SSIM 2025 - JJ Potter & J Wesseloo (eds)

doi:10.36487/ACG_repo/2535_46

Estimating borehole wall strength from televiewer data using machine learning

Eric Wilton [a,*], John Danielson [a]

[a] *BGC Engineering, Canada*

Abstract

Acoustic televiewers (ATVs) have become ubiquitous tools within the field of rock slope engineering. ATVs are geophysical probes which create an image of a borehole wall by emitting ultrasonic pulses and recording the reflected waves' amplitudes and travel times. As denser rocks generally reflect waves of higher amplitudes, amplitude data can serve as useful proxies for rock strength. However, signal quality and interpretation can be influenced by confounding factors such as opaque borehole fluids, the centralisation of the probe within the borehole and the presence of mud or cuttings on the borehole wall.

In this study, a Python-based convolutional neural network, a commonly used architecture in computer vision, was trained to estimate rock strength grades from ATV amplitude logs. These logs were converted to images and matched with manually logged field strength grades to generate the training data for the model. Rock strength grades were based on the International Society of Rock Mechanics (ISRM) 1978 scale, which ranges from R0 (extremely weak) to R6 (extremely strong). The training dataset included data from many different boreholes, probes and project settings. The model's prediction accuracy was assessed against a validation dataset and complete test surveys. The model reached a minimum mean square error value of 0.36 on the validation dataset and the model's predictions on average were within approximately one-half-strength grade of the manually logged grades for the test surveys.

By providing continuous, automated strength grade estimates – even in zones of no core recovery – this approach reduces subjectivity and improves coverage compared with traditional field tests. When benchmarked against manually logged strength grades, the model detected finer-scale changes in rock quality, particularly in faulted zones. This tool can assist in reviewing and corroborating traditional strength logs and forms part of BGC Engineering's artificial intelligence-powered televiewer data processing platform, Structura.

Keywords: *acoustic televiewer, machine learning, rock strength estimation, artificial intelligence, computer vision*

1 Introduction

Acoustic televiewers (ATVs) are downhole geophysical probes which record the amplitude and travel time of ultrasonic acoustic waves reflected from the borehole wall. The amplitude and travel time data are recorded along with depth and orientation data to generate an image of the borehole walls. Travel time values are dependent on the borehole radius and the borehole fluid velocity. Reflected wave amplitude values are correlated with the acoustic impedance (density and seismic wave velocity) contrast between the borehole wall and fluid (Cheung 1999). These borehole image logs provide continuous data across the survey length from which geological structures (faults, beds, joints, etc.), lithology contacts, weathering zones and more can be identified. As such, ATV data are now commonly collected and used in multiple fields such as rock slope engineering, mineral exploration and geotechnical site investigations.

[*] Corresponding author. Email address: ewilton@bgcengineering.ca

This paper aims to further expand upon the use of ATV data by developing a machine learning model to estimate rock strength grades from ATV amplitude logs. Rock strength grades in this paper are based upon the International Society for Rock Mechanics (ISRM) 1978 scale (ISRM 1978) as defined below in Table 1. Although the ISRM 1978 rock strength scale is discrete and non-linear (e.g. R2 rock is not defined as twice as strong as R1 rock), it is common practice to assign half-strength grades; as such, this paper includes half point R-values (R1.5, R2.5, etc.). Rock strength is a critical and defining parameter for geotechnical units used in engineered designs. It is therefore imperative to develop an understanding of in situ units' strengths and extents. Current methods of estimating and testing rock strength rely upon laboratory testing of individual samples (i.e. unconfined compressive strength testing), field tests for determining rock strength grades such as those outlined in Table 1, or tests done by portable instruments (i.e. Leeb hardness testers [Ulusay et al. 2025], Schmidt hammers and point load test machines [Palmström 1995]). Each of these methods has its benefits and drawbacks. Laboratory testing is accurate and has well-documented procedures which limit subjectivity but is time consuming, expensive and often requires shipping samples to specialised laboratories if the necessary equipment is not present on site. Field estimates of strength are quick to perform, and strength grade estimates can be made along the entirety of the recovered core length, but the methods are subjective by definition and will vary from logger to logger. Portable instruments provide quick and repeatable measurements along the length of recovered core but still require care to be taken by the logger to follow the standardised procedures for operating the tools. Portable instrument strength measurements can be difficult to directly relate to intact compressive strength values. Leeb hardness and Schmidt hammer test results are affected by the roughness of the sample's surface and point load test results require conversion factors which vary widely, depending on rock type and sample anisotropy (Shenggong et al. 2024).

Table 1 International Society for Rock Mechanics 1978 rock strength scale

Grade	Term	Field identification	Approximate range of compressive strength (MPa)
R0	Extremely weak rock	Indented by thumbnail	0.25–1
R1	Very weak rock	Crumbles under firm blows with point of geological hammer; can be peeled by a pocketknife	1–5
R2	Weak rock	Can be peeled by a pocketknife with difficulty, shallow indentations made by firm blow with point of geological hammer	5–25
R3	Medium strong rock	Cannot be scraped or peeled with a pocketknife; specimen can be fractured with single firm blow of geological hammer	25–50
R4	Strong rock	Specimen requires more than one blow of geological hammer to fracture it	50–100
R5	Very strong rock	Specimen requires many blows of geological hammer to fracture it	100–250
R6	Extremely strong rock	Specimen can only be chipped with geological hammer	>250

All the aforementioned methods of estimating rock strengths are tests completed on ex situ core samples. The process of removing core from the ground may disturb the integrity of the samples and there may be zones where little to no core is recovered, thus limiting the number of available samples to test with conventional methods. If only a small section of intact rock is recovered from within a larger extremely weak fault zone the field logger may assign a non-representative strength grade to the entire interval. The method of rock strength grade estimation proposed in this paper aims to address these limitations in existing methods. Other authors (Almalikee 2019; Scheppers et al. 2001; Zhou et al. 2005) have proposed methodologies which use downhole geophysical surveys to estimate rock strength, however, these either require multiple different geophysical log types as the model input, involve site-specific empirically derived relationships or fail to produce reliable results in heterogeneous rock units. The work presented here was completed as part of the development of Structura, BGC Engineering's artificial intelligence-powered televiewer data processing platform.

2 Methodology

ATVs emit ultrasonic pulses and record the amplitude of the reflected waves. This amplitude is a measure of the acoustic energy reflected to the televiewer; a quantity primarily governed by the acoustic impedance of the reflecting material. Acoustic impedance is a material property equal to the product of the material's seismic velocity and density (Cox et al. 2020). The acoustic impedance of a material is related to its strength; as shown by Bu et al. (2011), who demonstrated a strong, positive linear correlation with compressive strength for concrete samples. Danielson et al. (2023) demonstrated a similar result using ATV amplitude value distributions plotted against field estimated rock strength grades. The plots showed a clear correlation between the two parameters. This paper leverages this relationship to develop a methodology of deriving strength grade estimates from ATV amplitude data.

Given that the amplitude values are theoretically proportional to rock strength, the natural first attempt to estimate rock strength grades from televiewer data could be to derive a simple formulaic relationship, such as directly equating thresholds of amplitude scores to rock strength grades. However, several confounding factors present issues when relying on these simple approaches. While amplitude values are related to rock strength, they also depend upon the overall survey quality. Drilling mud and cuttings coating the borehole wall can obscure the signal clarity and reduce the acoustic impedance contrast between the borehole wall and fluid. Poor centralisation of the probe within the borehole can unevenly alter the amplitude values along the circumference of the survey. Additionally, differences between the types and ages of tools used, along with their firmware versions, may record amplitude scores differently. Table 2 provides examples of how the amplitude values can vary between sites, surveys and tools. The maximum recorded amplitude value from each survey varies widely. All surveys are cropped to exclude any sections of casing or steel drill rods so as not to affect the maximum recorded amplitude. Survey 3, which exhibits poor centralisation, is further examined in Figure 1, which shows the maximum amplitude value recorded at each azimuth along the length of the survey. If the amplitude values were directly mapped to strength values, Survey 3 would incorrectly and implausibly show weaker zones along the entire borehole length at the azimuth ranges of 050–135 and 225–310^{o} relative to high side.

Table 2 Factors affecting amplitude values (all surveys were recorded at different sites)

Survey file	Year recorded	Centralisation	Maximum strength recorded	Maximum amplitude value recorded
1	2024	Good	R4	897
2	2022	Good	R5	13,602
3	2024	Poor	R5	2,314
4	2012	Good	R5	1,303

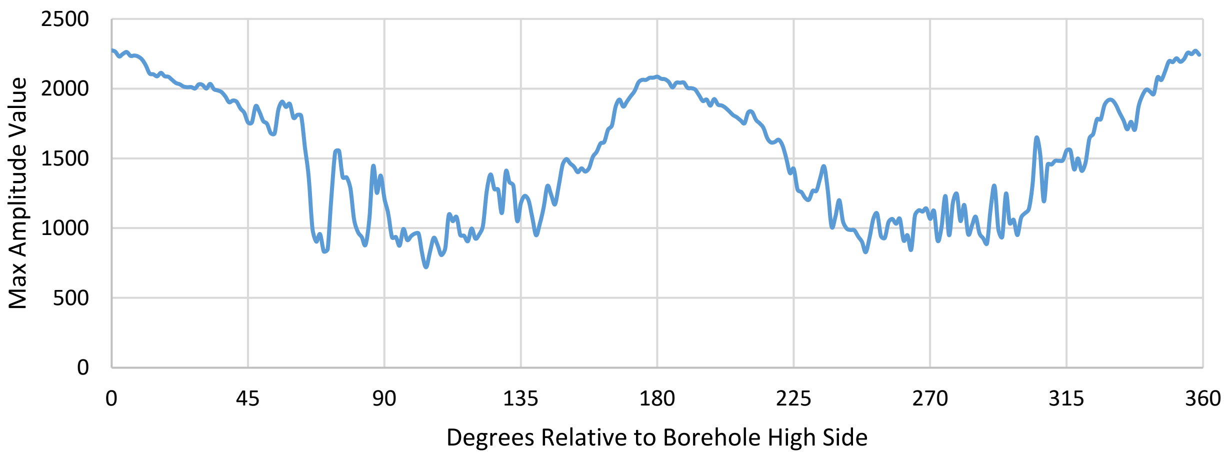

Figure 1 Maximum amplitude value recorded at each azimuth along the length of a survey. This figure corresponds to Survey 3 from Table 2. The effect of poor survey centralisation is evident

Due to these factors, amplitude scores cannot be directly related to strength grade values. Instead, the strength grade estimates must be derived empirically from the patterns within the amplitude data. Machine learning models excel at deriving relationships and discerning patterns from complex and noisy datasets. Given a sufficiently large training dataset, a machine learning model can learn which patterns are correlated with higher rock strengths and which patterns are effectively noise.

For this study, a convolutional neural network (CNN) was employed. CNNs are common model architectures for computer vision tasks, where the inputs to the model are image data. The foundational layer of the model is the convolutional layer, which passes a matrix (known as a kernel) over each pixel and extracts local features to create what is known as a 'feature map'. Each convolutional layer learns its kernel's parameter values during the training process and, as such, each layer extracts different features from the input image. Each additional convolutional layer that is incorporated into the model (e.g. how deep a model is) allows it to learn more complex patterns. The CNN used in this study has four convolution layers. Each convolution layer passes its feature map to a batch normalisation layer. This layer normalises the feature maps which helps the model to train more efficiently. The features are then passed to an activation layer which introduces nonlinearity to the model, thereby enabling it to approximate more complex relationships within the data. Finally, the data are then passed to a pooling layer which down-samples the data to prevent overfitting and removes background noise (O'Shea & Nash 2015). Each of these four-layer combinations (convolutional, batch normalisation, activation and pooling layers) forms a block within the model. There are four of these blocks in total within the model architecture. After the fourth convolutional block, the extracted features are passed to linear connection layers which output the final activation; a score from 0 to 6 representing the image's rock strength grade. Although this output is a continuous value it is rounded to the nearest half-grade to align with the resolution of the training dataset. This approach departs from the discrete ISRM 1978 rock strength scale but enables regression analyses and other numerical assessments required during model training. A high-level overview of the model's architecture is shown in Figure 2. The model's architecture is deep enough to learn the relationship between the amplitude data and rock strength grades, and it can process surveys rapidly, with an average processing time of two milliseconds per image (depending on the hardware used).

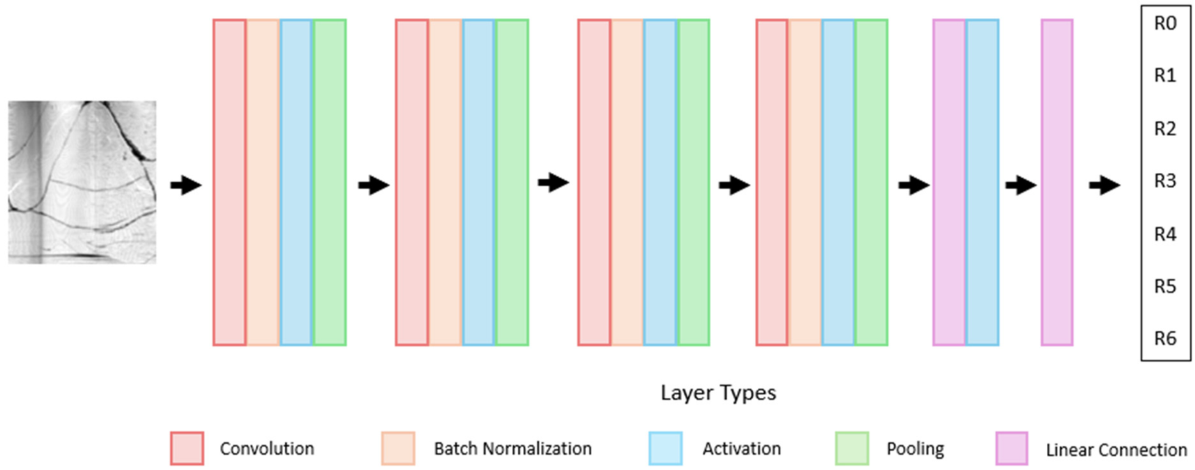

Figure 2 Convolutional neural network architecture

3 Data

The model was trained using a database of 15,042 example images (approximately 7,070 m of televiewer surveys), each representing approximately 0.5 m of survey length, with the exact length varying with the survey resolution. The images were labelled with the corresponding manually logged strength grades. The database of field estimated rock strength grades is based on the ISRM 1978 rock strength scale but includes half-strength grades. To allow for mathematical quantification of error during training, strength grades were converted to numeric values (R0 to 0, R0.5 to 0.5, etc.). The database includes examples from eight different project sites, which helps the model generalise across differing tools, survey qualities and geological settings. The distribution of training images per strength grade is shown in Figure 3 and the distribution per site is shown in Figure 4. The training data were programmatically generated using Python, but all training images were subjected to a manual review to remove any depth misalignments and erroneously assigned strength grades.

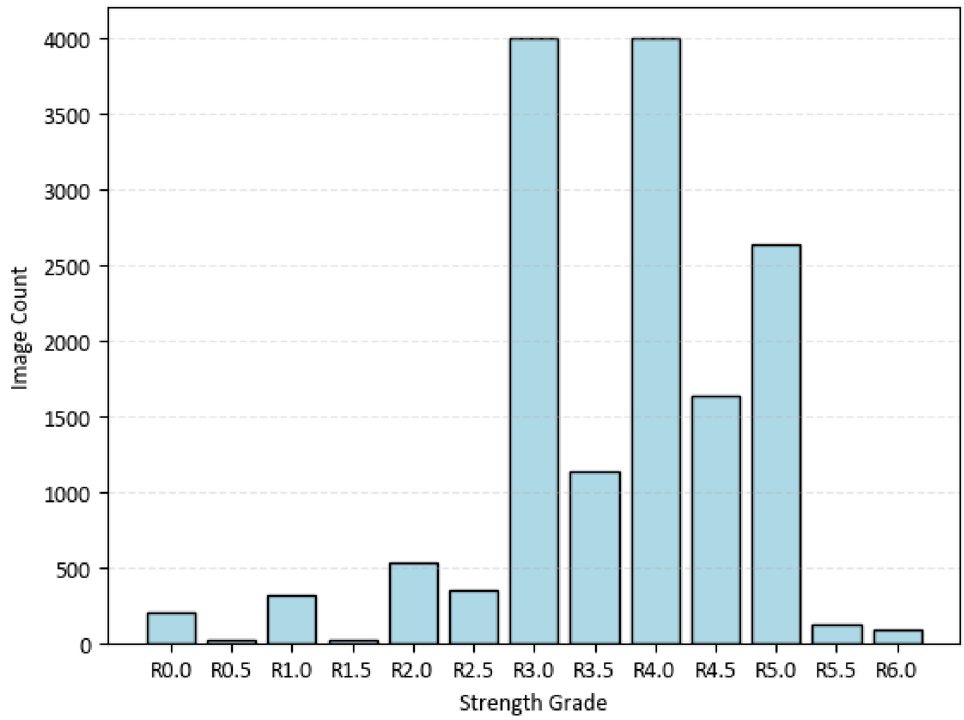

Figure 3 Distribution of images across rock strength grades within the training dataset

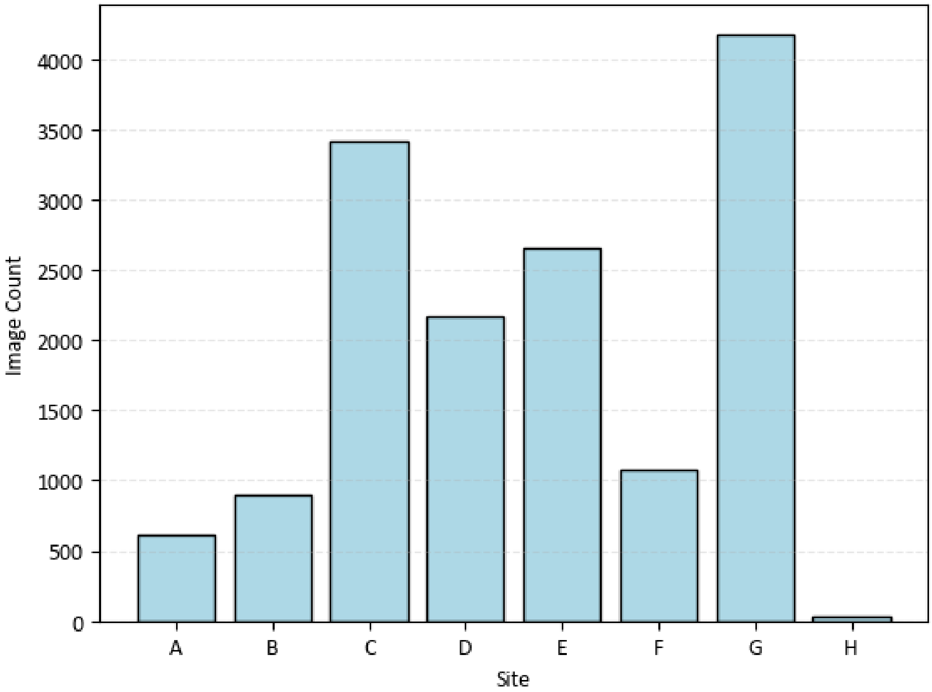

Figure 4 Distribution of images across sites within the training dataset

4 Results

The database of images was split into training and validation datasets for the training process, with 90% (13,538 images) going to training and 10% (1,504 images) for validation. During training, the model calibrated its parameters based on the training data and periodically evaluated its performance using the validation data to monitor for overfitting. The model's performance was measured using the mean square error (MSE) between the model's estimate and the true value of the segment's strength grade. The training process aims to minimise this MSE value. The formula for MSE is shown below in Equation 1. During training, the model achieved a minimum MSE value of 0.36 on the validation dataset.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad (1)$$

where:

n = number of images

y_i = true strength grade

$\hat{y}_i$ = predicted strength grade.

In addition to validation done during training, the model's performance was tested on three complete test surveys. Each of these surveys are greater than 70 m long. These test surveys were not included in the training dataset, but they did come from project settings which were included within the dataset. Example visual outputs can be seen in Figure 5. For each test survey, the mean absolute error (MAE) between the model's predictions and the logged strengths was calculated. The formula for MAE is shown below in Equation 2. The true strength grade in this equation is calculated by taking the average of the manually logged strength grades over the same interval as the input image. The results (shown in the bottom of the images in Figure 5) show that the model can estimate the strength values from a survey with an accuracy of around 0.6 ISRM rock strength grades, depending on the survey. Each survey's length and the time taken by the model to estimate the strength grades are also shown.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (2)$$

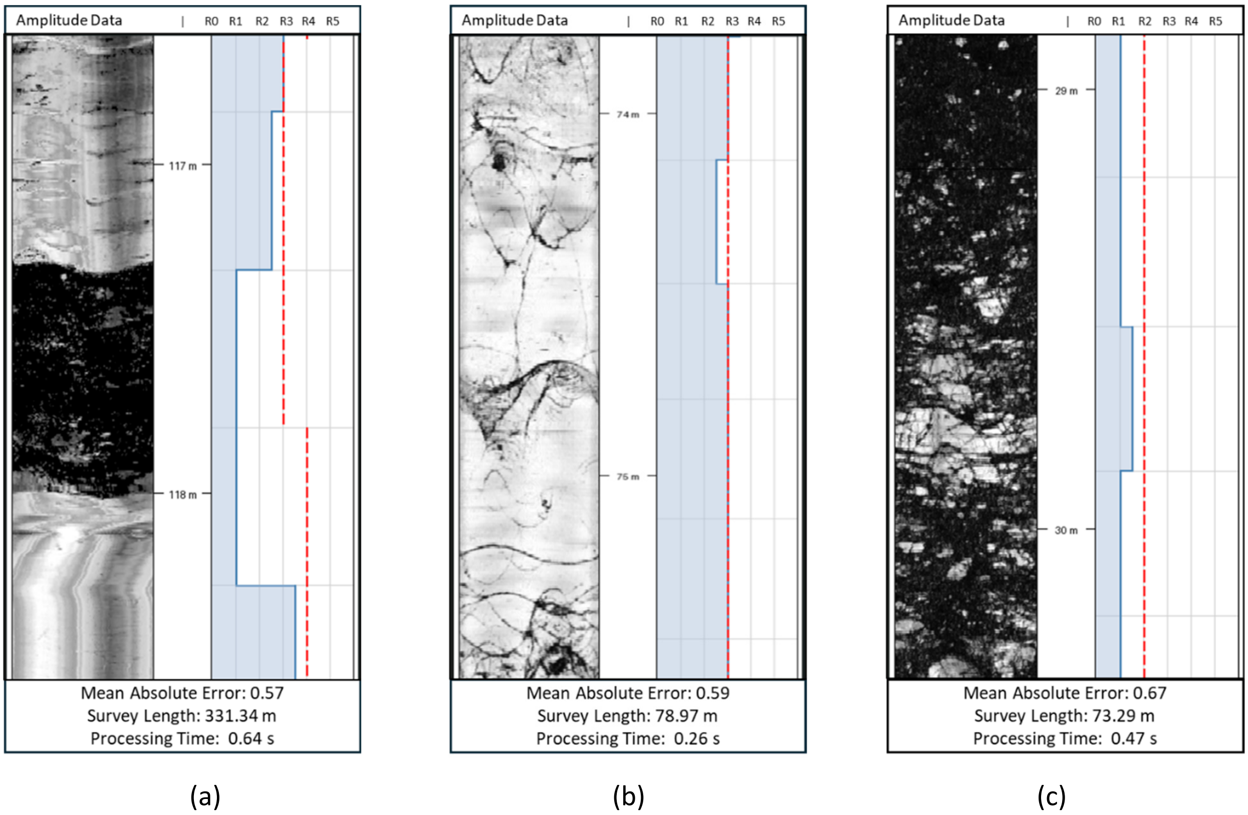

Figure 5 Visualisations of the model's predictions (blue bar graphs) compared to the manually logged strength grades from a field logger (red dashed vertical lines) for three test holes. (a) The model predicted a decrease in strength for a coal seam; (b) Agreement between the model and the logged strengths; (c) The model predicted a higher strength for the section containing intact rock within a fault zone

5 Discussion

The examples shown in Figure 5 further highlight the advantages of deriving strength estimates from televiewer data. The model objectively assesses each segment of the survey, potentially accounting for variations or structures that might be generalised or missed during manual logging. In Figure 5a, the logger recorded a strength grade of R3 to R4 across the entire interval, but the model lowered its estimate over the section containing a coal seam. Figure 5b shows an example of the model's estimates aligning closely with the manually logged data. Figure 5c emphasises how this method for deriving strength grades is especially helpful within faulted ground. Recovering core samples from major fault zones to test for strength estimation is difficult; often only sections of intact rock are recovered, leaving the logger to estimate the entire interval's strength based on limited data. If televiewer survey data are present, strength grade estimates can be made across the entire fault zone and at a finer resolution, as shown in Figure 5c by the estimated strength grade increasing over the section of intact rock (close to matching the manually logged strength) and decreasing for the surrounding heavily faulted materials. The processing time for each survey is less than one second, which shows how rapidly this method can produce meaningful rock strength grade estimations.

6 Conclusion

Acoustic televiewer (ATV) amplitude data, being generally proportional to borehole wall strength, is a valuable source for estimating rock strength. However, differences in tools, software and survey qualities prevent a simple, direct relationship. The strength grade values must be derived from the patterns within the amplitude data. As machine learning models excel at identifying patterns in complex datasets, a model was trained to estimate strength grade values from ATV amplitude data.

In this study, a CNN model was trained using 13,538 image segments derived from ATV amplitude logs, paired with manually logged strength grades. The trained model achieved a minimum MSE of 0.36 on the validation dataset of 1,504 images. On three test surveys excluded from the training dataset, the model was able to predict the rock strength grades within approximately 0.6 strength grades of the manually logged strength grades over the combined length of the test surveys.

This machine learning-based approach to estimating rock strength grades from televiewer data is a powerful supplement to traditional methods. It provides rapid, objective and continuous estimates of the strength grade along the borehole, in contrast to manual logging, which is time consuming, subjective and limited to the recovered core. Once a televiewer survey has been conducted, the model can generate strength grade estimations along the survey in often less than one second. A key advantage is the ability to assess rock strength in situ, even in zones with poor or no core recovery where representative conventional testing is impossible. As demonstrated, the model can detect finer-scale changes in rock strength, particularly in zones of strength heterogeneity, such as fault zones or bedded stratigraphy, compared with manual logging.

The developed model can assist in reviewing, corroborating, and supplementing field and laboratory strength estimates, improving the consistency and coverage of geotechnical characterisation. It is recommended that the model be calibrated with site-specific data for optimal results. However, as the training dataset is expanded, the model will be able to better generalise to new sites. With ongoing refinement and validation, this automated strength estimation could facilitate the derivation of other critical geotechnical parameters from televiewer data, such as fracture frequency and rock quality designation, which often depend on reliable strength assessments.

References

Almalikee, HS 2019, 'Predicting rock mechanical properties from wireline logs in Rumaila Oilfield, Southern Iraq', *American Journal of Geophysics, Geochemistry and Geosystems*, vol. 5, no. 2, pp. 69–77

Bu, Y, Song, W, Wang, M & He, Y 2011, 'The improvement of cementation quality rating method based on compressive strength for low density cement system', *Procedia Engineering*, vol. 18, pp. 289–294, https://doi.org/10.1016/j.proeng.2011.11.045

Cheung, PS 1999, 'Microresistivity and ultrasonic imagers: tool operations and processing principles with reference to commonly encountered image artefacts', in G Williamson, MA Lovell & PK Harvey (eds), *Borehole Imaging: Applications and Case Histories*, Geological Society of London, London, pp. 45–58.

Cox, D, Newton, A, Huuse, M 2020 'Chapter 22 - An introduction to seismic reflection data: acquisition, processing and interpretation', in N Scarselli, J Adam, D Chiarella, D Roberts & AW Bally (eds), *Regional Geology and Tectonics (Second Edition),* Elsevier

Danielson, J, Clayton, MA, Kelly, L & Kinakin, D 2023, 'Getting more out of drillhole televiewer data: geotechnical toolbox edition', in PM Dight (ed.), *SSIM 2023: Third International Slope Stability in Mining Conference*, Australian Centre for Geomechanics, Perth, pp. 421–434, https://doi.org/10.36487/ACG_repo/2335_26

ISRM 1978, 'International society for rock mechanics commission on standardization of laboratory and field tests: Suggested methods for the quantitative description of discontinuities in rock masses', *International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts*, vol. 15, no. 6, pp. 319–368, https://doi.org/10.1016/0148-9062(78)91472-9

O'Shea, K & Nash, R 2015, 'An introduction to convolutional neural networks', *CoRR*, http://arxiv.org/abs/1511.08458

Palmström, A 1995, *RMi – A Rock Mass Characterization System For Rock Engineering Purposes*, PhD thesis, University of Oslo, Oslo.

Schepers, R, Rafat, G, Gelbke, C & Lehmann, B 2001, 'Application of borehole logging, core imaging and tomography to geotechnical exploration', *International Journal of Rock Mechanics & Mining Sciences*, vol. 38, pp. 867–876.

Shenggong, G, Runqing, C, Yang, Z, Hu, N & Faquan, W 2024, 'Comparison and combination of Leeb hardness and point load strength for indirect measuring tensile and compressive strength of rocks', *Bulletin of Engineering Geology and the Environment*, vol. 83, no. 4, pp. 83–109, https://doi.org/10.1007/s10064-024-03608-x

Ulusay, R, Ersoy, H, Sünnetci, MO & Karahan, M 2025, 'The leeb (Equotip) hardness test for rock materials: An overview, assessments on the factors influencing test results, and prediction models based on a large database', *Bulletin of Engineering Geology and the Environment*, vol. 84, no. 3, https://doi.org/10.1007/s10064-025-04170-w

Zhou, B, Fraser, S, Borsaru, M, Aizawa, T, Sliwa, R & Hashimoto, T 2005, 'New approaches for rock strength estimation from geophysical logs', in *Proceedings of the Bowen Basin Symposium*, Yeppoon, pp. 151–164.