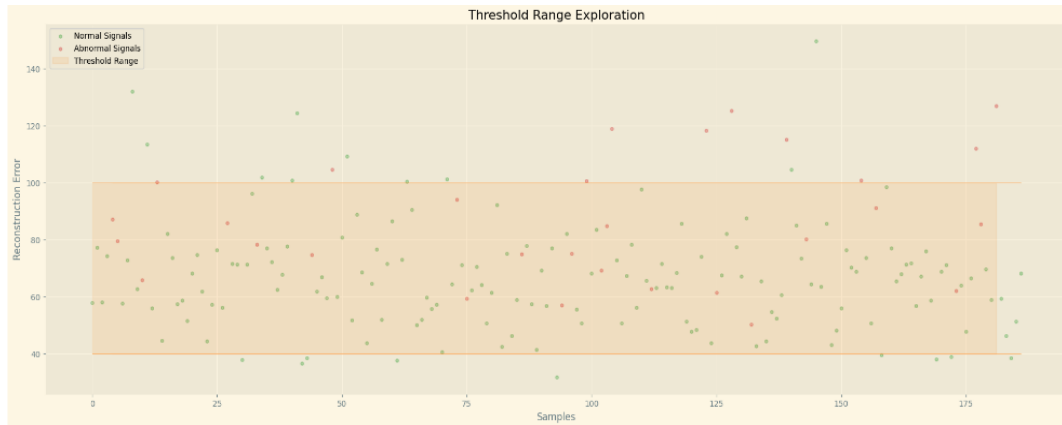


그림 3.8 오토인코더 모델 최적의 Threshold 범위탐색

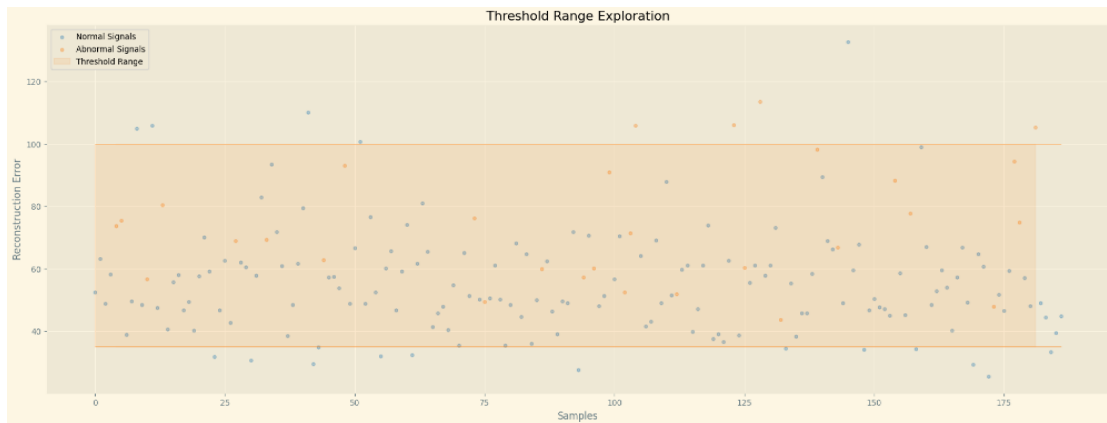


그림 3.9 다중오토인코더 모델 최적의 Threshold 범위탐색

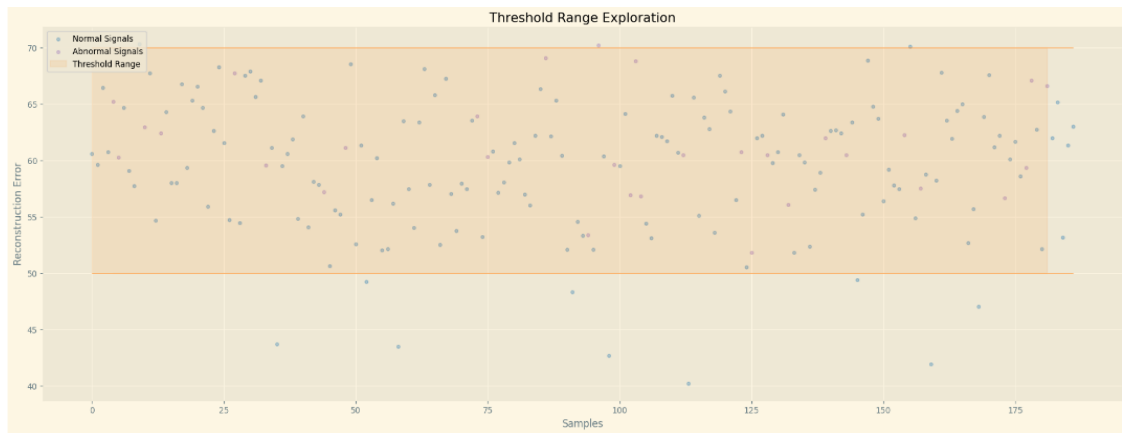


그림 3.10 컨볼루션 오토인코더 모델 최적의 Threshold 범위탐색

따라서, 최적의 Threshold 값을 찾는 것은 모델의 성능을 최적화하는데 중요한 역할을 한다. 본 연구에서는 다양한 Threshold 값을 실험하여, 모델의 성능을 최적화하는 최적의 Threshold 값을 찾았다. 이를 통해, 다중 오토인코더 모델이 이상 신호 감지에 뛰어난 성능을 보이는 것을 입증하였다.

이 과정은 본 연구의 가장 중요한 부분 중 하나로, 이를 통해 다중 오토인코더 모델이 기존 모델들에 비해 어떠한 장점을 가지는지를 명확히 확인할 수 있었다. 또한, 이 과정을 통해 본 연구에서 제안한 모델이 복잡하고 노이즈가 많은 음성 데이터에서 이상 신호를 효과적으로 감지할 수 있다는 것을 입증하였다. 이러한 결과는 다중 오토인코더 모델의 뛰어난 성능을 보여주는 것뿐만 아니라, 이 모델이 실제 응용 분야에서 높은 가치를 가질 수 있음을 보여준다.

### 3.2.3. 최적의 Threshold

재구성 오류의 분포와 Threshold 범위 탐색을 통해 얻은 결과를 바탕으로, 본 연구에서는 최적의 Threshold 값을 결정하였다. 이는 모델의 성능을 최적화하는데 핵심적인 단계로, 모델이 이상 신호를 정확하게 감지하는 데 중요한 역할을 한다.

최적의 Threshold 값은 재구성 오류가 최소가 되는 값으로 선택하였다. 이는 모델이 '정상' 데이터를 가장 잘 재구성하고, 이상한 패턴을 가진 데이터는 잘 재구성하지 못하는 구조를 바탕으로 결정되었다. 이를 통해, 본 연구에서 제안한 다중 오토 인코더 모델은 정상적인 데이터와 이상치를 더욱 명확하게 구분할 수 있었다.

이 결과를 그래프로 시각화 하여 보여주었다. 그래프에서는 Threshold 값에 따른 재구성 오류의 변화를 명확하게 볼 수 있다. 이를 통해, 본 연구에서 제안한 다중 오토 인코더 모델이 정상 데이터와 이상치를 어떻게 구분하는지, 그리고 최적의 Threshold 값이 어떻게 결정되는지를 명확하게 이해할 수 있다.



그림 3.11 일반 오토인코더 모델 최적의 Threshold 그래프



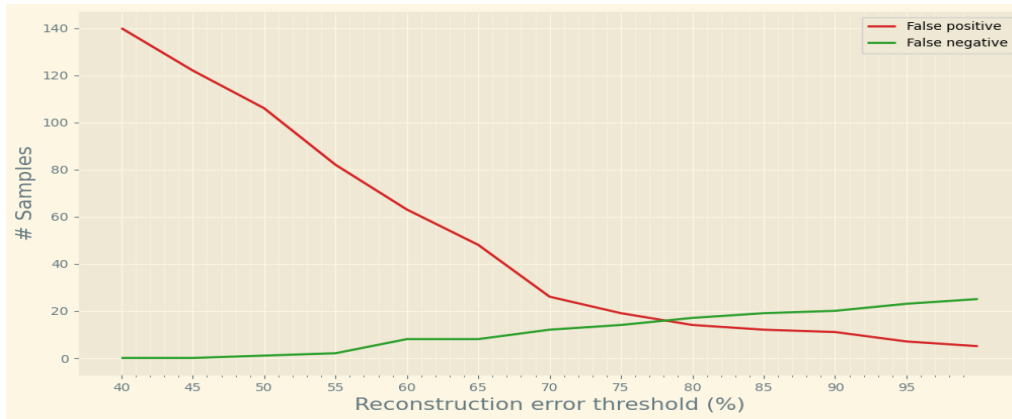


그림 3.12 다중오토인코더 모델 최적의 Threshold 그래프

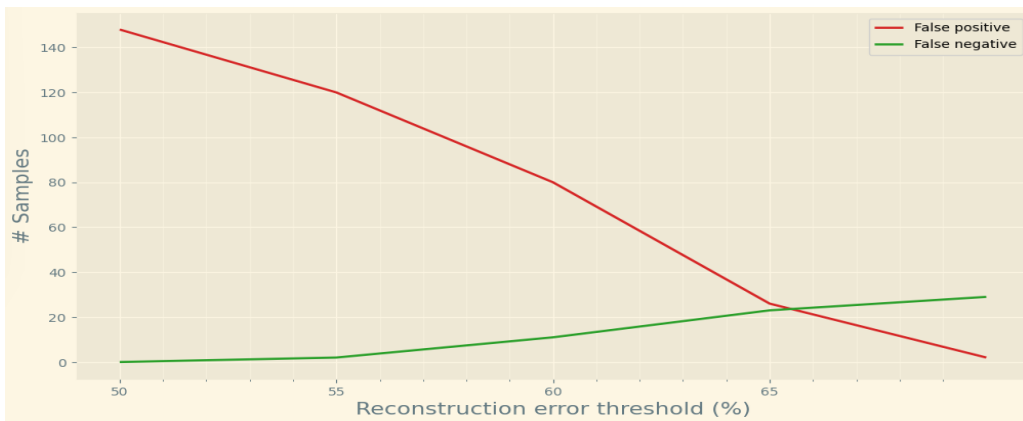


그림 3.13 컨볼루션 오토인코더 모델 최적의 Threshold 그래프

### 3.2.4. Confusion Matrix

본 연구에서는 재구성 오류 분포와 최적의 Threshold 값 탐색을 통해, 제안한 다중 오토 인코더 모델의 성능을 정량적으로 평가하였다. 그 결과를 Confusion Matrix를 이용하여 시각화 하였다.

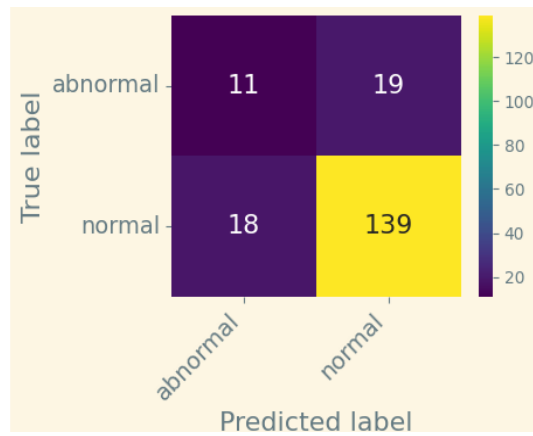


그림 3.14 오토인코더 모델 Confusion Matrix

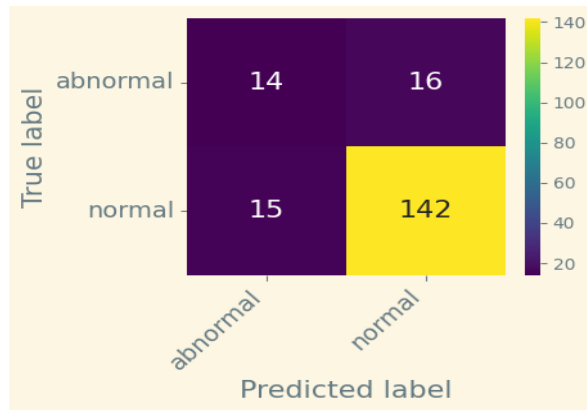


그림 3.15 다중오토인코더 모델 Confusion Matrix

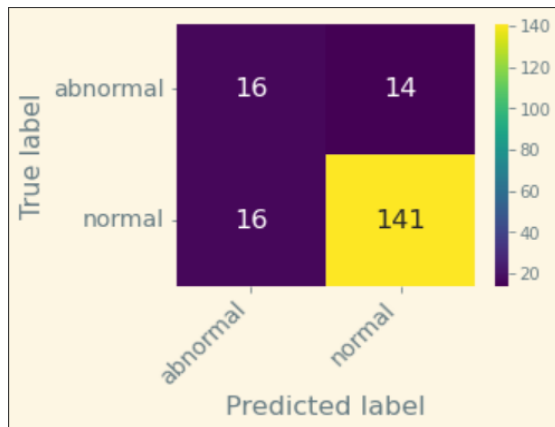


그림 3.16 컨볼루션 오토인코더 모델 Confusion Matrix

Confusion Matrix는 이진 분류 문제에서 모델의 성능을 평가하는 방법 중 하나로, True Positive(TP), False Positive(FP), True Negative(TN), False Negative(FN)의 네 가지 경우를 표현한 행렬이다. 이를 통해, 모델이 데이터를 어떻게 분류하는지, 그리고 그 분류가 얼마나 정확한지를 쉽게 이해할 수 있다.

Confusion Matrix의 핵심은 모델의 예측과 실제 값이 얼마나 일치하는지를 측정하는 것이다. TP와 TN은 모델이 올바르게 분류한 경우를 나타내며, FP와 FN은 모델이 잘못 분류한 경우를 나타낸다. 이를 통해, 모델의 정확도, 민감도, 특이도 등 다양한 성능 지표를 계산할 수 있다.

본 연구에서는 최적의 Threshold 값인 78을 기준으로 이상 신호를 감지하였다. 이를 통해 얻어진 예측 결과를 Confusion Matrix로 표현하였다. 이 과정을 통해, 제안한 다중 오토 인코더 모델이 '정상' 데이터와 '이상' 데이터를 얼마나 정확하게 분류하는지를 명확하게 확인할 수 있었다.

## 제 4 장 결론 및 향후 연구

### 4.1. 결론

본 연구에서는 다중 오토 인코더 모델을 제안하고, 그 성능을 기존의 오토 인코더와 컨볼루션 오토인코더 모델과 비교하였다. 성능 평가는 Confusion Matrix를 이용하여 수행하였으며, Precision, Recall, Accuracy, 그리고 F1 Score라는 네 가지 평가 지표를 사용하였다.

Confusion Matrix를 통해 성능평가를 진행한 결과는 다음과 같다.

	오토인코더	다중 오토인코더	컨볼루션 오토인코더
Precision	88.5%	91.1%	89.8%
Recall	88.0%	89.9%	91.0%
Accuracy	80.2%	84.5%	84.0%
F1 Score	88.3%	90.5%	90.4%

표 4.1 3가지 오토인코더 모델 결과 비교

다음으로 위의 결과들을 한눈에 보기 편하게 그래프로 시각화 하여 보여주었다.

세가지 모델을 비교하여 그래프로 나타내면 다음과 같다.

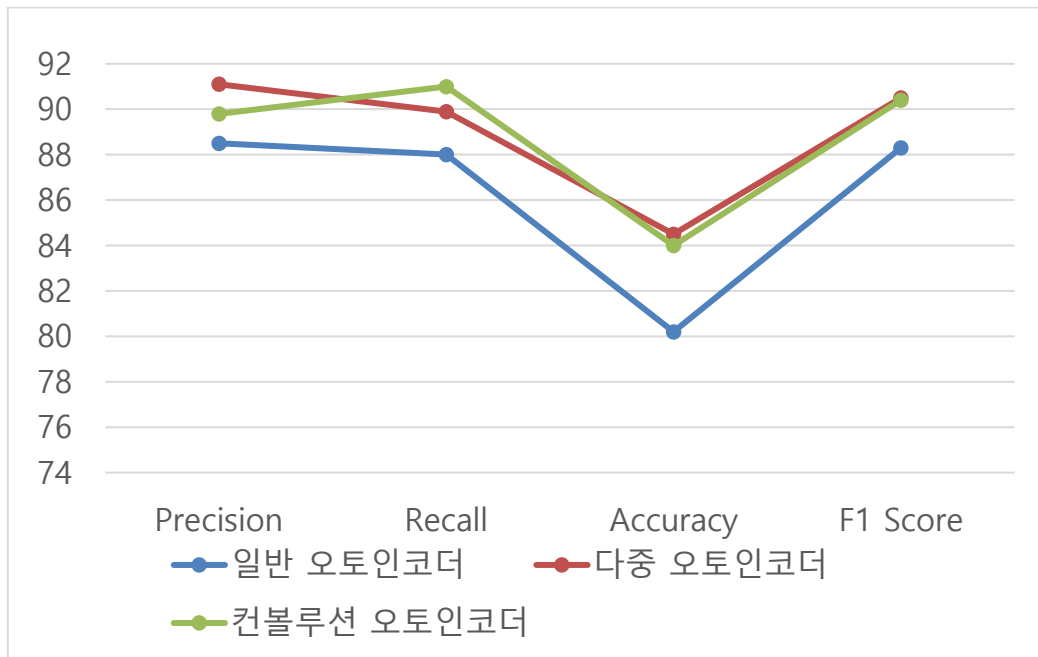


그림 4.1 3가지 오토인코더 모델 결과 그래프

이 결과를 통해, 제안된 다중 오토 인코더 모델이 기존의 오토 인코더에 비해 약 4%의 성능 향상을 보였음을 확인하였다. 또한, 컨볼루션 오토인코더와 비교하였을 때도, 미세하게나마 더 높은 성능을 보였다. 이는 다중 오토 인코더가 복잡하고 노이즈가 많은 데이터에서 이상치를 감지하는 데 있어서 더욱 효과적인 접근 방식임을 입증하는 결과이다.

## 4.2. 향후 연구

본 연구에서 제시된 다중 오토 인코더 모델은 이상 신호 감지와 노이즈 제거라는 두 가지 중요한 기능을 수행하는 데 성공하였다. 하지만, 항상 개선의 여지가 있으며, 이를 통해 더 높은 성능과 다양한 응용 분야로 확장될 수 있는 가능성을 갖추고 있다. 이제 우리는 향후 연구 방향에 대해 몇 가지 제안을 해 볼 것이다.

첫째로, 다중 오토 인코더의 각 부분을 개선하는 것은 이 모델의 성능을 향상시키는 중요한 방법일 수 있다. 이상 신호 감지 오토 인코더는 더 정교한 이상 패턴을 감지할 수 있도록 개선될 수 있으며, 노이즈 제거 오토 인코더는 더 다양한 유형의 노이즈를 제거할 수 있도록 설계될 수 있다. 이런 개선 작업은 모델의 복잡성을 높일 수 있지만, 적절한 학습 및 정규화 기법을 사용하면 과적합 문제를 방지하면서 더 높은 성능을 달성할 수 있을 것이다.

둘째로, 다중 오토 인코더를 다른 모델과 결합하는 것은 또 다른 흥미로운 연구 주제가 될 수 있다. 예를 들어, 다중 오토 인코더를 재구성 오류를 기반으로 하는 분류 모델과 결합하면, 이상 신호의 세부 유형을 구분하는 데 사용될 수 있을 것이다. 이를 통해, 이상 신호 감지의 정확성을 높이는 동시에, 이상 신호의 원인을 더 정확하게 파악하는 데 도움을 줄 수 있다.

셋째로, 다중 오토 인코더는 더 다양한 데이터 유형에 적용될 수 있다. 본 연구에서는 음성 데이터에 초점을 맞추었지만, 다중 오토 인코더는 이미지,

텍스트, 시계열 데이터 등 다양한 유형의 데이터에 적용될 수 있다. 각 데이터 유형마다 고유한 특성과 패턴이 있으므로, 다중 오토 인코더를 이러한 다양한 데이터 유형에 맞게 조정하는 것이 중요하다. 특히, 시계열 데이터는 이상 패턴 감지에 매우 중요한 데이터 유형이며, 다중 오토 인코더는 이러한 데이터에 대한 강력한 도구가 될 수 있다.

또한, 다중 오토 인코더는 다양한 실제 환경에서의 테스트와 유효성 검사를 거쳐야 한다. 현재까지, 이 모델은 특정 데이터셋에 대해 테스트되고 있지만, 실제 세계의 데이터는 종종 노이즈가 많고 복잡한 패턴을 가지고 있다. 따라서, 모델을 실제 환경에서의 도전 과제에 대해 더 잘 대비시키기 위해, 다양한 유형의 데이터셋과 실제 세계의 시나리오에서의 성능을 평가하고 검증하는 것이 중요하다.

마지막으로, 다중 오토 인코더의 해석 가능성을 향상시키는 것은 또 다른 중요한 연구 주제입니다. 복잡한 모델은 종종 '블랙 박스'라는 문제를 가지고 있으며, 이는 모델의 내부 작동 방식을 이해하거나 설명하는 것이 어렵다는 것을 의미한다. 따라서, 다중 오토 인코더의 내부 구조와 동작을 더 잘 이해하고 설명할 수 있는 방법을 개발하는 것은 이 모델을 실제 세계의 문제에 더 효과적으로 적용하는 데 중요한 단계가 될 수 있다.

이러한 연구 방향들은 다중 오토 인코더의 잠재력을 최대화하는 데 도움이 될 것이다. 본 연구는 이런 방향을 제시하며, 이 분야의 연구자들이 이 모델을 더욱 발전시키고 다양한 문제에 적용하는 데 기여하길 바란다.

## 참고문헌

- [1] 식품의약품안전평가원, 치매 등 퇴행성뇌질환 치료제 및 진단기술 동향 정보집.pdf, 14p
- [2] 티스토리 블로그 - [머신러닝] 오토인코더(Autoencoder)를 알아보자.  
<https://lewisxyz000.tistory.com/22>, (2018.03.11)
- [3] 깃허브 블로그 - Novelty Detection(이상치 탐지) - Overview  
[https://jayhey.github.io/novelty%20detection/2017/10/18/Novelty\\_detection\\_overview/](https://jayhey.github.io/novelty%20detection/2017/10/18/Novelty_detection_overview/)(2017.10.18)
- [4] 위키백과, <https://ko.wikipedia.org/wiki/음향학>
- [5] 위키백과, <https://ko.wikipedia.org/wiki/음성학>
- [6] Anomalous Sound Detection Using Multiple Autoencoders  
Yong-Geun Moon, Min-Seong Kwon, Jung-Hoon Noh, Byungju Lee Kumoh  
National Institute of Technology (2022.02)
- [7] “Residual Error Based Anomaly Detection Using Auto-Encoder in SMD Machine Sound” , MDPI, Dong Yul Oh, Il Dong Yun (2018.03)
- [8] “Abnormal signal detection based on parallel autoencoders” ,  
The Journal of the Acoustical Society of Korea.  
Kibae Lee<sup>1</sup>, Chong Hyun Lee (2021.07)



## ABSTRACT

### Detection of abnormal signals of dementia voice data through multiple autoencoders

Kim, Young Jin

Major in Computer Science and  
Engineering

The Graduate School of Engineering  
Hanyang University

In this work, we propose a multi-autocoder model for effective anomaly signal detection. We compare this model with conventional autoencoder and convolutional autoencoder models. Through this, it was proved that the multi-auto encoder model showed superior performance compared to the existing models.

The proposed multiple autoencoder was constructed by combining an anomaly signal detection autoencoder and a noise cancellation autoencoder. These two functions played an important role in accurately detecting abnormal signals in complex and noisy voice data.

After the learning progressed, the reconstruction error distribution was analyzed to confirm the learning state of the model. Through this, we analyzed how the model recognizes data and distinguishes outliers from normal values.

In the next step, several threshold values were tested to find a threshold value that optimizes the performance of the model. The optimal threshold value played an important role in predicting the performance of the model. Based on this value, the performance of the model was visualized and performance evaluation was conducted using Confusion Maxtrix.

Through these processes, the proposed multiple autoencoder model showed performance improvement of more than 4% compared to the existing autoencoder. It also showed slightly higher performance compared to convolutional autoencoder.

## 감사의 글

제가 이 논문을 작성하는 데 있어서 항상 지원해주시고 긍정적인 피드백을 주신 조인휘 교수님께 감사의 말씀을 드립니다.

대학원 생활을 하면서 제 자신을 조금씩 성장시킬 수 있는 계기가 되었습니다. 대학원 생활을 하는 동안 여러 동기분들 그리고 기장님께서 끝까지 포기하지 않고 함께 해쳐 나갈 수 있도록 학교생활 전반으로 도움을 많이 주셨습니다.

특히 많은 인원에도 불구하고 모든 인원들이 함께 어울릴 수 있도록 학교 생활 뿐만 아니라 학교의 여러 행사까지도 챙겨 주셨던 김유현 기장형님께 감사의 말씀을 드립니다. 대학원 생활이 처음이고 걱정도 많았지만 그때마다 어려운 학교 생활을 이겨낼 수 있도록 동기분들이 도움을 주셨고 힘들었지만 보람찬 대학원 생활이 되었습니다.

무사히 졸업할 수 있도록 응원을 해주시고 조언 및 도움을 주신 분들께 진심으로 감사의 말씀 드립니다.

2023년 8월

김영진