Romanian Journal of Petroleum & Gas Technology
Vol. VI (LXXVII) • No. 1/2025

A CONCRETE DROPOUT NEURAL NETWORK FOR SHEAR SONIC LOG PREDICTION

Ephraim Ojoajogu Enemali [1*]

Sulaiman Alhaji Dodo Ibrahim [1]

Ibraheem Salaudeen [1]

Mahmud Abubakar Abdulmalik [1]

[1] Department of Petroleum Engineering, Abubakar Tafawa Balewa University, Bauchi, Nigeria
email (corresponding author): ephene2@gmail.com

DOI: 10.51865/JPGT.2025.01.08

ABSTRACT

Assessing the risk associated with drilling and wellbore stability studies requires the shear sonic log. These logs apart from distinguishing formation fluid from lithology are needed to obtain geo-mechanical rock parameters required for the safe design of rock fracturing. Although sonic logs are of great importance, they are usually not obtained due to the limiting cost of acquisition. Neural networks have been used to generate these logs to save cost, but these networks are prone to overfitting. The dropout rate has been proposed to tackle this problem, however selecting the optimum dropout probability rate can be challenging manually or expensive computation wise.

This research therefore investigated concrete dropout, a dynamic technique for adapting the dropout rate of a neural network to the data. The concrete dropout technique was applied to an artificial neural network (ANN) and a convolutional neural network (CNN) model to predict the shear sonic log with Monte Carlo simulation. Comparison was also made with the deterministic ANN and CNN models which had no dropout layers added and a Bayesian-optimized multilayer perceptron (MLP) model. These models were trained and validated with four (4) wells from the Volve field, using features with the highest correlation. The Concrete dropout ANN was found to outperform both the deterministic versions and the MLP model with R^2, RMSE, MSE and MAE scores of 0.9548, 3.6415, 2.4433 and 0.0179 respectively.

The neural networks built in this study showed an enhanced predictive performance with concrete dropout addition over the networks with no dropout added, showing that the technique was able to adapt the dropout rate to fit the nature of data and improve performance, which finds great application in real time deployment. The findings of this study also proposed a cost-effective way of sampling and averaging multiple outputs from a single neural network model, leading to enhanced predictive performance as the addition of concrete dropout allowed the network output distributions rather than point predictions.

Keywords: Monte Carlo simulation, deterministic models, concrete dropout, sonic logs, neural networks

INTRODUCTION

Sonic logs are geophysical logs that measure the transit time of sound propagating in a porous medium. They contain compressional and shear sonic logs which can be used to obtain geo-mechanical rock parameters which are indispensable to studies of well characteristics such as wellbore stability, fracturing design, surface seismic data calibration and others [14], [3]. Shear wave velocity can be used to calculate the dynamic modulus of the formation as well as other petro-physical properties [24]. It also allows us to discriminate the formation fluids from the rock formation [32]. Their availability is essential, as it allows for the estimation of formation parameters that are critical to assessing and quantifying the risk associated with drilling and wellbore stability [3].

Although these logs are of great importance, they are usually not obtained for old wells due to their high cost of acquisition [2] and some of those obtained have missing sections. Different approaches have been proposed to address this problem. The early approach to shear sonic log prediction relied heavily on empirical correlation and mathematical models [2]. One of the earliest correlations was by Faust [15]. He introduced a simple relationship to estimate the compressional wave velocity from resistivity and depth in his work. [18] also estimated the compressional wave velocity from density. The more popular Greenberg & Castagna correlation [20] produced a relationship between compressional porosity, saturation, lithology and shear wave and its prediction method is constrained to porous, brine-saturated reservoirs or zones [4]. Although these models show promising results and are simple to implement, they can only be applied to a specific geological setting or rock type. The limitations of these simple correlations in capturing the complex, non-linear relationship that exists between the subsurface formation properties, the need for unexplainable constants and their need for assumption which can alter the reservoir properties, have led to the development of more advanced techniques and the adoption of data-driven models to generate synthetic responses in light of the advances in deep learning and artificial intelligence [2].

New methods of predicting sonic logs are to be embraced because of the high acquisition cost of these logs and the increasing difficulty of securing funding for hydrocarbon projects due to the energy transition. Various studies have implemented diverse machine-learning models to predict reservoir properties using wells [29],[30]. Artificial neural networks (ANN) can be a useful tool to aid the prediction and analysis of existing and emerging areas in the oil and gas industry and also in geophysical tasks like sonic log [30].

An artificial neural network (ANN) is a machine-learning model whose design was inspired by the network of biological neurons in the human brain [4]. These networks are powerful computational models capable of learning the intricate and complex patterns within data. However, over-fitting is a persistent problem of neural networks. The introduction of dropout [22], [35] has to a great extent addressed this significant challenge of over-fitting in deep learning techniques. Since then, it has also been widely adopted in Convolutional Neural Network (CNN) models [12]. The question arising then is, how does one choose the optimum dropout rate to be applied to each layer of the network ? This has traditionally been done in the following works using neural networks for shear sonic log prediction ([1-6], [19], [23], [27-28], [31], [33], [39]) through manual experimentation which is tedious or through automatic tuning which can be computationally expensive. In the aforementioned works, an attempt was not made to investigate the concrete dropout technique in order to select the optimum neural network dropout rate as a function of the data. Therefore, the main contribution of this work is to demonstrate the application of a concrete dropout one-dimensional Convolutional Neural Network (1D-CNN) and ANN models with Monte-Carlo simulation to predict the shear sonic log. A comparison is also made with both their deterministic architectures which had no dropout added and a Bayesian optimized multilayer perceptron (MLP).

CONCRETE DROPOUT

Standard deterministic deep neural networks operate on a one-input-one-output basis. Unlike single-point predictions of such models, Bayesian methods such as Bayesian Neural Networks (BNNs) and Gaussian process (GPs) give predictive distributions [11]. Bayesian neural networks rather than outputting a single fixed value, output a probability distribution (a prior) over the neural network weights. These network weights each have a mean and variance associated with them as they are no longer point values but distributions for which these properties can be found. The model prediction is then achieved by integrating the entire set of possible weights and at training, updating the prior to posterior. The prohibitive computation cost of Bayesian inference has resulted in other methods of training neural networks with low computational cost implications [26]. One such method is the Monte Carlo dropout technique by [16]. The authors showed that a neural network trained with dropout [22], [35], active during both training and at test time could approximate a Bayesian inference [26]. Therefore, with each forward pass through the network, a new set of predictions could be obtained for which the mean prediction can be computed [10]. An additional benefit is that the architecture of the network remains largely unchanged in the approach [16].

One of the criticisms of the Monte Carlo technique is that the dropout rate and weight regularization which affect the model output requires a grid search to be implemented in other to obtain optimum dropout rate which can be computationally expensive [17]. In the Monte Carlo dropout technique, discrete Bernoulli distributions which are parameterized by a dropout rate are randomly drawn following a Bernoulli distribution and in a subsequent work by [17], the authors proposed an extension to the Monte Carlo technique by imposing a continuous relaxation to this distribution, known as the concrete distribution relaxation, thus creating a variant where the dropout rate of each layer is learned as part of the optimization process, therefore, eliminating the need to grid search over the dropout probabilities. This method allows for dropout probabilities to be tuned thus allowing the model to dynamically adapt to the data.

Concrete dropout allows the dropout probability to adapt as more data is collected, instead of being set once and held fixed [17]. The trained model is essentially an infinite ensemble of neural networks where each instance has its weights drawn from the posterior distribution [26]. Sampling the predictions from many trained similar networks has been noted to show enhanced performance, and robustness as well as reduce the uncertainty of deep learning model [25], as it reduces the disadvantages of using only a single network for inference thus improving the model's predictive performance [21].

MATERIALS AND METHODS

DATASET

The data used in this paper comes from the publicly available Volve field. The field is located in block 15/9 of the Norwegian North Sea and lies 200km west of Stavanger, being bounded to the east by the Loke gas field [6]. Production on the field started from 2008 and ended in 2016 when it was decommissioned with its datasets open-sourced in 2018 by [13]. This paper used four wells containing both shear and compressional sonic logs, namely, wells 15/9-F1B, 15/9-F1A, 15/9-F11, and 15/9-F11T2.

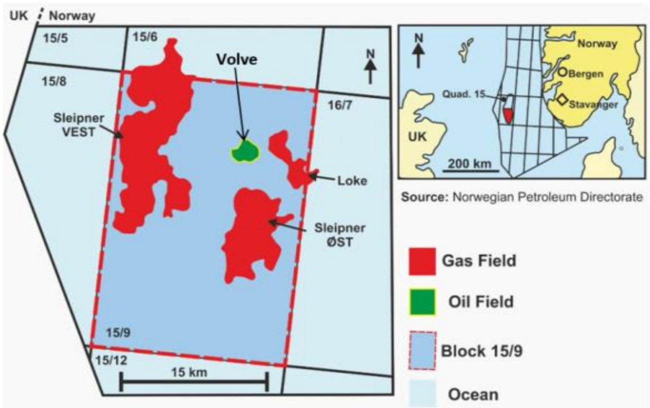


***Figure 1**. Location map of Volve field* [38]

PREPROCESSING

Domain knowledge and Statistical analysis using Pearson correlation coefficient and scatter plots were used to reduce the dataset to the features which had the highest correlation with the shear sonic log (DTS). They include: gamma ray (GR), bulk density (RHOB), photoelectric log (PEF), resistivity log (RT), neutron porosity (NPHI), and compressional sonic log (DT). The correlation of these features with the predictor can be seen in Figure 2.

***Figure 2**. Pearson Correlation Coefficient Matrix*

Logarithmic transformation enhances a model's prediction [2]. Therefore, the resistivity log was log-transformed and rows with missing data were dropped. The box plot was used to identify outlier points in the chosen features and Isolation Forest [36], an unsupervised machine learning model for outlier detection was used to identify these outliers and they were removed separately from each of the four wells. Data from three of the cleaned wells F1B, F11, and F11T2 were then merged into a single dataset for model training and dataset from well F1A was reserved as the blind validation dataset for final model evaluation. All models built in this paper were trained with 90% of the data (17,359 rows) and 10% (1,929 rows) for testing. Blind validation was done on well F1A (Table 2) with 8171 data points. The summary statistics for both training and blind validation datasets are found in Table 1 and Table 2. The summary statistics show that the data values for blind validation of the built models lies within the range of the values of the training data which allows the neural networks to be able to perform well.

Table 1. *Training Data Summary Statistics*

	DEPTH	**GR**	**NPHI**	**RHOB**	**PEF**	**RT**	**DT**	**DTS**
count	19288	19288	19288	19288	19288	19288	19288	19288
mean	3162.63	27.369	0.136	2.505	7.176	0.432	73.383	132.634
std	267.882	21.357	0.056	0.119	1.119	0.358	9.730	18.748
min	2625.8	0.852	0.032	2.133	4.511	-0.629	55.75	95.477
25%	2948.5	8.873	0.093	2.486	6.322	0.2835	66.509	120.252
50%	3164.2	24.938	0.128	2.546	7.593	0.481	71.467	129.739
75%	3356.4	41.157	0.170	2.583	8.0370	0.664	78.067	138.819
max	3721.6	126.895	0.323	2.737	9.965	1.394	107.99	261.496

Table 2. *Blind Validation Data Summary Statistics*

	DEPTH	**GR**	**NPHI**	**RHOB**	**PEF**	**RT**	**DT**	**DTS**
count	8171	8171	8171	8171	8171	8171	8171	8171
mean	3115.60	29.690	0.136	2.51	6.942	0.403	72.950	133.135
std	277.454	22.303	0.056	0.1188	1.036	0.373	8.894	17.715
min	2632.1	1.041	0.032	2.1608	4.511	-0.529	57.911	103.724
25%	2889.45	9.662	0.095	2.4898	6.259	0.248	67.094	120.925
50%	3098.3	29.466	0.129	2.5511	7.400	0.475	71.81	130.813
75%	3303.85	42.891	0.169	2.586	7.739	0.631	76.424	138.232
max	3641.8	126.895	0.323	2.737	9.965	1.2150	103.657	193.475

The dataset was not smoothed as this likely changes the nature of the data and is not recommended as described by [9]. Feature Scaling allows the machine learning algorithm not to be biased to the magnitude of the different data features in the dataset. Scaling of inputs can affect the gradients in deep learning by preventing the activation function from flattening towards 1 since this would make the gradient descent method ineffective [7], [39].

In this paper, experimentation with standardization produced better learning than normalization scaling which produced poor performance. This suggests that standardization better suits the well log data used in this study. Well log data is a sequential type of data. Therefore, this work followed the methodology of data splitting proposed by [37]. They demonstrated that the conventional methodology of shuffling and taking random samples of sequential geologic data to create a train-test dataset is flawed and it can produce unrealistically good results. This idea is also embraced here as having a model train on previous sections of a well log and predicting the next section has great application in real-time deployment.

MODEL ARCHITECTURE SECTION

The structure and hyper-parameters of the 1D-CNN and ANN models are presented in Table 3. The hyper-parameters of both the CNN and ANN models in Table 3 were determined using manual experimentation and those of the MLP model were obtained using Bayesian hyper-parameter optimization using hyper-opt.

Table 3. *Model architecture and hyper-parameters*

Model	Hyper-parameter	Value
CNN	convolution layers:	3
	filter sizer per layer:	128/64/32
	kernel size per layer:	3/2/2
	activation per layer:	relu/leakyReLU/LeakyReLU
	Pooling layers:	Global average pooling
	dense layers neurons:	1 (20 neurons)
	optimizer:	Adam
	learning rate:	0.001
	loss:	mse
	metric:	mae
	epochs (concrete dropout):	28
	epochs (deterministic):	27
ANN	hidden layers:	3
	output layers:	1
	neurons per hidden layer:	30/60/30
	learning rate:	0.001
	optimizer:	Adam
	activation function:	LeakyReLU
	loss:	mse
	metric:	mae
	epochs (concrete dropout):	99
	epochs (deterministic):	16

MLP	alpha: hidden_layer_sizes: learning_rate_init: max_iter: Shuffle: Solver:	8.527 2 1.973 500 False lbfgs

CONCRETE DROPOUT CNN

The CNN's architecture includes three (3) convolution layers. The first layer used a relu activation function while the two other layers had leakyReLU as activation function. The values from the convolution layers are passed to a single dense layer of 20 neurons. The hyper-parameters for this network were selected through an experimental process. The model used the Adam optimizer, with a learning rate of 0.001 mean absolute error (MAE) as the metric and mean squared error (MSE) as the loss function. MAE was chosen as it has the same unit as the target feature. The concrete dropout network has the length scale (l) and model precision (τ) as model parameters. The length scale values of the concrete dropout architecture imply that setting length scale l can show our belief or a priori assumption over function frequency characteristics of the data with small length-scale value meaning function output can change fairly rapidly, while large length-scale indicating function values changing slowly [34]. Also, the concrete dropout is used in all the convolutional layers in the CNN as well as in all the dense layers. The model was trained for 28 epochs with an early stopping of 10 epochs patience. The architecture of the concrete dropout 1D-CNN with the shaded neurons indicating neurons not active is shown in Figure 3.

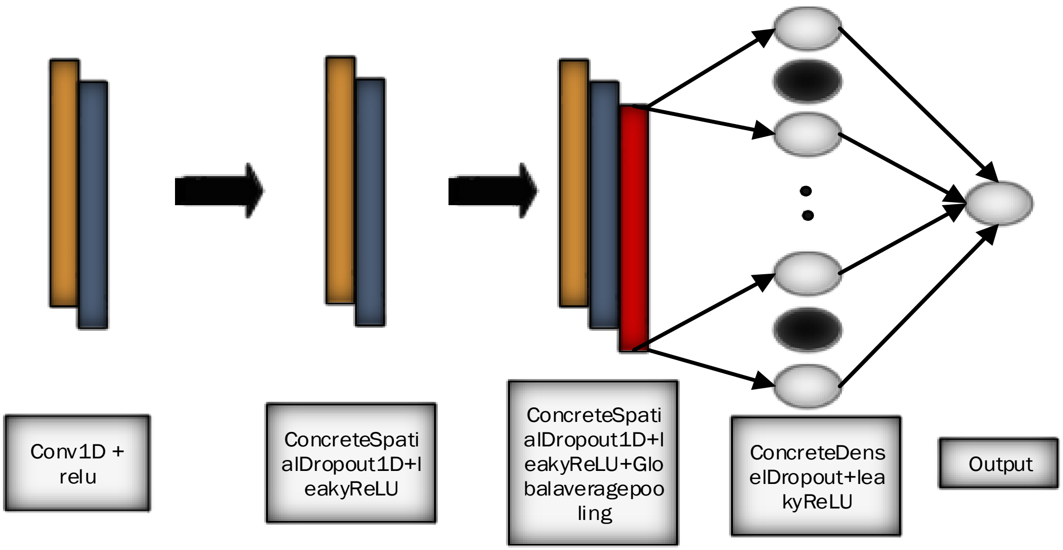


Figure 3. *Concrete dropout 1D-CNN model architecture*

CONCRETE DROPOUT ANN

A 3-hidden-layered ANN was also used in this study with 30, 60 and 30 neurons for each layer respectively. The same learning rate, optimizer and loss function as the 1D-CNN was used in the ANN model and training was completed at 99 epochs with early stopping. The same architecture and hyper-parameter were used for both deterministic and concrete dropout networks. The concrete dropout method, allows the dropout to be included in the convolution layers as well, while no dropout was added to the deterministic model. The MLP model was trained and tuned using hyper-opt [8], a Bayesian optimization library that directs its optimal hyper-parameter search in the direction of reducing loss function based on past trials which makes it more efficient than random search and grid search optimizations. Figure 4 shows the model architecture of the concrete dropout ANN model. The dark shaded neurons observed in Figure 4 shows the neurons turned off dynamically in the concrete dense layers during the particular forward pass through the network.

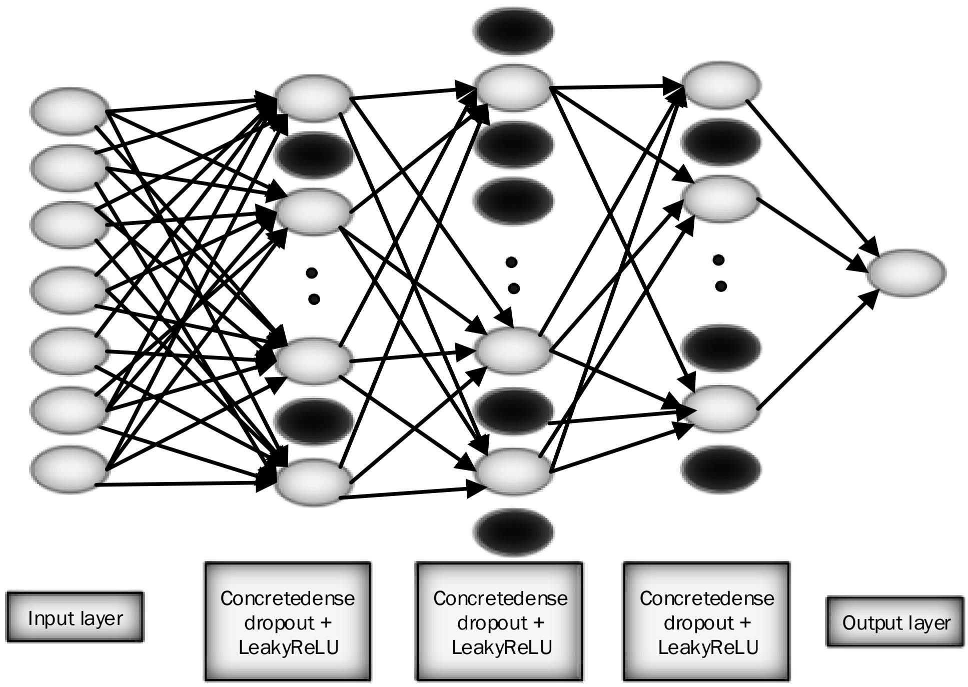


***Figure 4**. Concrete dropout ANN model architecture*

A forward pass through the network of 100 was chosen for both the concrete dropout ANN and CNN models as experimentation with a higher number of network sweeps did not produce significant positive change in the model performance, and for each prediction, the mean was computed. The CPU time for the 100-forward pass was less than 2 minutes for both models.

MODEL EVALUATION

This study chose the coefficient of determination (R^2), root mean squared error (RMSE), mean squared error (MSE) and MAE for the general model performance [2, 4].

1. Mean Absolute Error (MAE)

This is the average of the absolute difference between what the model predicts and the actual values of the DTS. Low values of MAE therefore imply that the model is performing well in its prediction and its unit is the same as that of the predictor DTS (us/hr). The predicted values of the model will lie within $ypredicted - MAE \leq ypredicted \leq ypredicted + MAE$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|DTS_i - DTS_{ipredicted}\right| \tag{1}$$

where N is the number of observations.

2. Root Mean Squared Error (RMSE)

It is the square root of the average of the square of residuals. An RMSE of zero (0) indicates a perfect fit. Low values indicate good model performance and therefore predictions.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(DTS_i - DTS_{ipredicted}\right)^2}{N}} \tag{2}$$

3. Coefficient of Determination (R^2)

This regression metric measures the goodness of the fit of the model predictions to the actual values of the DTS. The better the fit, the closer the value to one (1), with zero indicating no fit. It explains the extent to which the variance in one variable is explained by the second variable's variance.

$$R^2 = \frac{\sum_{i=1}^{N}\left(DTS_i - DTS_{ipredicted}\right)^2}{\sum_{i=1}^{N}\left(DTS_i - DTS_{mean}\right)^2} \tag{3}$$

4. Mean Absolute Percentage Error (MAPE)

This metric describes how off the prediction is from the actual values in percentages. A lower value closer to 0 indicates better prediction and a higher value means an increasing margin between actual and predicted values.

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|DTS_i - DTS_{ipredicted}\right|}{DTS_i} \tag{4}$$

RESULTS AND DISCUSSION

The addition of concrete dropout to a neural network architecture allows the selective turning off of the neurons in the layers of the neural network structure in Table 3. The 100 forward pass through the network produced different results with the mean of these model predictions computed to give the results of the concrete dropout versions of the CNN and ANN models in Table 4. The deterministic ANN and CNN models as well as the optimized MLP model are also shown in Table 4.

Table 4. *Prediction performance of each model*

		Deterministic CNN	Concrete Dropout CNN	Deterministic ANN	Concrete Dropout ANN	MLP
Train	R^2	0.8823	0.8950	0.9092	0.9301	0.9189
	RMSE	5.7781	5.6594	5.2994	4.7160	5.1297
	MAE	3.9220	3.8457	3.5089	3.0363	3.3731
	MAPE	0.0287	0.0284	0.0258	0.0225	0.0247
Blind data	R^2	0.9165	0.9359	0.9410	0.9548	0.9409
	RMSE	4.5830	4.1575	4.1222	3.6415	4.2840
	MAE	3.2355	2.9378	2.9770	2.4433	2.9118
	MAPE	0.0235	0.0216	0.0217	0.0179	0.0213

LEARNING AND PREDICTIVE PERFORMANCE OF CONCRETE DROPOUT CNN MODEL

Experimentation with the CNN model's parameters resulted in the training performance in Table 4. Deterministic CNN's prediction on blind data showed that the model learned the data as it could better predict the shear sonic log of the blind well. No dropout was applied to both the deterministic ANN and CNN models. The addition of concrete dropout increased the learning from 0.8823 in the deterministic case to 0.8950 in the concrete dropout case showing only a slight increase in learning compared to the deterministic CNN model scoring. However, this slight improvement produced a higher performance on the blind data thus signifying that it allowed the network to have better learning by turning off some of the neurons in the network thus preventing overfitting and leading to higher model improvement.

LEARNING AND PREDICTIVE PERFORMANCE OF CONCRETE DROPOUT ANN MODEL

The ANN model was built by experimentation with its hyper-parameters in Table 3. Compared to both the deterministic and concrete dropout CNN, the deterministic ANN model's performance on the training data achieves a 0.9092 R^2 score, outperforming even the concrete dropout CNN. ANN's model's prediction on blind data also had a higher performance of 0.9410 R^2 score. This performance exceeds that achieved by both the deterministic and concrete dropout CNN models. This performance increase could be attributed to the fact that the deterministic ANN architecture is better suited for the problem than the CNN model as both networks employed the same learning rate, loss metric and optimizer with the only difference being the number of dense layers and neurons. Retaining the same architecture, with the addition of concrete dropout to the ANN model, the concrete dropout ANN model showed a very pronounced increase in the model's learning, scoring an R^2 score of 0.9301 compared to the 0.9092 achieved in the deterministic ANN. Its prediction on the blind well had an RMSE of 3.6415, a notable decrease from the 4.716 recorded during the training. This shows that the model is not overfitting the training data but is able to generalize better to the blind validation data.

In comparison with the RMSE values of 4.5830, 4.1575 and 4.122 us/ft (microseconds per foot) recorded in the deterministic CNN, concrete dropout CNN and deterministic ANN respectively, RMSE of 3.6415 recorded in the concrete dropout ANN is a better performance. This showed that the addition of concrete dropout on a deterministic neural network architecture improves the model's learning performance as well as enhances its predictive capabilities.

COMPARISON WITH OPTIMISED MLP MODEL

The built ANN and CNN models were compared with a Bayesian optimized multilayer perceptron model using the hyper-opt library. The MLP model showed better learning on the training data compared with all the models except the concrete dropout ANN. Although the perceptron is a simple neural network with 2 hidden layers, it showed better learning than a deterministic deep ANN of 3 layers, it can be deduced that although the addition of concrete dropout improves model performance, as well as generalization, the model's overall predictive performance depends on the appropriate model chosen for the task.

The addition of the concrete dropout investigated in this study only serves to boost the performance of a model. The choice of an appropriate deterministic model must still be made first. This is evident from the fact that although simple, the optimized MLP model outperformed the deterministic CNN, concrete dropout CNN and deterministic ANN in both learning and blind prediction. The concrete dropout ANN model, however, outperformed the MLP model in all evaluation metrics investigated with an MAE of 2.4433 us/ft (microseconds per foot) compared to 2.9118 us/ft (microseconds per foot) indicating lower error during prediction.

IMPACT OF CONCRETE DROPOUT ON TRAINING EPOCH OF BOTH MODELS

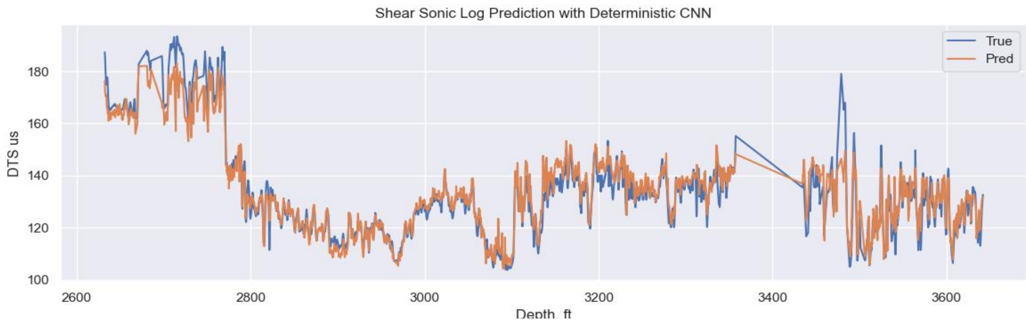
An interesting observation was that in the ANN case, the deterministic ANN network took 16 epochs to produce 0.9410 while with concrete dropout added, the same neural network took 99 epochs to train. The number of epochs from the findings of this study varied based on the neural network architecture, as deterministic CNN at 27 epochs had an R^2 of 0.9165 compared to the R^2 of 0.9359 achieved by the concrete dropout CNN at 28 epochs. Although the addition of concrete dropout was seen to increase training epochs, this is likely because longer epochs have been positively correlated with better model learning in capturing the relationship in the data. In this study, the concrete dropout addition to both the dense and convolutional layers of the CNN model produced improved model performance this is in opposition to the study by [11], who applied a variant called Monte Carlo dropout to convolution layers and reported negative performance.

SIGNIFICANCE OF THE FINDINGS

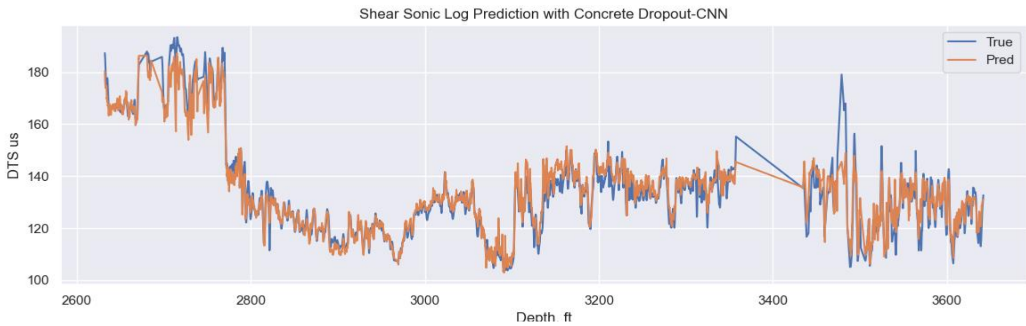
The findings of the study are significant in two major ways. It is proven to enhance a neural network's prediction and learning capability. This is because it automatically adapts the dropout rate of all layers to a value depending on the data. It also boosts performance because, at inference, it generates different output predictions for each forward pass through the network. As for the number of forward passes through the network, experimentation with 1000 iterations did not yield significant improvement over a 100 forward pass. The 100 forward pass generated 100 different model predictions and the mean value was computed. This makes the network's prediction more trustworthy and robust.

It is also significant in that it can be applied in real-time for shear sonic log prediction. In the traditional sense, models deployed in real-time are trained only on the training data, all other input data within or outside the range of training data can affect the model's performance. Although neural networks are great interpolators, when given data outside the range of training data, it might output poor predictions, leading to the model needing to be re-trained before re-deployment. The concrete dropout ANN models can adapt the dropout rate to the characteristics of the data at inference, selectively shutting off neurons to prevent overfitting and to better model the data.

Plots of the prediction performance of the deterministic and concrete dropout networks can be observed in Figure 5 and 6. In Figure 7, the optimized MLP model built using all training data with a 3-fold cross-validation is shown. The model although able to model the variation in the shear sonic log profile, struggles to do so in the shallow part of the log. This is improved in the concreted dropout ANN model.

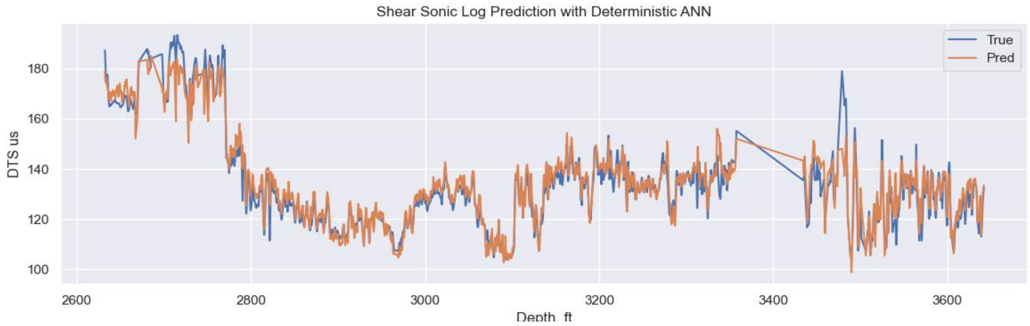


(a)

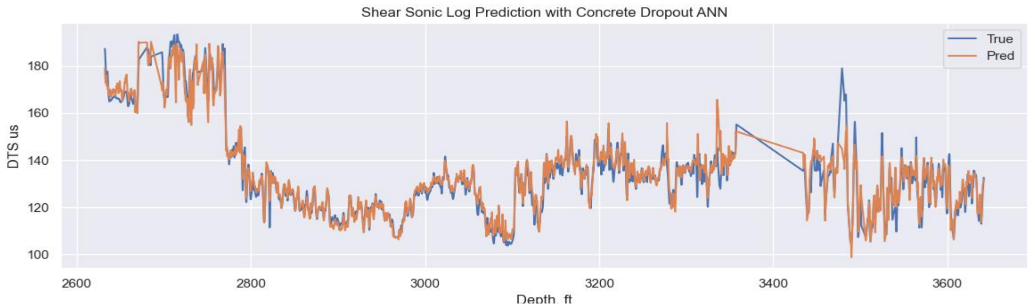


(b)

***Figure 5**. Comparison of CNN model results.*
(a) Deterministic CNN model, (b) Concrete Dropout CNN model



(a)



(b)

Figure 6*. Comparison of ANN model results.*
(a) Deterministic ANN model, (b) Concrete Dropout ANN model

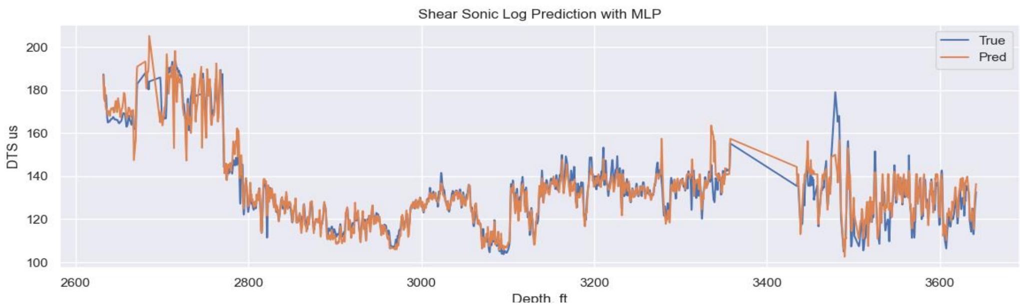


Figure 7*. Multilayer Perceptron Model results*

EXPERIMENTATION WITH LENGTH SCALE AND MODEL PRECISION CONCRETE DROPOUT HYPERPARAMETERS

The concrete dropout model architecture has two major hyper-parameters: length scale l and model precision τ. These parameters affect the model's performance and refer to the assumption about the data characteristics. Experimentation with these parameters is shown in Table 5. From the findings, for the concrete dropout CNN model, setting $l = e^{-2}$ and $\tau = 1$ produced the best model performance. Setting the length scale to higher values showed a predictive performance decline in the CNN model which was made worse by setting $\tau = 2$, with performance less than that seen in the deterministic CNN model. Varying both parameters, it can be deduced that increasing the length scale doesn't translate to higher model performance.

The concrete dropout ANN model however shows a fluctuating performance at increasing length scale and model precision, recording a high performance of 0.9501 at $\tau = 2$ and $l = e^{-2}$ and a low performance of 0.8909 at $\tau = 1$ and $l = e^{-3}$. The hyper-parameter range investigated in this study revealed that the best values for optimum performance for both CNN and ANN models were found to be $\tau = 1$ and $l = e^{-2}$. These values are the same as those recommended by [17]. These values produced better performance in this study regardless of the difference in data used for training.

Table 5. *Manual experimentation for both length scale and model precision parameters*

$\tau = 1$		Concrete dropout CNN				Concrete dropout ANN			
		$l = e^{-1}$	$l = e^{-2}$	$l = e^{-3}$	$l = e^{-4}$	$l = e^{-1}$	$l = e^{-2}$	$l = e^{-3}$	$l = e^{-4}$
Train	**R^2**	0.8905	0.8950	0.8760	0.8679	0.9014	0.9301	0.8867	0.9184
	RMSE	5.8290	5.6594	5.9263	6.1506	5.3815	4.7160	7.6195	5.0167
	MAE	4.0638	3.8457	4.0632	4.2807	3.4631	3.0363	4.6277	3.2301
	MAPE	0.0300	0.0284	0.0298	0.0316	0.0254	0.0225	0.0311	0.0237
Blind data	**R^2**	0.9328	0.9359	0.9202	0.9066	0.9351	0.9548	0.8909	0.9450
	RMSE	4.3158	4.1575	4.4782	4.8759	4.1706	3.6415	7.3037	3.8866
	MAE	3.0953	2.9378	3.1473	3.5071	2.8702	2.4433	4.3392	2.6428
	MAPE	0.0227	0.0216	0.0229	0.0256	00212	0.0179	0.0282	0.0192
$\tau = 2$		**Concrete dropout CNN**				**Concrete dropout ANN**			
		$l = e^{-1}$	$l = e^{-2}$	$l = e^{-3}$	$l = e^{-4}$	$l = e^{-1}$	$l = e^{-2}$	$l = e^{-3}$	$l = e^{-4}$
Train	**R^2**	0.8642	0.8312	0.8413	0.8279	0.9022	0.8886	0.9296	0.9067
	RMSE	6.1319	6.5962	6.4207	6.6278	5.3010	5.5184	4.9198	5.7596
	MAE	4.1614	4.5189	4.3964	4.5791	3.3889	3.5659	3.2170	3.8021
	MAPE	0.0306	0.0328	0.0321	0.0335	0.0248	0.0259	0.0238	0.0279
Blind data	**R^2**	0.9152	0.8699	0.8966	0.8753	0.9389	0.9225	0.9501	0.9481
	RMSE	4.5521	5.4781	4.9033	5.3641	3.9873	4.3829	3.9857	4.076
	MAE	3.1975	3.8629	3.4523	3.833	2.6300	2.9710	2.900	2.8595
	MAPE	0.0232	0.0277	0.0249	0.0276	0.0191	0.0216	0.0213	0.0208

CONCLUSIONS

In this work, we demonstrated the application of concrete dropout to both CNN and ANN models for shear sonic log prediction and compared their performance to the same network models built without dropout layers added. From the results obtained, it can be concluded that the addition of concrete dropout to the initial deterministic architecture improves the overall predictive performance of the models over its deterministic type as it leads to better training and enhanced prediction on blind data.

The concrete dropout ANN model in comparison with an optimized MLP model, deterministic models and the concrete dropout CNN was found to better model the shear sonic log profile in all evaluation metrics. The deterministic ANN model was also observed to perform better than the concrete dropout CNN model. This leads to the conclusion that although the addition of concrete dropout leads to significant improvement in a model's performance, a model coupled with concrete dropout will not always outperform other models on the same data. The appropriate model for the data must still be firstly chosen, before applying the concrete dropout to boost its performance.

The concrete dropout ANN investigated in this study can be applied in model deployment, as the dropout rate in traditional neural network models is fixed depending only on training data, the dropout rates are allowed to be learnt and dynamically adapted to even new data at prediction time allowing the model to be more robust and perform better as it selects the dropout for each layer based on the data.

It is recommended that a more comprehensive search be carried out to investigate the full effect of the length scale and model precision parameters on the data and performance of the models. In addition to boosting the performance of neural networks, the concrete dropout architecture also allows the quantification of epistemic and aleatoric uncertainty as it outputs a distribution rather than point estimates, this is an aspect that will be investigated in a subsequent paper.

REFERENCES

[1] Ahmed S.A., MonaLisa, Hussain, M., Khan, Z.U., Supervised machine learning for predicting shear sonic log (DTS) and volumes of petrophysical and elastic attributes, Kadanwari Gas Field, Pakistan. Frontier Earth Sciences 10:919130. DOI: 10.3389/fearst.2022.919130, 2022.

[2] Akhilesh, N., Reddy, K., Patra, S. Compressional and shear velocity prediction using artificial neural networks in Bhagyam Field, Barmer Basin, Rajasthan Geohorizons, vol. 29(1), 2024.

[3] Akinyemi, O.D., Elsaadany, M., Siddiqui, N.A., Elkurdy, S., Olutoki, J.O., Islam, M.M., Machine learning application for prediction of sonic wave transit time – A case of Niger Delta basin. Results in Engineering, 20. https://doi.org/10.1016/j.rineng.2023.101528, 2023

[4] Al Ghaithi A., Deep Learning Methods for Shear Log Predictions in the Volve Field Norwegian North Sea. Master's thesis, Colorado School of Mines, 2020.

[5] Ali, M., Jiang, R., Huolin, M., Pan, H., Abbas, K., Ashraf, U., Ullah, J. Machine learning – A novel approach of well logs similarity based on synchronization measures to predict shear sonic logs. Journal of Petroleum Science and Engineering 203, 108602, https://doi.org/10.1016/j.petrol.2021.108602, 2021.

[6] Atolagbe, J.M., Akindele, O.K., Log1DNet: A deep learning Architecture for sonic log prediction for seal rock identification in carbon capture and storage projects.

Asian Journal of Geological Research. vol 7 (3), pp. 232-244. https://journalajoger.com/index.php/AJOGER/article/view/167, 2024.

[7] Belyadi, H., Haghighat, A., Machine Learning Guide for Oil and Gas Using Python: A step-by-step Breakdown with Data, Algorithms, Codes and Applications. Gulf Professional Publishing (Elsevier Imprint): Cambridge, United States. 2021

[8] Bergstra, J., Yamins, D., Cox, D.D., Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. To appear in Proc. of the 30th International Conference on Machine Learning (ICML, 2013). 2013

[9] Bizhani, M, Kuru, E., Towards drilling rate of penetration prediction: Bayesian neural networks for uncertainty quantification. Journal of Petroleum Science and Engineering. 219, https://doi.org/10.1016/j.petrol.2022.111068, 2022

[10] Cajahuaringa, A., Palacios, R.A., Mauricio Villanueva, J.M., Morales-Villanueva, A., Machuca, J., Contreras, J., Rodríguez Bautista, K., Uncertainty Evaluation of a Gas Turbine Model Based on a Nonlinear Autoregressive Exogenous Model and Monte Carlo Dropout. Sensors, 24, 465. https://doi.org/10.3390/s24020465, 2023

[11] Choubineh, A., Chen, J., Coenen, F., Ma, F., Applying Monte Carlo Dropout to Quantify the Uncertainty of Skip Connection-Based Convolutional Neural Networks Optimized by Big Data. Electronics, 12, 1453. https://doi.org/10.3390/electronics12061453, 2023

[12] Contreras, J., Bocklitz, T., Enhancing decision confidence in AI using Monte Carlo dropout for Raman spectra classification, Analytica Chimica Acta, 1332, https://doi.org/10.1016/j.aca.2024.343346, 2024

[13] Equinor, Background disclosing all Volve data. https://www.equinor.com/en/news/2018-disclosing-volve-data, accessed 2 January 2025.

[14] Fanchi, J.R., Christiansen, R.L., Introduction to petroleum engineering. Hoboken, New Jersey: John Wiley & Sons, 2017

[15] Faust, L.Y., A velocity function including lithologic variation, Geophysics, 18(2), 271-288. https://doi.org/10.1190/1.1437869, 1953.

[16] Gal, Y., Ghahramani, Z., Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning" by (http://arxiv.org/abs/1506.02142), 2016

[17] Gal, Y., Hron, J., Kendal, A., Concrete Dropout. 31st Conference on Neural Information Processing Systems (NIPS) CA, USA, 2017

[18] Gardner G.H.F, Gardner L.W, Gregory A.R., Formation velocity and density - the diagnostic basics for stratigraphic traps. Geophysics, 39:770–780. https://doi.org/10.1190/1.1440465, 1974

[19] Gomaa, S., Shahat, J.S., Fotouh, T.M., Khaled, S., A neural network model for predicting shear wave velocity using well-logging data. Arabian Journal for Science and Engineering. https://doi.org/10.1007/s13369-024-09150-y, 2024

[20] Greenberg M.L., Castagna J.P., Shear-wave velocity estimation in porous rocks: theoretical formulation, preliminary verification and applications. Geophys Prospect 40:195–209. https://doi.org/10.1111/j.1365-2478.1992.tb00371.x, 1992

[21] Hermosilla, D.M., Codorniú, R. T., Baracaldo, R.L., Sagaró-Zamora, R., Rodriguez, D., Villarejo-Mayor, J.J., Núñez-Álvarez, J.R., Monte Carlo Dropout for Uncertainty Estimation and Motor Imagery Classification. Sensors, 21, 7241. https://doi.org/10.3390/s21217241, 2021

[22] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580, 2012

[23] Kim, J., Synthetic shear sonic log generation utilizing hybrid machine learning techniques. Artificial Intelligence in Geosciences, 3, pp. 53-70. https://doi.org/10.1016/j.aiig.2022.09.001, 2022

[24] Koomson, A.F. Sitsofe, A.P., Teddy, A.O., Africa, A., Estimating shear wave velocity using artificial networks: A case study of the Tano North Field. https://arxiv.org/pdf/2011.12109, 2020

[25] Lakshminarayanan, B., Pritzel, A., Blundell, C., Simple and scalable predictive uncertainty estimation using deep ensembles.In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, 2017

[26] Lambert, B., Forbes, F., Doyle, S., Dehaene, H., Dojat, M., Trustworthy clinical AI solutions: A unified review of uncertainty quantification in deep learning models for medical image analysis. Artificial Intelligence in Medicine 150. https://doi.org/10.1016/j.artmed.2024.102830. 2024

[27] Lee, J., Chen, Y., Dommisse, R., Huang, G., Savvaidis, A., Basin-scale prediction of S-wave sonic logs using machine learning techniques from conventional logs. Geophysical Prospecting. 72, pp. 2557-2579. DOI: 10.1111/1365-2478.13527, 2023

[28] Melo, F.F., Perez, V., Lee, C., Shear sonic log prediction with deep neural networks: An example from the Gulf of Mexico. Geoconvention: Calgary, Canada. 2022

[29] Nero, C., Aning, A.A., Dnuor, S.K., Mensah, V., Prediction of compressional sonic log in the western (Tano) sedimentary basin of Ghana, West Africa using supervised machine learning algorithms. Heliyon, 9. https://doi.org/10.1016/j.heliyon.2023.e20242, 2023

[30] Onalo, D., Adedigba, S., Khan, F., James, L.A., Butt, S., Data-driven model for sonic well log prediction. Journal of Petroleum Science and Engineering, 170, pp. 1022-1037. https://doi.org/10.1016/j.petrol.2018.06.072, 2018

[31] Onalo, D., Adedigba, S., Oloruntobi, O., Khan, F., James, L.A., Butt, S., Data-driven model for shear wave transit time prediction for formation evaluation. Journal of Petroleum Exploration and Production Technology. 10. pp. 1429-1447. https://doi.org/10.1007/s13202-020-00843-2, 2020

[32] Pendrel, J., Schouten, H., Facies – The drivers for modern inversions. The Leading Edge, 39(2), 102-109. 2020

[33] Rajabi, M., Hazbeh, O., Davoodi, S. et al., Predicting shear wave velocity from conventional well logs with deep and hybrid machine learning algorithms. Journal of Petroleum Exploration and Production Technology, 13, pp. 19-42. https://doi.org/10.1007/s13202-022-01531-z, 2022

[34] Seoh, R., Qualitative analysis of Monte Carlo dropout. arXiv:2007.01720v1 [stat.ML] 2019

[35] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R., Dropout: A simple way to prevent neural networks from overfitting. Journal of Machine Learning Research, 15, pp. 1929-1958. 2014

[36] Tony Liu, F., Ming Ting, K., Zhou, Z.-H., in 2008 Eighth IEEE International Conference on Data Mining. Isolation forest. https://doi.org/10.1109/ICDM.2008.17, 2008

[37] Tunkiel, A.T., Sui, D., Wiktorski, T., Reference dataset for rate of penetration benchmarking. Journal of Petroleum Science and Engineering 196, 108069, https://doi.org/10.1016/j.petrol.2020.108069, 2020

[38] Wong W.K., Nuwara Y., Juwono F.H., Motalebi F., Sonic waves travel-time prediction: when machine learning meets geophysics. International Conference on Green Energy, Computing and Sustainable Technology, 159-163. Available: https://doi.org/:10.1109/GECOST55694.2022.10010361, 2022

[39] Wong, Y., Wong, S., Using convolutional neural networks and Monte Carlo dropout to generate synthetic well logs with accurate uncertainty estimation. Proceedings of the 8th International Symposium on Geotechnical Safety and Risk (ISGSR) (pp. 584-593). Singapore. 2022

Received: March 2025; Revised: April 2025; Accepted: April 2025; Published: April 2025