geoconvention
Calgary • Canada • May 12-14 2025

Conventional to GenAI: Boosting Lithology Classification with Multimodal Learning

Ryan A. Mardani
DataEnergy

Summary

This paper proposes a multimodal approach for lithology classification that integrates numeric well logs (e.g., NPHI, RHOB, DTC) with descriptive text (e.g., formation descriptions, color, grain size, rounding). A fine-tuned BERT model handles the text data, while a feedforward neural network encodes the numeric logs; the two representations are then fused for final predictions. From 35 wells with about 30000 samples, one blind well is reserved for realistic evaluation. Gen-AI-Assisted method outperforms a conventional numeric-only neural network, resulting fewer misclassifications and better prediction of minor lithologies. The findings demonstrate that textual incorporation can noticeably enhance classification accuracy, offering a more robust solution for well-log prediction, though it is a computationally expensive operation.

1. Introduction

Lithology classification is an important step in subsurface characterization, drilling decisions, reservoir engineering, and hydrocarbon exploration. Historically, such classification tasks rely heavily on numerical well logs (e.g., density, neutron porosity, gamma-ray) combined with limited manual interpretation of geological descriptions. However, the availability of both data and machine learning techniques has enabled new ways to fuse multiple data modalities.

Recent innovations in Generative AI (GenAI), particularly large language models like BERT, have helped the scope of how textual information can be embedded into machine learning pipelines. This paper seeks to answer whether GenAI-driven methods can outperform conventional numeric-only models by incorporating text-based geological descriptions into a unified multimodal architecture for lithology classification.

2. Multimodal Theory

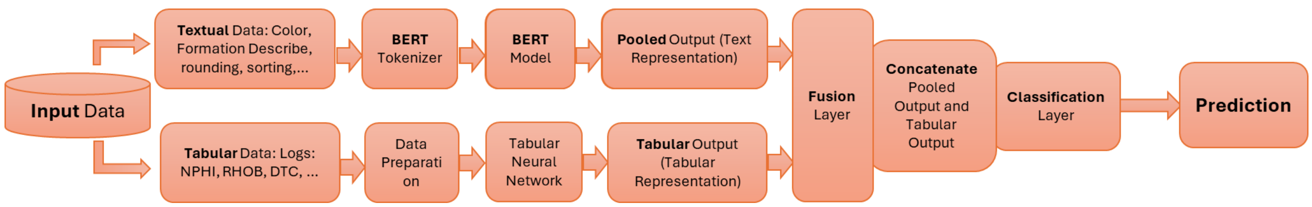
In many real-world examples, such as geological interpretation or well-log analysis, no single data modality provides the complete picture. Textual descriptions, resulted from core lab studies by geologists or drilling cutting description by well-site geologist may capture qualitative features (e.g., color, formation descriptions, sorting), while numerical well-logging sensor readings (tabular data) capture quantitative measurements (e.g., NPHI, RHOB, DTC). When these two data sources are viewed in isolation, important interrelationships can be missed, leading to suboptimal classification results.

Multimodal learning integrates different data modalities in a unified model. The motivation lies in the idea that textual data can provide contextual or semantic insights, while numerical logs capture physical measurements that may be less obvious from text alone. By jointly modeling both, the system can learn richer, more robust representations, resulting improved predictive performance.

To handle the textual component, we fine-tune a large language model (BERT, in this case) that has been pretrained on vast amounts of text. This pretrained foundation gives the model a strong grasp of language structure and semantics right from the start. By further training (fine-tuning) on domain-specific geological descriptions, BERT learns to recognize and encode specialized geological terminology and phrasing. Simultaneously, we train a feedforward neural network on tabular logs to capture subtle variations in numeric features. In a final fusion layer, text-based embeddings (from BERT) are combined with tabular embeddings, so that the classifier can draw on the full multimodal context for making predictions.

3. Workflow

Figure below illustrates the end-to-end pipeline.



1. **Textual Data**
   Geological descriptions (e.g., color, rounding, sorting, formation details) are collected into a single text field (e.g., "Combined Text").
2. **Tabular Data**
   Well logs such as NPHI, RHOB, DTC, and other numeric measurements are aggregated into a feature matrix.

3.2 Textual Processing and Encoding

1. **Tokenization**
   The combined text is passed through a BERT tokenizer, splitting the text into subword tokens and generating input IDs plus attention masks.
2. **Pretrained Model (BERT)**
   The tokenized sequences are fed into the BERT model, which outputs rich text embeddings (the "pooled output" from the [CLS] token).
3. **Fine-Tuning**
   Rather than using BERT merely as a feature extractor, we continue training it on our geological dataset so that it adapts to domain-specific vocabulary and context.

3.3 Tabular Processing and Encoding

1. **Preprocessing**
   Missing values in logs are imputed (e.g., replaced by mean), ensuring a consistent numeric feature space.
2. **Feedforward Network**
   The tabular features are normalized or standardized, then passed through a small

feedforward network. This produces a “tabular representation”—an embedding that captures numeric characteristics of each sample.

3.4 Fusion Layer

1. **Concatenation**
BERT’s pooled text embedding is concatenated with the tabular embedding, creating a combined feature vector that represents both linguistic and numeric dimensions.

3.5 Classification

1. **Fully-Connected Layer**
The fused feature vector is passed through a linear classification layer (with softmax output). Parameters of BERT, the tabular network, and the fusion layer are trained end-to-end to minimize cross-entropy loss on lithology classes.

2. **Prediction**
During inference, each new sample (text + logs) is run through the same pipeline to generate a single predicted lithology label.

3.6 Optimization and Early Stopping

1. **Training Epochs**
The model is trained over multiple epochs, with gradient updates applied to both the language and tabular components.

2. **Monitoring Performance**
Validation loss and accuracy are tracked each epoch, and early stopping is employed to prevent overfitting.

4. Data and Experimental Setup

Dataset (~30,000 Samples)

- **Blind Well**: One well is reserved as a blind test well, ensuring no information leakage during model development.
- **Train/Validation/Test Split**: The remaining samples are divided into training, validation, and test sets, ensuring a robust evaluation protocol.

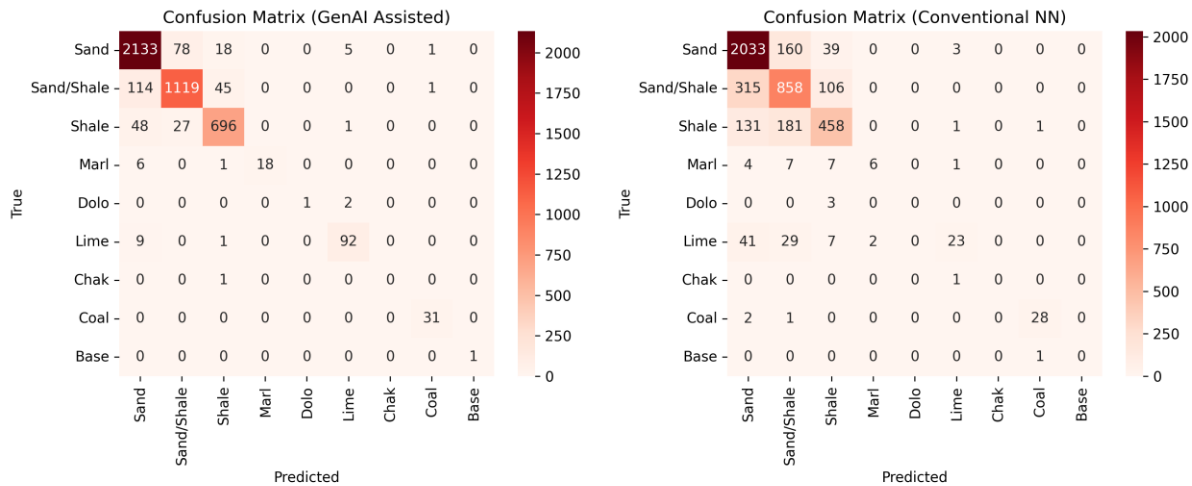
Two Modeling Approaches

1. **Conventional NN**
A deep neural network trained solely on numeric well-log features (e.g., NPHI, RHOB, DTC).

2. **GenAI-Assisted**
A multimodal model that takes both numeric logs and textual descriptions (color, formation description, rounding, sorting, etc.) as input. BERT is fine-tuned for the text component, and its output is fused with the tabular representation from a feedforward network.

5. Results

Two confusion matrices illustrate the performance of the GenAI-Assisted model vs. the Conventional NN.



GenAI-Assisted Model

In the confusion matrix for the GenAI-Assisted model, there is a clear predominance of correct classifications for major lithology classes like *Sand*, *Sand/Shale*, and *Shale*, as aligned on diagonal entries. This diagonal dominance highlights the model's ability to capture both numeric and textual cues that differentiate these commonly encountered lithologies. In addition to performing well on such frequently observed classes, the GenAI-Assisted approach also shows reduced confusion in smaller or underrepresented classes, such as *Dolo*, *Coal*, and *Base*. While numeric logs alone may fail to provide strong distinguishing signals for these classes—often due to overlapping log characteristics—textual information, such as specific color descriptors or formation identifiers, helps the model discern these lithologies more accurately.

Two illustrative examples: in the case of *Sand* vs. *Sand/Shale*, references to "clayey" or "silty" intervals in the accompanying descriptions enable the model to distinguish sediment mixtures from cleaner sands. Similarly, mentions of "organic layers" or "carbonate features" facilitate more precise recognition of *Coal* or *Limestone*, guiding the classifier beyond the limitations of numeric logs. Together, these findings underscore how textual data aids the model in capturing subtleties that would otherwise remain ambiguous if relying solely on numerical readings.

Conventional NN (Numeric-Only)

In contrast, the confusion matrix for the numeric-only network reveals a more substantial overlap between certain lithologies, most notably between *Sand* and *Sand/Shale*. The higher off-diagonal entries in these classes indicate that, without textual input, the model struggles to distinguish classes that share similar well-log signatures. This overlap becomes especially problematic for lithologies that differ primarily in minor features or mixed compositions, which numeric logs may not consistently capture.

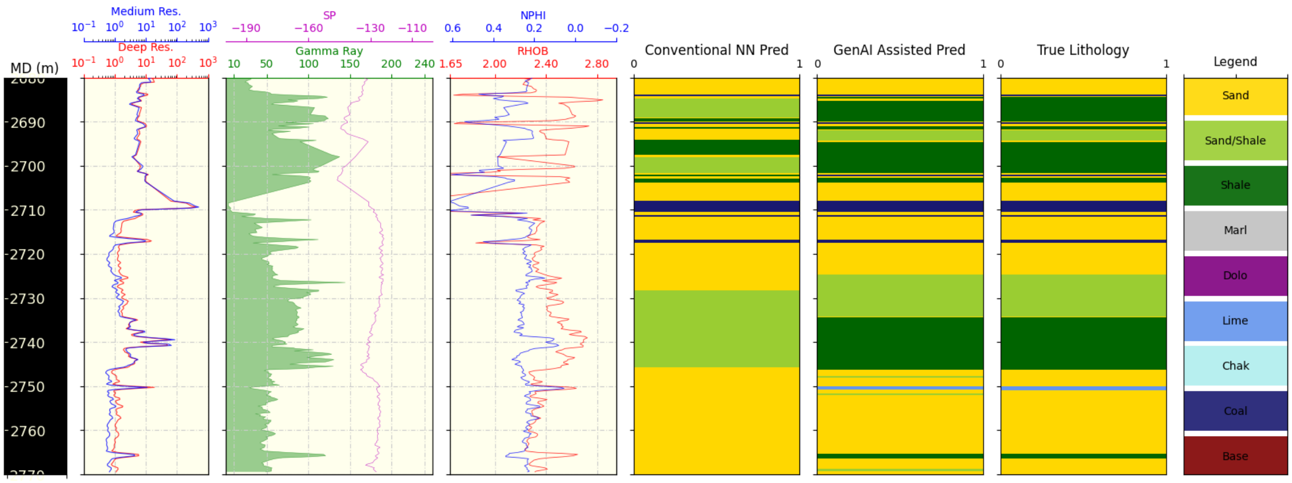
Moreover, minority classes like *Dolomite* and *Basement* appear more frequently misclassified in the numeric-only framework, suggesting that such classes rely heavily on contextual or descriptive features for accurate identification.

Key Observations

1. **Overall Accuracy Gain**: The GenAI-Assisted approach consistently outperforms the numeric-only network, particularly for classes with subtle distinctions.
2. **Textual Context Matters**: Clues like “calcareous,” “laminated,” or “organic layers” play a crucial role in enhancing classification.
3. **Reduced Misclassification**: Class confusions (e.g., Sand vs. Sand/Shale, Dolo vs. Lime) are alleviated by incorporating textual signals.
4. **Minor Classes Benefit**: While still challenging, the GenAI-Assisted model more accurately identifies classes with fewer samples, thanks to domain-specific language in the text.

Visual Comparison of Predictions

Side-by-side visual logs for *Conventional NN Pred* vs. *GenAI Assisted Pred* against the *True Lithology* is illustrated in the picture below.



We can see:

1. **Lithology Continuity**
   - *Conventional NN*: Abrupt transitions and mismatched intervals appear where numeric logs struggle to discriminate mixed lithologies.
   - *GenAI Assisted*: The text-enabled pipeline aligns more closely with ground-truth boundaries, producing smoother transitions.
2. **Minor Lithology Intervals**
   - *Conventional NN*: Thin beds of *Dolomite* or *Coal* often merge into neighboring classes with similar log readings.
   - *GenAI Assisted*: Mentions of carbonate cements or organic content help differentiate these minor intervals, improving alignment with actual lithology.
3. **Improved Class Boundaries**
   - *Conventional NN*: Can overextend or truncate formations where numeric logs gradually shift.
   - *GenAI Assisted*: Geological descriptors (e.g., “slightly clayey sand,” “calcareous interbeds”) lead to more precise boundary delineations.
4. **Overall Accuracy**
   - The GenAI-Assisted predictions visibly track the reference lithology, mitigating the inherent ambiguities of numeric logs alone.

6. Conclusion

In conclusion, this study demonstrates that fusing numeric well-log data with domain-specific textual descriptions yields more accurate and geologically consistent lithology predictions than a numeric-only approach. By fine-tuning a BERT model on geological text and combining its representations with tabular features, the proposed approach benefits from both semantic and quantitative signals. Experimental results on 30,000 samples and one blind well highlight notably improved performance, particularly in distinguishing subtle classes and thin intervals. These findings underscore the value of multimodal data integration, suggesting that future developments in GenAI-assisted analytics have considerable potential to refine reservoir characterization, reduce uncertainty, and guide better decision-making in geoscience applications. However, while the GenAI-Assisted model proves more accurate, it comes at a higher operational cost in terms of GPU requirements and computational overhead.

References

P. Baltrusaitis, C. Ahuja, and L. P. Morency, “Multimodal Machine Learning: A Survey and Taxonomy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 2, pp. 423–443, Feb. 2019.

J. Ngiam, A. Khosla, M. Kim, J. Nam, H. Lee, and A. Y. Ng, “Multimodal Deep Learning,” in *Proceedings of the 28th International Conference on Machine Learning (ICML)*, Bellevue, WA, USA, 2011, pp. 689–696.

P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, “Multimodal Fusion for Multimedia Analysis: A Survey,” *Multimedia Systems*, vol. 16, no. 6, pp. 345–379, 2010.

A. Zadeh, M. Chen, S. Poria, E. Cambria, and L.-P. Morency, “Tensor Fusion Network for Multimodal Sentiment Analysis,” in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, Copenhagen, Denmark, 2017, pp. 1103–1114.