Rapid Predicted Profiles of Elemental Contents Along Slabbed Cores from Probe Magnetic and Well Log Data Using a Machine Learning Approach Trained on XRF Data

Emmanuel Okwoli, *David K.* Potter[*] and *Toan H.* To

Department of Physics, University of Alberta, Edmonton, Alberta, T6G 2E1, Canada

Abstract. X-ray fluorescence (XRF) analysis of core samples can provide high-resolution profiles of elemental contents of core samples. However, there are some limitations, such as the high capital cost of XRF equipment, the measurements can take several minutes to perform, and strict training and safety procedures are required when working with X-rays. Therefore, a cost-effective, rapid, simple method to predict profiles of elemental contents from core screening or well log data would be useful. The aim of this study was to predict elemental contents of iron, potassium and aluminium (all important components of permeability controlling clays for the studied samples) within shale and oil sand intervals of two Albertan test wells from combinations of rapid, high-resolution probe magnetic susceptibility measurements on slabbed core (using a small hand held probe) and well log data. This involved a machine learning approach that first created a training dataset relating XRF measurements to the probe magnetic and well log data. Two neural network algorithms were tested and compared for training this large dataset and the subsequent prediction of elemental contents: multi-layer perceptron (MLP) artificial neural networks (ANNs) and 2D convolution neural networks (2D CNNs). ANNs have proved useful in many petrophysical applications, and 2D CNNs (which first convert the input data to images) have recently shown improved results over some other neural networks. The 2D CNNs and ANNs both produced excellent predictions of the elemental contents of iron, potassium, and aluminium, comparable to the measured XRF values, in large intervals of the two test wells (the 2D CNN predictors marginally out performed ANNs in some cases). Furthermore, predictions using much smaller training datasets also gave very similar results, indicating that limited representative training data can still provide excellent predictions, saving one time and cost. The presence or absence of the probe magnetic susceptibility data particularly influenced the prediction of the elemental iron contents. In the absence of the probe magnetics, the performances of the 2D CNN and ANN iron content predictors in both test wells was somewhat reduced. The probe magnetic susceptibility results are very sensitive to small amounts of minerals containing iron (such as siderite and magnetite) and demonstrated the usefulness of the probe magnetic data for predicting iron content. These machine learning predictors are complementary to XRF measurements, since we initially use some representative XRD data to train the predictors. However, once we have trained good predictors we can rapidly and cheaply predict the elemental contents throughout large intervals from probe magnetic and well log data.

1 Introduction

X-ray fluorescence (XRF) analysis of core samples has provided a high-resolution and less destructive means of characterizing the elemental contents of rocks, including unconsolidated core samples [1-4]. However, there are some limitations to XRF techniques. For instance, the method is expensive as the cost of XRF equipment ranges from $50,000 - $100,000 US. Also, each measurement takes several minutes to perform. Another limitation is that most commercial XRF devices require some adjustments to measure the elemental contents of earth materials with fewer than eleven elements [5]. As a result of these and other limitations, it would be useful to provide a cost-effective and rapid method that can establish a relationship between XRF elemental contents and probe core screening data or well log data. One way of doing this is to relate measured XRF elemental contents on some representative samples with probe core screening and/or well log data using neural networks (NNs), in order to train predictors that can subsequently provide profiles of elemental contents over large intervals from probe core screening and/or well log data rapidly and cheaply. This study aimed to produce predictors of elemental contents from high-resolution probe magnetic susceptibility measurements on slabbed core and/or well log data that were trained in combination with some initial XRF measurements. Well

---

[*] Corresponding author: dkpotter@ualberta.ca

log data are readily available from most wells, and magnetic probe data can be rapidly (each measurement can be made in a few seconds) and cheaply measured on rock cores using a small portable MS2E probe device [4]. The magnetic probe equipment (the probe sensor and recording meter) is relatively inexpensive, with both pieces of equipment costing about $4,500 US in total. Hence, if one can successfully predict elemental contents from a combination of high-resolution, rapid, cheap probe magnetic susceptibility core measurements and/or well log data, it would reduce the cost and time that would otherwise be spent making expensive and time consuming XRF measurements.

Two types of neural networks were used for data training (relating XRF data to probe magnetic and well log data) and subsequent prediction of elemental contents: two-dimensional convolution neural networks (2D CNNs) and multi-layer perceptron (MLP) artificial neural networks (ANNs). Zhong et al. [6] showed that CNNs could be applied to solve regression problems. They used five well log parameters as inputs to the network and their CNN architecture was comprised of two convolution layers, two fully connected layers, and one output layer representing the predicted permeability. Their findings showed that the CNN predictors performed better than the three other neural network methods applied in the study. Motivated by these results and other related research [7], we applied a similar approach to predict elemental contents from probe magnetic and well log data. The network architectures used in our research contain more convolution layers than those employed by [6], due to the complexities of the problem at hand. We chose 2D CNNs for various reasons. Firstly, our dataset is two-dimensional. Each row represents a training instance while each column represents a geologic parameter (probe magnetic and well log data). Secondly, 2D CNNs are more efficient and cost-effective than 3D CNNs. Multi-layer perceptron ANNs were also used for comparison, since they have previously been successfully applied to predict various petrophysical parameters, [e.g., 8-12].

Data from two oil sands wells (Well 02 and Well 03) in northern Alberta, Canada, was used in this study. Well 02 is located in the northern region of Cold Lake, while Well 03 is located in the northeastern region of the Athabasca oil sands near Fort McMurray.

2 Dataset and Methods

2.1 Dataset

A large dataset was used for this study comprising well log data, high-resolution low field probe volume magnetic susceptibility measurements on slabbed core, and high-resolution probe XRF measurements on slabbed core from the two wells. The available well log data included the total gamma ray, spectral gamma ray, and spontaneous potential. The magnetic susceptibility measurements were taken using a Bartington MS2E probe [4], and the XRF measurements using a a portable XRF analyzer, the Thermo Scientific Niton XL3t [13].

Three main lithological units were identified in Well 03: shale, inclined heterolithic stratification (IHS) beds comprising interbedded sand and clay, and clean sand [4]. Well 02 had two main lithological units: shale and clean sand [4]. Both wells were drilled in the McMurray formation in the Mannville Group of the Western Canadian Sedimentary Basin (WCSB). The data from both wells were randomized, and initially 70% of the data from each well were combined to train the predictors. The predictors were then tested on the remaining 30% of the data from Well 02 and Well 03. Subsequently, predictors were trained on just 30% of the data from each well, and tested on the remaining 70%. The purpose of this was to see whether a smaller (but still representative) training dataset can provide good predictions comparable to using the much larger training dataset. The advantage of using a smaller training dataset is that it is more time and cost-effective, both in terms of initially obtaining the measurements for the training data, and subsequently for running the neural network software.

2.2 2D CNN Training and Architecture for Predicting Elemental Contents

The input data to a CNN has a grid-like topology and can be presented as 1D, 2D or 3D data pixels. For this reason, CNNs are mainly used to solve classification problems, however they can also be used to solve regression problems as in the present study. One way to do this is to first convert the input data to images. 2D CNNs were most appropriate for the present study, as mentioned above, and so the probe magnetic and well log data used to train the networks were converted to black and white images before being fed into the networks. These images effectively capture spatial patterns by sliding a 2D filter (called an image kernel) over the image and detecting local features. This is particularly useful for black and white images, where each pixel only represents one value, as opposed to colour images where each pixel is composed of three separate colour channels. The steps to convert discrete input variables to images are described below:

1. The input variables were converted from decimal to binary strings (feature strings). The IEEE 754 floating-point format (single-point precision) was used to convert these input variables to binary strings [14]. IEEE 754 single precision is a standard for representing floating-point numbers in computers. It uses 32 bits to represent a floating-point number, with 24 bits representing the mantissa and 8 bits representing the exponent. Figure 1 shows the details of the conversion process. Figure 1 (a) shows the probe magnetic and well log values (decimal numbers). Figure 1 (b) shows the 32-bit binary string for each input variable.

2. Next, we converted the binary strings to black and white images (Figure 1 (c)) using a Python code.

3. Figure 1 (c) shows the generated images stacked to create feature images, and Figure 1 (d) shows a schematic of the feature images at different depths.

The network architecture used in this research was inspired by the VGG16 neural network [15]. This consists of input data (images), convolution layers, pooling layers, fully connected layers (each fully onnected layer comprising a certain number of neurons), and the output parameter that is predicted. The VGG16 neural network has recently been used across different research areas [e.g., 16, 17]. Due to the dimensions of our images (maximum of 32 x 6 pixels, Figure 1 (c)), we only implemented a part of the VGG16 architecture in this research. We used four convolution layers. The first convolutional layer had 64 filters with a 3×3 pixel image kernel (e.g., the second figure from the left in Figure 1 (e)) and a stride distance of 1, which means it moves the kernel one place to the right over the input image. The input data were fed into the convolution layer as dimension $m \times n \times c$ pixel arrays, where the image width $m = 32$, the image height $n = 6$, and the "colour" channel $c = 1$ for a black and white image. As the image kernel scans over the input image, its values are multiplied by the pixel values of the normalized image (third and fourth figures from the left in Figure 1 (e)). This process is referred to as elementwise multiplication. The products are then summed and stored as feature strings in a specific location. This process is repeated until the kernel scans the entire image, and the values are stacked together to produce a feature image (Figure 1 (c)). This generated 64 feature images per each instance of the training sample. To avoid losing information along the borders, we padded our images with zeros on all four borders (a schematic example for a smaller image is shown in the far-left part of Figure 1 (e)). The second convolution layer had the same features as the first, i.e. 64 filters with a 3×3 pixel image kernel and a stride distance of 1. A pooling layer (i.e., downsampling layer) with a 2×2 pixel maximum pooling image kernel, and a stride distance of 2 and no padding, was then applied after the second convolution layer. The function of the pooling layer was to reduce the dimensions of the feature images. This is important as it increases efficiency by reducing the training time as the data size is reduced, and it helps to prevent overfitting of the training data. This reduced the dimensions of the images from 32 x 6 to 16 x 3 pixels (for each of the 64 feature images). The third and fourth convolution layers had 128 filters, each with a 3×3 pixel image kernel and a stride distance of 1, and produced 128 feature images each. A pooling layer with a 2×2 pixel maximum pooling image kernel with a stride distance of 2 and no padding was applied after the fourth convolution layer, reducing the dimensions to 8 x 1 pixels (for each of the 128 feature images). The output was then fed into two fully connected layers, consisting 120 and 70 neurons respectively. The optimum number of neurons in each layer was determined experimentally by trial and error. The fully connected layers work like a multilayer perceptron (MLP) model in which the output neurons from one layer are connected to the next, and so on. The output from each convolution layer and fully connected layer was achieved using a rectified linear unit (ReLU) [18]. The final output used a linear activation function to give a continuous output of iron, potassium or aluminium content.

Several different cases (different combinations of the probe magnetic susceptibility and well log data) were considered for training the predictors. Python was used to write the codes, along with other open source supporting deep learning libraries TensorFlow [19] and Keras (https://keras.io/), to train the 2D CNN predictors, while the Scikit Learn library [20] was used to carry out data analysis, like randomization and metrics evaluation.

The K-fold cross-validation approach was used to train the 2D CNN predictors. It involves randomly splitting the dataset into different folds. In this study, each training dataset was divided into 10 different folds with an equal number of data points (probe XRF, probe magnetics and well log values) per fold. Predictors were initially trained on 70% of the entire dataset, and subsequently on 30% of the entire dataset to test whether a smaller but representative training dataset could also perform well. We included a random seed in the code to ensure that the results were repeatable. For each round of training, the predictors were trained on 9 folds, while 1 fold was used to validate the predictor's performance. At the end of the training, the predictors obtained for the 10 different folds were averaged across all folds, and this was used to make predictions on the test datasets. Early stopping was also used during training to prevent overfitting. The early stopping algorithm monitors the training of each fold and stops the training if the validation loss continues to increase for five consecutive iterations (while the training loss decreases). The coefficient of determination (R^2), mean absolute error (MAE), root mean squared error (RMSE), standard deviation of the error (residual standard deviation (RSD)) and the mean error (ME) (also called the standard error of the mean) between the measured and predicted parameters for the test datasets, were used to evaluate the performance of the predictors on the test datasets.

2.3 ANN Training and Architecture for Predicting Elemental Contents

The data from Wells 02 and 03 were also used to train multi-layer perceptron ANNs. The main difference was that the inputs to the ANN predictors were decimal data instead of images for the 2D CNN predictors. Python was again used to write the codes, along with TensorFlow [19] and Keras, to train the ANN predictors, while the Scikit Learn library [20] was used to carry out data analysis. The ANNs were trained using identical cases of magnetic probe and well log data to the 2D CNN cases. As before, the datasets from both wells were randomized and divided into training and test datasets similar to the 2D CNN training and testing. Different architectures were used to train the ANNs, using multi-layer feed-forward networks. There is no rule of thumb for choosing the number of hidden layers and artificial neurons. However, care must be taken as too many hidden layers can cause overfitting of the training data. We experimented with different numbers of hidden

layers and discovered that two or three hidden layers worked best for all the predictions. 10, 7, and 5 artificial neurons were used where three hidden layers were required to make predictions. 10 and 7 artificial neurons were used with two hidden layers in some instances, and 7 and 5 artificial neurons in other training instances. The output layer had one neuron corresponding to the predicted elemental contents (iron, potassium, or aluminium). The ANN predictors were also trained using the K-fold cross-validation approach. The data were randomized and divided into 10 folds. The ANN predictors were trained on 9 folds per each round of training, while 1 fold was used to validate the predictor's performance. Again, the predictors trained for the 10 different folds were averaged across all folds, and this was used to make predictions on the test datasets. Similar evaluation metrics to those applied to the 2D CNNs were also used to evaluate the performance of the ANN predictors on the test datasets.

3 Results and Discussion

3.1 2D CNN Results and Discussion

Table 1 summarizes the results obtained for predicting iron content (cases 1–5). All "a" cases (in green type in Table 1) represent cases where 70% of the entire dataset was used to train the predictors, whereas all "b" cases (in black type in Table 1) represent cases where 30% of the entire dataset was used to train the predictors. The table gives the values of R^2, MAE, RMSE, RSD, and ME between the measured and predicted parameters for the training and test datasets. As expected, the predictors performed excellently on all cases involving the training dataset with high R^2 values and low errors.

Significantly, the predictors performed almost equally well on the test data in Wells 02 and 03. For iron, the predictors gave R^2 values ranging from 0.83 to 0.99 for cases 1–4. Case 5, which was trained only on the well log data, showed somewhat lower R^2 values of 0.64 to 0.89 compared to cases 1–4, whilst the MAE and RMSE were slightly higher. The lower performance of the predictors in the absence of probe magnetics as an input is likely due to the strong magnetic susceptibility of iron containing minerals (such as the ferrimagnetic mineral magnetite, Fe_3O_4, and the paramagnetic carbonate mineral siderite, $FeCO_3$). Hence, the absence of the probe magnetic data for cases 5a and b likely impacted the performance of the predictors.

For all cases, the results of the predictors trained with 70% of the entire dataset marginally outperformed those trained with 30%. The R^2 values were either the same or marginally higher, and the MAE and RMSE values were generally slightly lower for the predictors trained with 70% of the entire dataset. Nevertheless, the predictors trained on 30% of the entire dataset still performed very well, demonstrating that these elemental contents can be predicted from minimal representative training data. Note, however, that they are not identical situations, as the predictors trained on 70% of the entire dataset only test the remaining 30%, whereas the predictors trained

on 30% of the entire dataset test the remaining 70%, so the R^2 and error values are derived from different numbers of data points.

Figure 2 (a) displays a crossplot of the measured XRF (vertically averaged every 0.3048 m = 1ft) versus the 2D CNN predicted elemental iron contents for case 1a (test Well 03) trained on 70% of the entire dataset. The coefficient of determination R^2 of 0.99, MAE of 0.10, and RMSE of 0.13 demonstrated excellent predictions in this test well. Also, the predicted and measured iron content followed the same trend with depth (Figure 2 (b)) for the entire study interval. One notable feature of the depth plot is that the 3 main lithological units (shale, IHS beds, and clean sand interval) are clearly apparent. The shale interval generally has the highest iron contents. The IHS beds have intermediate iron content values, lower than the shale and slightly higher than the clean sand interval below. Figures 2 (c) and (d) show the crossplot and the depth profile comparing the XRF measured and 2D CNN predicted iron content values for case 1b (test Well 03) trained on 30% of the entire dataset. The predictors were again very good, with an R^2 value of 0.97, MAE of 0.13, and RMSE of 0.20. Note that since for Figures 2 (a) and (b) 30% of the total dataset is tested, whereas for Figures 2 (c) and (d) 70% of the total dataset is tested, there are a larger number of data points in Figures 2 (c) and (d).

We also made detailed comparison tables and figures for aluminium and potassium contents, but due to the 12 page limit of this paper we're unable to present everything here and instead will summarise the key results and show some representative examples. The performance of the predictors on the test wells for aluminium was excellent for all cases, with R^2 values ranging from 0.91 to 0.96 for test Well 02 and 0.85 to 0.96 for Well 03. The error values were again low (MAE ranged from 0.28–0.42 in test Well 02 and 0.22–0.40 in test Well 03, whilst RMSE ranged from 0.37–0.56 in Well 02 and 0.27–0.50 in Well 03). The absence of probe magnetic data for cases that were just trained on well log data did not have a detrimental effect on predicting aluminium elemental contents in this case, which was expected since aluminium does not have a high magnetic susceptibility signal (compared to iron). Figure 3 (a) shows a crossplot of the XRF measured (vertically averaged every 0.3048 m = 1ft) versus the 2D CNN predicted aluminium contents trained on probe magnetics and all available well log data for test Well 03 (trained on 70% of the entire dataset). A coefficient of determination R^2 of 0.96, MAE of 0.22 and RMSE of 0.27 were obtained. Figure 3 (b) shows the associated depth profile of the XRF measured and 2D CNN predicted elemental aluminium contents. The predicted aluminium contents follow the trend of the measured values in most of the studied interval. Figures 3 (c) and (d) show the corresponding crossplot and depth profile comparing the XRF measured and 2D CNN predicted values for an identical case in test Well 03, with the only difference being that the predictor was trained on just 30% of the entire dataset. The predictors again did well

in predicting the aluminium content, with an R^2 value of 0.88, MAE of 0.33, and RMSE of 0.45.

Similarly, the performance of the predictors on the test wells for potassium was excellent for all cases, with R^2 values ranging from 0.91 to 0.96 for test Well 02 and 0.83 to 0.93 for Well 03. The error values were also low (MAE ranged from 0.07–0.12 in test Well 02 and 0.08–0.12 in test Well 03, whilst RMSE ranged from 0.10–0.16 in both Well 02 and Well 03). Figure 4 (a) shows a crossplot of the XRF measured (vertically averaged every 0.3048 m = 1 ft) versus the 2D CNN predicted potassium contents trained on probe magnetics and all available well log data for test Well 02 (trained on 70% of the entire dataset). A coefficient of determination R^2 of 0.96, MAE of 0.07 and RMSE of 0.10 were obtained. The depth plot (Figure 4 (b)) showed a good agreement between the XRF measured and predicted potassium contents for the study interval, and clearly differentiated the two main lithological units (shale and clean sand) in Well 02. The shale interval generally has higher potassium content, whereas the clean sand interval has lower potassium content The interval between 380.50–389.60 m was missing core. Figures 4 (c) and (d) show the corresponding crossplot and depth profile comparing the XRF measured and 2D CNN predicted values for an identical case in test Well 02 with the only difference being that the predictor was trained on 30% of the entire dataset. An R^2 value of 0.93, MAE of 0.10, and RMSE of 0.14 were obtained.

3.1 ANN Results and Discussion

Overall, the performance of the ANN predictors was either equally as good or very slightly lower than that of the 2D CNN predictors for identical cases. For example, iron content predictions for equivalent cases 1-4 gave ranges for the following parameters: R^2 0.86–0.93, MAE 0.28–0.41, and RMSE 0.44–0.85 for test Well 02, and R^2 0.93–0.98, MAE 0.12–0.18, and RMSE 0.16–0.29 for test Well 03. For case 5 where the predictors were trained solely on well log data (without input from probe magnetic susceptibility) the R^2 values were lower and the errors higher as follows: R^2 0.54–0.69, MAE 0.61–0.71, and RMSE 0.94–1.18 for test Well 02, and R^2 0.74–0.86, MAE 0.29–0.38, and RMSE 0.39–0.57 for test Well 03. This again demonstrated the importance of including probe magnetic susceptibility training data for prediction of iron contents.

Aluminium content predictions for equivalent cases 1-5 gave ranges for the following parameters: R^2 0.92–0.94, MAE 0.27–0.39, and RMSE 0.38–0.50 for test Well 02, and R^2 0.70–0.92, MAE 0.29–0.45, and RMSE 0.43–0.68 for test Well 03. The absence of probe magnetic susceptibility data for case 5 did not have any significant effect on the results for Well 02 compared to the prediction of iron contents, although the lowest R^2 value and highest error for aluminium content prediction was for case 5 in Well 03 (but lower than the equivalent values for case 5 for iron prediction).

Potassium content predictions for equivalent cases 1-5 gave ranges for the following parameters: R^2 0.89–

0.95, MAE 0.09–0.14, and RMSE 0.11–0.17 for test Well 02, and R^2 0.80–0.92, MAE 0.09–0.12, and RMSE 0.10–0.15 for test Well 03. Again, the absence of probe magnetic susceptibility data for case 5 did not have as significant effect on the potassium results compared to the prediction of iron contents. Figure 5 (a) shows a crossplot of the XRF measured (vertically averaged every 0.3048 m = 1 ft) versus the ANN predicted potassium contents trained on probe magnetics and all available well log data for test Well 02 (trained on 70% of the entire dataset). A coefficient of determination R^2 of 0.95, MAE of 0.09 and RMSE of 0.11 were obtained. The depth plot (Figure 5 (b)) showed a good agreement between the XRF measured and predicted potassium contents. Figures 4 (c) and (d) show the corresponding crossplot and depth profile comparing the XRF measured and ANN predicted values for an identical case in test Well 02 with the only difference being that the predictor was trained on 30% of the entire dataset. An R^2 value of 0.94, MAE of 0.11, and RMSE of 0.14 were obtained. Comparisons between Figures 4 and 5 show that both the 2D CNN and ANN predictors give very similar results for potassium contents.

4 Conclusions

The following overall conclusions can be drawn:

1. Two types of neural network predictors (2D CNN and ANN) were both capable of producing excellent predictions of elemental contents of iron, potassium, and aluminium (comparable to measured XRF values) from input data comprising probe magnetic susceptibility and different combinations of well log data in two test wells.
2. Cases where the predictors were trained on 70% of the entire dataset appeared to give slightly better predictions than those trained on 30% of the entire dataset in terms of the coefficient of determination (R^2), mean absolute error (MAE), and the root mean squared error (RMSE) between the neural net predicted and XRF measured elemental contents. Nevertheless, the predictors trained on just 30% of the dataset still gave excellent predictions, demonstrating that limited representative training data can still provide excellent predictions, saving one time and cost. Note that direct comparisons between the predictors trained on the different amounts of data should be treated with some caution, since the predictors trained on 70% of the dataset were tested on the remaining 30%, whereas the predictors trained on 30% of the dataset were tested on the remaining 70%.
3. The 2D CNN predictors marginally outperformed the ANN predictors in some, but not all, cases.
4. The presence or absence of the probe magnetic susceptibility data influenced the prediction of the elemental iron contents in both test wells. In the absence of the probe magnetics, the performance of the 2D CNN and ANN predictors in test Well 02 and 03 was significantly reduced. The probe magnetic susceptibility results are very sensitive to small amounts of minerals containing iron (such as magnetite and siderite, where magnetic susceptibility is high) and demonstrated the

usefulness of the probe magnetic data for predicting iron content.

5   References

1. U. Kramar. Advances in Energy-Dispersive X-Ray Fluorescence. Journal of Geochemical Exploration, **58**(1), 73–80 (1997).

2. K. E. Young, C. A. Evans, K. V. Hodges, J. E. Bleacher, J. E. and Graff, T. G. A Review of the Handheld X-Ray Fluorescence Spectrometer as a Tool for Field Geologic Investigations on Earth and in Planetary Surface Exploration. Applied Geochemistry, **72**, 77-87 (2016).

3. T. D. T. Oyedotun. X-Ray Fluorescence (XRF) In the Investigation of the Composition of Earth Materials: A Review and an Overview. Geology, Ecology, and Landscapes, **2**(2), 148-154 (2018).

4. T. H. To and D. K. Potter. Comparison of High-Resolution Probe Magnetics, X-ray Fluorescence and Permeability on Core with Borehole Spectral Gamma Ray and Spontaneous Potential in an Oil Sand Well. *Proceedings of the 34th International Symposium of the Society of Core Analysts (online symposium 13-16 September)*, Paper SCA2021-035, 12 pages (2021).

5. J. E. Taggart Jr, J. R. Lindsay, B. A. Scott, D. V. Vivit, A.J. Bartel, and K. C. Stewart. Analysis of Geologic Materials by Wavelength-Dispersive X-Ray Fluorescence Spectrometry. *In: Methods for Geochemical Analysis*, Denver, CO, USA: US *Geological Survey Bulletin*, **1770**, E1-E19 (1987).

6. Z. Zhong, T. W. Carr, X. Wu, and G. Wang. Application of a Convolutional Neural Network in Permeability Prediction: A Case Study in the Jacksonburg-Stringtown Oil Field, West Virginia, USA. Geophysics, **84**(6), B363-B373 (2019).

7. S. Kwon, G. Park, Y. Jang, J. Cho, M. G. Chu, and B. Min. Determination of Oil Well Placement using Convolutional Neural Network Coupled with Robust Optimization Under Geological Uncertainty. Journal of Petroleum Science and Engineering, **201**, 108118 (2021).

8. D. A. Osborne. Neural Networks Provide More Accurate Reservoir Permeability. Oil and Gas Journal, **90**(39), 80-83 (1992).

9. S. J. Rogers, H. C. Chen, D. T. Kopaska-Merkel, and J. H. Fang. Predicting Permeability from Porosity using Artificial Neural Networks. AAPG Bulletin, **79**(12), 1786-1797 (1995).

10. H. B. Helle, A. Bhatt, and B. Ursin. Porosity and Permeability Prediction from Wireline Logs using Artificial Neural Networks: A North Sea Case Study. Geophysical Prospecting, **49**(4), 431-444 (2001).

11. H. Al Khalifah, P. W. J. Glover, and P. Lorinczi. Permeability Prediction and Diagenesis in Tight Carbonates using Machine Learning Techniques. Marine and Petroleum Geology, **112**, 104096 (2020).

12. F. Male, J. L. Jensen, and L. W. Lake. Comparison of Permeability Predictions on Cemented Sandstones with Physics-Based and Machine Learning Approaches. Journal of Natural Gas Science and Engineering, **77**, 103244 (2020).

13. B. Backman, K. Lindqvist, and E. Hyvönen. Manual for measuring soil samples with the Niton XL3 GOLDD+ XRF-instrument. Geologian Tutkimuskeskus, pp. 14 (2016).

14. M. Brain, C. Tinelli, P. Rümmer, and T. Wahl. An Automatable Formal Semantics for IEEE-754 Floating-Point Arithmetic. *In: IEEE 22nd Symposium on Computer Arithmetic*, Lyon, France, 22-24 June, 160-167 (2015).

15. K. Simonyan and A. Zisserman. Very Deep Convolutional Networks for Large-Scale Image Recognition. Available online at: https://doi.org/10.48550/arXiv.1409.1556 (2014).

16. T. R. Tavares, L. C. Nunes, E. E. N. Alves, E. D. Almeida, L. F. Maldaner, F. J. Krug, H. W. Carvalho, and J. P. Molin. Simplifying Sample Preparation for Soil Fertility Analysis by X-Ray Fluorescence Spectrometry. Sensors, **19**(23), 5066 (2019).

17. S. Debnath, R. Roy and S. Changder. A Novel Approach Using Deep Convolutional Neural Network to Classify the Photographs Based on Leading Line by Fine-Tuning the Pre-Trained VGG16 Neural Network. *Multimedia Tools and Applications*, 1-26 (2022).

18. V. Nair and G. E. Hinton. Rectified Linear Units Improve Restricted Boltzmann Machines. *In: Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel, 21-24 June, 807–814 (2010).

19. M. Abadi, et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. Software available from: https://www.tensorflow.org/ (2015).

20. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg and J. Vanderplas. Scikit-learn: Machine Learning in Python. The Journal of Machine Learning Research, **12**, 2825-2830 (2011).
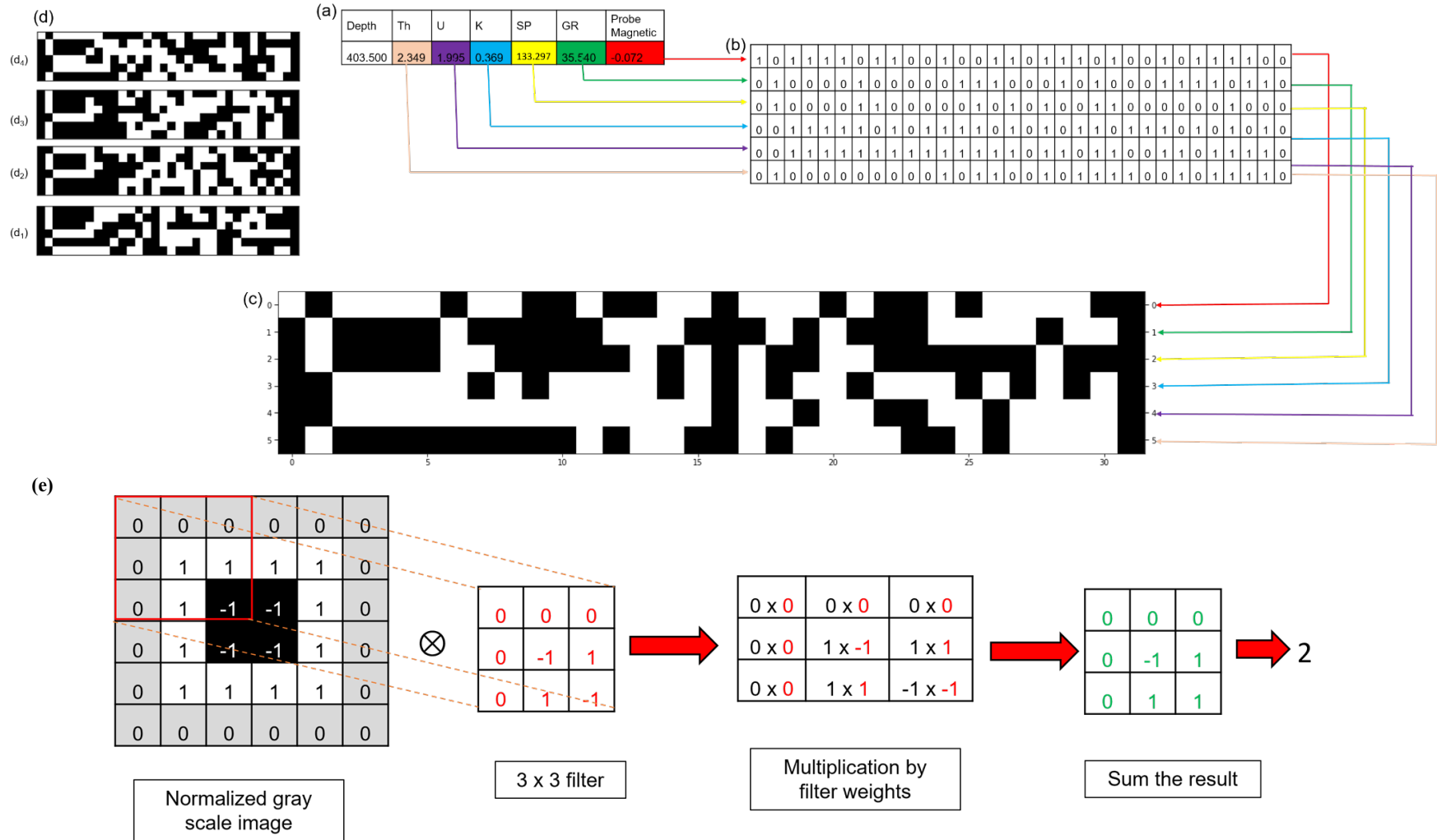
Fig. 1. Schematic diagram of the data conversion process, and 2D convolution process: (**a**) decimal data, (**b**) the 32-bit string for each input variable, (**c**) conversion of the binary string into black and white images, which are stacked together to generate a feature image, (**d**) the feature images at different depths, (**e**) schematic diagram of the 2D convolution process.

Table 1. Summary of the performances of the 2D CNN iron content predictors using 70% (green type) and 30% (black type) training data from the entire dataset, and tested on Well 02 and Well 03 (using 30% of the data from each test well when the predictors were trained on 70% of the entire dataset, and using 70% of the data from each test well when the predictors were trained on 30% of the entire dataset). The predictors were based on combinations of the different probe (vertically averaged every 0.3048 m = 1ft) and well log measurements.

| Input parameters	Training dataset			Testing on Well 02					Testing on Well 03				
	R^2	MAE	RMSE	R^2	MAE	RMSE	RSD	ME	R^2	MAE	RMSE	RSD	ME
Case 1a: probe magnetic susceptibility and all available well logs	**0.99**	0.09	0.12	**0.95**	0.24	0.36	0.36	0.026	**0.99**	0.10	0.13	0.12	0.011
Case 1b: probe magnetic susceptibility and all available well logs	**0.99**	0.11	0.15	**0.90**	0.33	0.55	0.56	0.026	**0.97**	0.13	0.20	0.20	0.012
Case 2a: probe magnetic susceptibility, total GR and SP	**0.99**	0.11	0.16	**0.93**	0.26	0.38	0.37	0.027	**0.98**	0.09	0.13	0.12	0.011
Case 2b: probe magnetic susceptibility, total GR and SP	**0.99**	0.12	0.16	**0.86**	0.39	0.66	0.66	0.030	**0.94**	0.16	0.28	0.26	0.016
Case 3a: probe magnetic susceptibility and total GR	**0.98**	0.13	0.20	**0.91**	0.27	0.41	0.41	0.030	**0.97**	0.14	0.21	0.19	0.018
Case 3b: probe magnetic susceptibility and total GR	**0.94**	0.25	0.40	**0.83**	0.43	0.72	0.71	0.033	**0.90**	0.21	0.34	0.32	0.020
Case 4a: probe magnetic susceptibility and spectral GR	**1.00**	0.09	0.11	**0.93**	0.25	0.41	0.41	0.030	**0.97**	0.11	0.17	0.16	0.016
Case 4b: probe magnetic susceptibility and spectral GR	**0.99**	0.13	0.18	**0.87**	0.40	0.66	0.66	0.030	**0.95**	0.15	0.23	0.22	0.013
Case 5a: all available well logs	**0.99**	0.12	0.17	0.84	0.39	0.63	0.63	0.046	0.89	0.21	0.33	0.33	0.032
Case 5b: all available well logs	**0.97**	0.19	0.30	**0.64**	0.55	1.03	1.03	0.047	**0.68**	0.31	0.61	0.60	0.036
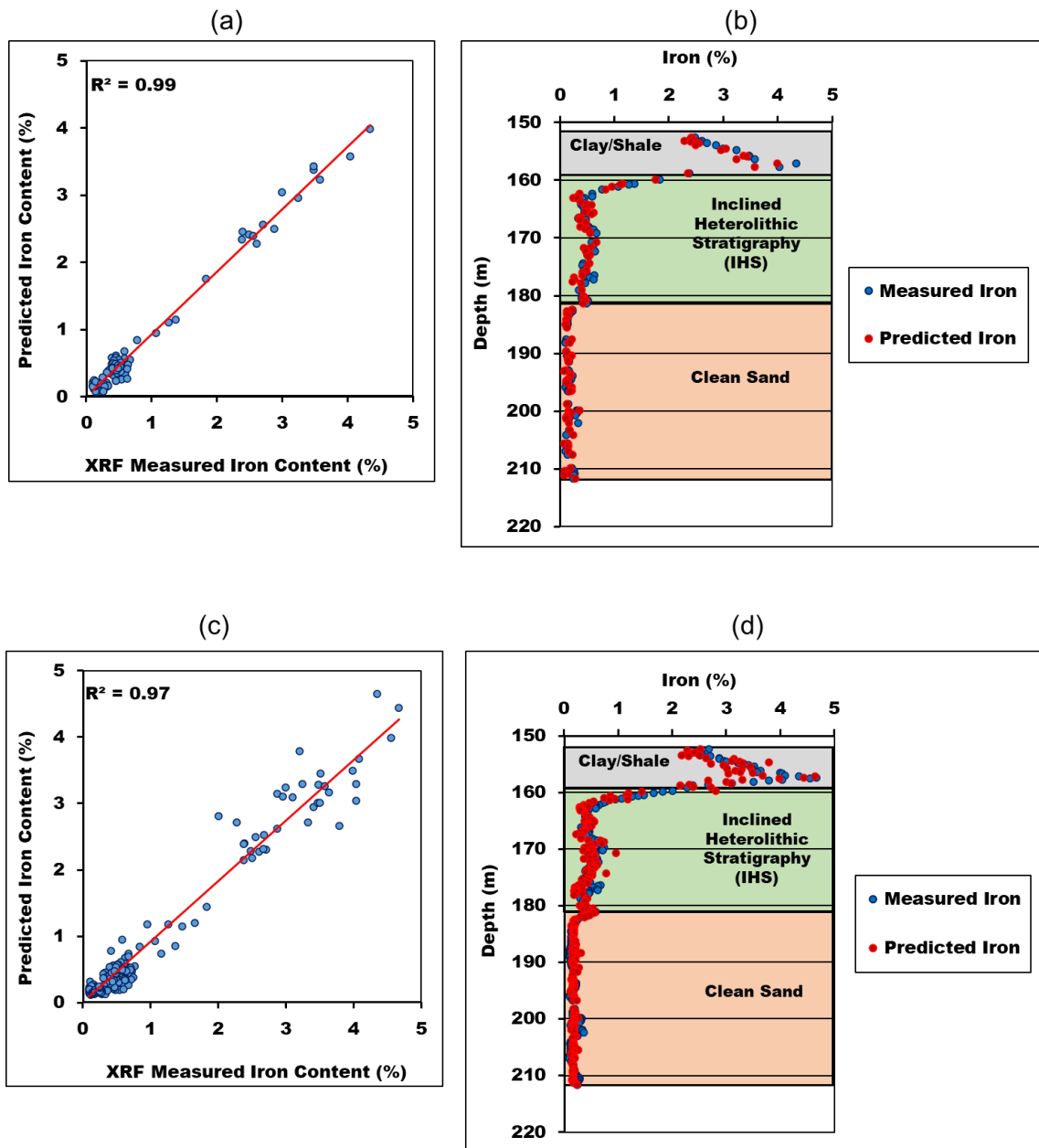
Fig. 2. (**a**) and (**c**) Crossplots of the XRF measured (vertically averaged every 0.3048 m = 1 ft) versus the 2D CNN predicted iron contents trained using the probe magnetics and all available well log data and tested on Well 03 (**Cases 1a** and **1b** in **Table 1** trained using (**a**) 70% and (**c**) 30% of the total dataset respectively). (**b**) and (**d**) Corresponding variations with depth of the XRF measured and 2D CNN predicted iron contents. Note that for (**a**) and (**b**) 30% of the total dataset is tested, whereas for (**c**) and (**d**) 70% of the total dataset is tested, hence the larger number of points in (**c**) and (**d**). Uncertainties in the measured XRF values are less than the plotted points.
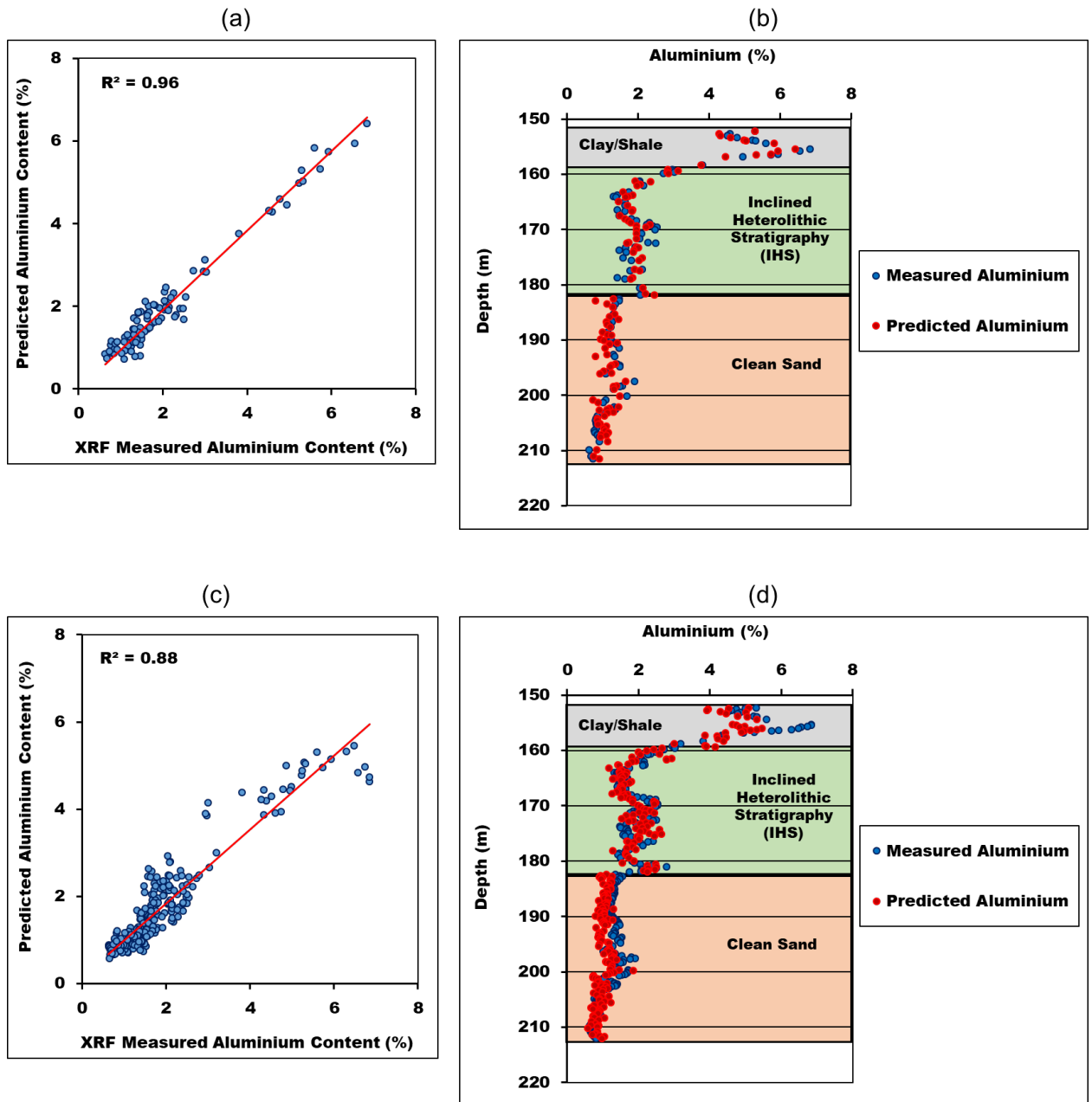
Fig. 3. (a) and **(c)** Crossplots of the XRF measured (vertically averaged every 0.3048 m = 1 ft) versus the 2D CNN predicted aluminium contents trained using the probe magnetics and all available well log data and tested on Well 03 (cases trained using **(a)** 70% and **(c)** 30% of the total dataset respectively). **(b)** and **(d)** Corresponding variations with depth of the XRF measured and 2D CNN predicted aluminium contents. Note that for **(a)** and **(b)** 30% of the total dataset is tested, whereas for **(c)** and **(d)** 70% of the total dataset is tested, hence the larger number of points in **(c)** and **(d)**. Uncertainties in the measured XRF values are less than the plotted points.
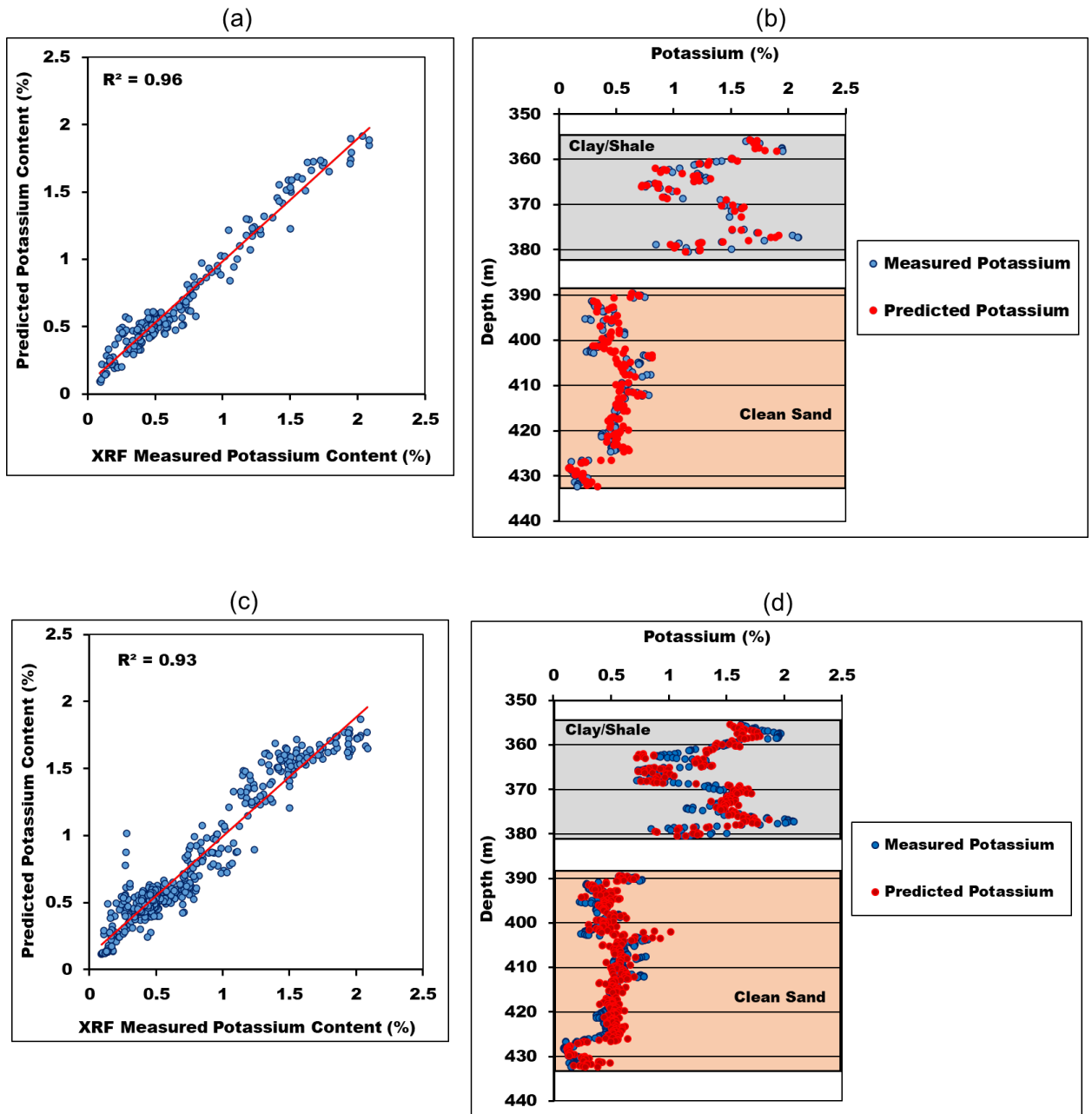
Fig. 4. (**a**) and (**c**) Crossplots of the XRF measured (vertically averaged every 0.3048 m = 1 ft) versus the 2D CNN predicted potassium contents trained using the probe magnetics and all available well log data and tested on Well 02 (cases trained using (**a**) 70% and (**c**) 30% training data respectively). (**b**) and (**d**) Corresponding variations with depth of the XRF measured and 2D CNN predicted potassium contents. Note that for (**a**) and (**b**) 30% of the total dataset is tested, whereas for (**c**) and (**d**) 70% of the total dataset is tested, hence the larger number of points in (**c**) and (**d**). Uncertainties in the measured XRF values are less than the plotted points.
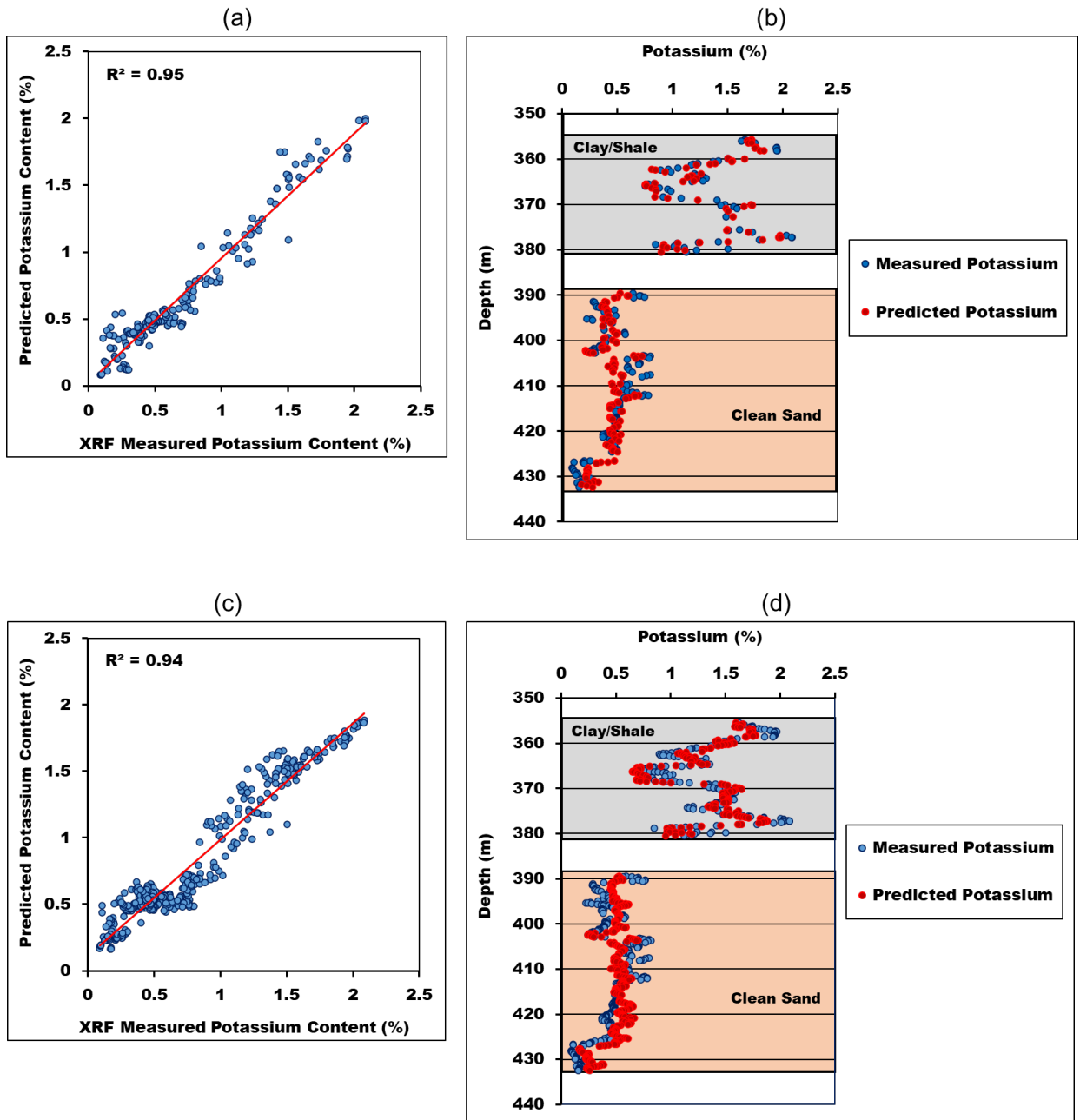
Fig. 5. (**a**) and (**c**) Crossplots of the XRF measured (vertically averaged every 0.3048 m = 1 ft) versus the ANN predicted potassium contents trained using the probe magnetics and all available well log data and tested on Well 02 (cases trained using (**a**) 70% and (**c**) 30% of the total dataset respectively). (**b**) and (**d**) Corresponding variations with depth of the XRF measured and ANN predicted potassium contents. Note that for (**a**) and (**b**) 30% of the total dataset is tested, whereas for (**c**) and (**d**) 70% of the total dataset is tested, hence the larger number of points in (**c**) and (**d**). Uncertainties in the measured XRF values are less than the plotted points.