creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

공학석사 학위논문

인적 오류로 인한 결함 식별 누락 방지를 위한 디지털 방사선투과검사 데이터 품질 평가 및 AI 알고리즘 연구

Research on data quality evaluation and AI algorithm of Digital Radiographic Testing to prevent defect identification missing as human error

울산대학교 산업대학원

산업경영공학

김 균 엽

인적 오류로 인한 결함 식별 누락 방지를 위한 디지털 방사선투과검사 데이터 품질 평가 및 AI 알고리즘 연구

Research on data quality evaluation and AI algorithm of Digital Radiographic Testing to prevent defect identification missing as human error

지도교수 이 수 동

이 논문을 공학석사 학위논문으로 제출함

2022 년 08 월

울산대학교 산업대학원

산업경영공학

김 균 엽

김균엽의 공학석사 학위논문을 인준함
2022 년 08 월

심사위원장 정 기 효 (인)

심사위원 황 규 선 (인)

심사위원 이 수 동 (인)

울산대학교 산업대학원

2022 년 08 월

인적 오류로 인한 결함 식별 누락 방지를 위한 디지털 방사선투과검사 데이터 품질 평가 및 AI 알고리즘 연구

김 균 엽

울산대학교 산업대학원 산업경영공학 전공

(지도교수: 이 수 동)

(국문 초록)

방사선투과검사는 용접이 적용된 철강 제품의 신뢰성을 검증하는 주요한 검사 방법으로, 인정된 비파괴 검사원이 검사 결과물(필름 또는 디지털 이미지)을 판독하여 용접부 결함을 식별 및 분류한다. 검사원에 의한 평가 오류는 물적 · 인적 사고와 직결되므로, 이를 예방하기 위하여 다른 검사원에 의한 샘플링 검증 절차가 수행되고 있으나, 용접 결함이 식별 및 수리되지 않고 고객에게 납품되거나 현장에 설치/사용되는 경우가 있다.

이와 같은 리스크와 손실을 예방하기 위해, 용접부 결함 식별 및 분류를 위한 AI 연구가 활발히 이루어지고 있다. 하지만 지금까지의 연구는 GDX-ray 등의 공유된 소수의 용접부 방사선투과검사 이미지 데이터베이스 또는 필름으로 촬영된 이미지를 디지털화하여 AI 학습 및 성능 평가가 이루어지고 있다.

본 연구에서는 적용되는 산업표준 및 고객 요구사항에 따라 제조 및 검사되어 납품된 실제 제품의 다양한 디지털 방사선투과검사 이미지를 활용하여 이미지 데이터의 품질을 평가하고 AI 모델의 성능을 평가하였다. 평가결과를 바탕으로 향후 연구개발을 추진하면, 디지털 방사선투과검사 결과 평가를 지원하는 상용적인 AI 모델이 개발될 것으로 판단된다.

목차

국문초록 v

제 1 장 연구 배경 1

1.1 연구의 필요성 1

1.2 연구 내용 2

1.3 연구 현황 3

1.4 연구의 독창성 5

제 2 장 관련 작업 7

2.1 용접 강관 및 디지털 방사선투과검사 7

2.2 디지털 방사선투과검사 데이터 9

제 3 장 연구 내용 10

3.1 데이터 품질 평가 10

3.2 데이터 전처리 11

3.3 AI 성능 평가 13

제 4 장 결과 및 고찰 16

4.1 데이터 품질 평가 16

4.2 AI 성능 평가 17

제 5 장 결론 25

5.1 요약 25

5.2 기대효과 26

5.3 향후 연구개발 27

참고문헌 28

Abstract 31

감사의 글 32

그림 목차

[그림 1] UT, RT, rVT 및 ET 에서의 Risk 우선순위 등급 3
[그림 2] 용접 결함 자동 검출 절차 4
[그림 3] 데이터 셋 샘플 5
[그림 4] 용접 강관 제조 공정도 7
[그림 5] 디지털 방사선투과검사 설비 8
[그림 6] 디지털 방사선투과검사 이미지 9
[그림 7] 이미지 데이터 품질 평가 샘플 10
[그림 8] 이미지 데이터 분할 처리 개략도 11
[그림 9] 데이터 셋 구성 요약 12
[그림 10] ResNet50 Hyper-parameter 13
[그림 11] ResNet50 구조 및 데이터 셋 14
[그림 12] Orange Data Mining 학습 네트워크 14
[그림 13] 고에너지 X-ray 에서의 MTF 및 NPS 특성 17
[그림 14] 학습 시간(초) 18
[그림 15] 학습 과정 중 Epoch 에 따른 정확도 19
[그림 16] 모델별 예측 Confusion Matrix 23
[그림 17] 모델별 예측 ROC Curve 24
[그림 18] 향후 연구개발 계획 27

표 목차

[표 1] Performance metrics of DCNN models with processing time 4
[표 2] 학습 데이터 양에 따른 모델별 학습 성능 19
[표 3] 모델별 예측 정확도 20
[표 4] 모델별 예측 성능 지표 22
[표 5] 모델별 예측 AUC 24

제 1 장 연구 배경

1.1 연구의 필요성

용접 강관은 구조용 말뚝, 천연가스 및 석유 배관, 풍력 타워 등으로 사용되고 있으며, 수소 배관으로의 사용을 위해 지속적인 연구개발이 진행되고 있는 주요 철강 제품이다. 이와 같은 용접 강관의 용접부 건전성은 강관의 주요한 품질 특성이며, 강관이 사용되는 곳의 안전성을 보장하기 위한 최소한의 품질 특성이다. 적용되는 국내 · 외 산업표준 및 법규에 따라 용접부의 건전성을 검증하는 여러 비파괴검사가 요구되고 있으며, 그 중에서 방사선투과검사가 일반적으로 요구되고 있다.

현재, 방사선투과검사 설비 관련 설비 자동화 기술 및 ICT 의 발전으로 용접 강관의 검사 준비 및 진행 등의 기본적인 설비 운영은 자동 및 원격으로 이루어지고 있으며, 생성된 데이터(디지털 이미지 및 판정 정보)는 자동으로 저장 · 공유되고 있다.

하지만 용접 강관의 용접부 결함 유무 및 결함 종류의 판정은 국내 · 외 적용 법적 · 규제적 요구사항에 따라 인정된 검사원에 의해 수행되고 있다. 인정된 검사원에 의해 판정된 결과는 식별 오류 리스크를 예방하기 위해 다른 검사원에 의해 샘플링 재검토되고 있으나, 용접 결함이 식별되어 수리되지 않고 고객에게 납품되거나 현장에 설치 · 사용되는 경우가 있다.

2006 년 4 월 4 배의 방사성 물질이 환경으로 배출된 월성원전의 용접 결함부 중수누출 사고, 2018 년 12 월 백석역 온수관 용접부 파열로 1 명이 사망하고 30 여명이 화상을 입은 사고, 1980 년 3 월 용접 결함부 피로 파괴로 노르웨이

석유 굴착용 해양구조물이 전복되어 123 명이 사망하는 등, 용접 결함이 식별되지 못하고 설치 및 사용되어 심각한 인적 · 물적 피해가 발생된 사례이다.

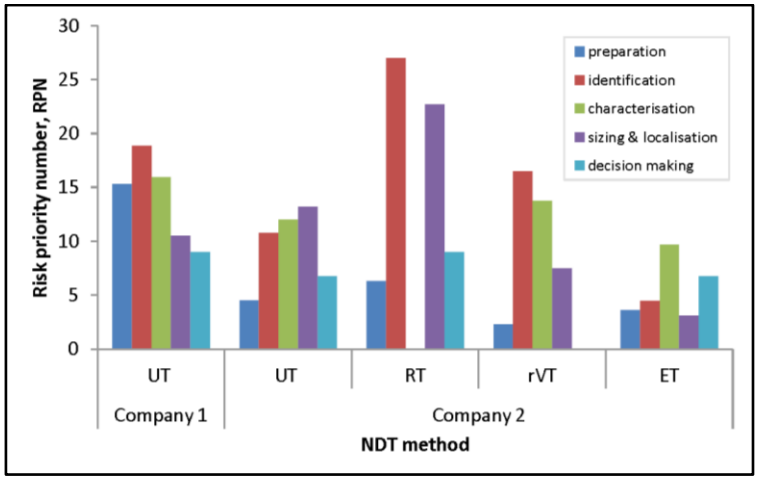
이와 같은 리스크를 예방하기 위해 검사원의 결함 판정을 지원하기 위한 AI 기술 적용이 필요하다.

1.2 연구 내용

Vorgelegt Von 등의 연구[1]에 따르면 [그림 1]과 같이 방사선투과검사(RT)의 경우에는 검사자의 결함 식별 오류 리스크가 가장 높음을 알 수 있으며, 반면에 소프트웨어에 의해 자동으로 식별이 지원되는 와전류탐사검사의 경우에는 식별 리스크가 가장 낮음을 알 수 있다. 현재 방사선투과검사의 준비(preparation)와 치수 및 위치(sizing & localization) 측정은 자동화된 장비에 의해 지원되고 있지만, 결함 유무는 검사원에 의해서만 식별(identification)된다.

특이하게, 결함의 종류를 특징짓는 오류는 주요한 리스크가 아님을 확인할 수 있었다. 이는 식별된 결함에 대해 결함의 종류를 육안으로 특징짓는 것은 어렵지 않기 때문이라 생각된다.

본 연구에서는 적용 국내 · 외 산업표준에 따라 제조 및 납품된 용접 강관의 디지털 방사선투과검사 이미지 데이터 품질을 검증하고, 인정된 비파괴 검사원에 의해 판정되고 공인기관의 검사원에 의해 재검토된 이미지 데이터 기반으로 “인적 오류로 인한 결함 식별 누락 방지를 위한 디지털 방사선투과검사 데이터 품질 평가 및 AI 알고리즘 연구” 하고자 한다.



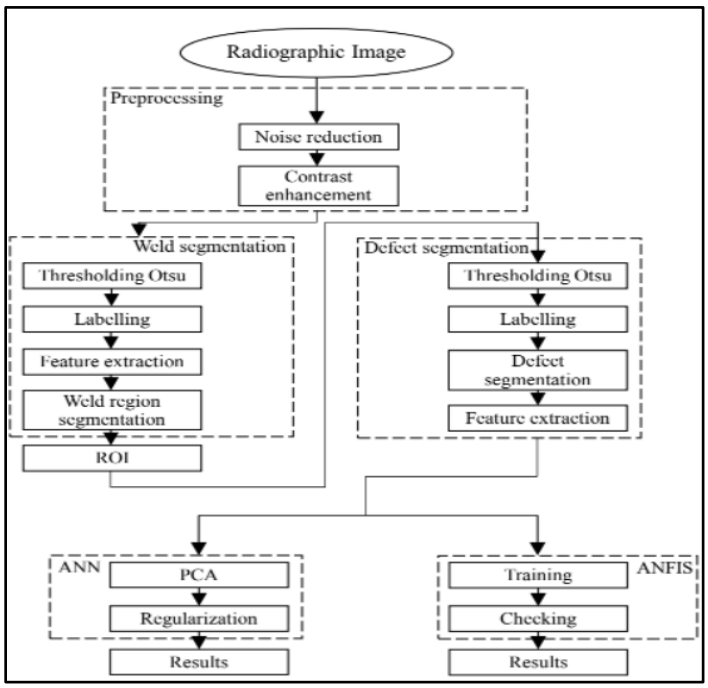
[그림 1] 초음파탐상검사(UT), 방사선투과검사(RT), 원격육안검사(rVT) 및 와전류탐상검사(ET)에서의 Risk 우선순위 등급

1.3 연구 현황

1) 국외 연구개발 현황

2000 년 Nacereddine Naffaa 외[2]는 ANN(Artificial Neural Network)를 활용한 방사선투과검사 필름 상의 용접 결함을 분류하였다. 또한 2002 년 I.M. Elewa 외[3]는 이미지 처리 소프트웨어를 활용하여 방사선투과검사 필름 상의 용접 결함을 식별하였다. 이때 사용된 이미지 데이터는 모두 필름으로 촬영된 방사선투과검사 이미지를 디지털화하여 적용하였다. 2019 년 Haizing Zhu 외[4]는 CNN 을 활용하여 용접부 표면 결함을 분류하였다. 2020 년 Wenhui Hou 외[5]는 [그림 2]와 같은 용접 결함 자동 검출 절차를 제안하였다.

그리고 GDX-ray 데이터 셋이 공유됨에 따라, Lu Yang 외[6]는 용접 결함을 5 가지 종류로 분류하였으며, Topias Tyystjarvi 외[7]는 우주항공산업의 용접부 결함 자동 검출 등을 연구하였다. Chiraz Ajmi 외[8]는 [표 1]과 같이 관련된 AI 모델별 성능을 평가하였다.



[그림 2] 용접 결함 자동 검출 절차

[표 1] Performance metrics of DCNN models with processing time

Pretrained model	Accuracy	Error	Sensitivity	Specificity	Precision	False positive rate	Time
AlexNet	100	0	100	100	100	0	47 s
VGG-16	95	0.05	100	85	92	0.145	3 min 34 s
VGG-19	97.8	0.0216	100	93.75	96.8	0.0625	3 min 55 s
GoogLeNet	99.3	0.0072	98.9	100	100	0	1 min 38 s
ResNet50	100	0	100	100	100	0	3 min 12 s
ResNet101	100	0	100	100	100	0	6 min 52 s

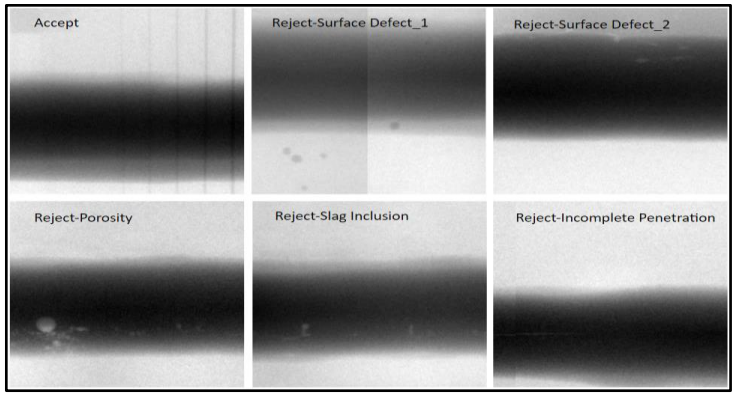
2) 국내 연구개발 현황

2021 년 이기동 외[9]의 CNN 모델 평가와 시각화를 하였다. 오준택 외[10]는 압출재의 표면 결함을 연구하였으며, 최원준 외[11]는 ResNet50 을 활용하여 냉연 강판의 표면결함 자동분류를 연구하였다.

1.4 연구의 독창성

국내 · 외 연구현황에 기술하였듯이 디지털 방사선투과검사 이미지는 제조 현장에서 발생되는 대외비성의 검사 데이터이다. 따라서 학습을 위한 대량의 데이터 셋을 준비하는 것이 어렵기 때문에 공유 데이터베이스인 GDX-ray 에서 제공하는 제한적인 데이터를 이용하거나 필름을 디지털화 하여 학습 및 성능 평가가 이루어졌다.

본 연구에서는 국내 · 외 적용 산업표준에 따라 제조 및 검사되어 납품된 용접 강관의 실제 데이터를 활용하여 [그림 3]과 같이 용접부 결함 유무를 분류하고 학습 및 테스트를 위한 데이터 셋을 구축하였다.



[그림 3] 데이터 셋 샘플

그리고 Wenhui Hou 외[5]의 용접 결함 자동 검출 절차에 따라 이미지 데이터의 품질을 평가하고 검사 조건을 최적으로 조절하였으며, Chiraz Ajmi 외[8]의 모델별 성능을 참조하여 ResNet50 등을 적용하였다.

또한 스마트공장 추진으로 관련 데이터베이스를 구축하고 있고 이를 활용할 의지가 분명한 중소기업이 데이터 엔지니어의 도움 없이 AI 적용 가능성을 검토할 수 있도록 하기 위해 코드리스(Codeless) 머신러닝 도구인 Orange Data Mining 을 활용하여 ResNet50 과 성능을 비교 평가하였다.

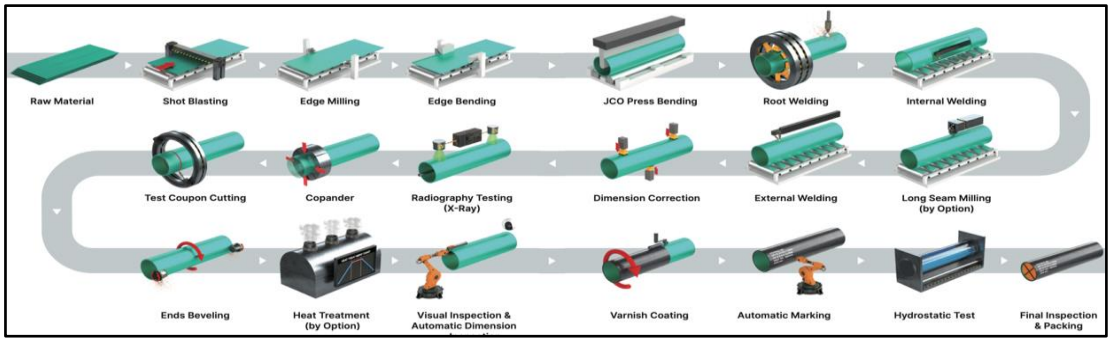
제 2 장 관련 작업

2.1 용접 강관 공정 및 디지털 방사선투과검사

1) 용접 강관 제조 공정

용접 강관은 [그림 4]와 같이 강판을 원형으로 성형하여 이음부를 용접하고, 용접 변형 등을 교정한 후, 구매자가 요구한 2 차 가공 공정(관단 가공, 열처리, 표면 처리, 도장, 표시 등)을 거쳐서 최종 제품으로 만들어진다. 공정 중 또는 최종 검사는 육안, 치수, 물리적 · 화학적 특성 평가 시험, 수압 시험과 비파괴검사가 적용된다.

용접은 SAW(Submerged Arc Welding) 또는 ERW(Electric Resistance Welding)가 적용되는데, 저온 · 고압 환경에서 사용되는 강관은 SAW 가 적용된다. 일반적인 용접 강관의 표준 치수는 외경 400~2,000 mm, 두께 6~60 mm, 길이 4,000~13,000 mm 이다. 본 연구에서는 두께 10 mm 이하의 용접 강관의 방사선투과검사 이미지를 적용하여 수행하였다.

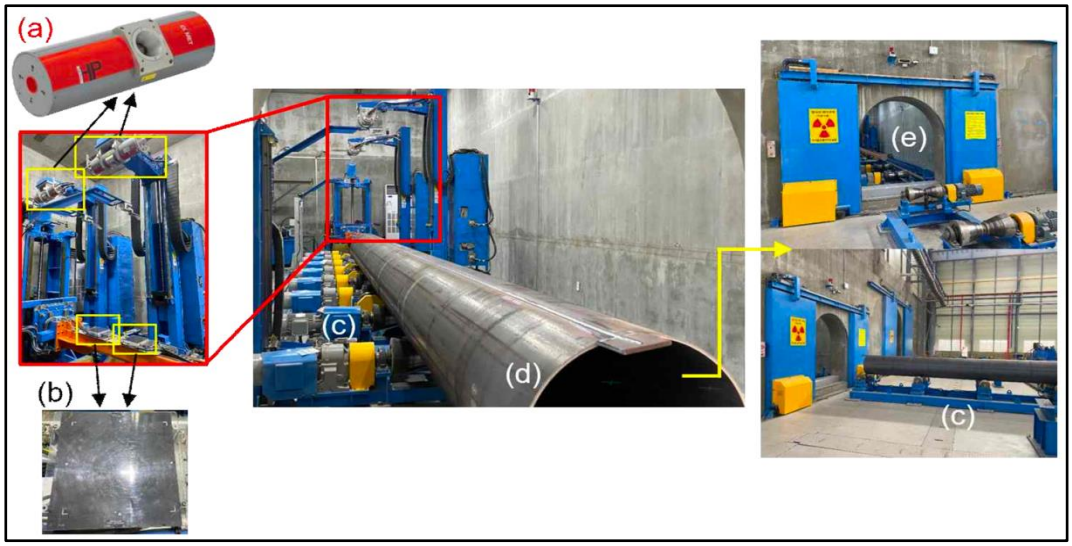


[그림 4] 용접 강관 제조 공정도

2) 디지털 방사선투과검사

디지털 방사선투과검사 설비는 [그림 5]와 같으며, X-ray 및 디텍터가 용접 강관의 용접부를 약 80mm 씩 촬영하여 이미지를 생성한다. 1 일 평균 용접 강관 약 50 본(本, 강관의 셈 단위)에 대해 방사선투과검사가 수행되며, 생산량 대비 용접 부적합률은 약 5 % 이하이므로, 1 일 2~3 본 부적합이 포함된 강관이 생산 및 수리되어 재검사된다. 대표 길이 12 m 용접 강관 단품 기준으로 약 150 장(images)의 방사선투과검사 이미지가 생성되며, 생성된 이미지 장수 대비 부적합률은 약 1 % 이하이므로, 1 일 50~80 장의 부적합 이미지가 생성된다.

본 연구에서 사용된 X-ray 발생 장치는 최대 450 kV-15 mA 이며, X-ray 를 빛으로 변환하는 디지털 디텍터의 최소 픽셀 크기는 0.119 mm 이고, 16 bit grayscale 이며, 유효 면적은 149 x 149 mm^2 이다.



[그림 5] 디지털 방사선투과검사 설비 (a) X-ray 튜브 (b) 디지털 디텍터 (c) 용접 강관 이송 장치 (d) 용접 강관 (e) 방사선 차폐 시설

2.2 디지털 방사선투과검사 데이터

디지털 방사선투과검사 데이터인 이미지는 [그림 6]과 같다. 약 80 mm 간격으로 촬영된 방사선투과검사 이미지를 판독 및 결함 위치 추적을 위해 제품의 길이와 동일하게 이어 붙인다. 개별 이미지를 이어 붙이는 과정에서 [그림 6]에 표시한 것과 같은 “이미지 연결부”가 생기는데, 추가적인 프로그램 수정이 필요하여 본 연구에서는 현상태로 사용하였다. 하나의 이미지로 붙여진 길이 12 m 용접 강관 단품의 이미지 데이터 크기는 약 250 MB 이며 .tif 파일로 생성되어 서버에 저장이 된다. 용접부는 강관의 길이방향으로 위치하며, 강관의 양쪽 끝은 용접 시작/완료 탭피스가 포함된다. [그림 6]의 (c)와 같이 용접부 이미지에는 용접부 이외에 모재부 및 IQI(Image Quality Indicator)를 포함한다. 용접부의 폭은 용접 강관의 두께에 따라 다르며, 두께가 두꺼울수록 넓어지나 대략 15~25 mm 정도이다.

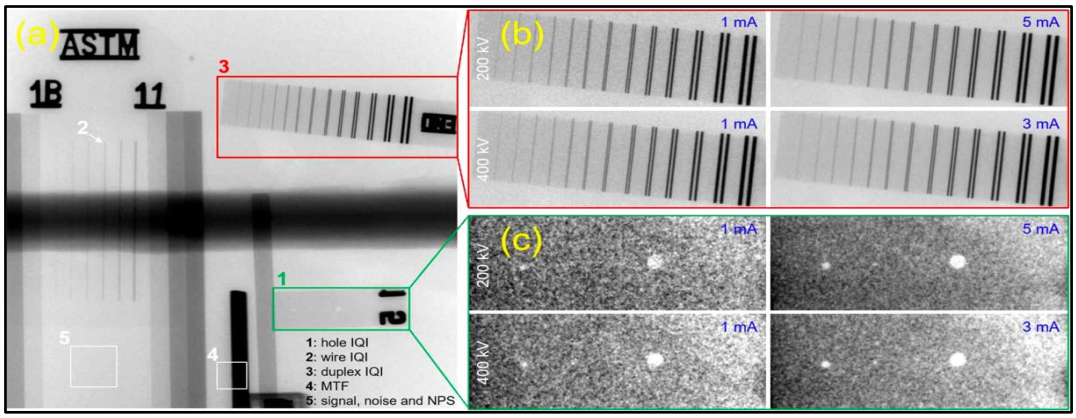
[그림 6] 디지털 방사선투과검사 이미지(a) 용접 강관, (b) 방사선투과검사 이미지 (전 길이), (c) & (d) 양쪽 관단 방사선투과검사 이미지

제 3 장 연구개발 내용

3.1 데이터 품질 평가

AI 기술을 적용하여 용접 강관의 다양한 두께에서의 여러 가지 결함을 검출하기 위해서는 이미지 데이터의 콘트라스트 및 공간 해상도와 같은 데이터 품질이 요구되나, 고에너지 X-ray 영역에서는 콤프턴 산란 반응 등의 영향으로 콘트라스트가 낮아지고 노이즈가 높아져서 이미지 데이터의 품질이 저하되게 된다[12]. 이와 같은 이유로 결함 판정을 위한 AI 기술 적용을 위해 원본 이미지 데이터 품질을 평가하고 개선하는 것이 선행되어야 한다.

이를 위해, 국제 산업표준인 ASME 및 ASTM[13][14]에 따른 IQI 등의 이미지 데이터 품질 측정 기구를 사용하여 용접 강관의 대표 두께인 9.53 mm 에서 [그림 7]과 같이 이미지 데이터 품질을 평가하였다.



[그림 7] 이미지 데이터 품질 평가 샘플 (a) 용접 강관의 방사선투과검사 이미지 (b) duplex IQI (c) hole IQI

또한 이미지 데이터 품질을 정량적으로 표현하는 MTF(Modulation Transfer Function) 및 NPS(Noise Power Spectrum)를 구하였다[15][16]. 여기서, MTF 는 어떤 영상시스템의 주파수응답을 나타내는 것으로써 피사체대조도(Subject Contrast)에 대한 영상대조도(Image Contrast)의 비로 정의된다. NPS 는 잡음의 분산 값의 분포를 공간주파수 상에 표현한 것으로 영상에서 픽셀간의 변동요인인 잡음의 공간주파수에 대한 의존도를 나타낸다.

3.2 데이터 전처리

길이 약 80 mm 씩 촬영하여 생성된 이미지를 판독 및 결함 위치 추적을 위해 하나의 이미지로 이어 붙여 생성된 길이 12 m 용접 강관의 디지털 방사선투과검사 이미지의 크기는 약 250 MB(125,000 x 1,250 pixel) 정도이다. 학습 및 테스트를 위해 [그림 8]과 같이 관심영역인 용접부 및 일부 모재부를 포함하도록 400 x 400 pixel 로 이미지를 분할하였다.

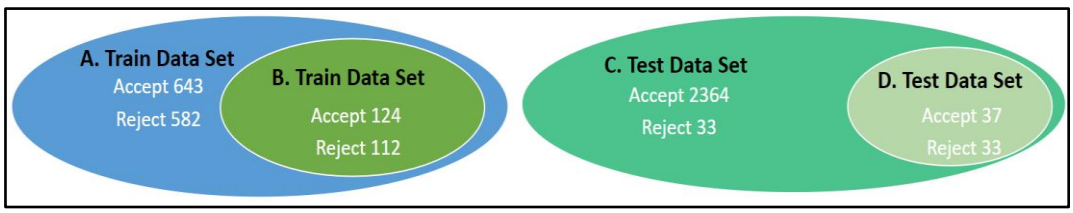
[그림 8] 이미지 데이터 분할 처리 개략도

분할된 이미지는 검사원의 판정 결과에 따라 합격과 부적합으로 분류하고 공인기관의 검사원이 재검

토하여 [그림 9]와 같이 데이터 셋을 구축하였다.

학습용 데이터 셋은 합격과 부적합 비율이 유사하도록, 합격 643 images 및 부적합 582 images 를 준비하였다. 데이터 양에 따른 성능 비교를 위하여, 준비된 학습 데이터 셋에서 임의로 236 images 를 추출하였다.

동일한 테스트 데이터 셋을 이용하여 최종적으로 각 모델별 예측 성능을 평가하기 위해, 학습 데이터 셋과는 별도로 실제 제품의 부적합률(약 1.0~1.5 %)이 반영된 테스트 데이터 셋을 별도 준비하였다. 또한 학습 데이터 셋의 부적합 비율(합격 53 %: 부적합 47 %)과 동일한 비율을 가지는 데이터 셋을 준비된 테스트 셋에서 임의로 70 images 를 추출하였다.



[그림 9] 데이터 셋 구성 요약

SAW 용접 강관의 용접결함의 종류는 PO(Porosity), SI(Slag Inclusion), SD(Surface Defect), IP(Incomplete Penetration), LF(Lack of Fusion), CR(Crack)있다. 이 중에서 PO, SI 및 SD 가 전체 발생 결함의 90 % 이상을 차지한다. 결함의 종류별 데이터 양의 편차가 있으므로 본 연구에서는 결함의 종류를 고려하지 않고 데이터 셋을 준비하였다. 또한 약 80 mm 씩 생성된 이미지를 제품 길이와 동일하도록 하나의 이미지로 이어 붙이는 과정에서 생긴 이미지 상의 세로줄을 제거할 수 없어 그대로 사용하였다.

3.3 AI 성능 평가

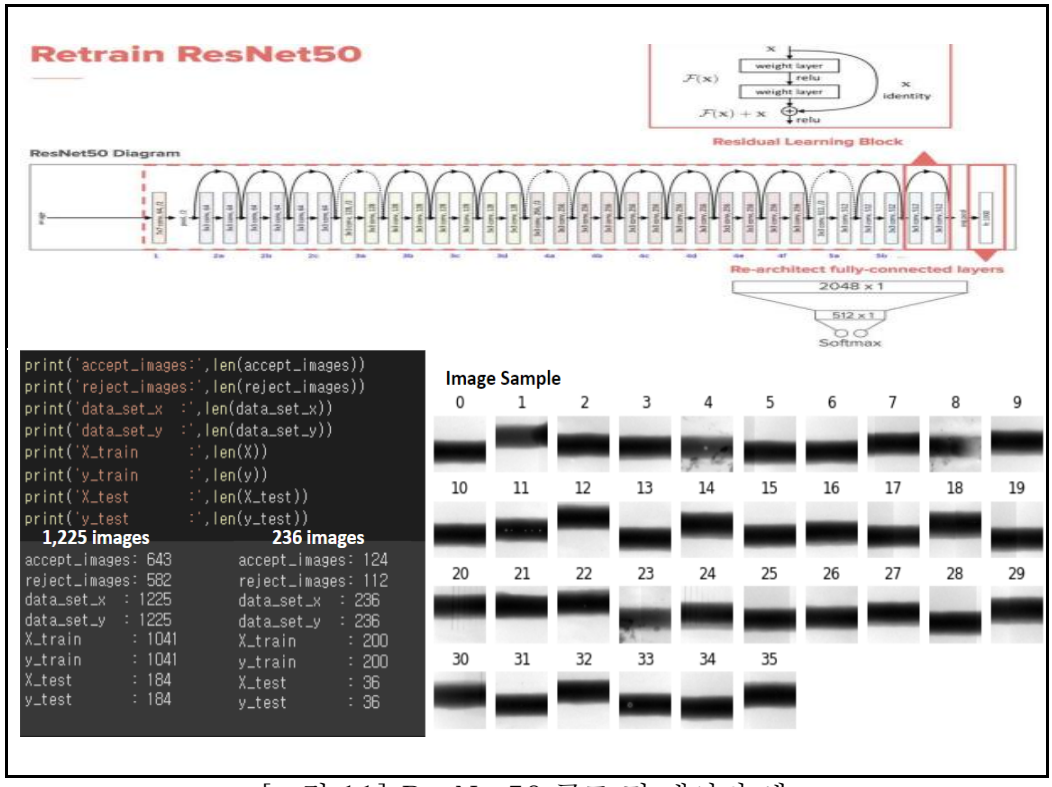
ResNet50 과 Orange Data Mining 의 Convolution Neural Network(CN2), Neural Network(NN), Gradient Boosting(GB) 및 Support Vector Machine(SVM) 모델에 대해, 학습 데이터 셋 236 및 1,225 images 로 각각 적용하여 학습 데이터 양에 따른 성능을 평가하고, 동일한 테스트 데이터 셋을 적용하여 예측 테스트를 진행하고 모델별 예측 성능 비교 평가를 진행하였다.

ResNet50 은 [그림 10]과 같은 조건으로 학습하였으며, 0.5 를 기준으로 초과를 부적합(1)으로 분류하고 이하를 합격(0)으로 분류하도록 하였다. [그림 11]에 ResNet50 의 구조와 학습에 사용된 데이터 셋 및 이미지 샘플을 나타내었다.

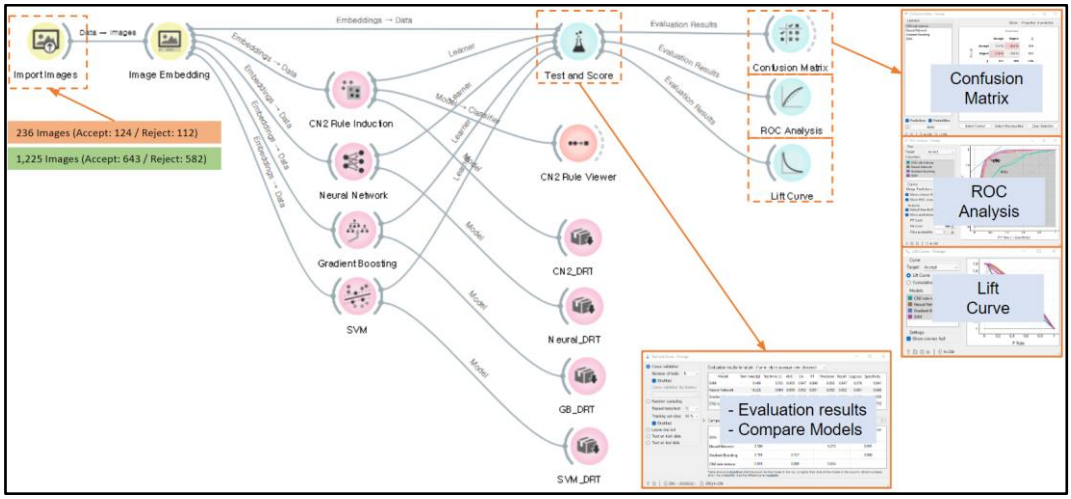
```
stf = StratifiedKFold(n_splits=5)
lr = ReduceLROnPlateau(monitor="val_loss", factor=0.9, patience=6, verbose=1)
es = EarlyStopping(monitor="val_loss", patience=32, verbose=1, mode="min", restore_best_weights=True)

opt = keras.optimizers.Adam(learning_rate=0.01)
model.compile(optimizer=opt, loss='binary_crossentropy', metrics=['accuracy'])

epochs=200, batch_size=32
```

[그림 10] ResNet 50 Hyper-parameter

ResNet50 과의 성능 비교 평가를 위해, [그림 12]와 같이 Orange Data Mining 네트워크를 구축하여 학습하였다. Orange Data Mining 의 경우, 학습 조건을 설정할 필요 없으며 ResNet50 과 동일하게 학습하기 위하여 Stratified-KFold (n_splits=5) 만을 적용하였다.



[그림 11] ResNet50 구조 및 데이터 셋



[그림 12] Orange Data Mining 학습 네트워크

학습 및 테스트는 HP Z4 G4 Workstation(Intel(R) Xeon(R) W-2104 CPU @ 3.20GHz, Ram 32.0 GB)로 수행하였다. 단, ResNet50 은 Jupyter Notebook 에 GPU(RTX3060)를 연결하였으며, Orange Data Mining 은 GPU 지원 기능이 없어 적용되지 않았다.

성능은 Classification Accuracy (CA), Confusion Matrix, Receiver Operating Characteristic(ROC) Curve 등으로 비교 평가하였다.

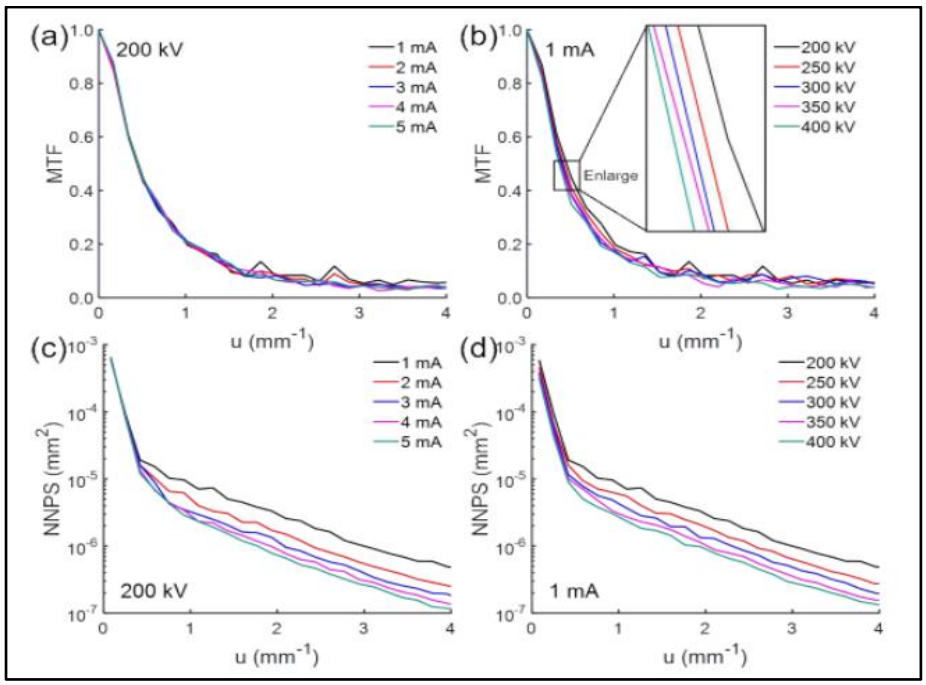
제 4 장 결과 및 고찰

4.1 데이터 품질 평가

ASME 에 따른 IQI 평가 결과, 요구사항에 따른 이미지 데이터 품질 수준인 기하학적 불선명도(Geometric Un-sharpness, Ug) 0.50 mm 및 선형과 구형 IQI 식별 기준에 만족함을 확인할 수 있었다.

수평(Horizontal) 방향의 MTF 결과를 [그림 13] (a)와 (b)에 나타내었다. (a)는 mA 증가에 따른 MTF 결과로 mA 가 증가할수록 오차가 줄어든 것을 확인할 수 있다. (b)는 kV 증가에 따른 MTF 결과로 kV 가 증가할수록 전반적인 MTF 가 저하되고 있는데 u = ~5 mm^{-1} 부근에서 확대한 결과를 보면 더욱 두드러지게 관찰된다. 이는 엑스선의 에너지가 증가할수록 물질과의 반응 중 산란선에 의한 반응이 지배적이기 때문에 그 산란선이 엑스선 영상의 해상도를 저하시키는 것으로 판단된다.

방사(Radial) 방향의 NPS 결과를 [그림 13] (c)와 (d)에 나타내었다. mA 및 kV 가 증가할수록 노이즈가 줄어드는 것을 확인할 수 있다. 이는 X-ray 튜브의 전류(mA) 또는 전압(kV)의 증가는 X-ray 의 포톤(Photon) 수의 증가이기 때문에 발생되는 랜덤 노이즈를 줄이기 때문이라 판단된다.

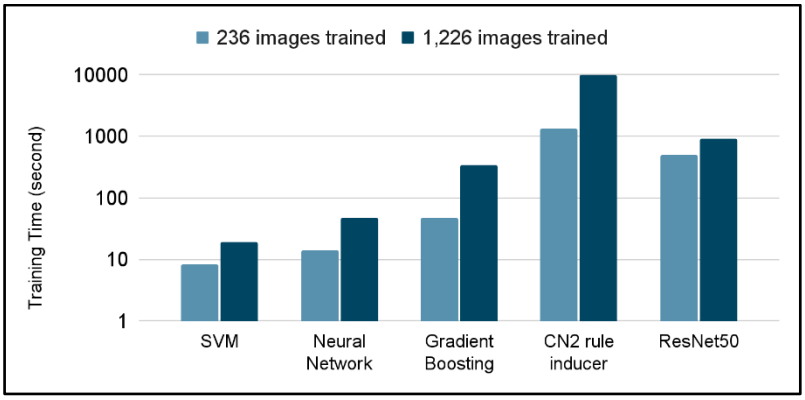


[그림 13] 고에너지 X-ray 에서의 MTF 및 NPS 특성 (a) mA 변화에 따른 MTF (b) kV 변화에 따른 MTF (c) mA 에 따른 NPS (d) kV 에 따른 NPS

4.2 AI 성능 평가

1) 모델별 학습 시간 비교

학습 데이터 양에 따른 학습 시간 비교를 위해 학습 데이터 셋을 2 가지 (236 images 및 1,226 images)로 적용하여 각 모델별 학습 시간을 [그림 14]에 나타내었다. 학습 데이터 1,226 images 에 대한 Orange Data Mining 의 CN2 모델의 학습 시간은 10,075 초로 GPU 가 지원되고 Early Stopping 이 적용된 ResNet50 의 학습 시간 910 초 대비 약 11 배의 시간이 소요되었다.



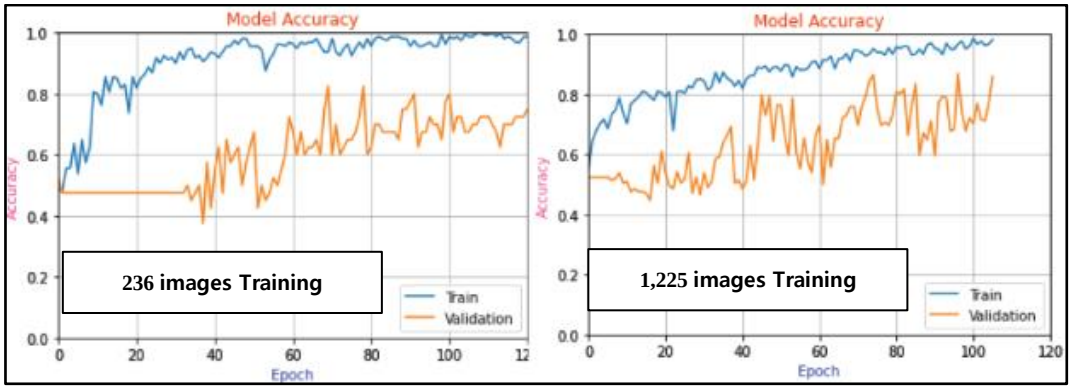
[그림 14] 학습 시간(초)

학습 데이터 셋의 양이 약 3,000 images 가 넘지 않는다면, 코드리스 도구인 Orange Data Mining 의 Neural Network 모델은 시간적으로 적용가능 할 것으로 판단된다.

2) 모델 학습 성능 평가

학습 데이터 양을 236 및 1,225 images 로 다르게 적용하여, ResNet50 의 학습 과정에서의 Epoch 에 따른 정확도(Accuracy)를 [그림 15]에 나타내었다. 학습 성능은 대략 100~120 epoch 에서 더 이상 증가하지 않았으며, 학습 중 epoch 에 따른 정확도가 변화가 급격함이 확인되었다.

학습 데이터 양에 따른 각 모델별 학습 성능(분류 정확도; Classification Accuracy, CA)을 [표 2]에 나타내었다. 데이터의 양이 약 5.2 배가 증가하였으나, ResNet50 은 4.6 %P 증가에 그쳤으며 Orange Data Mining 모델들은 반대로 약 6~24 %p 감소됨이 확인되었다.



[그림 15] 학습 과정 중 Epoch 에 따른 정확도

[표 2] 학습 데이터 양에 따른 모델별 학습 성능

Number of Training Data [Note 1)]	ResNet50	CN2	Neural Network	Gradient Boosting	SVM
236 images	80.0%	**71.6%**	**85.2%**	**84.3%**	**84.7%**
1,225 images	**84.6%**	58.4%	79.6%	76.2%	60.5%
Δ Accuracy	**4.6%p**	**-13.2%p**	**-5.6%p**	**-8.1%p**	**-24.2%p**

Note 1) Composition of training data: Accept 53% and Reject 47%

Orange Data Mining 의 경우, 여러 번 반복 학습을 수행하였지만 적은 양으로 학습하였을 때 성능이 더 좋게 나타났다.

Orange Data Mining 은 샘플링 방법만 설정 가능하며 성능 향상을 위한 Hyper-parameters 를 조절을 할 수 없기 때문에 데이터 양의 증가에 따른 성능 저하 원인은 확인할 수 없지만, 본 연구에 적용된 이미지 데이터 양의 증가에 따라, 학습에 적용된 이미지 데이터와 전혀 다른 유형의 결함과 개별 이미지를 하나의 이미지로 이어 붙이는 과정에서 생긴 세로줄 및 결함이 있으나 적용 요구사항에 만족되어 합격으로 판정되는 지시(Indication) 가 포함된 이미지 데이터가 증가되었기 때문이라 판단된다.

따라서, 학습 데이터 양이 증가에 따라 성능 향상을 기대하기 위해서는 학습 데이터의 전처리가 중요하다고 판단된다. 즉, 본 연구로 확인된 적정한 방사선 조사 조건으로 기본적인 원본 이미지 데이터의 품질을 보장하고, Multi-scale Contrast Enhancing 등과 같은 이미지 데이터 품질 향상 처리 알고리즘이 적용이 필요하며, 판정과 무관한 지시 등에 대한 처리가 필요하다 판단된다. 또한, 지속적인 성능 향상을 위해 전이학습 및 앙상블 모델 적용과 다양한 이미지의 지속적인 재학습이 가능한 네트워크 구축도 필요하다고 판단된다.

3) 모델별 예측 성능 평가

별도로 준비된 예측 성능 평가용 테스트 셋을 적용하여 모델별 예측 성능을 [표 3]에 나타내었다. 여기서, 테스트 셋의 구성(Composition of Test Data Set) 은 학습시 적용된 합격과 부적합 분율과 동일한 70 images 및 실제 제품의 합격과 부적합 분율과 동일한 2,369 images 를 의미한다. 모델별 성능 비교 평가를 위한 성능 지표, Confusion Matrix, ROC Curve 및 AUC 를 [표 4], [그림 16], [그림 17] 및 [표 5]에 각각 나타내었다.

[표 3]모델별 예측 정확도

Composition of Test Data Set	Model Note 1)	ResNet50	CN2	Neural Network	Gradient Boosting	SVM
70 images Accept 53% Reject 47%	236	**74.3%**	62.9%	**74.3%**	58.6%	**70.0%**
	1225	64.3%	**74.3%**	70.0%	**81.4%**	54.3%
2,369 images Accept 99% Reject 1%	236	**34.3%**	38.3%	37.7%	46.6%	38.6%
	1225	32.1%	**40.9%**	**88.5%**	**89.1%**	**93.1%**

Note 1) 236: 236 images trained Model, 1225: 1,225 images trained Model

학습 시의 성능과 비교하여 예측 성능은 전반적으로 낮았다. 특히, [표 3]과 같이 ResNet50 의 경우, 약 1 %의 실제 부적합률이 적용된 테스트 데이터 셋에 대한 예측 정확도가 Orange Data Mining 의 모델들과 비교하여 현저하게 낮았으며, 학습 성능과 다르게 236 images 로 학습된 모델의 예측 정확도가 1,225 images 로 학습된 모델의 예측 정확도 보다 높았다. ResNet50 과 같은 깊은 신경망을 가지는 Orange Data Mining 의 CN2 모델도 ResNet50 과 유사한 경향성이 확인되었다.

ResNet50 또는 CN2 와 같이 깊은 신경망으로 구성된 AI 모델의 예측 성능은 학습 데이터 양 및 데이터 전처리와 Hyper-parameter 의 적절한 조정에 영향을 받는다. 하지만 이와 같은 전처리 기술 및 Hyper-parameter 조정에 따른 성능 향상은 본 연구에서는 다루지 않았다.

합격(0)/부적합(1) 분류 학습의 경우, 학습을 위한 데이터 셋의 합격과 부적합 비율을 비슷하게 하는 것이 일반적이다. 하지만 합격과 부적합 비율이 비슷하게 구성된 학습 데이터 셋으로 학습하는 경우, [표 4] 및 [그림 16] 과 [그림 17] 및 [표 5] 와 같이 부적합을 잘 분류하는 모델 보다는 합격을 잘 분류하는 모델이 예측 성능이 높음을 알 수 있다. 이는 실제 부적합률(1%)로 합격과 부적합이 구성된 테스트 셋에 대한 예측 시, 합격 비율이 극단적으로 많기 때문에 전체를 합격으로 처리해도 99 %의 예측 성능으로 보일 수 있기 때문이라 생각된다.

이와 같은 부적합 비율에 따른 분류 학습에서의 과적합을 방지하기 위해서는 실제 예측할 데이터의 유형 및 합격과 부적합의 비율 등을 고려하여 학습 데이터 셋 구축해야 됨을 본 연구를 통하여 확인할 수 있었다.

[표 4] 모델별 예측 성능 지표

a) Prediction Performance for 256 images Trained Model
Test Data Set: 70 images (Accept 53%:Reject 47%)

[ResNet50]	precision	recall	f1-score	support
Accept	0.69	0.92	0.79	37
Reject	0.86	0.55	0.67	33
accuracy			0.74	70
macro avg	0.78	0.73	0.73	70
weighted avg	0.77	0.74	0.73	70

[CN2]	precision	recall	f1-score	support
Accept	0.92	0.32	0.48	37
Reject	0.56	0.97	0.71	33
accuracy			0.63	70
macro avg	0.74	0.65	0.60	70
weighted avg	0.75	0.63	0.59	70

[NN]	precision	recall	f1-score	support
Accept	0.83	0.65	0.73	37
Reject	0.68	0.85	0.76	33
accuracy			0.74	70
macro avg	0.76	0.75	0.74	70
weighted avg	0.76	0.74	0.74	70

[GB]	precision	recall	f1-score	support
Accept	0.75	0.32	0.45	37
Reject	0.54	0.88	0.67	33
accuracy			0.59	70
macro avg	0.64	0.60	0.56	70
weighted avg	0.65	0.59	0.55	70

[SVM]	precision	recall	f1-score	support
Accept	0.90	0.49	0.63	37
Reject	0.62	0.94	0.75	33
accuracy			0.70	70
macro avg	0.76	0.71	0.69	70
weighted avg	0.77	0.70	0.69	70

b) Prediction Performance for 256 images Trained Model
Test Data Set: 2,397 images (Accept 99%:Reject 1%)

[ResNet50]	precision	recall	f1-score	support
Accept	0.98	0.34	0.51	2364
Reject	0.01	0.55	0.02	33
accuracy			0.34	2397
macro avg	0.50	0.44	0.26	2397
weighted avg	0.97	0.34	0.50	2397

[CN2]	precision	recall	f1-score	support
Accept	1.00	0.37	0.54	2364
Reject	0.02	0.97	0.04	33
accuracy			0.38	2397
macro avg	0.51	0.67	0.29	2397
weighted avg	0.99	0.38	0.54	2397

[NN]	precision	recall	f1-score	support
Accept	0.99	0.37	0.54	2364
Reject	0.02	0.85	0.04	33
accuracy			0.38	2397
macro avg	0.51	0.61	0.29	2397
weighted avg	0.98	0.38	0.53	2397

[GB]	precision	recall	f1-score	support
Accept	1.00	0.46	0.63	2364
Reject	0.02	0.88	0.04	33
accuracy			0.47	2397
macro avg	0.51	0.67	0.34	2397
weighted avg	0.98	0.47	0.62	2397

[SVM]	precision	recall	f1-score	support
Accept	1.00	0.38	0.55	2364
Reject	0.02	0.94	0.04	33
accuracy			0.39	2397
macro avg	0.51	0.66	0.29	2397
weighted avg	0.98	0.39	0.54	2397

c) Prediction Performance for 1,225 images Trained Model
Test Data Set: 70 images (Accept 53%:Reject 47%)

[ResNet50]	precision	recall	f1-score	support
Accept	0.80	0.43	0.56	37
Reject	0.58	0.88	0.70	33
accuracy			0.64	70
macro avg	0.69	0.66	0.63	70
weighted avg	0.70	0.64	0.63	70

[CN2]	precision	recall	f1-score	support
Accept	0.83	0.65	0.73	37
Reject	0.68	0.85	0.76	33
accuracy			0.74	70
macro avg	0.76	0.75	0.74	70
weighted avg	0.76	0.74	0.74	70

[NN]	precision	recall	f1-score	support
Accept	0.71	0.73	0.72	37
Reject	0.69	0.67	0.68	33
accuracy			0.70	70
macro avg	0.70	0.70	0.70	70
weighted avg	0.70	0.70	0.70	70

[GB]	precision	recall	f1-score	support
Accept	0.80	0.86	0.83	37
Reject	0.83	0.76	0.79	33
accuracy			0.81	70
macro avg	0.82	0.81	0.81	70
weighted avg	0.82	0.81	0.81	70

[SVM]	precision	recall	f1-score	support
Accept	0.54	0.95	0.69	37
Reject	0.60	0.09	0.16	33
accuracy			0.54	70
macro avg	0.57	0.52	0.42	70
weighted avg	0.57	0.54	0.44	70

d) Prediction Performance for 1,225 images Trained Model
Test Data Set: 2,397 images (Accept 99%:Reject 1%)

[ResNet50]	precision	recall	f1-score	support
Accept	0.99	0.31	0.48	2364
Reject	0.02	0.88	0.03	33
accuracy			0.32	2397
macro avg	0.51	0.60	0.26	2397
weighted avg	0.98	0.32	0.47	2397

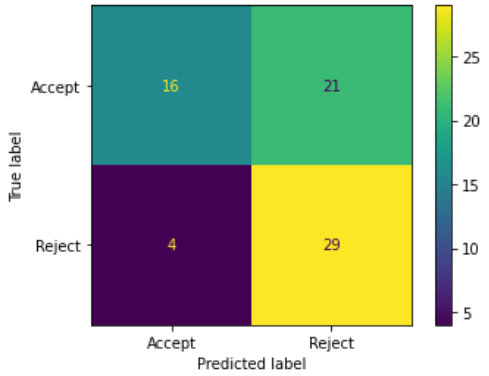
[CN2]	precision	recall	f1-score	support
Accept	0.99	0.40	0.57	2364
Reject	0.02	0.82	0.04	33
accuracy			0.41	2397
macro avg	0.51	0.61	0.31	2397
weighted avg	0.98	0.41	0.57	2397

[NN]	precision	recall	f1-score	support
Accept	0.99	0.89	0.94	2364
Reject	0.08	0.67	0.14	33
accuracy			0.88	2397
macro avg	0.54	0.78	0.54	2397
weighted avg	0.98	0.88	0.93	2397

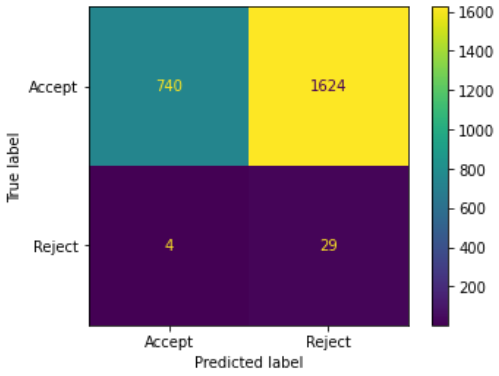
[GB]	precision	recall	f1-score	support
Accept	1.00	0.89	0.94	2364
Reject	0.09	0.76	0.16	33
accuracy			0.89	2397
macro avg	0.54	0.83	0.55	2397
weighted avg	0.98	0.89	0.93	2397

[SVM]	precision	recall	f1-score	support
Accept	0.99	0.94	0.96	2364
Reject	0.02	0.09	0.04	33
accuracy			0.93	2397
macro avg	0.50	0.52	0.50	2397
weighted avg	0.97	0.93	0.95	2397

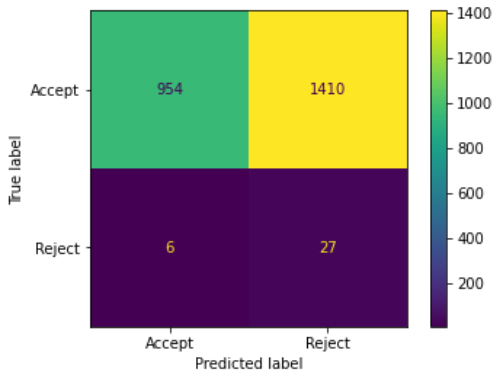
ResNet50_Test Data 70 images



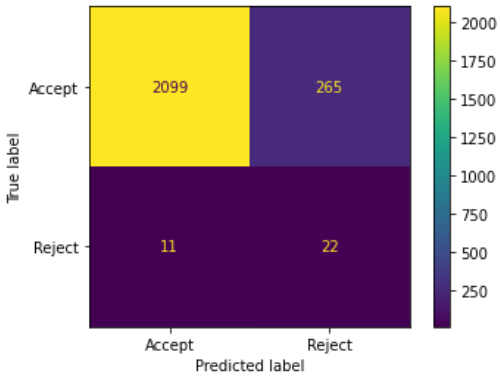
ResNet50 _Test Data 2369 images



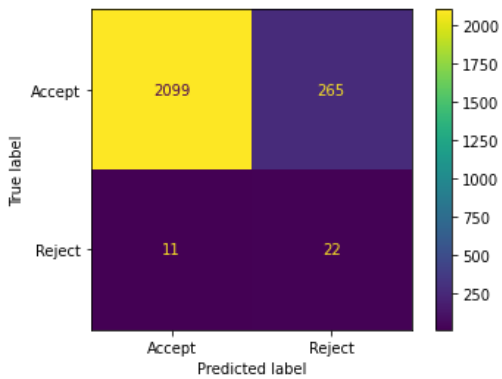
Convolution Neural Network



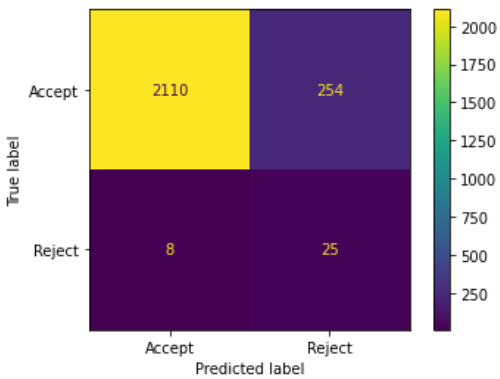
Neural Network



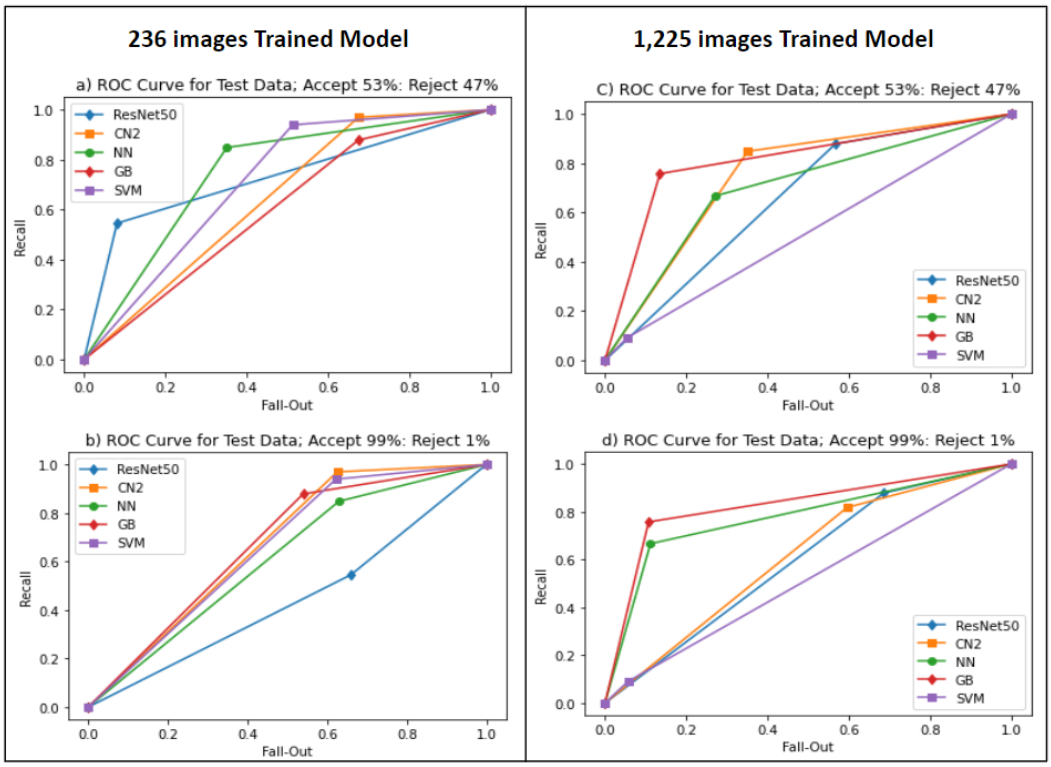
Gradient Boosting



Support Vector Machine



[그림 16] 모델별 예측 Confusion Matrix



[그림 17] 모델별 예측 ROC Curve

[표 5]모델별 예측 AUC

Composition of Test Data Set	Model Note 1)	ResNet50	CN2	Neural Network	Gradient Boosting	SVM
70 images Accept 53% Reject 47%	236	**0.73**	0.65	**0.75**	0.60	**0.71**
	1225	0.66	**0.75**	0.70	**0.81**	0.51
2,369 images Accept 99% Reject 1%	236	0.44	**0.67**	0.61	0.67	**0.66**
	1225	**0.60**	0.61	**0.78**	**0.83**	0.52

Note 1) 236: 236 images trained Model, 1225: 1,225 images trained Model

제 5 장 결론

5.1 연구 결과 요약

먼저, AI 학습을 위한 이미지 데이터의 품질이 적용 요구사항을 만족하고 더 나아가서는 이미지 데이터의 품질을 향상시켜 AI 성능을 높이기 위해, 디지털 방사선투과검사 이미지 데이터의 품질을 평가하였다. 평가 결과, 국제 산업표준에 따른 이미지 데이터 품질 수준에 만족됨을 확인할 수 있었으며, 이미지 데이터의 노이즈 영향을 최소화하기 위해서는 피검사체인 용접 강관의 두께를 투과할 수 있는 전압(kV)에서 전류(mA)를 최대로 하는 것이 이미지 데이터 품질 향상에 도움됨을 확인하였다. 이는 전류(mA)를 최대로 하는 조건으로 방사선을 조사하는 것이 이미지 데이터의 노이즈를 감소시키며, 전압(kV) 증가에 따른 분해능 저하를 최소화하기 때문이다.

다음으로, ResNet50 과 코드리스 도구인 Orange Data Mining 의 모델을 적용하여 용접 강관의 결함 자동 분류 모델을 학습하고 성능을 평가하였다. 학습 시간과 성능을 고려하면, ResNet50 이 Orange Data Mining 모델 보다 짧은 시간에 학습되고 성능도 우수하였으나, 별도로 준비된 동일한 테스트 셋에 대한 예측 성능은 대체적으로 낮았다. 이는 학습 데이터에 과적합되어 합격을 부적합으로 판정하기 때문이 었으며, 과적합 방지를 위해 추가적인 Hyper-parameter 조정이 필요한 것으로 판단된다.

학습 데이터의 양에 따른 모델별 성능은 큰 차이를 확인할 수 없었으며, 오히려 Orange Data Mining 모델은 학습 데이터 양이 증가한 경우 성능이 낮아졌다. 이는 양의 증가에 따른 데이터의 결함 종류 및 합격되는 지시들의 증가로 인한

것으로 판단된다. 그리고 합격과 부적합이 약 50 %씩 되도록 준비된 학습 데이터 셋으로 학습된 모델은 실제 부적합 비율(1 %)로 준비된 테스트 셋에 대한 예측 정확도가 전반적으로 낮았다.

성능이 우수한 AI 모델 및 코드리스 도구들은 지속적으로 개발되고 있지만, 가장 중요한 것은 데이터인 것이 본 연구에서 확인되었다. 디지털 방사선투과검사 AI 판정을 위해서는 이미지 데이터의 전처리 방법을 표준화하고 실제 부적합 비율을 고려하여 다양한 이미지 데이터 셋을 구축하고 지속적으로 재학습할 수 있는 네트워크 구축 및 AI 모델 적용이 필요하다 판단된다.

5.2 기대효과

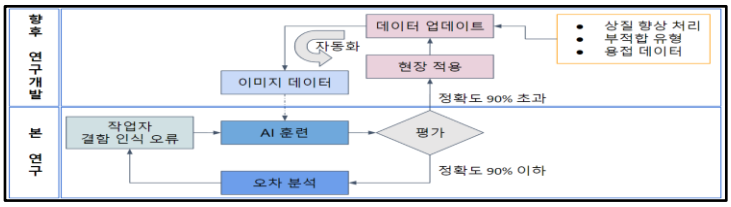
방사선투과검사는 용접이 수행되는 모든 산업에서 적용되는 주요한 검사 방법이다. 주로 용접을 수행하고 검사하는 국내의 많은 중소기업에서는 관련 빅 데이터를 보유하고 있지만 활용하지 못하는 경우가 많다. 본 연구는 이와 같은 중소기업들이 자신들이 소유한 데이터에 가치를 부여할 수 있는 계기가 될 것으로 기대한다.

또한 향후 추가적인 연구개발을 통하여 현장에 적용될 경우에는, 생산성 향상 및 방사선투과검사 결과 판정에서의 식별 오류로 인한 리스크를 예방할 수 있을 것으로 기대한다. 이와 같은 효과는 잠재적으로 국내 관련 제조업의 수출 경쟁력 향상에 도움이 될 것으로 기대한다.

5.3 향후 연구개발

본 연구에 사용된 방사선투과검사 이미지 데이터는 Multi-scale Contrast Enhancing(MCE) 등과 같은 이미지 데이터 품질 향상 처리 알고리즘이 적용되지 않고, 개별로 생성된 이미지를 하나의 이미지로 합치면서 생긴 세로줄을 제거하지 않은 원본 이미지를 적용하여 학습 및 테스트를 진행하였다. 그렇게 때문에 여러 지시가 있으나 적용 규격을 만족하여 합격으로 처리된 이미지 및 미세하여 육안으로 쉽게 보이지 않는 결함 등 다양한 경우의 정제되지 않은 이미지들이 혼재되어 데이터 셋이 준비되었다. 향후, 여러가지 이미지 데이터 품질 향상 처리 알고리즘 및 이미지 상의 세로줄을 제거한 이미지를 적용하고, 부적합으로 분류한 이미지 데이터에 대해 결함 유형을 추가하여 데이터 셋을 준비하고, 전이학습, 앙상블 모델 등 여러 모델의 성능 평가를 진행할 계획이며, [그림 18]과 같이 지속적인 재 학습 모델을 위한 네트워크 구축할 계획이다.

부적합 발생을 예측하고 부적합을 사전에 예방하기 위해서는 부적합의 원인과 결과 데이터의 분석이 필요하지만, 본 연구에서는 용접 공정의 부적합 결과 데이터인 방사선투과검사 이미지 데이터를 합격과 부적합으로 분류하는 인공지능 학습 및 적용 가능성을 연구하였다. 이를 바탕으로 향후에는 부적합 원인인 용접 데이터와 방사선투과검사 결과 간의 분석을 통하여, 용접 공정에서 부적합 발생을 예측하고 예방 조치할 수 있는 방법을 찾도록 연구개발을 진행할 계획이다.



[그림 18] 향후 연구개발 계획

참고 문헌

[1] Human Factors in Non-Destructive Testing (NDT): Risks and challenges of Mechanised NDT (Vorgelegt Von, 1 September 2015)

[2] Weld Defect Extraction and Classification in Radiographic Testing Based Artificial Neural Networks (Nacereddine Nafaa, October 2000)

[3] Assessment of welding defects for gas pipeline radiographs using computer vision (I.M. Elewa, December 2002)

[4] Deep Learning-Based Classification of Weld Surface Defects (Haixing Zhu, 12 August 2019)

[5] Review on Computer Aided Weld Defect Detection from Radiography Image (Wenhui Hou, 10 March 2020)

[6] Weld defect classification in radiographic images using unified deep neural network with multi-level features (Lu Yang, May 2020)

[7] Automated defect detection in digital radiography of aerospace welds using deep learning (Topias Tyystjarvi, 23 February 2022)

[8] Deep Learning Technology for Weld Defects Classification Based on Transfer Learning and Activation Features (Chiraz Ajmi, 14 August 2020)

[9] Review on the Recent Welding Research with Application of CNN-Based Deep Learning Part II: Model Evaluation and Visualizations (Kidong Lee, January 18, 2021)

[10] 결함 분류를 위한 CNN 분석 (오준택, 2021.07)

[11] ResNet50 을 이용한 냉연강판에서의 표면결함 자동분류 (최원준, 2021)

[12] Understanding X-ray cargo image: Nuclear Instruments and Methods in Physics Research Section B, Beam Interactions with Materials and Atoms, Vol. 241, No.1-4, pp.810-815 (G.Chen, 2005)

[13] An International Code ASME Boiler & Pressure Vessel Code (July 1, 2017)

[14] ASTM E2002, Standard Practice for Determining Total Image Un-sharpness in Radiology (Reapproved 2009)

[15] 대형 강관 엑스선 검사 시스템의 신호 및 노이즈 분석 (김준우, 김균엽, November 8, 2021)

[16] CMOS 영상센서 기반 디지털 X-선 영상시스템의 구현 및 영상특성 평가에 관한 연구 (정만희, June 2004)

Research on data quality evaluation and AI algorithm of Digital Radiographic Testing to prevent defect identification missing as human error

Gyun-Yeop Kim
Department of Industrial Management Engineering
Graduate School of Industry, Ulsan University
(Advisor: Professor Lee su-dong)

(Abstract)

Radiography is a major inspection method that verifies the reliability of steel product to which welding is applied. A qualified NDE inspector evaluates the NDE results (film or digital image) to detect and classify weld defects. Since evaluation errors are directly related to financial and human accidents, sampling verification procedures are being carried out by the other inspector to prevent these errors, but some nonconformity welded pipes which welding defects are not identified and repaired, have been delivered to customers or installed/used on site.

such as sampling re-evaluation by other inspector are being carried out to prevent the risk of evaluation errors by NDE inspector.

To prevent such risks and losses, AI researching for defect identification and classification is being actively conducted. However, research so far has conducted AI learning and performance evaluation by a small number of shared radiography image databases such as GDX-ray or digitizing images taken from film.

In this study, the quality of image data and the performance of the AI model was evaluated using various digital radiography images of actual products manufactured, inspected, and delivered according to applicable industry standards and customer requirements. Further R&D based on this study, the commercial AI model to support the evaluation of digital radiography testing results will be developed.

감사의 글

금속 및 용접 지식을 기반으로 하여 중소기업의 품질 및 생산기술 업무에서 16 년을 일하면서, 애정과 열정으로 지속적인 품질 및 생산성 개선을 위해 노력하였습니다. 하지만 어느 시점에서 더 이상 개선되지 않음을 느꼈으며, 개선을 위해 취해진 조치가 프로세스화되어 시스템으로 구축되기 않고 일회성으로 추진되거나, 전사적인 동참 및 협력 없이 개인적 개선 수준에서 그치는 경우가 많아 한계와 회의를 느꼈습니다.

이런 시기에 배움이 필요함을 느껴 산업대학원에 진학하였습니다. 4 차 산업혁명 기술이 확산되고 에너지 패러다임이 바뀌는 시점에서 많은 것을 생각할 수 있는 계기가 되었습니다.

새로운 지식과 변화하는 세상을 볼 수 있게 도와주신 모든 교수님들께 감사드립니다. 특히, 논문 쓰고 싶다는 부탁에 적극적으로 지원해주신 박주철 대학원장님께 감사드리며, 현업의 경험을 바탕으로 공감과 친절히 지도 해주신 이수동 교수님께 감사드립니다.

저와 애증의 관계인 ㈜현대알비와 동료에게 감사드리며, 특히, 저에게 일 할 수 있는 기회와 묵묵히 지원해주시는 이수창 회장님과 이상철 사장님께 감사드립니다.

마지막으로, 자기계발에 적극적으로 지원해 준 저의 사랑하는 아내가 있어 석사 졸업장을 가지게 되었습니다. 아내에게 졸업장을 바칩니다.

본 논문은 중소벤처기업부의 “SAW 용접 강관의 방사선투과검사 판독 및 식별 오류 방지를 위한 AI 영상 판독 및 식별 시스템; S2970554” 과제의 연구비로 수행된 연구결과임을 밝히며, 연구비 지원에 감사드립니다.