

연구논문

머신러닝을 이용한 이미지로그의 압축파쇄대 깊이 검출모델 개발

염주열¹ · 김하영¹ · 이경북^{1*} · 장찬동² · 조영욱³¹공주대학교 지질환경과학과, ²충남대학교 지질환경과학과, ³한국지질자원연구원 국토우주지질연구본부 심층처분환경연구센터

Detection of Borehole Breakout Depth in Image Logs Using Machine Learning Algorithms

Juyeol Yeom¹, Hayoung Kim¹, Kyungbook Lee^{1*}, Chandong Chang² and Yeonguk Jo³¹Department of Geoenvironmental Sciences, Kongju National University, Kongju, Korea²Department of Geological Sciences, Chungnam National University, Daejeon, Korea³Deep Subsurface Storage and Disposal Research Center, Geology & Space Research Division, Korea Institute of Geoscience and Mineral Resources, Daejeon, Korea

*Corresponding Author. Kyungbook Lee, kblee@kongju.ac.kr, Department of Geoenvironmental Sciences, Kongju National University, Kongju, Korea

Received

7 August 2023

Final version Received

21 August 2023

Accepted

28 August 2023

Abstract

This study developed a classification model for the detection of borehole breakout depths in image logs using machine learning algorithms. Approximately 1 km of image log data acquired in Gyeongju was preprocessed to make it suitable for applying machine learning algorithms. This included removing missing values and data within casing intervals and aligning the resolution among image log data — that is, vertical and horizontal resolutions were set at 0.01 m and 2.5°, respectively. Decision tree-based algorithms, random forest, and XGBoost were then applied to the preprocessed 99,090 data entries, which was divided into training (80%) and testing (20%) groups. The trained models were evaluated based on the false negative (FN) of the confusion matrix. The XGBoost model showed a superior predictive performance with 87 FNs when compared to the random forest model's 162 FNs. The developed model can enable experts to perform detailed reviews of detected depths via the model.

Key words : breakout, image log, machine learning, confusion matrix

요약

본 연구는 머신러닝을 이용해 이미지로그의 압축파쇄대 깊이를 검출하는 분류모델을 개발하여 전문가 분석효율을 높이는 것이 목적이다. 경주지역 내 시추공에서 취득된 약 1 km 이미지로그에 대해 케이싱 구간 및 결측치를 제거하고 이미지로그 자료간 해상도(수직 및 수평 해상도가 각각 0.01 m와 2.5°)를 일치시키는 전처리과정을 수행하였다. 이를 통해 얻어진 총 99,090개 자료를 8:2 비율로 학습과 테스트 자료로 설정한 후, 의사결정나무 기반의 랜덤포레스트와 XGBoost를 적용하였다. 모델평가는 테스트자료에 대한 혼동행렬의 false negative(FN)를 통해 이뤄졌으며, XGBoost 모델은 FN이 87개로 랜덤포레스트 모델의 FN 162개에 비해 향상된 예측성능을 보였다. 이렇게 개발된 모델은 전문가의 압축파쇄대 상세 검출을 도와 분석 효율성을 제고할 수 있을 것으로 기대된다.

주요어 : 압축파쇄대, 이미지로그, 머신러닝, 혼동행렬

서 론

지각에 작용하는 응력의 특성(크기, 방향, 이방성 등)은 지하공간개발의 안정성과 지진 및 단층활동의 기작을 이해하는 데 중요한 요소이다(Heidbach *et al.*, 2018). 최근 1 km 미만의 천부지각을 이산화탄소 지중저장, 고준위 핵폐기물 처리시설 등 다양한 목적으로 개발하고 있어 지하공간의 응력규명은 안정성과 직결된다.

천부지각의 응력장은 수압파쇄법, 오버코어링법, 단층조선 역산법, 시추공 응력지시자 활용법과 같은 다양한 지질공학적 방법으로 측정된다. 이들 중 시추공 응력지시자 활용법은 시추공벽에서 관찰될 수 있는 압축 또는 인장파쇄를 통해 비교적 빠르고 연속적으로 응력장 정보를 획득할 수 있어 많이 활용되고 있다(Ekstrom *et al.*, 1986; Lagraba *et al.*, 2010; Zoback, 2010).

시추공벽에서 나타나는 응력지시자(stress indicator)는 인장파쇄대(drilling-induced tensile fractures)와 압축파쇄대(borehole breakout)가 있다. 이러한 파쇄대는 시추 과정에서 시추공 주변 응력이 이완되거나 집중되는 곳에서 발생하며, 현장 응력의 방향과 크기에 대한 정보를 얻을 수 있다(Zoback *et al.*, 1985; Barton *et al.*, 1988; Shamir and Zoback, 1992; Barton *et al.*, 1997). 이 중 압축파쇄대는 최대수평주응력이 작용하는 방향과 직교하는 방향에서 응력이 집중되어 그 지점의 응력 크기가 암석의 강도보다 클 때 발생한다. 압축파쇄대는 초음파 영상검층을 통해 취득되는 진폭(amplitude)과 주시(travel time) 자료를 분석하여 발생여부를 판단할 수 있다. 주응력 특성상 180° 이격을 두고 압축파쇄대가 발생하며, 이때 진폭은 작아지고 주시는 크게 측정된다. 따라서 검층자료를 통해 깊이별 최대수평주응력 방향을 유추할 수 있다.

일반적으로 압축파쇄대는 이미징로그에 대해 전문가가 수작업을 통해 판별하게 된다. 이러한 작업은 전문가의 숙련도에 따라 분석결과가 달라지며 주관에 배제하기 어렵다. 또한 주어진 시간과 비용에 따라 분석 정밀도가 달라질 수 있으며, 자료의 수가 많은 고밀도 시추 또는 고심도 자료의 경우 분석에 많은 시간이 소요된다는 문제점이 존재한다. 이러한 문제점을 개선하고자 최근에는 다양한 머신러닝 기법을 이용한 연구가 제안되었다.

Dias *et al.*(2020)은 수작업으로 진행되는 이미징로그 분석 작업의 효율성 제고를 위해 딥러닝기반 Fast-RCNN(fast-region convolution neural network) 기법을 이용한 모델을 개발하였다. 해당 모델은 이미징로그의 압축파쇄대와 균열(fracture)을 자동으로 검출한다. Yang *et al.*(2022)은 기존 압축파쇄대 검출방법이 수작업으로 진행되어 효율이 낮고 큰 비용이 발생한다는 점을 개선하기 위해 U-Net

기법을 활용한 압축파쇄대 검출모델을 개발하였다. 특히, ATV(acoustic televiewer) 자료분석에 앞서 잡음제거방법을 고찰하였다. 두 연구 모두 다수의 시추공자료를 활용할 수 있어 딥러닝모델 학습이 용이했으며, 이미징로그의 픽셀단위 압축파쇄대 예측모델을 개발한 공통점이 있다. 하지만 가용한 자료가 적은 경우 딥러닝모델 적용이 제한되며, 이 경우 픽셀단위의 상세분석모델(2차원) 보다 깊이단위의 분석모델(1차원)로 문제를 단순화할 필요가 있다.

본 연구에서는 이미징로그의 압축파쇄대가 발생한 깊이를 검출하는 분류모델을 개발하고자 한다. 모델의 신뢰도를 향상시키기 위해 도메인 지식을 활용한 자료 전처리 과정 및 머신러닝 기법 선정 과정을 제시하고자 한다. 이를 바탕으로 개발된 압축파쇄대 검출모델의 결과와 실제 자료간의 일치 여부를 확인하여 모델의 정밀도를 검증하고자 한다. 최종 개발된 모델은 전문가가 선별된 깊이를 중심으로 압축파쇄대 상세 검출을 수행하여 분석 효율성을 제고할 수 있을 것으로 기대된다.

연구방법

연구자료

본 연구에서 사용된 자료는 한국지질자원연구원에서 수행중인 ‘한반도 동남권 심부 복합지구물리 모니터링 시스템 구축’ 사업 중 두 관측소(A1, A2) 부지 내에 심부 지질 특성화를 위해 굴착된 조사 시추공(test borehole)에서 획득한 초음파 영상검층자료이다. 해당 자료는 공벽에 초음파 펄스를 방출한 후 되돌아오는 파형을 기록하는 작업을 모든 깊이에 걸쳐 수행하였으며, 취득된 자료는 시추공 내 깊이와 방위에 따른 진폭으로 구성되어 있다.

A1과 A2공은 경상북도 경주시 일대에 위치하는 시추공으로 전문가 분석결과 A1공은 모든 깊이에서 압축파쇄대가 관찰되지 않는 것이 확인되었다. A2공은 주로 840 m보다 깊은 곳에서 압축파쇄대가 관찰되며, 특히 940 m 이하부터 본격적으로 압축파쇄대가 관찰된다(Fig. 1).

A1과 A2공의 자료는 측정 장비의 해상도 차이에 따라 깊이와 방위 해상도가 상이하다(Table 1). A1공은 1.67~1,009 m 깊이에서 0.01 m 간격으로 자료가 취득되었고, A2공은 8.280~1,003.768 m 깊이에서 0.002 m 간격으로 자료가 취득되었다. 또한 A1공은 1.25° 방위 간격으로 진폭이 취득되어 열이 총 288개 이고, A2공은 2.5° 간격으로 진폭이 취득되어 144개의 열을 갖는다.

기초통계분석을 위해 히스토그램을 활용하여 두 시추공 영상검층자료를 분석하였다(Fig. 2). A1공은 최소 -2,938.24, 최대 2,615 그리고 평균 1,861.49의 진폭을 갖는다. A2공은 최소 -999, 최대 2,912 그리고 평균 1,842.42의 진폭을

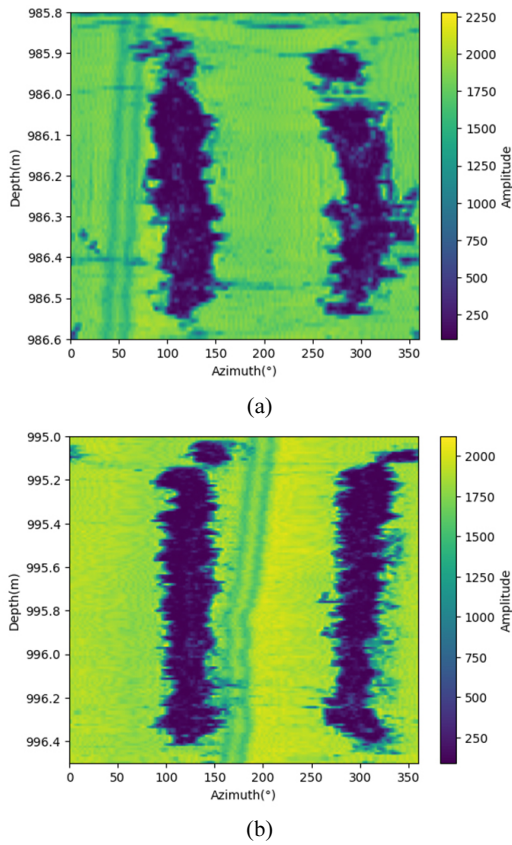


Fig. 1. Typical examples of breakout in well A2: (a) 985.8~986.6 m and (b) 995~996.5 m.

Table 1. Statistical analysis of well A1 and well A2

Parameters	Well A1	Well A2
Resolution	100,734×288	497,745×144
Depth	Data point	100,734
	Interval (m)	0.01
Angle	Start/End (m)	1.67/1009
	Data point	288
Amplitude	Interval (°)	1.25
	Min/Max	-2938.24/2615
	Average	1861.49

갖는다. 이때 음수로 나타나는 진폭은 장비에서 측정되지 않은 결측치(non-value)이다. A1공은 결측치가 나타난 자료수가 39개, A2공은 37개로 분석되었다. 해당 자료의 진폭은 모든 방위에 대해 측정되지 않아 각 시추공은 11,232개와 5,328개의 결측치를 갖는다.

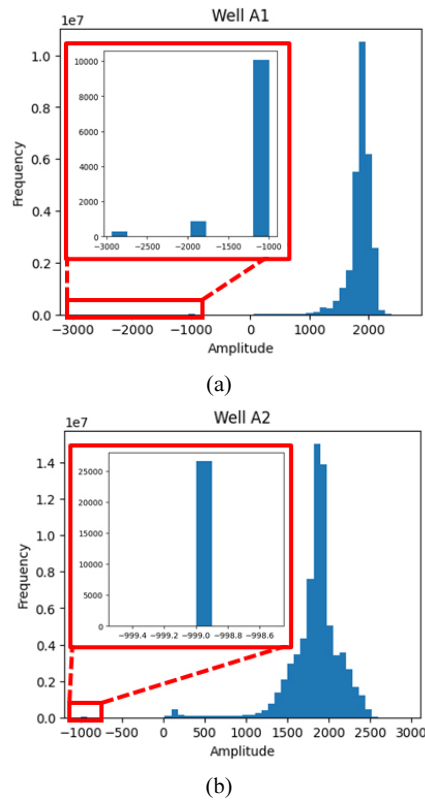


Fig. 2. Histograms of amplitude: (a) well A1 and (b) well A2.

자료 전처리

자료 전처리는 머신러닝 기법을 효과적으로 적용하기 위해 원본자료를 가공하는 작업이며, 모델 예측결과 성능을 결정하는데 매우 중요하다. 본 연구에서는 도메인 지식을 활용해 아래와 같이 총 4가지의 자료 전처리를 수행하였다.

1. 방위 해상도 일치: A1과 A2공은 해상도가 상이하어 모델 입출력 자료 구성을 위해 동일한 해상도로 재구성이 필요하다. 두 시추공의 방위 간격을 일치시키기 위해 두 자료를 A2공의 방위 간격과 동일한 2.5° 간격을 갖도록 재구성하였다. 이를 위해 A1공의 홀수열을 삭제하여 144개 열을 갖도록 구성하였다. 해당 과정을 통해 각 시추공은 100,734×144와 497,745×144의 해상도를 갖는다.

2. 깊이 해상도 일치: A1과 A2공의 깊이 간격이 같도록 자료를 재구성하였다. A1과 A2공의 깊이 간격은 각각 0.01 m, 0.002 m이다. A2공의 중복된 깊이 정보를 제거하여 0.01 m의 깊이 간격을 갖도록 구성하였다. 해당 과정을 통해 A2공은 99,512×144의 해상도를 갖는다.

3. 결측치 포함 자료 제거: 자료상에 음수 값인 결측치는 실제로 측정된 값이 아니며, 모델학습에 혼동을 발생시킬

수 있는 깊이므로 제거가 필요하다. 결측치가 나타난 자료 수는 A1공의 경우 39개, A2공은 37개가 존재한다. 이러한 결측치를 포함하는 자료를 모두 제거하여 각 시추공은 100,695×144와 99,512×144의 해상도를 갖는다.

4. 케이싱(casing) 상부 구간 깊이 제거: 두 시추공은 케이싱과 나공(open-hole) 구간이 포함되어 있다. 케이싱 끝단 위치를 기준으로 진폭이 달라지는 것을 확인할 수 있으며 (Fig. 3), 두 구간의 진폭 분포가 상이하여 모두 학습자료에 사용하는 경우 부적절한 학습이 진행될 수 있다. A1공은 9.10 m 상부 구간을, A2공은 12.50 m 상부 구간의 자료를 제거하였다.

앞선 자료 전처리과정을 통해 A1공은 99,952×144, A2

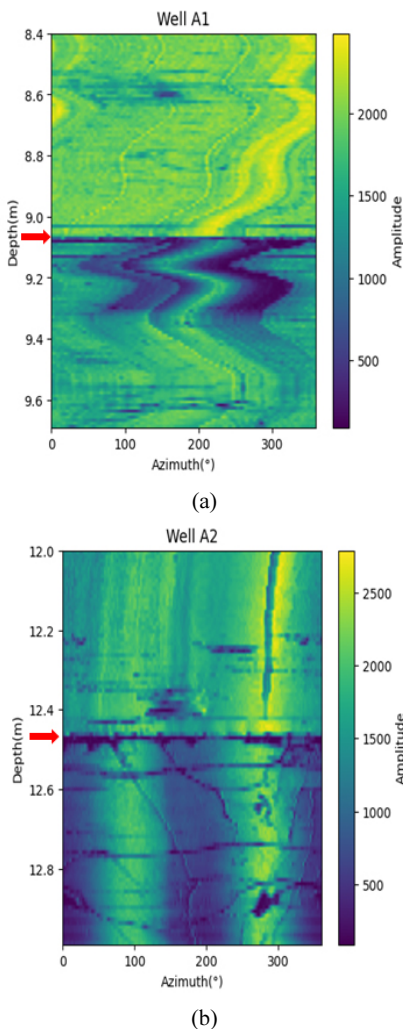


Fig. 3. Nearby boundary of casing and open-hole change in amplitude: (a) well A1 and (b) well A2.

공은 99,090×144의 해상도를 갖는 자료로 재구성된다.

머신러닝 기법 선정

본 연구에서는 압축파쇄대 깊이 검출모델 개발을 위해 의사결정나무(decision tree)기반 머신러닝 기법인 랜덤포레스트(random forest)와 XGBoost(extreme gradient boosting)를 선정하여 이용하였다. 의사결정나무는 신경망기반 딥러닝 모델에 비해 모델구조가 비교적 단순하여 학습자료가 제한되는 경우에도 적용과 성능이 뛰어나다. 또한 분류와 회귀에 모두 적용할 수 있는 장점을 갖고 있다. 하지만 나무의 단계(depth)가 많아질수록 최종 노드의 수가 증가하여 과적합(overfitting)이 발생할 수 있으며, 새로운 자료에 대한 예측정밀도가 낮다는 단점이 있다(Woo et al., 2018).

이를 개선하기 위해 Breiman(2001)은 여러 개의 의사결정나무로 구성된 앙상블 기반의 랜덤포레스트를 제안하였다(Fig. 4a). 랜덤포레스트는 의사결정나무와 달리 설명변수를 무작위로 선택한 후 최적 결과를 도출하는 방법을 사용한다(Kwon, 2013). 이때 의사결정나무는 원본 학습자료로부터 복원추출을 통해 학습자료를 새롭게 구성하는 기법인 부트스트랩(bootstrap)을 사용한다. 이를 통해, 서로 다른 자료로 구성된 의사결정나무는 독립적인 특성을 갖게 되어 과적합을 최소화할 수 있으며, 검증자료를 별도로 구성하지 않아도 되는 장점이 있다.

XGBoost는 각종 데이터 분석 대회 우승팀이 사용했던 알고리즘으로 효율과 성능이 입증되었다. XGBoost는 랜

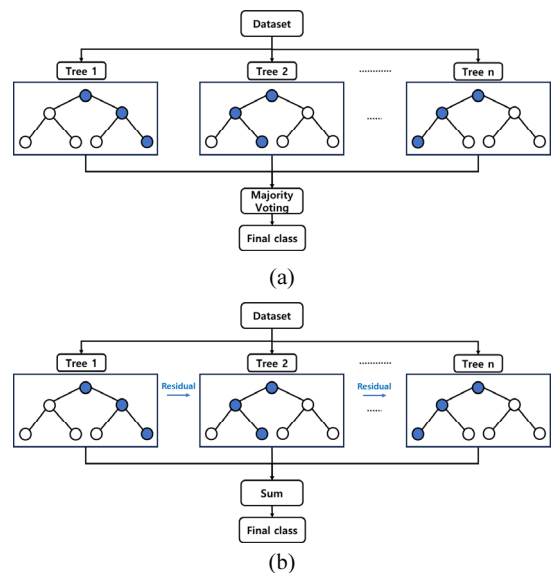


Fig. 4. Concept of tree-based machine learning algorithms: (a) random forest and (b) XGBoost.

덤프레스트와 달리 이전 의사결정나무의 오류를 순차적으로 보완해 가는 잔차(residual) 과정이 큰 특징이다(Fig. 4b). XGBoost는 먼저 약한 예측 모델을 구성한 후 학습자료와 일치도(similarity score)를 계산한다(Lee *et al.*, 2019). 이후 경사하강법을 이용하여 일치도가 높아지는 방향으로 새로운 의사결정나무를 구성하여 실제값과 예측값의 차이인 잔차를 계산한다. 계산된 잔차를 다음 예측모델에 활용하여 최소의 잔차를 갖는 최종 예측모델을 구성한다. 이러한 과정을 통해 학습 시 오류를 최소화하며, 과적합을 방지할 수 있게 된다(Hwang *et al.*, 2018).

연구결과

모델학습

모델을 학습하기에 앞서, 모델의 입력력 자료를 구성하는 과정을 수행하였다. 랜덤포레스트와 XGBoost는 동일한 2차원 배열의 입력자료를 사용할 수 있어 사용된 자료 구성은 동일하다. A1공은 압축파쇄대가 관측되지 않아, 전처리된 A2공 자료를 사용하여 모델의 입력력 자료를 구성하였다. A2공의 특정 깊이의 144개 방위의 진폭을 하나의 입력자료(1, 144)로 구성하였으며, 동일 깊이의 압축파쇄대 여부를 출력자료(1, 1)로 구성하였다. 입력력자료는 각각 (99,090, 144), (99,090, 1)의 배열을 가지며, 두 모델은 깊이별 144개의 진폭을 이용하여 압축파쇄대 여부(0 또는 1)를 예측하게 된다.

구성된 입력력자료는 학습자료와 테스트자료로 분리되었다. 전체 자료 99,090개 중 학습자료는 80%인 79,272개, 테스트자료는 20%인 19,818개로 구성하였다. 전체 자료 중 압축파쇄대가 확인된 깊이는 총 1,896개로 1.9%에 해당한다. 이러한 라벨자료의 편향은 모델학습에도 영향을 미칠 뿐만 아니라 성능지표 선정에도 주의가 필요하다.

랜덤포레스트를 이용한 분류모델은 사이킷런(scikit-learn)의 RandomForestClassifier를 이용해 구현하였다. 랜덤포레스트의 경우 의사결정나무의 수가 많을수록 과적합 방지에 유리하므로 기본값인 10개보다 큰 1,000개를 사용하도록 설정하였다. 또한 모델 학습과정에서 같은 결과를 도출하기 위해 random_state값을 고정하여 사용하였으며, 병렬 학습을 통해 학습시간을 단축했다. 그 외 나무의 깊이와 나무 및 잎의 노드에 사용되는 자료수 등과 같은 하이퍼파라미터는 기본값을 사용하였다. 랜덤포레스트 모델은 학습에 2분 1초가 소요되었으며, 학습자료와 테스트자료의 정확도(accuracy)는 각각 1과 0.9916이다.

XGBoost를 이용한 분류모델은 XGBClassifier를 이용해 구현하였다. 랜덤포레스트 모델과 성능비교를 위해 의사결정나무 수와 random_state값을 동일하게 설정했으며,

병렬학습을 진행하였다. XGBoost와 동일하게 부스팅을 활용하는 gradient boost는 순차적으로 학습을 진행하지만, XGBoost는 병렬학습을 지원하여 학습에 소요되는 시간을 줄일 수 있다. 그 외 부스터의 종류 및 부스터 세부 설정 등과 같은 하이퍼파라미터는 기본값을 사용하였다. XGBoost 모델은 학습에 5분 37초가 소요되었으며, 학습자료와 테스트 자료의 정확도는 각각 1과 0.9947이다.

모델평가

본 연구에서 개발된 모델은 깊이별 압축파쇄대 여부를 검출하는 이진분류 모델로 0의 값을 가지는 자료가 절대적으로 많은 불균형 자료라는 특성을 고려하였을 때 0.99 이상 정확도만 비교하는 것은 부적합하다. 따라서 개발된 모델의 객관적인 평가를 위해 혼동행렬(confusion matrix)을 활용하였다.

혼동행렬은 분류모델의 성능을 시각화한 표로 TP(true positive), FP(false positive), TN(true negative), FN(false negative)으로 구성된다. 본 연구에서는 머신러닝 모델에 의해 압축파쇄대로 예측되는 경우를 positive, 그렇지 않은 경우를 negative로 정의했으며, 실제값과 예측값의 일치 여부에 따라 true와 false로 정의하였다. 예를 들어, TP는 실제 압축파쇄대가 있는 깊이를 분류모델이 압축파쇄대가 있다고 예측한 경우를 의미한다.

혼동행렬 결과 중 FP와 FN은 트레이드오프 관계로 분류 문제에서 해결하려는 문제 특성에 따라 어느 것을 핵심 평가지표로 할지 결정할 수 있다. 일반적으로 한 시추공에서 압축파쇄대가 없는 깊이가 대부분인 불균형 자료임을 고려했을 때, FN의 개수를 핵심 평가지표로 설정하는 것이 모델 평가에 더욱 적합하다. 왜냐하면 FP의 경우 압축파쇄대로 예측된 깊이의 자료(TP+FP)에 대해서는 소수이므로 전문가가 검토하여 FP를 선별할 수 있다. 그에 반해 FN의 경우 압축파쇄대로 예측되지 않은 깊이의 자료(FN+TN)에 대한 검토를 해야만 오류를 찾아낼 수 있다. 결국 대부분의 시추공 자료를 검토해야 하므로 머신러닝 모델을 사용해 전문가 검토의 효율성을 높이는 목표를 달성할 수 없게 된다.

혼동행렬을 이용한 평가지표는 수식 (1)과 같이 정밀도(precision)와 재현율(recall)이 있다. 정밀도와 재현율은 각각 FP와 FN에 가중치를 둔 평가지표로, 서로 트레이드오프 관계를 갖는다.

$$Precision = \frac{TP}{TP + FP}, Recall = \frac{TP}{TP + FN} \quad (1)$$

Fig. 5는 랜덤포레스트 모델을 통해 학습자료와 테스트 자료를 예측한 결과를 나타낸 혼동행렬이다. 테스트자료

에서 FP는 3개, FN은 162개로 큰 차이를 보인다. 이는 해당 모델이 압축파쇄대가 존재하는 깊이에서 대부분 압축파쇄대가 존재하지 않는다고 예측함을 의미한다. 테스트자료의 혼동행렬 결과를 바탕으로 정밀도는 0.99, 재현율은 0.57로 재현율이 낮은 값을 갖는다(Table 2). 따라서, 본 연구의 목적인 높은 재현율(낮은 FN) 달성에는 적합하지 않은 모델이라 볼 수 있다.

Fig. 6은 XGBoost 모델을 통해 학습자료와 테스트자료를 예측한 결과를 나타낸 혼동행렬이다. 테스트자료에서 FP가 17개, FN은 87개로 랜덤포레스트 모델 대비 FP 개수가 증가했지만, FN 값이 감소한 것을 확인할 수 있다. 정밀도는 0.94, 재현율은 0.77이며 랜덤포레스트 모델 대비 재현율이 증가한 것을 확인하였다(Table 2).

앞서 테스트자료에 대한 혼동행렬을 기반으로 랜덤포레스트와 XGBoost 모델의 성능 차이를 확인할 수 있었다. FN의 개수가 각각 162개(랜덤포레스트)와 87개(XGBoost)

로 XGBoost 모델의 재현율이 35% 개선됨을 확인하였다.

Fig. 7은 A2공 전체자료에 대해 학습된 두 모델을 적용한 압축파쇄대 깊이를 나타낸다. XGBoost 모델은 약 750 m 구간을 비롯하여 압축파쇄대가 아닌 깊이를 압축파쇄대로 예측하는 경향을 보인다. FP가 크고 정밀도가 낮은 결과와 일치한다. 비록 전문가가 검토해야 할 깊이가 증가하지만 FN을 최소화하기 위해 감수해야 하는 단계라 볼 수 있다. 랜덤포레스트 모델의 경우, FN보다 FP를 중요시하는 모델

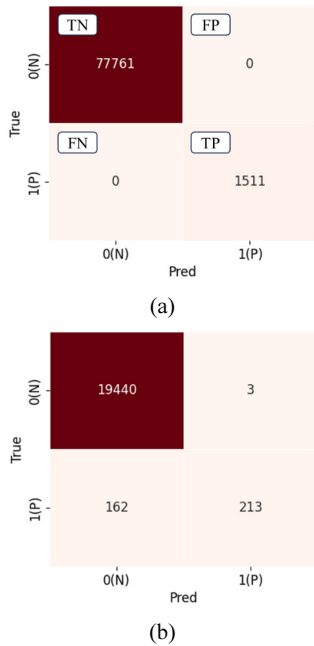


Fig. 5. Confusion matrix for the trained random forest model: (a) training dataset and (b) testing dataset.

Table 2. Comparison of the two machine learning models based on precision and recall

Models	Precision	Recall
Random forest	0.99	0.57
XGBoost	0.94	0.77

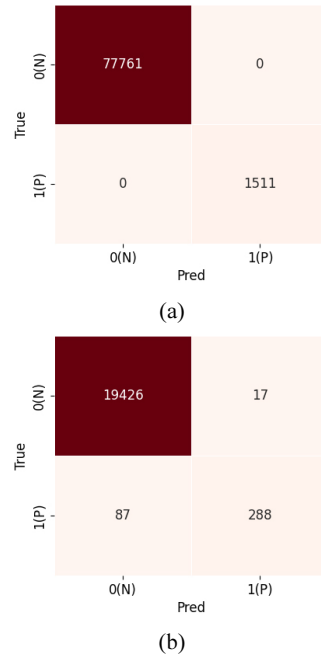


Fig. 6. Confusion matrix for the trained XGBoost model: (a) training dataset and (b) testing dataset.

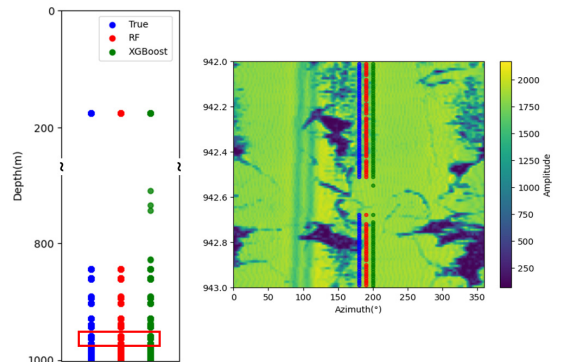


Fig. 7. Prediction result of well A2 using the trained machine learning algorithms.

이므로 Fig. 7의 942~943 m 구간 예시와 같이 실제 압축파쇄대 일부를 압축파쇄대 깊이로 예측하지 못하면서 불연속적인 압축파쇄대로 예측하는 문제점을 보인다. 이러한 분석결과를 바탕으로 XGBoost를 이용한 모델이 압축파쇄대 깊이 검출에 적합하다는 것을 확인하였다.

연구결론

기존의 압축파쇄대를 활용하여 응력을 규명하는 방법은 연구자가 압축파쇄대 검출을 수작업으로 진행하여 주관성이 개입되고 자료 검토에 소요 시간이 크다는 문제점이 있다. 이러한 문제점을 개선하기 위해 본 연구에서는 머신러닝을 이용하여 이미지로그의 압축파쇄대 깊이를 자동 검출하는 모델을 개발하였다. 이를 위해 영상검층자료의 기초통계 분석을 진행하고, 자료 전처리 과정을 수행하였다. 또한 두 가지 머신러닝 기법을 선정하여 모델을 구성하고, 연구목적에 맞는 평가지표(FN 또는 재현율)를 활용하여 압축파쇄대 깊이 검출에 적합한 모델을 선정하였다. 일련의 연구 과정을 통해 Fig. 8과 같은 이미지로그 자료의 압축파쇄대를 검출하는 머신러닝 적용으로 연구 흐름도를 정립하였다.

XGBoost 모델을 통해 테스트자료를 예측한 결과, FN 개수가 87개로 랜덤포레스트 모델 예측결과인 162개보다 적어 35% 개선된 것을 확인하였다. 물론 FP의 경우 랜덤포레스트 모델은 3개, XGBoost 모델은 17개와 같은 결과를 보였지만 FP와 FN은 트레이드오프 관계로 본 연구에서는 FN은 최소화하는 모델이 선호된다.

전체 시추공 자료가 아닌 본 연구에서 개발된 모델이 선별한 깊이를 중심으로 전문가가 압축파쇄대 상세 검출을 수행한다면 분석 효율성을 제고할 수 있을 것으로 기대된다. 다만 모델학습에 사용된 자료의 수가 제한적이므로 타 시추공 자료 적용을 위해선 전이학습(transfer learning)이

필요한 것으로 사료된다. 또한 개발된 모델은 압축파쇄대 깊이 정보만을 제공하므로 응력규명에 필요한 압축파쇄대의 길이와 폭 등 주요 정보는 전문가가 여전히 해석해야 하는 한계가 있다. 따라서 압축파쇄대의 속성정보를 도출하고 응력을 규명할 수 있는 연구가 필요하다.

사 사

이 성과는 산업통상자원부의 재원으로 한국에너지기술평가원의 지원(No. 20212010200020, 40%)과 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구(No. 2021R1C1C1004460, 40%)이며, 또한 2021년 정부(산업통상자원부)의 재원으로 해외자원개발협회의 지원을 받아 수행된 연구(데이터사이언스 기반 석유·가스 탐사 컨소시엄, 20210160001, 20%)이다.

References

- Barton, C.A., Moos, D., Peska, P., and Zoback, M.D., 1997. Utilizing wellbore image data to determine the complete stress tensor: application to permeability anisotropy and wellbore stability. *The Log Analyst*, 38(6), p.21-33.
- Barton, C.A., Zoback, M.D., and Burns, K.L., 1988. In-situ stress orientation and magnitude at the Fenton Geothermal Site, New Mexico, determined from wellbore breakouts. *Geophysical Research Letters*, 15(5), p.467-470.
- Breiman, L., 2001. Random Forests, *Machine learning*, 45(1), p.5-32.
- Dias, L.O., Bom, C.R., Faria, E.L., Valentin, M.B., Correia, M.D., de Albuquerque, M.P., de Albuquerque, M.P., and Coelho, J.M., 2020. Automatic detection of fractures and breakouts patterns in acoustic borehole image logs using fast-region convolutional neural networks, *Journal of Petroleum Science and Engineering*, 191, 107099.
- Ekstrom, M.P., Dahan, C.A., Chen, M.-Y., Lloyd, P.M., and Rossi, D.J., 1986. *Formation Imaging With Microelectrical Scanning Arrays*. In: Proceedings of S-PWLA 27th Annual Logging Symposium, Houston, Texas, June 9-13.
- Heidbach, O., Rajabi, M., Cui, X., Fuchs, K., Müller, B., Reinecker, J., Reiter, K., Tingay, M., Wenzel, F., Xie, F., Ziegler, M.O., Zoback, M.L., and Zoback, M., 2018. The World Stress Map database release 2016: Crustal stress pattern a-cross scales, *Tectonophysics*, 744, p.484-498.
- Hwang, H.J., Kim, S.H., and Song, G.W., 2018. XGBoost Model to Identify Potential Factors Improving and Deteriorating Elderly Cognition, *The Journal of Korean Institute of Next Generation Computing*, 14(3), p.16-24.

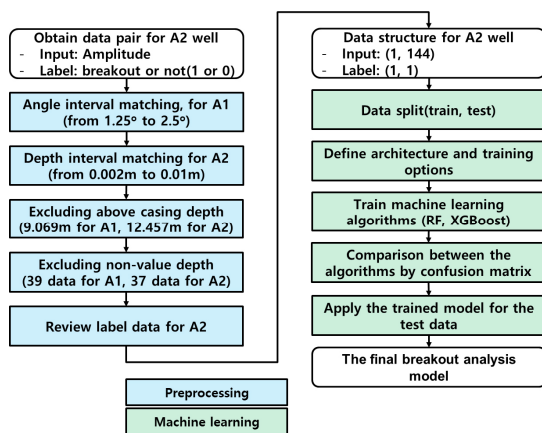


Fig. 8. Flow chart of the proposed method.

- Kwon, A.N., 2013. *Variable Selection Using Random Forest*, MS Thesis, Inha University, Korea, 28p.
- Lagraba, J.O.P., Hansen, S.M., Spalburg, M., and Helmy, M., 2010. Borehole Image Tool Design, Value of Information, and Tool Selection, *Dipmeter and Borehole Image Log Technology*, 92, p.15-38.
- Lee, Y.R., Kim, H.J., Lee, D.H., Lee, C.J., and Lee, D.H., 2019. Validation of Forecasting Performance of Two-Stage Probabilistic Solar Irradiation and Solar Power Forecasting Algorithm using XGBoost, *The Transactions of the Korean Institute of Electrical Engineers*, 68(12), p.1704-1710.
- Shamir, G. and Zoback, M.D., 1992. Stress orientation profile to 3.5 km depth near the San Andreas Fault at Cajon Pass, California, *Journal of Geophysical Research: Solid Earth*, 97(B4), p.5059-5080.

- Woo, S.Y., Jung, C.G., Kim, J.U., and Kim, S.J., 2018. Assessment of climate impact on aquatic ecology health indices in Han river basin using SWAT and random forest, *Journal of Korea Water Resources Association*, 51(10), p.863-874.
- Yang, J., Goodfellow, S.D., and Harrison, J.P., 2022. *Automated extraction of borehole breakout properties from acoustic televiewer (ATV) data*, American Rock Mechanics Association, USA.
- Zoback, M.D., 2010. *Reservoir Geomechanics*, Cambridge University Press, UK.
- Zoback, M.D., Moos, D., Mastin, L., and Anderson, R.N., 1985. Well bore breakouts and in situ stress, *Journal of Geophysical Research: Solid Earth*, 90(B7), p.5523-5530.



염 주 열

현재 공주대학교 지질환경과학과 학사과정
(E-mail; jy0506kr@smail.kongju.ac.kr)



김 하 영

2023년 공주대학교 자연과학대학 지질환경과학과 이학석사

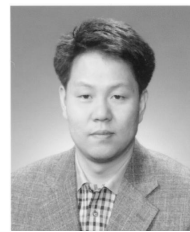
현재 공주대학교 지질환경과학과 박사과정
(E-mail; nicehy1023@gmail.com)



이 경 복

2014년 서울대학교 공과대학 에너지시스템공학부 공학박사

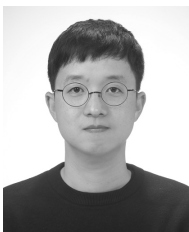
현재 공주대학교 지질환경과학과 조교수
(E-mail; kblee@kongju.ac.kr)



장 찬 동

2001년 University of Wisconsin-Madison, Geological Engineering, 공학박사

현재 충남대학교 지질환경과학과 교수
(E-mail; cchang@cnu.ac.kr)



조 영 욱

2019년 충남대학교 대학원 우주지질학과 이학박사

현재 한국지질자원연구원 국토우주지질연구본부 심층처분환경연구센터
(E-mail; joy@kigam.re.kr)