

2021년 지질자원 데이터 활용 및 인공지능 경진대회 최종 제안서

인공지능 기반 추론 모형을 활용한 S파 속도 예측

김도완¹⁾, 김도경²⁾, 유정훈³⁾, 박광용⁴⁾

I. 서론

1. DTS 추론 모형 개발의 필요성

S파 역속도 혹은 속도 검층(이하 DTS)은 전단 계수 및 포아송비와 같은 탄성물성 계산 및 AVO (Amplitude Variation with Offset) 역산 등 다양하게 활용되어 저류층 특성화에 많은 도움을 준다 (Liu et al., 2021). 하지만 종종 검층 비용 절감을 위해서 혹은 열악한 시추공 환경으로 인해 **DTS 검층이 수행되지 않는 경우가 발생**한다. 최근 기계학습의 발전으로 누락 혹은 이상치(outlier) 영역의 검층자료를 추정하는 연구가 활발히 진행되고 있다. 따라서 이 대회에서는 북해 Volve 유전 내 다섯 개의 시추공(15_9-F-1A, 15_9-F-1B, 15_9-F-1C, 15_9-F-11A, 15_9-F-11B)을 이용하여, **주어진 자료에 최적화된 인공지능 기반 DTS 추론 모형을 제안**한다.

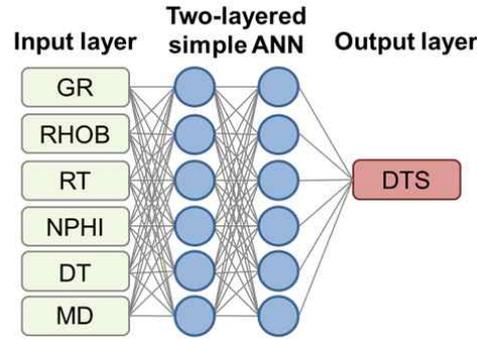
2. DTS 추론을 위한 특성 선정의 필요성

특성 선정(feature selection)은 추론 모형 개발 시 사용할 입력 변수를 선정하는 작업을 말한다. 적절한 특성 선정은 추론 모형의 계산 비용을 줄여주는 동시에 추론 성능을 향상시키기 때문에, 특성 선정에는 비중 있는 분석을 요한다. 특히 이 대회에서 주어진 시추공 자료의 경우 measured depth (MD), neutron porosity (NPHI), bulk Density (ROHB), gamma ray (GR), true formation resistivity (RT), delta time (DT), bit size (BS), caliper (CALI) 등 **다양한 검층 자료가 존재하기 때문에 특성 선정 작업을 통해 최적의 입력자료를 조합하는 것이 중요하다.**

3. 기존 연구 소개 및 한계점

Ghalthi and Prasad (2020)은 Volve 유전 내 시추공에서 DTS를 예측하기 위해 그림 1과 같은 추론 모형을 제안한 바가 있다. 해당 모형에서는 입력자료로 GR, RHOB, RT, NPHI, DT, MD 총 다섯 개의 특성을 조합하여 사용할 것을 제안하고 있다. 해당 조합을 사용하여 두 개의 층을 갖는 간단한 ANN (Artificial Neural Network) 모형으로 높은 성능의 추론을 수행하였다.

1) 한양대학교/자원환경공학과/물리탐사연구실 (kdw9492@hanyang.ac.kr)
2) 한양대학교/자원환경공학과/물리탐사연구실 (goodkdk96@hanyang.ac.kr)
3) 한양대학교/자원환경공학과/물리탐사연구실 (jhyoo9312@gmail.com)
4) 한양대학교/자원환경공학과/물리탐사연구실 (kwang482@naver.com)



[그림 1] Ghaithi and Prasad (2020)이 제안한 DTS 추론 모형.

하지만 해당 대회에의 경우, DTS 자료가 존재하지 않는 15_9-F-1C와 15_9-F-11B 시추공에 DT 검증 자료가 존재하지 않아 해당 조합을 사용하는 것이 불가능하다. 또한 MD는 시추공의 기울어짐 정도를 고려하지 않아, 심도에 따른 저주파 특성(low-frequency trend)을 올바르게 학습시키기 어렵다. 특히 이 과제에서는 기울어짐 정도가 서로 다른 다섯 개의 시추공을 사용하기 때문에 각 시추공의 기울어짐 정도를 고려하는 것이 더욱 중요하다.

4. DTS 추론 성능 개선 전략 및 목표

특성 탐색(feature discovery)은 의미있는 새로운 입력 특성을 생성하여 추론 모형의 성능을 향상시키는 작업을 말한다. 이 대회에서 주어진 검증자료의 경우, DTS 예측에 중요한 DT자료가 누락되었으며 MD는 기울어짐 정도가 고려되지 않아, 우리는 DT와 MD 자료를 대체할 새로운 특성을 생성하고자 한다. 새롭게 생성된 특성들을 이용하여 DTS 예측 정확도를 향상시키는 것을 목적으로 한다.

II. DTS 추론을 위한 특성 탐색

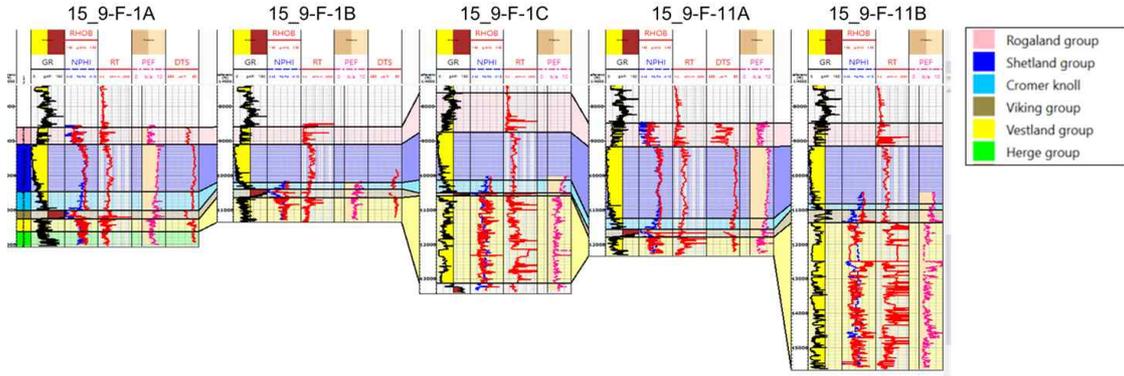
1. 데이터 분석 개요

(1) 시추공 내 층서 구조

주어진 다섯 개의 시추공 15_9-F-1A, 15_9-F-1B, 15_9-F-1C, 15_9-F-11A, 15_9-F-11B는 북해 Volve oil field에 위치하며, 2,750-3,120 m TVDSS 깊이의 중기 쥐라기 Hugin formation에서 저류층을 가지는 유전이다. 시추공 내 층서는 암층서 단위로 Rogaland, Shetland, Cromer knoll, Viking 그룹과 Hugin formation이 포함된 Vestland, Herge 그룹으로 총 여섯 개의 그룹으로 이루어져 있다 (그림 2).

(2) 시추공 내 검증 자료

시추공 내 검증 자료는 표 1과 같이 존재한다. 다섯 개의 시추공 중 15_9-F-1C, 15_9-F-11B 내에는 DT와 DTS 자료가 존재하지 않는다. 따라서 DTS가 존재하는 15_9-F-1A, 15_9-F-1B, 15_9-F-11A 검증 자료를 이용하여 추론 모형 학습을 진행하며, 15_9-F-1C, 15_9-F-11B에서의 DTS 예측을 최종 목표로 한다.



[그림 2] 시추공 내 층서 구조.

[표 1] 시추공 내 검층 자료 목록

검층 자료	설명	단위
ABDCQF01-04	density	g/cc
BS	bit size	inches
CALI	caliper	inches
DRHO	bulk density correction	g/cc
DT	compressional slowness	us/ft
DTS	shear slowness	us/ft
GR	gamma ray	API
NPHI	neutron porosity	v/v
PEF	photoelectric factor	b/elec
RACEHM	deep resistivity from attenuation	ohm.m
RACELM	shallow resistivity from attenuation	ohm.m
RHOB	bulk density	g/cc
ROP	rate of penetration	m/hr
RPCEHM	deep phase shift resistivity	ohm.m
RPCELM	shallow phase shift resistivity	ohm.m
RT	true formation resistivity	ohm.m

* 15_9-F-1C와 15_9-F-11B는 DT와 DTS가 존재하지 않음.

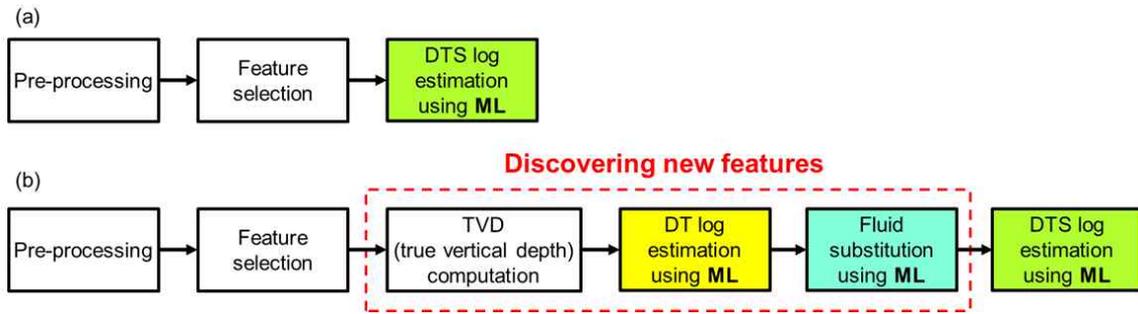
2. DTS 추론 워크플로우 개발

(1) DTS 추론 워크플로우 개요

우리는 DTS 예측에 중요한 DT와 심도정보를 생성 및 보완하기 위해 그림 3과 같이 특성 탐색을 포함한 워크플로우를 제안한다. 총 여섯 단계에 걸쳐 DTS를 예측하였으며 각 단계의 자세한 설명을 아래 기술하고자 한다.

(2) 검층자료 전처리

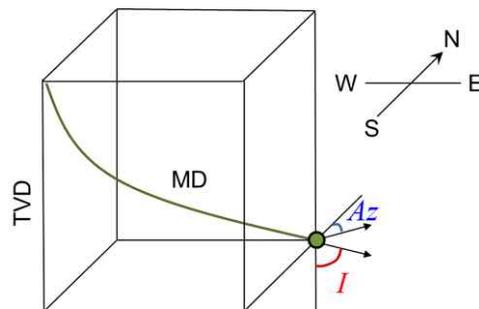
인공지능 기반 추론 모형에서 잡음제거, 자료변환 등의 전처리는 학습속도 개선, 과적합 방지 등의 효과를 얻을 수 있다. 우리는 1) CALI와 BS 검층 자료를 비교하여 bad hole에 의한 이상치 제거를 수행하였으며, 2) 물리적으로 타당하지 않은 DT가 다수 관측된 최상단 Rogaland group을 학습에서 제외시켰다. 또한 3) RT, RACEHM, RACELM 등 넓은 관측 범위를 갖는 비저항 검층은 로그를 취한 값을 사용하였다.



[그림 3] (a) 일반적인 DTS 예측 워크플로우. (b) 개발된 특성탐색을 포함한 워크플로우.

(3) 특성 선택(feature selection)

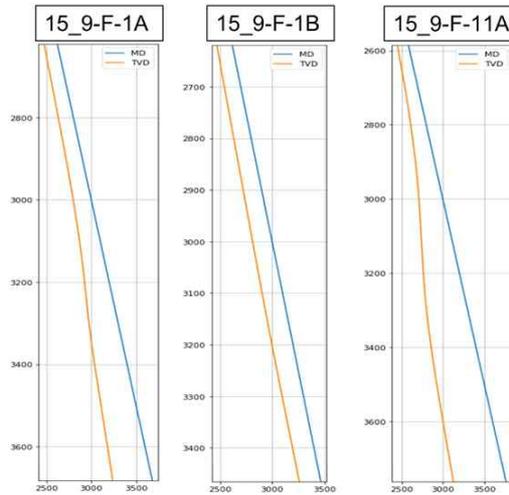
표 1에 기재된 검층 자료 중 추론 모형의 입력자료로 활용할 특성을 선별하였다. 특성 선택은 세 단계에 걸쳐 이루어졌다. 가장 먼저 이루어진 1) 이론기반 선별은 DTS와 이론적으로 연관이 없는 검층을 제외한다. 이 작업에서 BS, CALL, DRHO, ROP 검층이 제외된다. 다음으로 이루어진 통계기반 선별은 상관도 분석을 통한 선별과 p-value 기반 선별, 총 두 단계의 선별 과정을 거친다. 2) 상관도 분석을 통한 선별은 중복되는 특성들을 제외하여 특정 특성에 대해 과하게 과적합되는 것을 방지하는 선별기법이다. 우리는 0.85 이상의 상관관계를 가지는 검층들 중 하나의 검층자료만을 선택하여 사용하였다. 이 작업에서 RHOB와 상관도가 높은 ABDCQF01,2,3,4 그리고 RT와 상관도 높은 RACELM, RACEHM, RPCEHM, RPCELM가 제외되었다. 3) p-value 기반 선별은 DTS와 상관관계가 존재하는 특성만을 선별하기 위해 p-value 값이 0.005 미만의 특성들만을 선별하는 작업이다. 특성 선택 결과 최종적으로 GR, NPHI, PEF, RHOB, RT 검층이 p-value 0.005 미만으로 선별되었다.



[그림 4] MD와 TVD의 관계. I : inclination, Az : azimuth.

(4) TVD 계산

MD의 경우 시추공의 기울어짐을 고려하지 않는다. 하지만 이 과제와 같이 여러 개의 시추공을 동시 사용하는 경우 심도에 따른 DTS의 저주파 특성(low-frequency trend)을 올바르게 학습시키기 위해서는 각 시추공의 기울어짐 정도를 고려해야한다. 따라서 우리는 시추공의 기울어짐이 고려된 TVD (true vertical depth) (그림 4)를 계산하여 입력 특성으로 사용하였다. 이 때 TVD 계산은 well report 내 inclination 정보를 참조하여 이루어졌다. 기록된 MD와 계산된 TVD의 차이는 그림 5와 같다.



[그림 5] 기록된 MD와 계산된 TVD의 비교.

(5) DT 추정

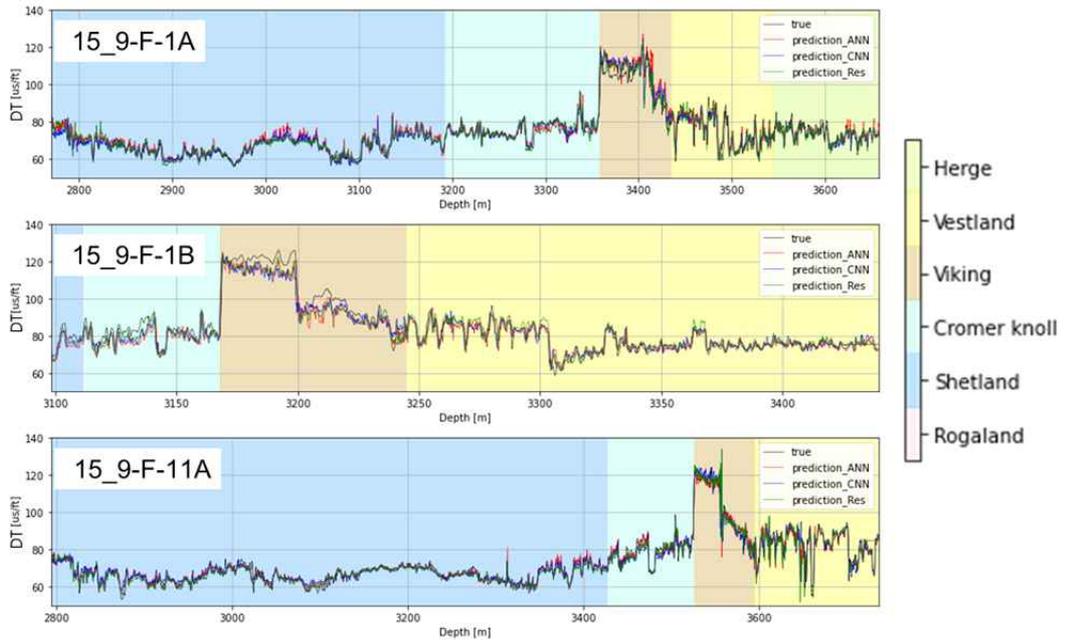
DT는 DTS와 높은 상관도를 갖는 검층이다. 따라서 DT를 입력특성으로 사용하는 경우 DTS의 추론 성능을 향상시킬 수 있다. 하지만 이 대회에의 경우 DTS를 추론해야 하는 15_9-F-1C, 15_9-F-11B 시추공 내 DT가 존재하지 않는다. 따라서 우리는 DT를 우선 추정 한 뒤, 추정된 DT를 입력자료로 다시 DTS를 계산하는 워크플로우를 제안한다.

DT 추정 모형에는 가장 간단한 형태의 ANN, 공간적인 변화를 고려하는 CNN (Convolutional Neural Network), skip-connection을 포함하는 복잡한 구조의 ResNet (Residual CNN) 총 세 개의 모형을 사용하여 모형의 의존도도 함께 비교해 보았다. 각 모형 정보는 표 2와 같다.

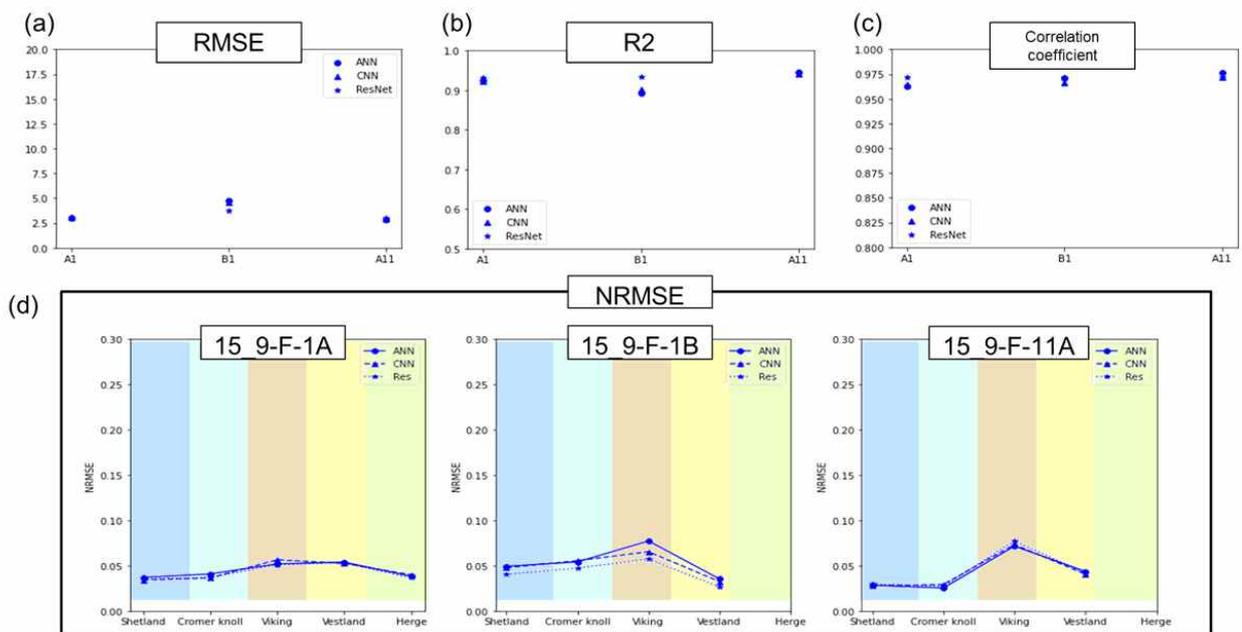
[표 2] 사용된 인공지능 모형 정보.

	ANN	CNN	ResNet
num(hidden layers)	3	3	3
num(neurons)	16,16,16	16,32,64	16,16,16
Batch normalization	O	O	O
Activation	ReLU	ReLU	ReLU
Optimization	Adam	Adam	Adam
Scaler	Standard	Standard	Standard
Patch size	-	7	7

또한 DT 예측의 검증 신뢰도를 향상시키기 위해 well cross-validation 기법을 사용했다. well cross-validation 기법은 학습에 사용된 세 개의 시추공 15_9-F-1A, 15_9-F-1B, 15_9-F-11A를 각각 한 번씩 blinde well로 설정하여 세 번의 blind well test를 수행하는 검증 기법이다. 성능 평가에는 RMSE (root mean squared error), NRMSE (normalized RMSE), correlation coefficient, R2 (coefficient of determination) 총 네 개의 평가지표를 사용하였다. cross-validation 결과는 그림 6에 성능 평가 결과는 그림 7에 도시하였다.



[그림 6] 모형에 따른 DT 예측 결과 비교.



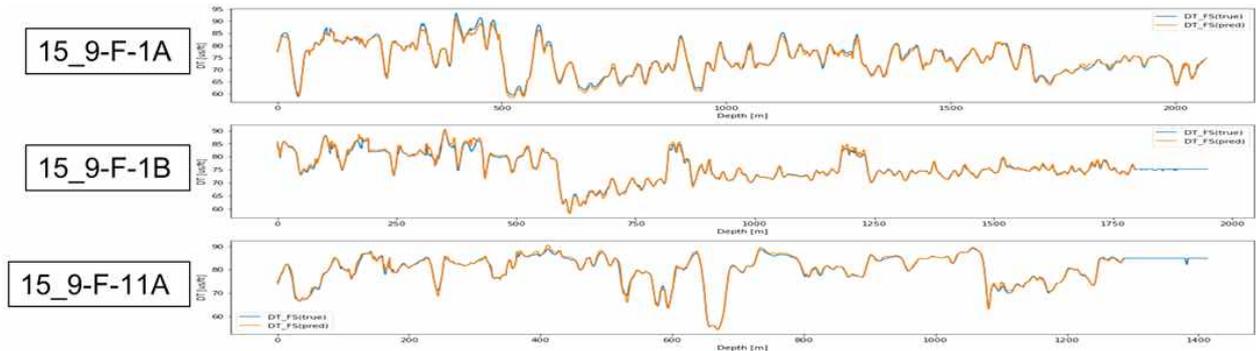
[그림 7] DT 예측 결과의 (a) RMSE (b) R2 (c) correlation coefficient (d) NRMSE 비교.

그림 7을 보면 모든 평가지표에서 높은 성능을 기록하였고, 모형에 따른 의존도는 낮은 것으로 확인된다. 따라서 우리는 DT가 충분히 신뢰할 수 있는 수준으로 예측된다고 판단하여 예측된 DT를 입력자료로 활용하였다.

(6) 유체 치환(fluid substitution) 수행

대부분의 유체(오일, 가스 등)는 전단 변형에 저항력이 거의 존재하지 않기 때문에 일반적으로 유체에서의 전단계수는 0으로 가정된다. 따라서 S와 속도는 P와 속도와 다르게 유체에 의한 영향을 거의 받지 않으며, 이로 인해 공극 내 유체의 변화가 존재하는 경우 DTS와 DT의 상관도가 낮아지

게 된다. 따라서 유체 변화에 따른 DT의 변화가 발생하지 않도록 100% water saturation으로 치환된 상황을 가정하여 DT를 보정해 주는 작업을 유체 치환이라고 한다. 유체 치환을 수행하는 경우 관심 영역인 저류층에서의 DTS와 상관도를 향상시켜주기 때문에 저류층에서의 DTS 예측 성능을 높일 수 있다. 따라서 우리는 표 2에 기재된 ANN 모형을 이용하여 유체 치환을 수행하였다. 이 때 학습자료는 15_9-F-1A, 15_9-F-1B, 15_9-F-11A에서 관측된 DT를 사용하였으며, 레이블은 Hampson-Russell 소프트웨어 내 Biot-Gassman equation (Biot 1941; Gassmann 1951)을 이용하여 생성하였다. ANN을 이용한 유체 치환 결과는 그림 8과 같으며, 예측 성능 평가는 표 3과 같다.



[그림 8] 유체 치환 결과.

[표 3] 유체치환 모형의 cross-validation 결과.

	ANN	CNN	ResNet
Correlation coefficient	0.996	0.995	0.990
R2	0.987	0.989	0.979

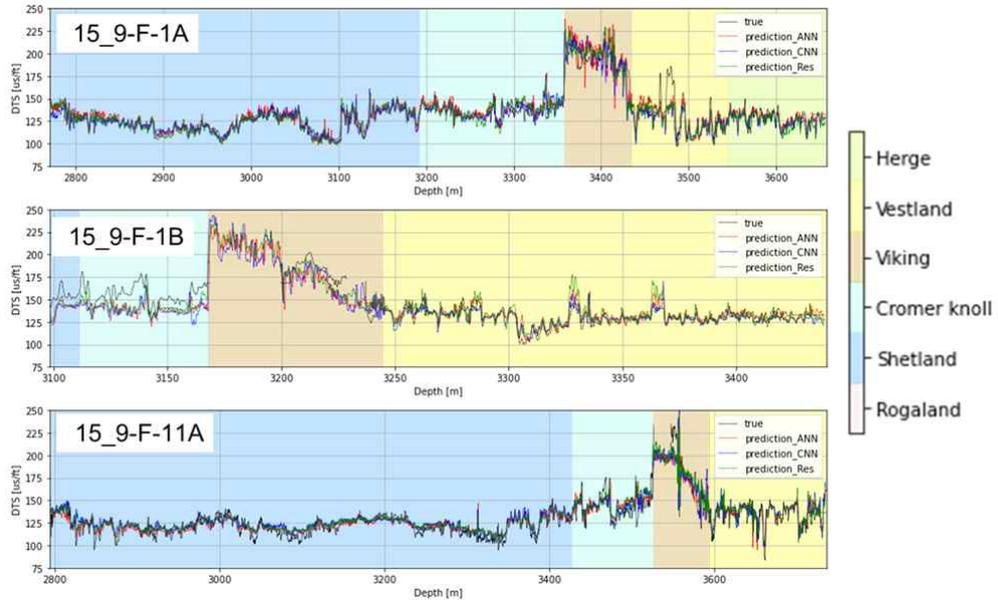
(7) DTS 예측을 위한 최종 입력 특성

우리는 관측된 검층자료에서 특성 선택을 수행하여 GR, NPHI, PEF, RHOB, RT를 선별했다. 다음으로 MD와 inclination 정보를 이용하여 TVD로 계산했다. 마지막으로 DT를 예측하고 저류층에서의 유체 치환을 수행하여 유체 치환된 DT (DT_FS)를 예측했다. 따라서 최종적으로 GR, NPHI, PEF, RHOB, RT, TVD, DT_FS 총 일곱 개의 입력 특성을 획득하였다. 다음 장에서는 획득한 일곱 개의 특성을 입력자료로하여 DTS를 산출하고 결과를 비교분석 하고자한다.

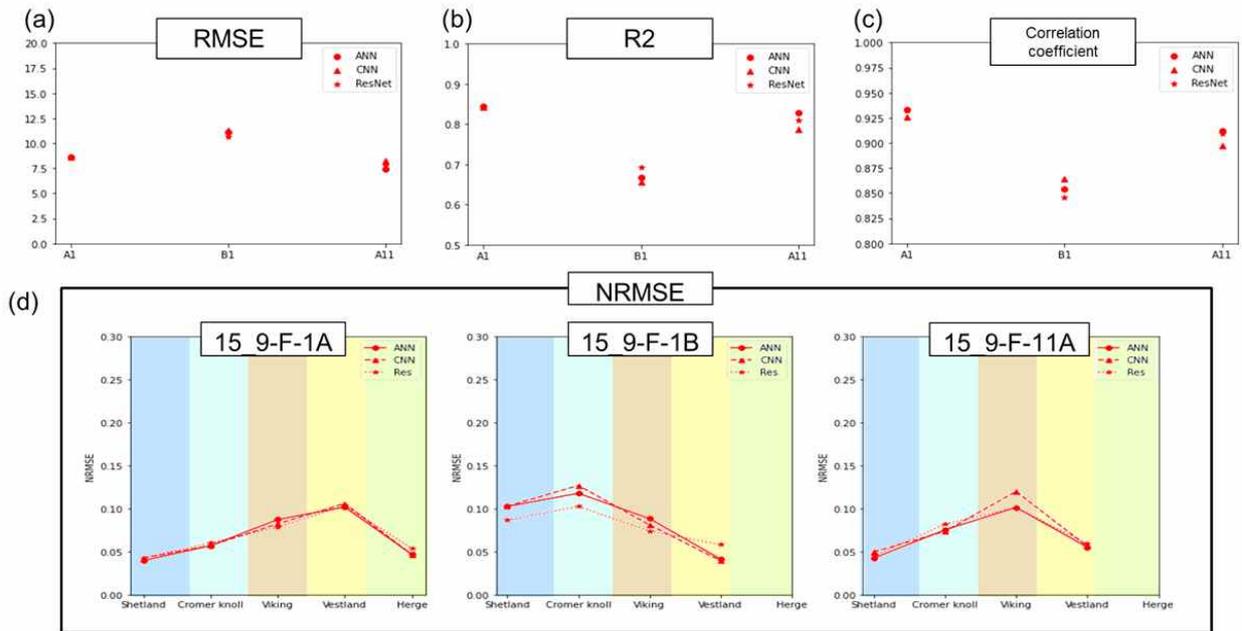
III. DTS 추론 결과

1. DTS 추론의 well cross-validation 결과

DTS 예측에는 DT 예측에 사용된 세 개의 모형(표 2)을 동일하게 사용하였다. DTS 추론의 well cross-correlation 결과는 그림 9와 같으며, 각 모형의 RMSE, NRMSE, correlation coefficient, R2 값은 그림 10과 같다.



[그림 9] 특성 탐색을 수행한 후 모형에 따른 DTS 예측 결과 비교.



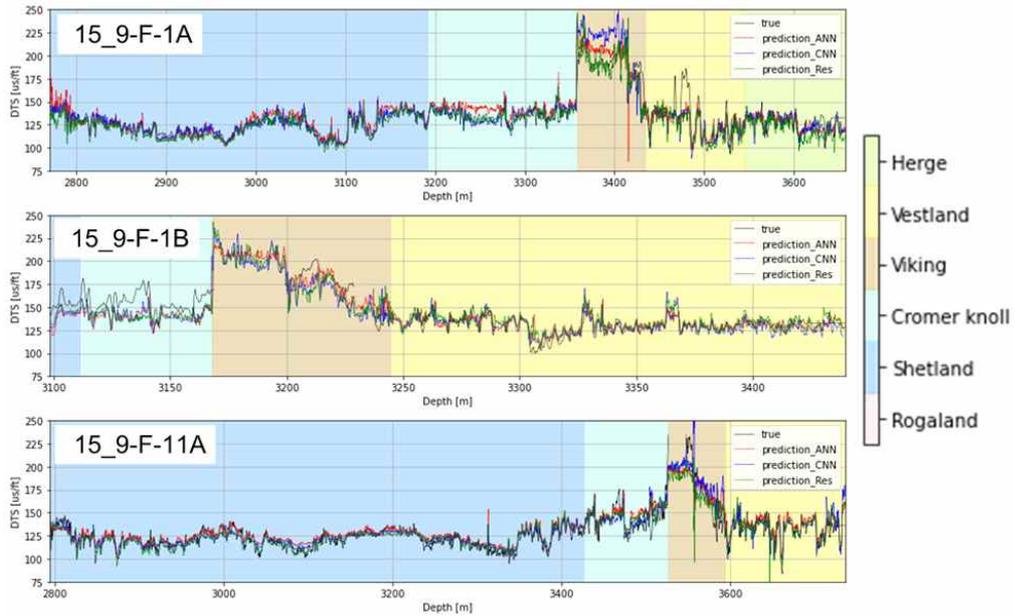
[그림 10] DTS 예측 결과의 (a) RMSE (b) R2 (c) correlation coefficient (d) NRMSE 비교.

전체적으로 좋은 성능을 확인할 수 있다. 다만 모형에 따른 의존도가 낮으며 오히려 복잡한 모형을 사용하는 경우 부분적으로 더 낮은 성능을 보이는 경우도 있기 때문에 계산 비용 측면에서 ANN 모형을 사용하는 것이 바람직하다고 판단된다.

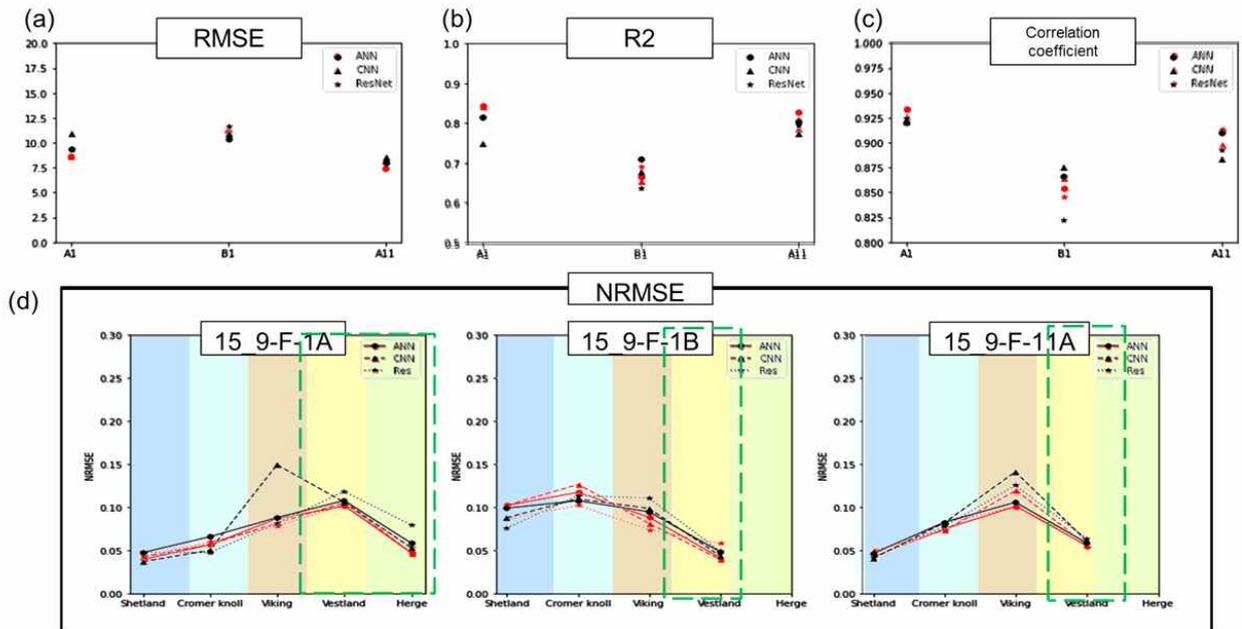
2. 특성 탐색을 수행하지 않은 경우의 DTS 추론 결과와의 비교

우리는 앞 장에서 수행한 특성 탐색이 DTS 추론 성능을 향상시켰는지 확인하기 위해 특성 탐색을 수행하지 않고 DTS를 추정 한 뒤(그림 3a) 이를 특성 탐색을 수행하였을 때의 결과와 비교하였다. 특성 탐색을 수행하지 않은 경우의 입력 특성으로는 GR, NPHI, PEF, RHOB, RT, MD를 사용하였

다. 특성 탐색을 수행하지 않은 경우 DTS 예측 결과는 그림 11과 같으며, 그림 12는 특성 탐색을 수행한 경우와 수행하지 않은 경우 DTS 추론 성능의 비교이다.



[그림 11] 특성 탐색을 수행하지 않은 경우 모형에 따른 DTS 예측 결과 비교.



[그림 12] 특성 탐색을 수행한 경우와 수행하지 않은 경우 DTS 추론 결과의 (a) RMSE (b) R2 (c) correlation coefficient (d) NRMSE 비교. 검은색 점은 특성 탐색을 수행하지 않은 경우를, 빨간색 점은 특성 탐색을 수행한 경우의 결과. 관심영역인 저류층 영역이 초록색 점선으로 표시됨.

전체 영역에 대해서는 큰 성능의 향상을 확인하기 어렵지만 관심영역에 해당되는 저류층 구간에서는 뚜렷한 성능 향상을 확인할 수 있다. 표 4에 관심영역인 Vestland group의 NRMSE 변화를 기재하

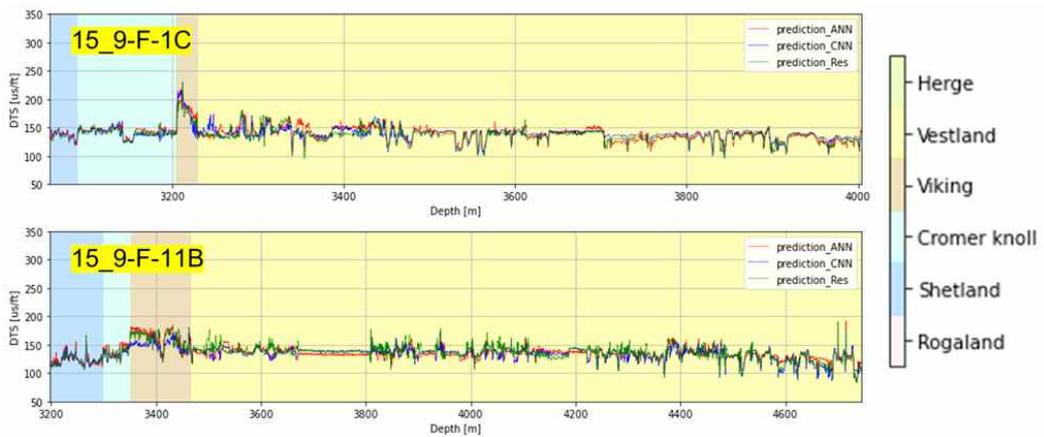
였다. ResNet을 사용한 경우 15_9-F-1B에서의 추론 성능이 하락한 것을 제외하고 모든 경우에서 추론 성능이 향상된 것을 확인할 수 있다. 이는 DT를 예측하고 저류층 영역에서의 유체 치환을 수행했기 때문으로 추정된다.

[표 4] 특성 탐색을 수행하지 않은 경우와 특성 탐색을 수행한 경우 Vestland group의 NRMSE 비교.

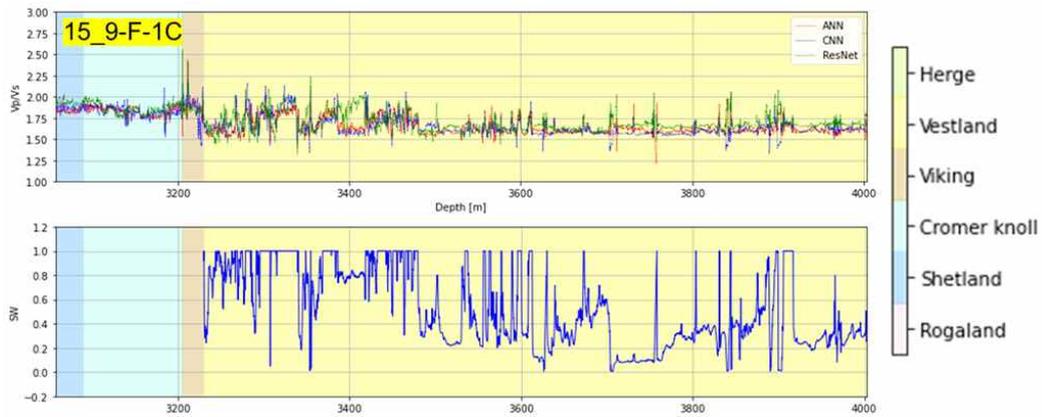
	ANN		CNN		ResNet	
	conventiol method	proposed method	conventiol method	proposed method	conventiol method	proposed method
15_9-F-1A	0.107	0.101	0.105	0.105	0.118	0.103
15_9-F-1B	0.044	0.041	0.043	0.039	0.048	0.057
15_9-F-11A	0.058	0.055	0.060	0.056	0.063	0.058

3. 15_9-F-1C와 15_9-F-11B에서의 DTS 추정

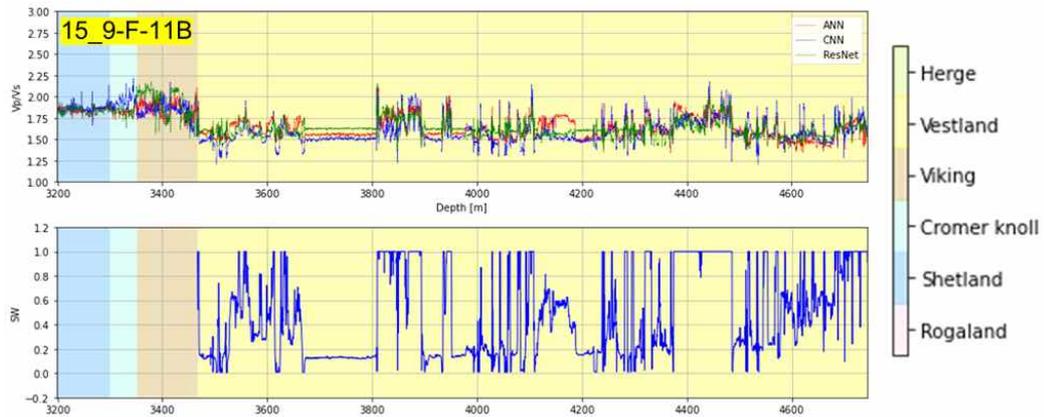
최종적으로 우리는 15_9-F-1C와 15_9-F-11B에서의 DTS를 추정하였다(그림 13). 해당 시추공들에는 DTS 검층자료가 전혀 존재하지 않아, DTS 추정 결과를 정량적으로 평가하는 것이 불가능하다. 하지만 우리가 제안한 워크플로우로 DTS를 예측하는 경우 DT도 함께 예측할 수 있다는 장점이 있어, 수포화도와 속도비(V_p/V_s)를 함께 도시하여 정성적 분석을 수행하였다. P파는 S파와 다르게 저류층 구간에서 유체의 영향으로 속도 변화가 일어나는 특징을 가진다. 그림 14와 그림 15는 각각 15_9-F-1C와 15_9-F-11B에서 추정된 속도비와 수포화도를 비교한 그림이다. 전반적으로 수포화도가 낮은 영역에서 속도비가 감소하는 경향을 잘 보여준다. 하지만 비교적 간단한 모형의 ANN과 CNN이 복잡한 구조의 ResNet보다 좋은 결과를 보여, ResNet은 계산상의 이점이 떨어지는 것으로 판단된다.



[그림 13] 15_9-F-1C와 15_9-F-11B에서의 DT와 DTS 추정



[그림 14] 15_9-F-1C에서 추정된 속도비와 수포화도 비교.



[그림 15] 15_9-F-11B에서 추정된 속도비와 수포화도 비교.

IV. 결론

인공지능 기반 추론 모형을 구성할 때에는 가장 먼저 사용될 입력 특성을 올바르게 구성하는 것이 중요하다. 따라서 우리는 다양한 특성 선택 기법들을 이용하여 최적의 검층 자료들의 조합을 결정하였다. 또한 시추공 마다 다른 기울어짐을 고려하기 위해 기존 MD를 이용하여 TVD를 산출하였다. 마지막으로 누락된 DT 자료를 대체하기 위해 DT 추정 모형을 개발하였고 저류층에서 유체에 따른 P와 속도를 보상하기 위해 유체 치환된 DT를 산출하였다.

최종적으로 획득한 일곱 개의 입력 특성들을 활용하여 DTS를 추정한 결과, 저류층 구간에서의 DTS 추정 성능이 향상되는 것을 확인할 수 있었다. 이는 저류층을 고려한 유체 치환된 DT를 산출한 것이 유효했음을 알 수 있다. 따라서 개발된 워크플로우는 오일 저류층에서 유용하게 사용될 수 있다고 기대한다. 또한 유체에 의한 영향이 더욱 심한 가스 저류층의 경우 보다 뚜렷한 성능 향상이 기대된다.

또한 우리는 DT와 DTS 예측에 서로 다른 세 개의 인공지능 모형을 사용하여 모형에 대한 의존도를 평가하였다. 그 결과 ANN, CNN, ResNet 모두 well cross-validation 기법을 이용한 정량적 평가에서 세 결과의 차이가 크지 않고, 일부 영역에서는 오히려 ResNet의 성능이 떨어지는 것을 확인하였다. 또한 test well인 15_9-F-1C와 15_9-F-11B에서의 정성적 평가에서는 ResNet을 사용한 경우, 속도비와 수포화도의 관계가 가장 잘 들어나지 않았다. 따라서 복잡한 모형을 사용할 때의 계산상의 이점

이 적은 것으로 판단되어, 비교적 간단한 모형인 ANN 혹은 CNN 모형을 사용하는 것이 바람직해 보인다.

V. 모형 실행 환경

모형 실행 환경은 표 5와 같다.

표 5. 사용된 패키지 버전

Package	Version	Package	Version	Package	Version
Tensorflow-gpu	2.5.0	lasio	0.29	scikit-learn	0.24.2
numpy	1.19.5	pandas	1.2.5	jupyter lab	3.0.16

* 기반 코드 : Python 3.8.10

참고문헌

- Al Ghaithi, A., and Prasad, M. 2020. Machine learning with artificial neural networks for shear log predictions in the Volve field Norwegian North Sea. In SEG Technical Program Expanded Abstracts 2020, 450-454.
- Biot, M. A. 1941. General theory of three dimensional consolidation. *Journal of applied physics*, **12**(2), 155-164.
- Gassmann, F. 1951. Uber die elastizitat poroser medien. *Vierteljahrsschrift der Naturforschenden Gesellschaft in Zurich*, **96**, 1-23.
- Liu, S., Zhao, Y., and Wang, Z. 2021. Artificial interlligence method for shear wave travel time prediction considerin greservoir geological continuity. *Mathematical Problems in Enginerring*.