"Predictive Analysis of Well Penetration in the Rio Grande Basin (Closed Basin) Aquifer in Saguache County, Colorado, Final Report"

Cammie Newmyer

June 20, 2023

[Visualization Presentation Link]

Predictive Analysis of Well Penetration in the Rio Grande Basin (Closed Basin) Aquifer in Saguache County, Colorado, Final Report

Abstract

This final report presents the successful completion of a project aimed at developing a deep learning model for assessing the likelihood and depth of well penetration through the smectite clay layer in the Rio Grande Basin (Closed Basin) Aquifer, situated in Saguache County, Colorado. The project encompassed the design and implementation of an intuitive interface, comprehensive data collection, dataset construction, deep learning model development, and interface deployment. The deep learning model, trained using a combination of historical well data and theoretical information on the smectite clay layer, demonstrates a high degree of accuracy in predicting well penetration likelihood and providing depth estimations. The user-friendly interface integrates real-time predictions and visualization capabilities, empowering stakeholders with valuable insights for informed decision-making in sustainable water management practices. The successful outcome of this project contributes to the responsible utilization of water resources in the Rio Grande Basin, supporting the region's long-term water supply and management goals.

Introduction

The Rio Grande Basin (Closed Basin) Aquifer, situated in Saguache County, Colorado, holds significant importance in ensuring the water supply of the region. Gaining insights into the probability of well penetration through the smectite clay layer and accurately predicting the depth of penetration are essential for maintaining sustainable water management practices. This project proposal aims to devise an interface, gather pertinent data, construct a comprehensive dataset, and employ a deep learning model to determine the likelihood of a well site, based on township range and section, penetrating the smectite clay layer. Furthermore, the model will provide an estimation of the depth at which the penetration may occur.

Primary Objectives

1. Design an intuitive and user-friendly interface to facilitate data collection and visualization.
2. Collect comprehensive data on well sites, township range, section, and associated geological characteristics.

3. Build a dataset combining collected data and theoretical information on the smectite clay layer.
4. Develop a deep learning model capable of predicting the likelihood of well penetration through the smectite clay layer.
5. Provide accurate depth estimations for wells that are likely to penetrate the clay layer.

Methodology

- Interface Design: Design and develop an interactive interface for data collection, allowing users to input wellsite information, township range, and section. Incorporate visualization capabilities to display relevant geological information, including the smectite clay layer and other relevant data layers.
- Data Collection: Gather historical well data, including well locations, township range, section, and associated geological characteristics. Acquire theoretical information on the smectite clay layer, such as thickness, permeability, and other relevant properties. Ensure the data collection process adheres to ethical guidelines and regulations.
- Dataset Creation: Combine the collected historical well data with theoretical information on the smectite clay layer. Cleanse and preprocess the dataset, handling missing values, outliers, and inconsistencies. Conduct feature engineering to extract meaningful features from the data, such as proximity to known geological formations and aquifer characteristics.
- Deep Learning Model Development: Implement a deep learning model architecture, such as a convolutional neural network (CNN) or a recurrent neural network (RNN), suitable for the predictive task. Split the dataset into training, validation, and testing sets. Train the model using the training set and optimize its performance through iterative experimentation. Validate and fine-tune the model using the validation set to achieve the best possible predictive performance. Include appropriate data visualizations.
- Model Evaluation and Deployment: Evaluate the trained model using the testing set, measuring performance metrics such as accuracy, precision, recall, and depth estimation error. Deploy the trained deep learning model into the interface for real-time predictions. Continuously monitor the model's performance and make necessary updates and improvements based on user feedback and additional data.

Deliverables

- A user-friendly interface for well site information input and visualization of geological data layers.
- A comprehensive dataset combining historical well data and theoretical information on the smectite clay layer.
- A trained deep learning model capable of predicting the likelihood of well penetration and providing depth estimations.
- Documentation outlining the project methodology, data collection process, model architecture, and instructions for using the interface.

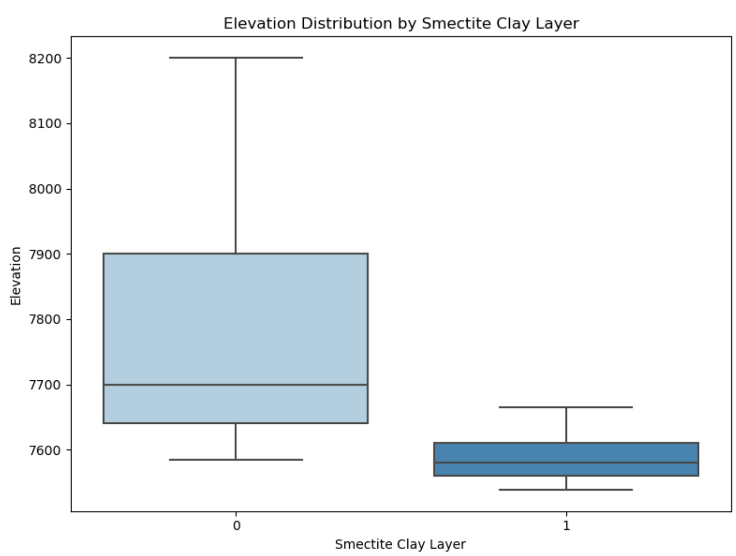
Tasks

- Interface design and development.
- Data collection and dataset creation.
- Deep learning model development and optimization.
- Model evaluation and interface deployment.
- Finalizing project documentation and presentation.

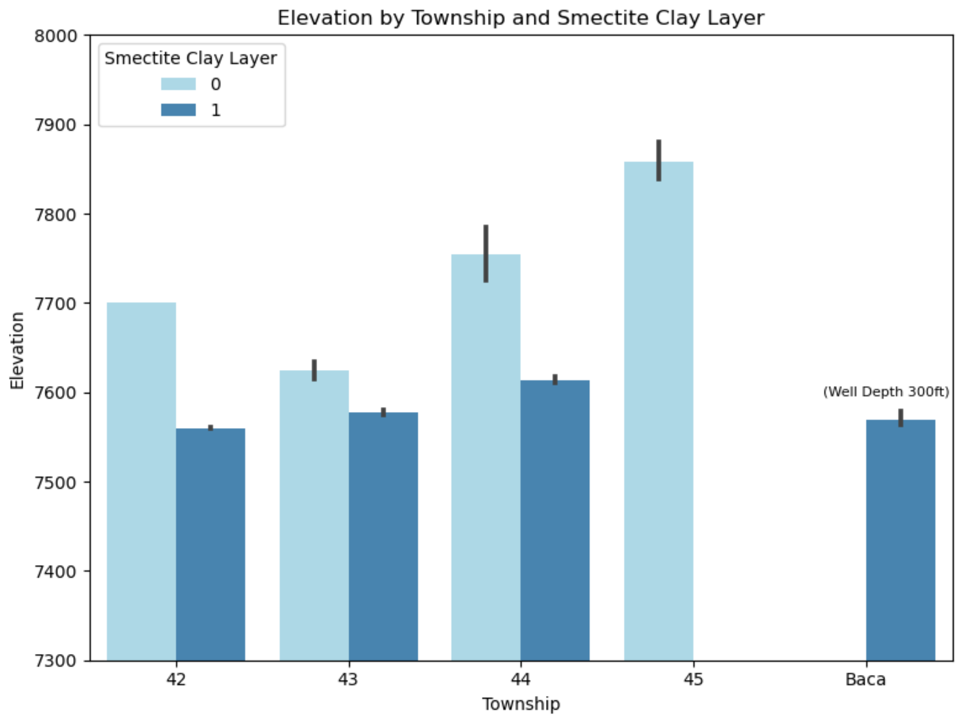
Resources

- Access to historical well data in Saguache County, including township range, section, and geological information.
- Theoretical data on the smectite clay layer, including thickness, permeability, and other relevant properties.
- Computational resources (e.g., CPUs, GPUs) for data processing, model training, and inference.
- Relevant software tools and libraries, such as Python, TensorFlow, or PyTorch, for deep learning model development.

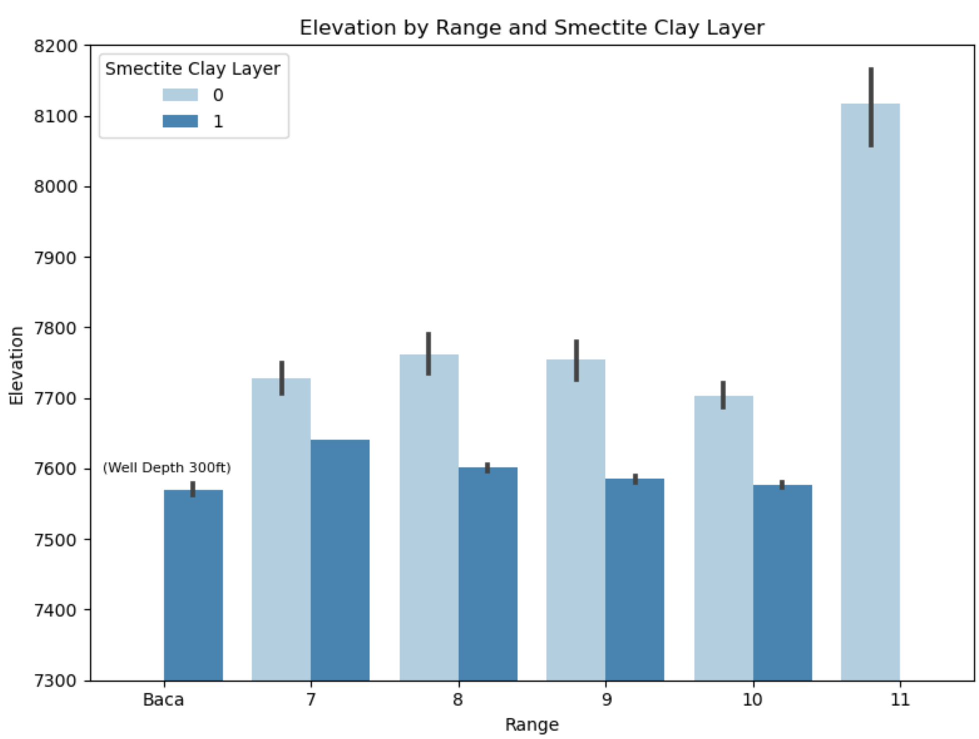
Visualization



Exploratory data analysis illustrates the differences in the Elevation Distribution by Smectite Clay Layer. The elevation of the clay layer with those wells predicted to reach the confined aquifer (1) is less than the elevation of those wells where the smectite clay layer is not predicted (0).

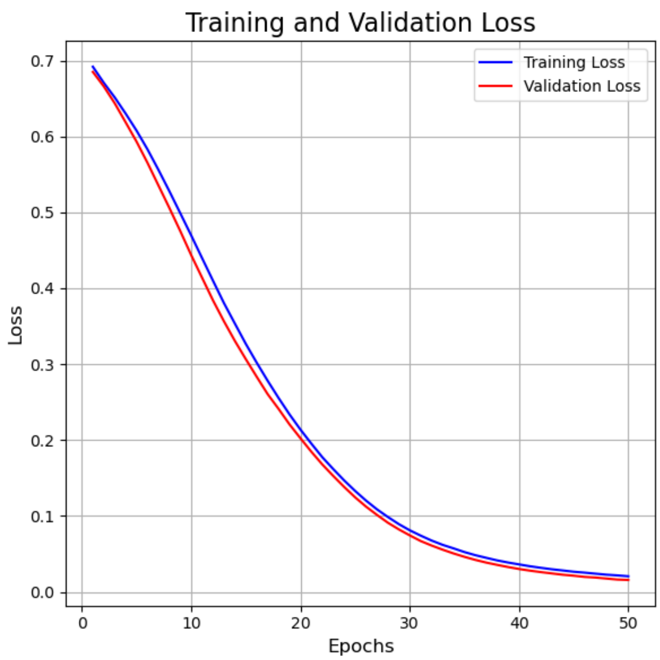


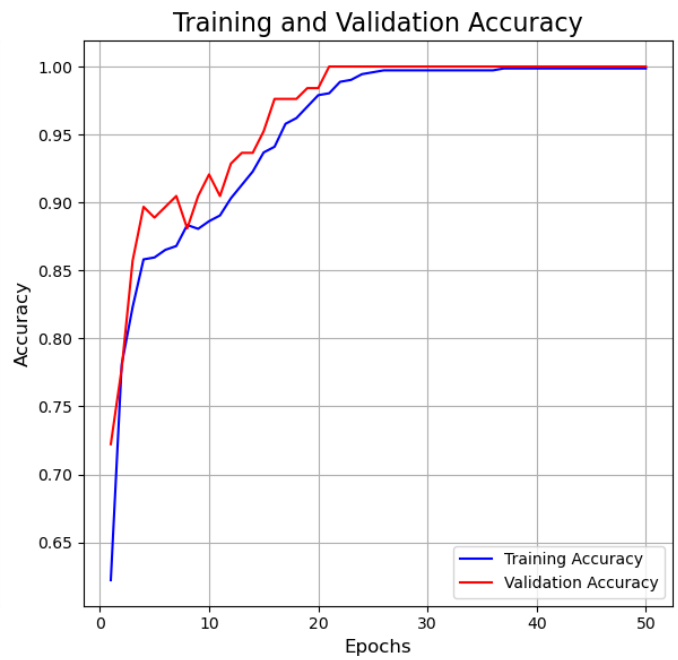
Townships are the largest geographical division in this study. The chart, Elevation by Township and Smectite Clay Layer shows that irrespective of the elevation in the Township (light blue), the elevation of the clay layers seems fairly consistent across Townships where the clay layer is detected.



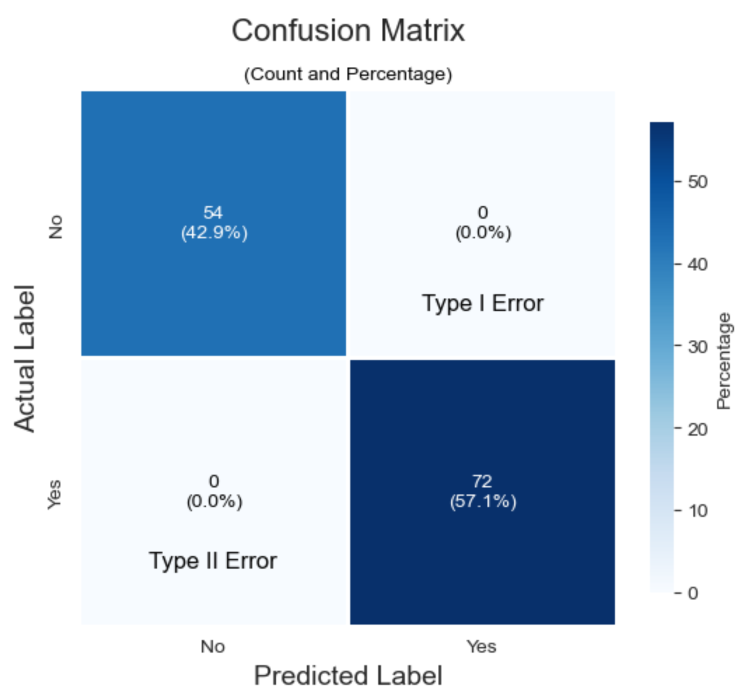
The Range is a division of Townships. In the chart, Elevation by Range and Smectite Clay Layer, the information is similar to the previous graph emphasizing that the depth of the confined aquifer is fairly consistent in elevation. This is an important predictor for the model.

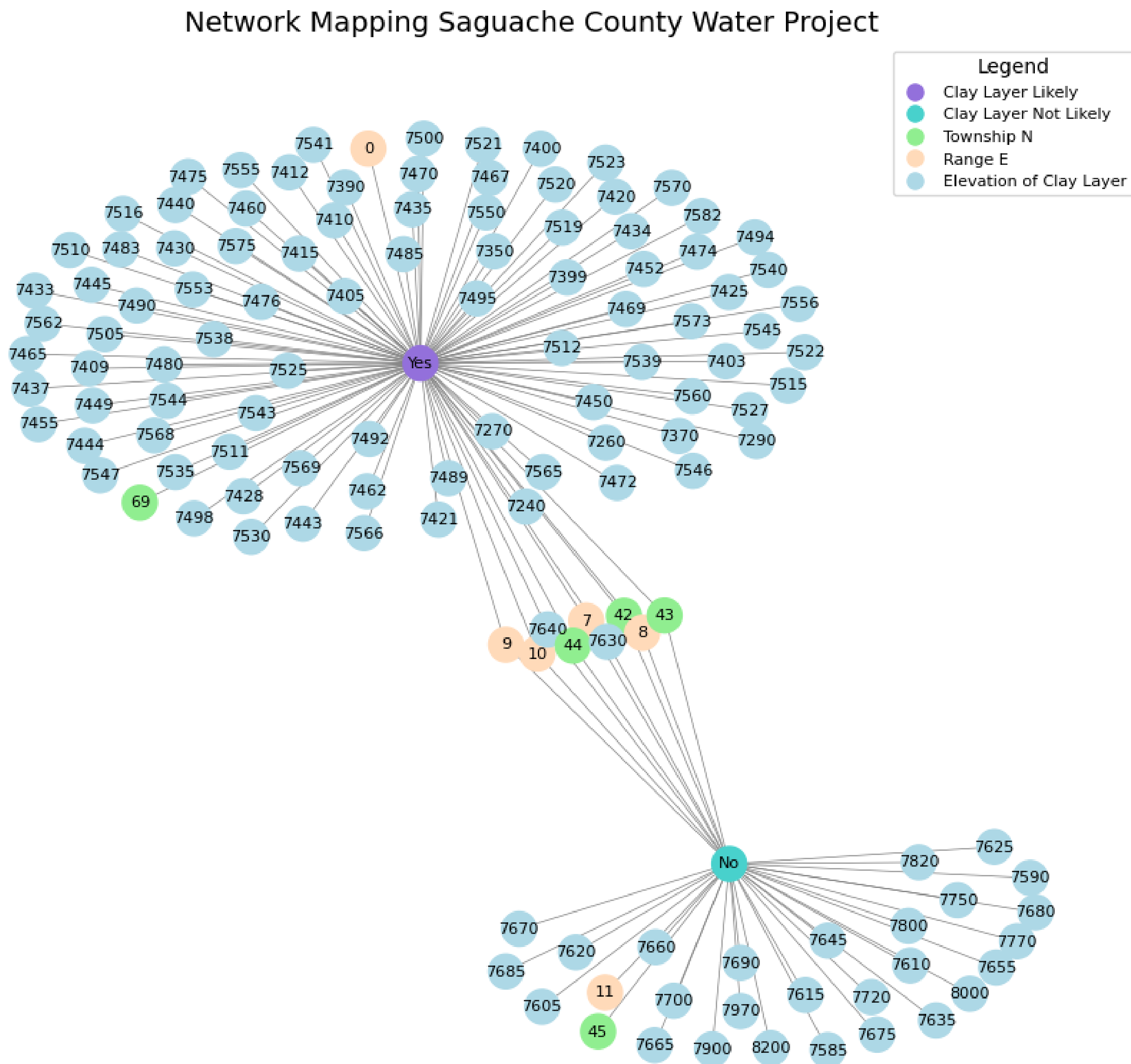
The following visualizations illustrate the performance of the deep learning model during training and testing. This visualization pair illustrates the performance of the training model. The training and validation loss are closely aligned decreasing from 70% to less than 5% by the end of the full run of 50 epochs. The accuracy of the model increased from 65% to 100% over 25 epochs. The training accuracy increases steadily while the validation accuracy varies more often.



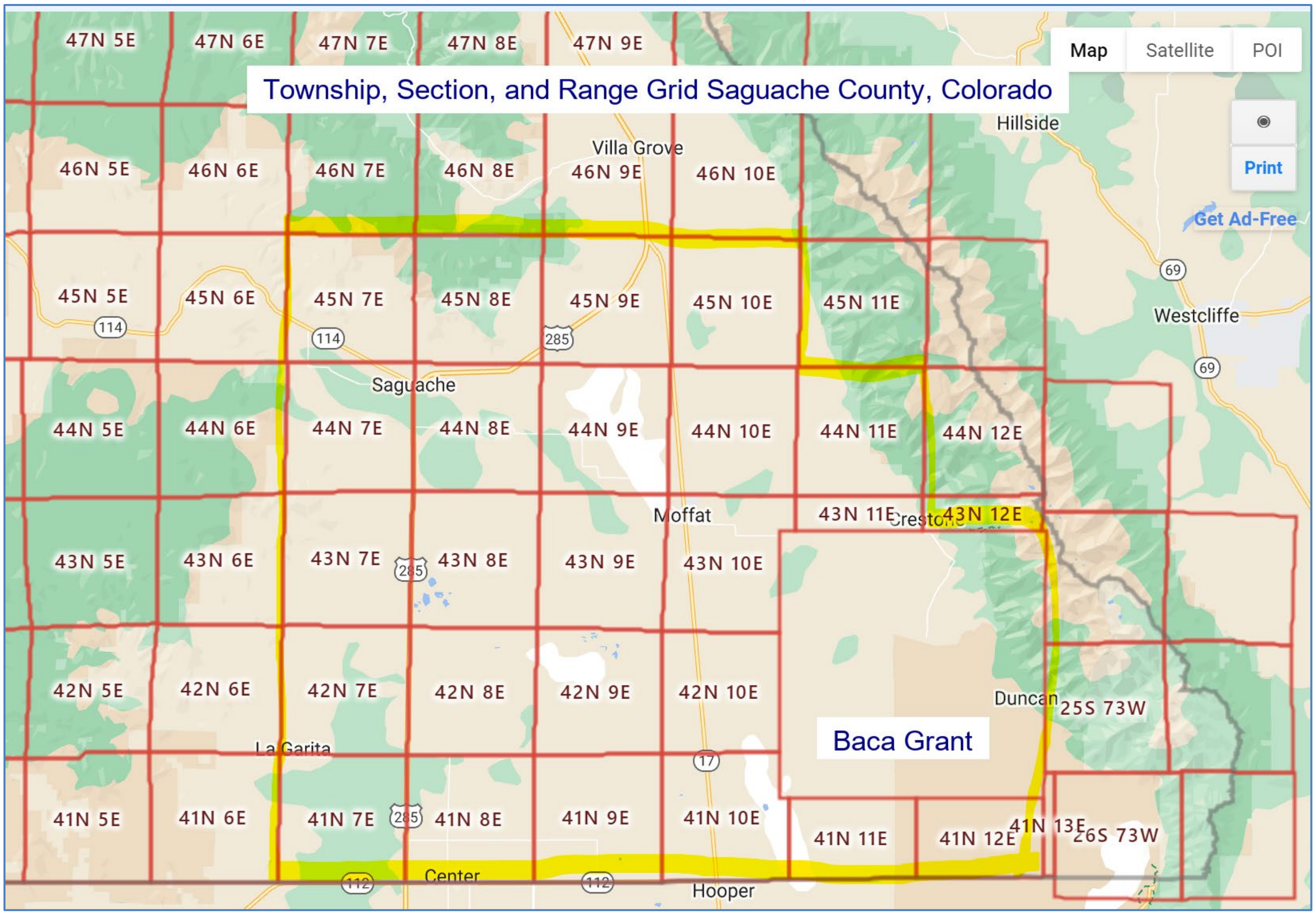


The confusion matrix is also a model performance visualization. After model tuning the predictive model has a high accuracy. The confusion matrix illustrates that when the model has been tested on the alternate data, all the predictions that the well would not penetrate the smectite clay layer were predicted correctly and that all the predictions that the well would not penetrate the smectite clay layer were also predicted correctly.





The Network Mapping Saguache County Water Project presents a visualization of the data which allows the user to intuitively understand the behavior of the data and how the deep learning model created a patterned structure for prediction. This graph represents a network mapping of the Saguache County Water Project. It visualizes the relationships between different nodes related to the project, including the presence of a smectite clay layer, township numbers, range values, and elevation of the clay layer. Each node is represented by a distinct color, enabling easy identification. The legend provides a key for interpreting the colors and node types in the graph. The graph highlights the connections between the smectite clay layer and other factors, such as townships, range values, and elevation. The visualization helps to understand the complex relationships within the water project, providing insights into the geographical distribution and connections between different elements involved in the project.



This map provides necessary information to the user in order to interact with the model. The area of the study is outlined in yellow. The Baca Grant is represented by the number “69” for the Township variable, the others are as shown. The range is the second number on each square.

```
Enter the township: 42
Enter the range: 7
Enter the section: 15
1/1 [==============================] - 0s 12ms/step
No smectite clay layer predicted.
Enter '1' to re-enter data or any other key to quit: 1
Enter the township: 42
Enter the range: 8
Enter the section: 15
1/1 [==============================] - 0s 12ms/step
Smectite clay layer likely at elevation: 7435 ft.

Enter '1' to re-enter data or any other key to quit:
```

Conclusion

The final model used in the Saguache County Water Study is a deep learning model. The architecture of the deep learning model consists of multiple layers designed to process the input data and make predictions. Here is a description of each layer in the model:

<u>Conv1D Layer:</u> This layer performs 1D convolution on the input data. It has 32 filters and a kernel size of 3. The activation function used is ReLU, which introduces non-linearity to the model. The input shape is determined based on the number of features in the scaled training data.

<u>MaxPooling1D Layer:</u> This layer performs max pooling operation to downsample the output from the previous convolutional layer. It uses a pool size of 2, which means it takes the maximum value within a sliding window of size 2.

<u>Flatten Layer:</u> This layer flattens the output from the previous layer into a one-dimensional vector. It reshapes the data from a 2D representation to a 1D representation, preparing it for the subsequent fully connected layers.

<u>Dense Layers:</u> Two dense layers follow the flatten layer. The first dense layer has 128 units with a ReLU activation function, which helps to capture complex relationships in the data. The second dense layer has 64 units, also using the ReLU activation function.

<u>Output Layer:</u> The final dense layer consists of a single unit with a sigmoid activation function. It produces the output prediction for the presence of the smectite clay layer. The sigmoid activation function squeezes the output between 0 and 1, representing the probability of the positive class (presence of the clay layer).

The model utilizes the Adam optimizer with a learning rate of 0.0001 and is compiled with the binary cross-entropy loss function, which is suitable for binary classification tasks. Early stopping is employed as a callback, with a patience of 3 epochs and monitoring the validation loss. The model architecture leverages convolutional layers to capture local patterns and spatial dependencies in the input data, followed by dense layers to learn higher-level representations and make predictions. This combination of convolutional and dense layers enables the model to effectively process the well completion report data and predict the presence of the smectite clay layer.

After training the model, it is evaluated on the test set to assess its performance. The test loss and test accuracy are calculated and displayed. Additionally, the test accuracy is also presented as a percentage for better interpretation. The accuracy of the model is illustrated by the Confusion Matrix where all test data was correctly identified. To utilize the trained model, users have the option to input Township, Range, and Section values. These values are used as input to the model, which predicts the presence of the smectite clay layer at the specified location. If the model predicts the presence of the clay layer, the average elevation of the layer at that location is returned as an output.