Martian Core Analysis Report

Abstract

This report summarizes the results, challenges faced, and methodologies used in completing the Martian Core Analysis, a high prep problem statement released by Equinox Club.

1. Methodologies Used

In this section, we outline the different methods used and models implemented for different modules of the problem statement.

1.1. Module 6

Feature Extraction: Preprocessed seismic waveforms undergo feature extraction, including Maximum Amplitude, RMS Amplitude, and dominant frequency (derived via FFT's highest spectral peak). The Hilbert Transform is applied to compute the instantaneous phase, aiding in wave propagation and timing analysis.

Synthetic Dataset Generation: We have created a synthetic dataset using two models - TauPyModel and Syngine. TauPyModel model is used to compute seismic wave travel times and estimate ray parameters for various seismic phases. Syngine is used to generate realistic synthetic seismic waveforms based on predefined Mars geophysical parameters.

Dataset Composition and Balance: The balanced dataset contains 1,500 shadow zone and 1,500 non-shadow zone events (3,000 total). Each event is processed separately: TauPyModel determines classification, while Syngine generates waveforms for feature extraction.

Classification of shadow zone: Preprocessing is done on the dataset which includes feature selection, normalization, and missing value handling. A Random Forest Classifier, optimized via RandomizedSearchCV for hyperparameter tuning, is trained on the stratified dataset to maintain class balance and ensure robust performance.

1.2. Module 7

Dataset Preparation : The initial data had a lot of redundant columns and they were removed to ensure the relevant columns are there as asked in the Module.

Data Preprocessing and Cleaning : The data needed slight preprocessing the terms of removing nan values.

Model Used : The final model utilized an ensemble learning approach, comprising three base regressors stacked together. Initially, these three models - Random Forest Regressor, AdaBoost Regressor, and LightGBM were trained independently. Their predictions were subsequently combined to construct a meta-model, which was further refined by applying Linear Regression as the final layer.

1.3. Module 8

Feature Extraction with ResNet50: Seismic images were passed through a pre-trained ResNet50 model. The model was modified to use feature maps from an intermediate layer, capturing high-level patterns. These extracted features formed the basis for clustering instead of raw images.

Dimensionality Reduction with UMAP: The high-dimensional feature vectors from ResNet50 were reduced using Uniform Manifold Approximation and Projection (UMAP). UMAP was chosen for its ability to preserve local and global structures in data. The number of dimensions was reduced to 2 for visualization and further clustering.

Clustering with K-Means: K-Means was applied to the UMAP-reduced features to group seismic images into clusters. The resulting cluster labels were assigned based on similarity in the reduced feature space.

1.4. Module 9

Model Architecture: The Physics-informed neural network (PINN) model is designed to predict scalar potential (ϕ or A) based on spatial (x, y) and temporal (t) inputs, as well as velocity (V_p or V_s) using PyTorch. The network consists of seven fully connected layers with Swish activation functions, which help the network learn complex nonlinear relationships efficiently.

Loss Functions: PINNs use loss functions to enforce both data-driven learning and physical constraints. The primary loss function in this implementation is based on the wave equation. Boundary condition loss ensures that the

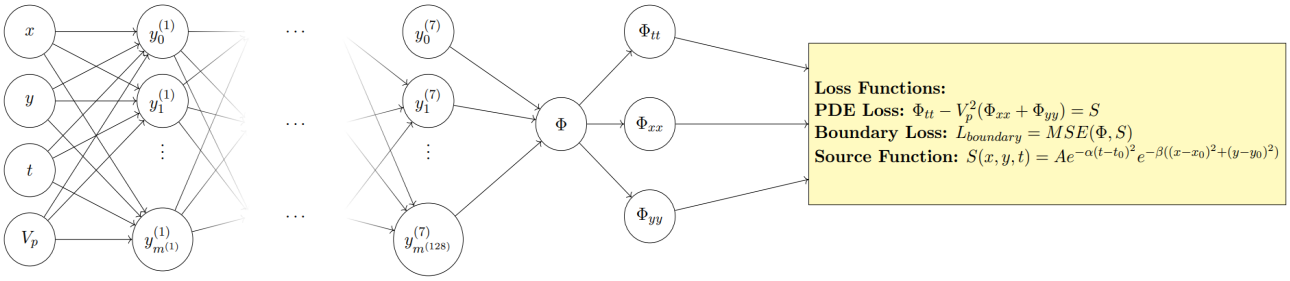

Figure 1. Overview of our proposed PINN framework for seismic wave propagation including the network architecture and the loss functions based on physical laws.

wave simulation follows physical constraints, particularly by enforcing a pulse at the epicenter
Loss function for p-waves and s-waves :

$$\frac{\partial^2 \Phi}{\partial t^2} = V_p \nabla^2 \Phi$$
$$\frac{\partial^2 \mathbf{A}}{\partial t^2} = V_S \nabla^2 \mathbf{A}$$

Boundary condition loss function is given mathematically by :

$$\phi(X_e, Y_e, t) = S(X_e, Y_e, t)$$

where ϕ(x,y,t) represents the predicted wave solution, and S(x, y, t) is the source function that models the pulse at the epicenter.

Training Process: The neural network initializes with random weights. Adam optimizer is used which updates weights by computing gradients from the loss function. The equation loss and boundary loss guide the model to learn physically consistent solutions. After several epochs, the loss stabilizes, indicating that the model has learned the correct physics.

Calculation of shadow zone : Additionally, to determine the shadow zone, we calculate the circumference of the planet and analyze the maximum disturbance at these points. By graphing out circumference versus disturbance we can accurately find out the shadow zones by analyzing dips in the graph.

2. Results

In this section, we present the results obtained for different modules of the problem statement.

2.1. Module 6

Amplitude distribution: The amplitude-time graph represents how seismic wave energy evolves over time. Peaks in amplitude indicate the arrival of different seismic wave modes (e.g., P-waves, S-waves).

Phase Shift Information: AThe instantaneous phase shift provides insight into wave propagation delays and changes in phase relationships. By analyzing phase shifts, we can determine the relative timing of seismic wave arrivals, which is crucial for event localization and subsurface imaging.

Martian core properties: Seismic classification of shadow zones provides insight into Mars' core composition: If classified as a shadow zone, it suggests a liquid outer core, as seismic waves cannot propagate through it. If classified as a non-shadow zone, it implies the presence of solid or partially molten material influencing wave transmission.

Important Features: A feature importance plot highlights key predictors, such as Ray Parameter and Epicentral Distance(deg), revealing their impact on classification accuracy.

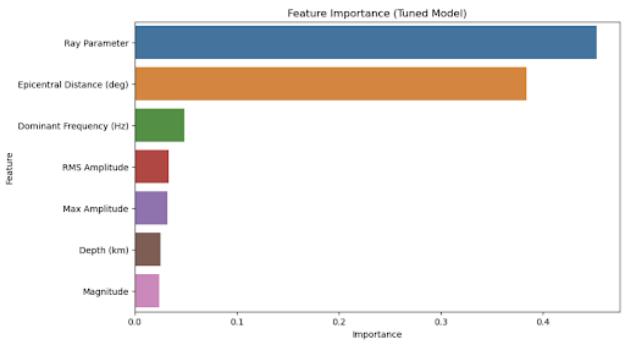

Figure 2. Important features in our seismic data

Model Evaluation: The optimized Random Forest classifier was evaluated using various metrics to assess its performance in seismic shadow zone classification. The model achieved an overall accuracy of 0.92 and an AUC score of 0.95.

Class	Precision	Recall	F1-Score
Non-Shadow (0)	0.98	0.85	0.91
Shadow (1)	0.87	0.98	0.92
Macro Avg	0.92	0.92	0.92
Weighted Avg	0.92	0.92	0.91

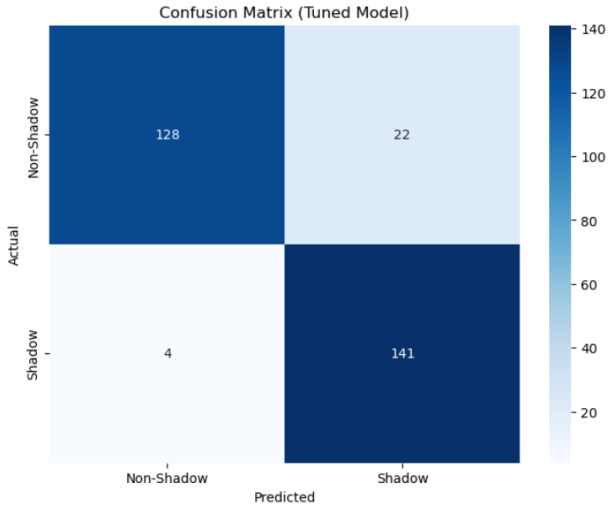

Figure 3. Confusion matrix for our model

2.2. Module 7

Radius: The core radius of Mars is coming out to be approximately 1800 km on average.

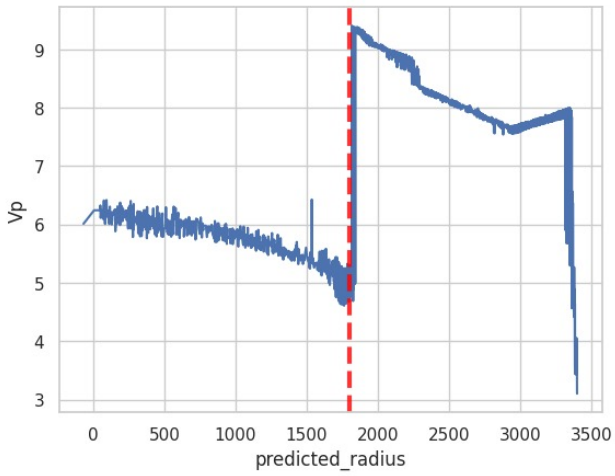

Figure 4. The above graph is between P-wave velocity vs depth. The discontinuity in the graph denotes the change from mantle to core. The red dotted line denotes the core radius.

2.3. Module 8

Clustering Results : The clustering analysis distinguished between distinct seismic patterns, including fault horizons (linear patterns) and chaotic regions (disordered geological structures).

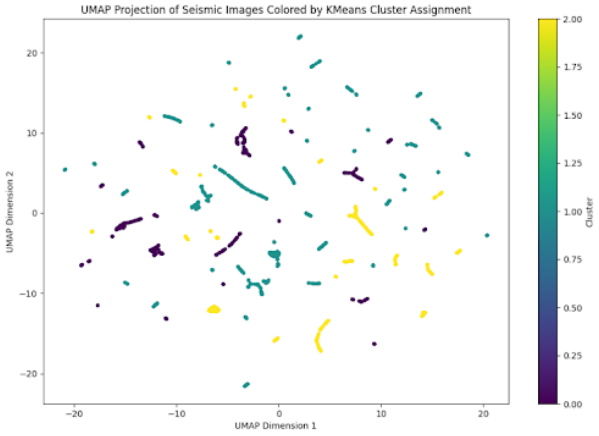

Figure 5. UMAP Projection of seismic images by true labels

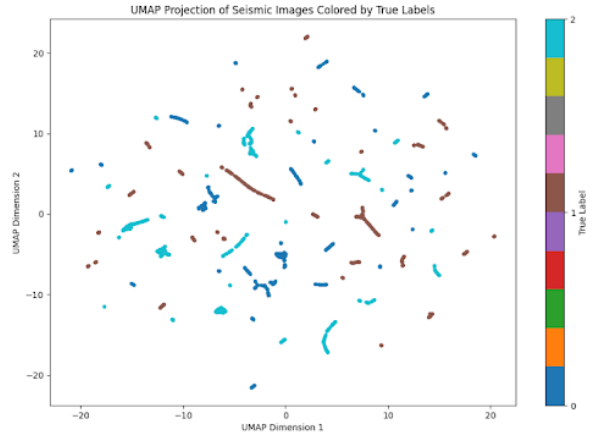

Figure 6. UMAP Projection of seismic images by K-Map Clustering

Fault horizons formed more distinct clusters in UMAP space, reflecting their structured nature.

Chaotic regions exhibited scattered distributions in both UMAP projections and clustering results, aligning with their irregular geometries.

Cluster Overlap and Transitional Regions Partial overlap between clusters was observed, which may indicate: Transitional seismic zones with mixed geological characteristics. Limitations in feature extraction or clustering resolution.

Model Performance and Geological Relevance: The

use of ResNet50-derived features validated the ability to differentiate fault horizons from chaotic regions, confirming that deep learning captures geologically meaningful structures. Misclassification or persistent overlaps suggest Potential gaps in feature representation and opportunities for domain-specific model adjustments or feature engineering to enhance separation.

2.4. Module 9

Simulation of seismic waves: A 2-D slice is taken out of the interior of mars which is then split into a 50 by 50 grid of pixels for visualization. 40 timeframes across 2000 seconds are taken to form the animation. Each pixel is colour graded according to the magnitude of u, the displacement vector at that point.

Shadow Zone Detection: To determine the shadow zones, we have plotted a graph of the maximum u observed at each point on the circumference of the cross section of mars, versus the angular position of that point with respect to the epicentre. The shadow zone is determined by analyzing where seismic waves fail to reach due to refraction. Shown below are the results obtained for P waves and S waves respectively.

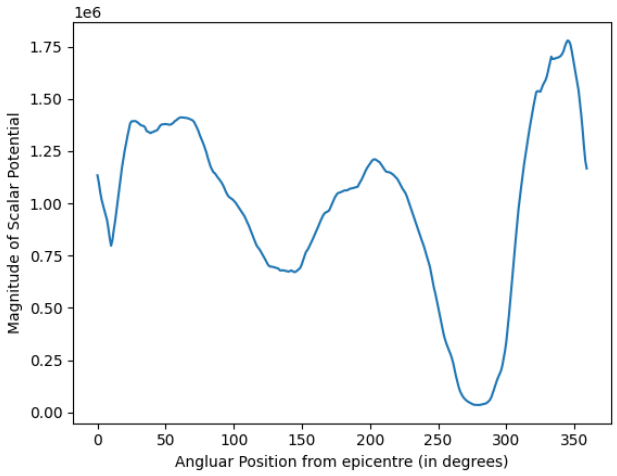

Figure 7. Graph of scalar potential vs epicentral degree for P-waves

For P waves, we can observe in the graph (Figure 7) two valleys, one approximately from 120 to 150 degrees and one approximately from 250 and 300 degrees. These are the P wave shadow zones that are formed due to the failure of P waves to reach these regions due to refraction. The formation of two separate shadow zones instead of one continuous shadow zone is attributed to the fact that P waves can pass through the core. Note that the disturbances in these regions are not exactly zero because of some finite amount of reflection of the p waves directly reaching the surface (without passing through the core).

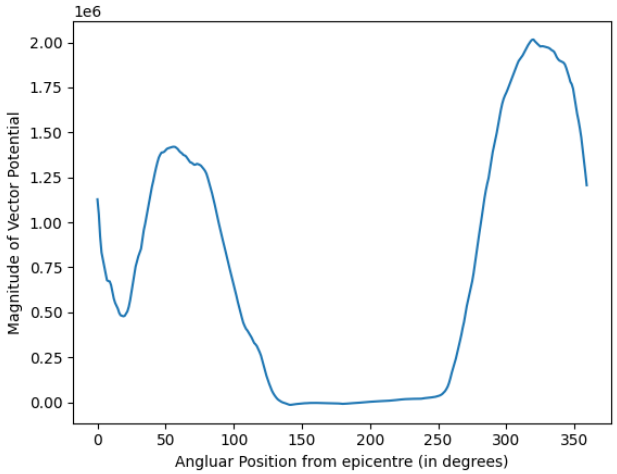

Figure 8. Graph of scalar potential vs epicentral degree for S-waves

For S waves, we can observe in the graph (Figure 8), a single broad valley, stretching from 130 degrees to 256 degrees. This the S wave shadow zone. Note that there is no break in the shadow zone, unlike the P wave shadow zone, because unlike P waves, S waves cant travel through the core.

3. Challenges Faced

Several challenges were encountered during the execution of the project, which are outlined below:

Complex Domain Expertise: Certain modules demanded in-depth knowledge of specialized domains such as seismology and Physics-Informed Neural Networks (PINNs), requiring significant time to adapt and apply effectively.

Extensive Literature Review: A substantial amount of time was dedicated to reviewing and analyzing numerous research papers to gain insights and develop specific modules.

Dataset Identification: One of the primary challenges was sourcing an appropriate dataset that met the required criteria and contained the necessary features for the task.

Resource-Intensive Model Training: The training of large-scale models proved to be time-consuming and resource-intensive, posing a significant hurdle in the project timeline.

Validation with Real-World Data: Aligning the predicted results with real-world data presented a considerable challenge, requiring rigorous analysis and refinement.