Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 11, 2025

MACHINE LEARNING-BASED PREDICTION OF SHEAR WAVE VELOCITY IN CARBONATE FORMATIONS OF ZAGROS BASIN IN IRAN

Shakir Ullah*[1], Sana Kashaf [2]

*[1]College of Geophysics, Lab: Earth Exploration and Information Technology, Chengdu University of Technology, 610059 China.
[2]College of Material, Chemistry & Chemical Engineering, Chengdu University of Technology 610059

*[1]shakirhayankhan365@gmail.com, [2]china.alikashaf107@gmail.com

DOI:https://doi.org/10.5281/zenodo.17568252

Keywords
Rock physical property, Shear wave Velocity. Reservoir characterization, Well log data, Compressional wave velocity, Reservoir data analysis, Geological data prediction

Article History
Received: 11 September 2025
Accepted: 21 October 2025
Published: 10 November 2025

Corresponding Author: *
Shakir Ullah

Abstract
The measurement of shear wave velocity is widely regarded as a critical rock physical property, and it may be effectively determined using the dipole sonic imager (DSI) tool. This parameter is essential for assessing geological properties like porosity, permeability, lithology, and fractures, though data is not universally available across wells. Accurate estimation is crucial for reservoir characterization. This study employs regression, multi-layer perceptron artificial neural network (MLPANN), adaptive neuro-fuzzy inference system (ANFIS), and multi-gene genetic programming (MGGP) to estimate shear wave velocity using well log data, while also reviewing existing empirical correlations in the literature. The dataset comprises various measurements obtained from the Bangestan Group Formation at a field located in the southwestern region of Iran. These measurements include depth, effective porosity, VP, gamma ray logs (both natural and spectral), neutron log, density log, and caliper log. This study compares all the employed methodologies based on their respective coefficients of determination (R2), root mean square errors (RMSE), mean squared errors (MSE), average absolute relative errors (AARE), and average relative errors (ARE). One of the approaches utilised in this study is the Modified Generalized Geometric Programming (MGGP) approach. This method was specifically designed to leverage its advantageous features, such as sensitivity analysis and correlation analysis. The input data is subjected to sensitivity analysis utilising the MLP-ANN and MGGP methods. The MGGP approach provides a correlation for predicting shear wave velocity using the input factors. Results show that the MLP-ANN technique is more accurate, reliable, and efficient than the other methods analyzed in this study. The R-squared values for the train, validation, and test phases are 0.9973, 0.9901, and 0.9898, respectively .The sensitivity analysis reveals that compressional wave velocity has the greatest impact on shear wave velocity.

INTRODUCTION

Shear wave velocity is commonly used in research on Earth's layers, especially when studying rock properties. Unlike other waves that can travel through liquids, shear waves cannot. This research helps us understand rocks better, particularly in identifying fluids within them and classifying their geological makeup. The link between rock type, porosity, and shear wave velocity offers key insights into underground water or oil resources (Khatibi and Aghajanpour, 2020). This data can be used to create a "mechanical earth model," a computer tool for analyzing the geological features and pressures beneath rock formations.

According to Plumb et al. (2000), the mechanical earth model helps us understand how rocks respond to external forces and their mechanical strength. Geomechanics, which studies rock behavior, is important in resource exploration, drilling, and reservoir management. Understanding how rocks react to stress is vital for ensuring safe and efficient drilling. Improving geomechanical models helps with issues like borehole stability, rock fracturing, drilling fluid pressure, reservoir changes, and gas injection effects. Plumb et al. (2004) note that geomechanics applies throughout the entire project, from exploration to completion.

The primary procedure in geomechanical investigations involves the examination and interpretation of data in order to forecast the elasticity of rocks, pore pressures, and inherent stresses inside rock formations (Shukla & Solanki, 2020). Nevertheless, the estimation of geomechanical characteristics poses challenges due to the intricate nature of reservoirs and the limited availability of dependable data. In instances where wells lack specific data, such as shear wave velocity, scholars employ well log data to generate approximations, as demonstrated by Shukla and Solanki (2020). There exist two primary approaches for this task: employing established relationships based on actual correlations and utilizing machine learning techniques.

Several studies have employed established correlations for determining Vs (Pickett, 1963; Castagna and Backus, 1993; Behnia et al., 2017). Nevertheless, the study conducted by Olayiwola and Sanuade (2020) state that it is crucial to acknowledge that there is no universally applicable methodology for establishing Vs in every reservoirs (Olayiwola & Sanuade, 2020). Several scholarly investigations (Heidari et al., 2010; Olayiwola and Sanuade, 2020) have identified issues associated with these established associations. The application of recognised correlations for estimating shear wave velocity in various reservoirs may not consistently produce trustworthy outcomes. Therefore, it is imperative to appropriately adjust or calibrate them in order to ensure their validity within the given context. The utilization of machine learning and optimization techniques has the potential to facilitate the prediction of shear wave velocity (Vs) by leveraging data pertaining to rock parameters. Numerous experts have employed various methodologies in an attempt to ascertain this velocity. In a study conducted by Eskandari et al. (2003), the forecasting of Vs values was conducted by employing a mixture of multiple regression modelling and artificial neural networks (ANN), utilizing rock property log data. In their study, Rezaei et al. (2007) employed artificial neural networks (ANN), neuro-fuzzy systems, and fuzzy logic methodologies to forecast the velocity of seismic waves (Vs) within the Carnau reservoir. This prediction was based on the analysis of well log data. In their study, Rajabi et al. (2010) employed neuro-fuzzy logic and genetic algorithms to estimate the velocities of various waves, specifically Vs, within an Iranian carbonate deposit. This estimation was conducted by utilizing rock property data as input.

In their study, Zoveidavianpoor et al. (2013) utilised both regression analysis and the ANFIS approaches in order to estimate the propagation velocity (VP). In their study, Asoodeh and Bagheripour (2013) employed Neural Networks (NN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and Fuzzy Logic (FL) methodologies for the purpose of predicting velocities of Vs., VP, and Vst. In their study, Nourafkan and Kadkhodai-Ilkhchi (2015) integrated ant colony optimization with a fuzzy inference system in order to forecast Vs values based on rock property data. In their study, Hadi and Nygaard (2018) employed artificial neural networks (ANN) and regression techniques to forecast the shear wave velocity (Vs) by utilizing well log data obtained from a well located in the southern region of Iraq. In their study, Anemangely et al. (2019) used a thorough approach that combined multiple methods, such as the least square support vector machine (LSSVM), particle swarm optimization (PSO), Cuckoo optimization algorithm (COA), and genetic algorithm (GA). The goal of their study was to figure out the Vs. values in the oil field in Ahvaz, which is in the southwestern part of Iran.

Behnia et al. (2017) various models have been developed for the purpose of predicting. The shear wave velocity (Vs.) in limestone. The researchers utilized data obtained from multiple reservoirs in Iran and employed gene expression programming (GEP) and the ANFIS as modeling techniques. In their study, Olayiwola and Sanuade (2020) employed ANN, ANFIS and LSSVM as computational models. The findings of their study shown that the utilization of LSSVM yielded highly precise forecasts for shear wave velocity.

The majority of these established relationships involve the estimation of shear wave velocity through the utilization of compressional wave velocity. However, it is vital to bear in mind that numerous variables can influence this particular metric. The models observed in scientific studies are mathematical constructs employed to depict and analyze correlations between variables. The data utilized in constructing these models is frequently derived from a particular rock type or geological formation.

Geomechanics is crucial for assessing oil reservoir development processes. In the oil industry, especially in complex drilling and enhanced oil recovery (EOR), inadequate knowledge of rock properties can lead to serious and potentially irreversible risks (Ali et al., 2003; Ezebialu et al., 2020). Developing mechanical models can help mitigate these risks. Incorporating data from acoustic logs—which include compressional and shear wave measurements, along with density information—is essential in building these models. Analyzing the elastic and strength properties of rocks allows for the assessment of stresses in the subsurface, whether from the rock itself or external influences. Key elastic properties, such as Young's modulus, Bulk modulus, and Poisson's ratio, can be predicted by examining wave velocities and the related stress conditions (Zoback, 2010).

Understanding properties like Poisson's ratio, Young's modulus, and Shear Modulus is vital for thorough geomechanical evaluations of reservoirs. These parameters are important for assessing well stability, planning fractures, modeling geomechanics, and plotting well trajectories. The dynamic modulus of a formation is largely determined by shear wave velocity (Vs), which is also crucial for analyzing the physical characteristics of the reservoir (Goodman, 1989).

The application of recognized correlations for estimating shear wave velocity in various reservoirs may not consistently produce trustworthy outcomes. Therefore, it is imperative to appropriately adjust or calibrate them in order to ensure their validity within the given context. The utilization of machine learning and optimization techniques has the potential to facilitate the prediction of shear wave velocity (Vs) by leveraging data pertaining to rock parameters. Numerous experts have employed various methodologies in an attempt to ascertain this velocity. In a study conducted by Eskandari et al. (2003), the forecasting of Vs values was conducted by employing a mixture of multiple regression modelling and artificial neural networks (ANN), utilising rock property log data. In their study, Rezaei et al. (2007) employed artificial neural networks (ANN), neuro-fuzzy systems, and fuzzy logic methodologies to forecast the velocity of seismic waves (Vs) within the Carnau reservoir. This prediction was based on the analysis of well log data. In their study, Rajabi et al. (2010) employed neuro-fuzzy logic and genetic algorithms to estimate the velocities of various waves, specifically Vs, within an Iranian carbonate deposit. This estimation was conducted by utilizing rock property data as input.

In their study, Zoveidavianpoor et al. (2013) utilised both regression analysis and the ANFIS approaches in order to estimate the propagation velocity (Vp). In their study, Asoodeh and Bagheripour (2013) employed Neural Networks (NN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and Fuzzy Logic (FL) methodologies for the purpose of predicting velocities of Vs, Vp, and Vst. In their study, Nourafkan and Kadkhodai-Ilkhchi (2015) integrated ant colony optimization with a fuzzy inference system in order to forecast Vs values based on rock property data. In their study, Hadi and Nygaard (2018) employed artificial neural networks (ANN) and regression techniques to forecast the shear wave velocity (Vs) by utilizing well log data obtained from a well located in the southern region of Iraq. In their study, Anemangely et al. (2019) used a thorough approach that combined multiple methods, such as the least square support vector machine (LSSVM), particle swarm optimization (PSO), Cuckoo optimization algorithm (COA), and genetic algorithm (GA). The goal of their study was to figure out the Vs values in the oil field in Ahvaz, which is in the southwestern part of Iran.

Behnia et al. (2017) various models have been developed for the purpose of predicting. the shear wave velocity (Vs) in limestone. The researchers utilized data obtained from multiple reservoirs in Iran and employed gene expression programming (GEP) and the ANFIS as modeling techniques. In their study, Olayiwola and Sanuade (2020) employed ANN, ANFIS and LSSVM as computational models. The findings of their study shown that the utilization of LSSVM yielded highly precise forecasts for shear wave velocity.

Table 1.1 Prior research on the estimate of Vs.

Input Variables	Methodology	Most Effective Technique	Reference
NPHI, GR, Vp, LLD, RHOB, Sw	Castagna, Brocher, Carroll, MLP, EPSO, and MLPPSO ANNs, and Elman	The Elman network is a type of recurrent neural network (RNN) that was proposed by Jeffrey Elman in 1990	Mehrgini et al. (2019)
RHOB- Vp- NPHI	ANFIS–Neuro Genetic–GEP	Adaptive Neuro-Fuzzy Inference System (ANFIS)	Behnia et al. (2017)
RHOB (Resistivity from the Borehole), MD (Measured Depth), GR (Gamma Ray), NPHI (Neutron Porosity), RES-DEP (Resistivity-Deep), RES-MID (Resistivity-Mid), and RESSHT (Resistivity-Shallow)	LSSVM, ANFIS, ANN, REG	Least Squares Support Vector Machines	(Olayiwola and Sanuade, 2020)
NPHI, Depth, GR, RHOB	The LSSVRegression technique encompasses linear regression, multiple linear regression, as well as artificial neural networks for both single variable and multiple variable analysis.	Artificial Neural Network-Multiple Variable	Adjei et al. (2020)
Resistivity of Hydrocarbon Bearing Formation (RHOB), Gamma Ray (GR), Resistivity (Res), Neutron Porosity, Density Porosity, Longitudinal Wave Acoustic Velocity, Stratigraphic Member Index.	TOB network	TOB Network	Wood (2020)
Vp, RHOB, GR, NPHI, PEF, LLD	Virtual Intelligence, EM, RPM, Statistical Regression	Clustering, classification, and regression data mining	Du et al. (2019)

Neutron Porosity, GR, RHOB, Vp, LLD	Multiple Regression, ANN	Artificial Neural Network	Eskandri et al. (2004)
Gamma Ray (GR) log, Compensated Acoustic Log, Bulk Density (RHOB) log, Photoelectric Absorption Crosssection Log, Neutron Porosity (NPHI) log, and Deep Lateral Resistivity Log.	Long Short-Term Memory	Long Short-Term Memory	Zhang et al. (2020)
NPHI (neutron porosity), Depth, RHOB (bulk density), PEF (photoelectric factor), GR (gamma ray), Caliper, Rt (resistivity), and Vp (compressional wave velocity).	LSSVM-COA, LSSVM-PSO, LSSVM-GA	The Least Squares Support Vector Machine with Centered Output Adjustment (LSSVM-COA)	Anemmangely et al. (2019)
Caliper- SP – GR – Vp – RS -Rt – DEN –CNL	ANN – ELM – SVR – CNN	ANN	(Wang and Peng, 2019)

Table 1.1 displays the investigations conducted in the domain of Vs estimation, together with most optimal way provided in each respective study. The risk analysis pertaining to the prediction of Vs using well logs was examined by some researchers. In these instances, a time period was chosen where the well lacked the pertinent log data. Initially, machine learning techniques were employed to forecast the well logs within the specified intervals. Following the prediction, a study of uncertainty was performed on the predicted outcomes, so offering supplementary insights and enhancing the dependability of the predictions. Additional information can be found in the investigations conducted by Feng et al. (2021).

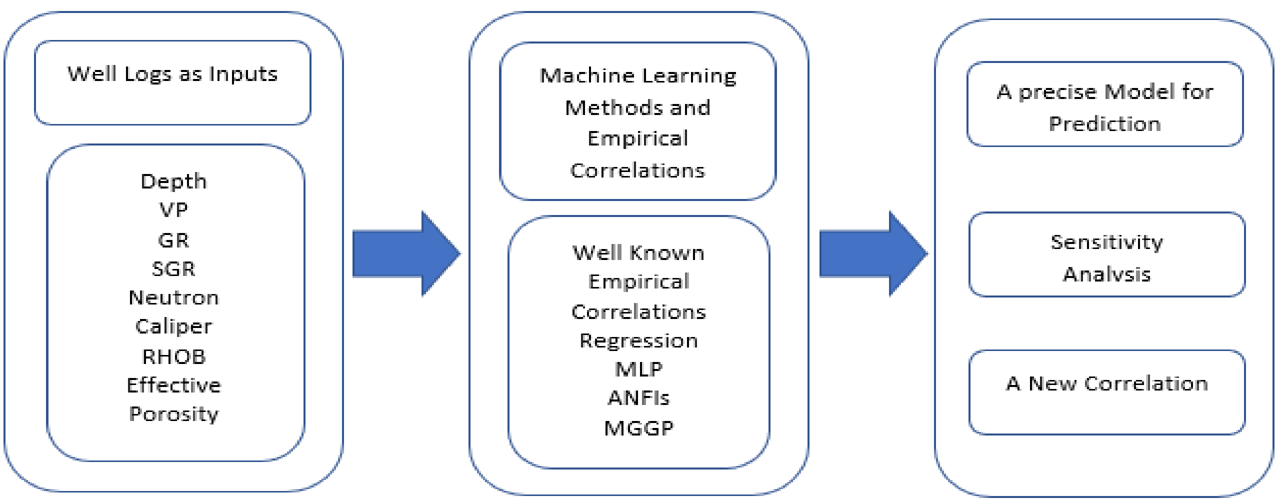


Figure 1.1 the procedural steps involved in utilizing artificial intelligence techniques for the Estimation of Vs.

Figure 1.1 illustrates the methodology employed in this study. In this research, the estimation of shear wave velocity is conducted by the utilization of empirical correlations and machine learning techniques. This estimation is based on log data and certain petrophysical parameters of a fractured carbonate reservoir located in the southwestern region of Iran. Initially, the application of commonly observed empirical correlations is employed to forecast the variable Vs. Subsequently, the application of straightforward and sequential linear regression techniques is employ

The prediction of parameters is accomplished by the utilization of sophisticated machine learning techniques such as Multilayer Perceptron Artificial Neural Networks (MLP-ANN), Adaptive Neuro-Fuzzy Inference Systems (ANFIS), and Multi-Gene Genetic Programming (MGGP). The selection of these methodologies was predicated upon a comprehensive examination of extant research. It is noteworthy that the variables employed for the estimation of Vs encompassed variables as depth, effective porosity, Vp, as well as both natural and spectral gamma ray logs, neutron log, density log, and calliper log. The results indicate that the MLP-ANN model outperforms other methods in terms of predictive accuracy for this particular parameter. Furthermore, the relationship demonstrates the estimation of the Vs value based on the aforementioned input elements.

Following the use of the MLP-ANN and MGGP methodologies, which exhibited superior accuracy in comparison to alternative approaches, an examination was conducted to assess the sensitivity of the input components. This study facilitated the identification of the most significant criteria in accurately predicting Vs. Subsequently, the Vs and Vp values were utilized to determine certain elastic parameters that are fundamental to the geomechanical methodology. Both current and future data were utilized in this study, and subsequent analysis was conducted on the obtained data.

This study presents a robust methodology for estimating shear wave velocity based on data obtained from an oil field located in Iran. The objective of this study was to conduct a comparative analysis of various widely-used machine learning techniques and conventional methods, utilizing historical data and regression models, with a specific focus on shear wave velocity. It has been determined that conventional methodologies exhibit certain limitations. Moreover, to the best of our knowledge, this study is the initial instance in which a correlation has been established between the estimation of shear wave velocity and the assessment of the input components' sensitivity. The proposed methodology demonstrates efficacy in forecasting the velocity of shear waves, as evidenced by an R2 coefficient of 0.8, indicating a high level of accuracy.

2. Theoretical Approach

2.1. Empirical Association

Numerous scholarly investigations have been conducted to examine the methodologies employed in ascertaining shear wave velocity (Vs) through the utilization of compressional wave velocity (Vp) across various geological formations. Several studies have been conducted on this topic, including those by Pickett (1963), Carroll (1969), Castagna and Backus (1993), Eskandari et al. (2003), and Brocher (2005). The primary results derived from these investigations are documented in Table 2.1. It is noteworthy to emphasize that the coefficients employed in the study conducted by Castagna et al. are specifically tailored for dolomite rocks. The coefficients have varying values when applied to sandstones.

Table 2.1 prevalent empirical and regression methods utilized for the estimate of Vs...

MODEL'S NAME	RELATIONSHIP	REFERENCES
Pickett (1963)	$Vs = Vp / 1.9$	Pickett (1963)
Carroll (1969)	$Vs = 1.09913326 \times V^{0.9238}p$	Carroll (1969)
Castagna and Backus, (1993)	$Vs = 0.58321Vp - 0.07775$	Castagna and Backus, (1993)
Eskandari et al. (2003)	$Vs = -0.1236V^2p + 1.612Vp - 2.3057$	Eskandari et al. (2003)

Brocher (2005)	Vs = 0.07858 − 1.2344Vp + $0.7949V^2p$ − $0.1238V^3p$ + $0.0064V^4p$	Brocher (2005)

2.2. Regression analysis

Regression analysis is a statistical methodology employed to construct a mathematical model that elucidates the relationship between two variables, drawing upon empirical or laboratory data. Once constructed, this model possesses the capability to forecast novel values or elucidate the manner in which the two variables interrelate. Multivariate regression analysis is a statistical technique that examines the relationship between numerous independent variables and a single dependent variable, with the objective of identifying the most optimal equation that provides the best fit. This approach utilizes the available data to make an estimation of the desired parameter by means of a mathematical connection, commonly referred to as the estimated function. The formula for this estimate is as follows:

$$h(x) = b0 + b1x1 + \ldots + bk\ xk \quad (2.1)$$

Where x1 is input variables and b1 is weight parameter. These are acquired through the process of minimizing the objective function (OF). The objective function (J), mathematically characterized as:

$$J=12\Sigma\ (h\ (xn) - yn)\ 2mn=1 \quad (2.2)$$

Within the given framework, the symbol "m" denotes the aggregate quantity of data input, whereas "yn" signifies factual value of the coefficient under consideration for prediction. In the context of multi-variable linear regression, the coefficients within the linear model are ascertained through the utilization of an objective function in conjunction with the values assigned to the variables. This study employs fundamental and systematic linear regression methods to construct the ultimate method and predict the Vs. In statistical analysis, stepwise regression is used to iteratively modify a linear model by including or excluding variables.

In each iteration the model makes determination regarding the inclusion or exclusion of characteristics based on a predefined criterion .The primary metric employed is SSE.The F-statistic and p-valued were employed to evaluate method at each iteration ,thereby assessing the statistical significances of each effect. Stepwise regression is a statistical technique that enables the creation of several models by iteratively selecting and removing parameters from a given collection. The models generated are contingent upon the initial inclusion of parameters and the sequence in which they are changed. This approach iteratively refines the model until further enhancements are no longer achievable. Nevertheless, the efficacy of these models in consistently yielding optimal solutions remains uncertain. In essence, although stepwise models may exhibit superior performance within their immediate context, they may not necessarily achieve the highest level of overall optimization.

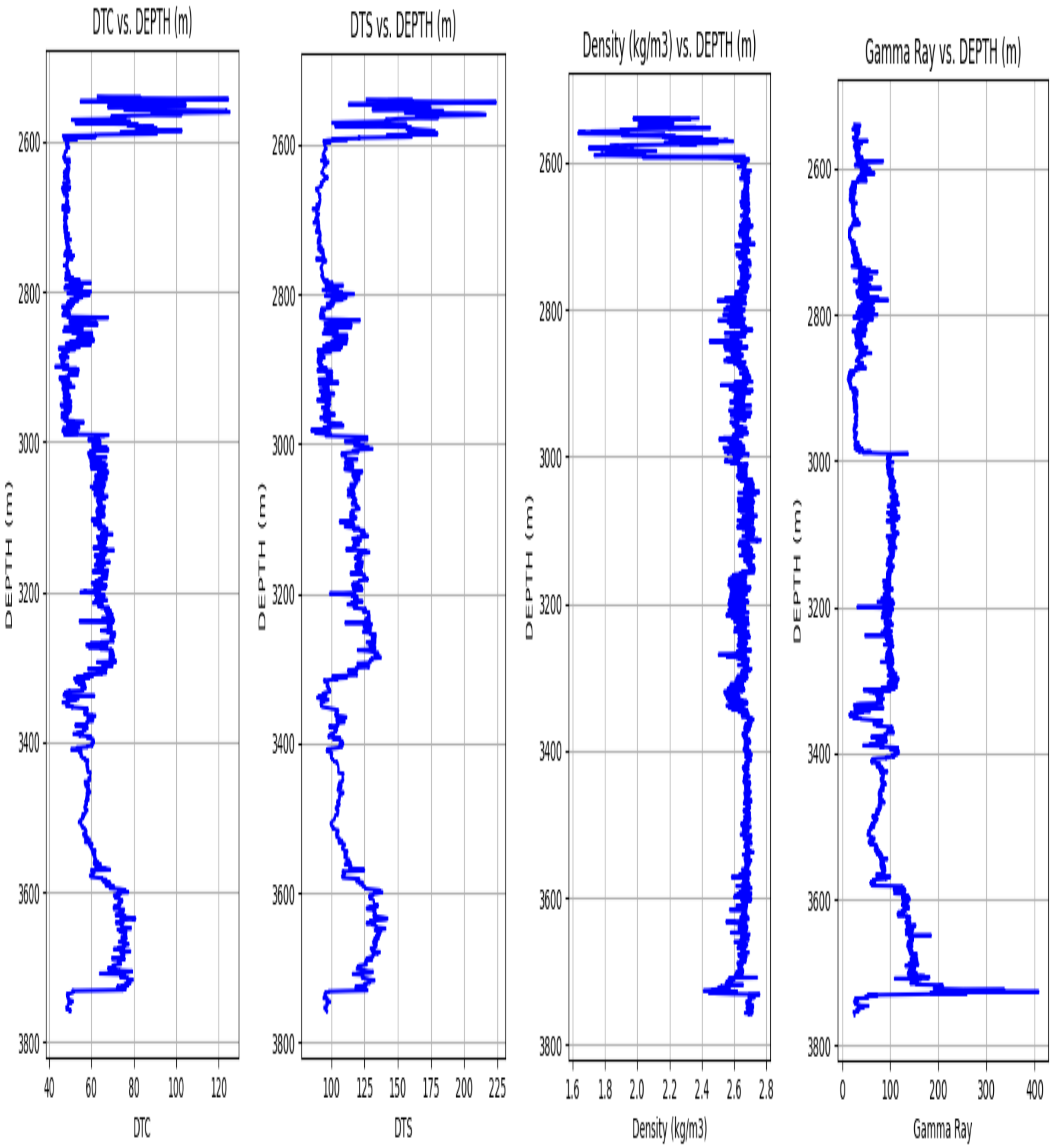
Table 2.2 Values of input parameters

Variables	Starting	Ending
Depth (m)	2022.61	2940
Vp (KM/SEC)	0.38008	0.840535
GR (GAPI)	4.771408	36.63161
SGR	6.70	82.6
Neutron	0	0.3813
Caliper	11.82	18.97
Eff. Por	1.00×10^6	0.197414
RHOB	2.1023	2.95036

Vs (km/s)	0.0046321	0.0130042

2.3. Artificial Neural Network (ANN)

Artificial Neural Networks (ANN) consist of interconnected units that operate concurrently. Similar to the natural world, the manner in which these units are interconnected plays a crucial role in determining the functioning of the network. It is imperative to establish a network wherein the interconnections among units are dynamically adapted in accordance with their relative significance or magnitude. The network continuously adapts its parameters in order to minimize the discrepancy between its predictions and the observed outcomes. In the context of supervised learning, the neural network undergoes training by utilizing a multitude of input-output pairings that are already known. Artificial neural networks (ANNs) are specifically engineered to replicate the fundamental learning mechanisms observed in biological systems. The utilization of these methods has been extensive across various disciplines within the field of earth sciences (Hertz et al., 1991).


DTC vs. DEPTH (m)
DTS vs. DEPTH (m)
Density (kg/m3) vs. DEPTH (m)
Gamma Ray vs. DEPTH (m)
DEPTH (m)
DTC
DTS
Density (kg/m3)
Gamma Ray

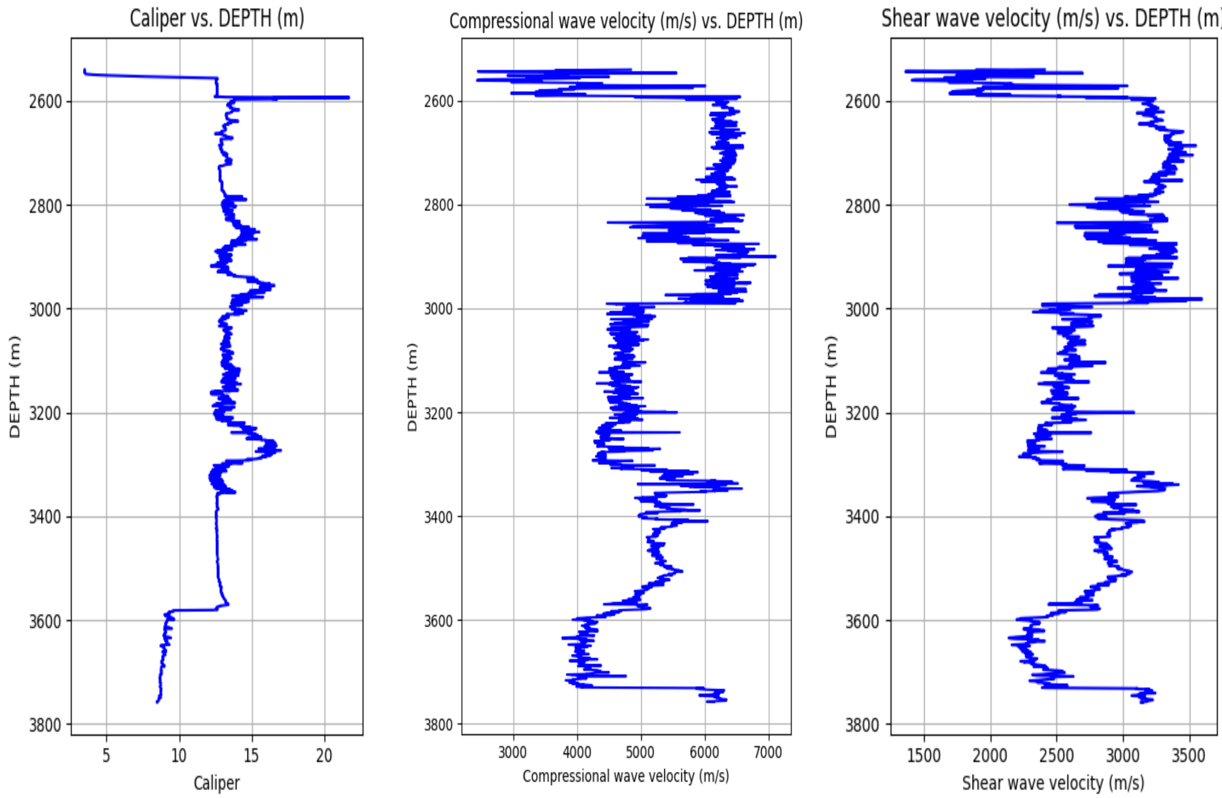


Figure 2.1 (A), (B), (C), (D), (E), (F) and (G) Data for developing intelligent models.

Neural networks have been recognized as highly effective instruments for identifying intricate and non-linear associations among variables (Haykin, 1999). Individuals possess the cognitive ability to comprehend complex nonlinear patterns, especially in instances where the data exhibits disorder or lacks clarity. In contrast to alternative approaches, it is not necessary to possess prior knowledge of the input variables in order to discern their relationship with the outputs. Neural networks possess inherent advantages over certain conventional empirical and statistical methodologies. The utilization of these tools is particularly advantageous within industrial contexts due to their ability to identify intricate patterns within data, hence facilitating the identification of significant associations in both input and output variables (Akhundi et al., 2014; Maghsoudian et al., 2021).

ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 11, 2025

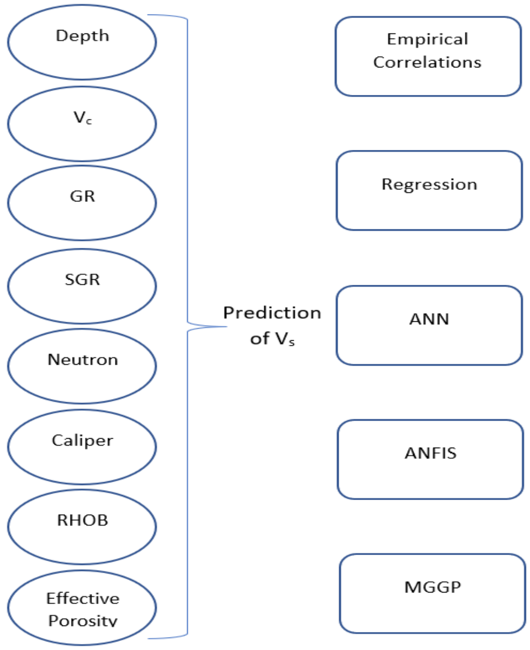


Figure 2.2 Data for developing intelligent models.

The training of neural network systems frequently involves the utilization of the Multi-layered Back-Propagation technique. The back propagation networks, also known as multi-layered networks, possess a distinctive characteristic wherein they employ a non-linear propagation function in conjunction with the Windrow-Hoff learning approach. The utilization of input and target vectors enables the computation of a function, facilitates comprehension of the relationship between input and output, and facilitates the categorization of inputs according to predetermined norms established by the developer. The back-propagation network with bias exhibits a sigmoidal activation function in its hidden layer and a linear activation function in its output layer, enabling it to effectively estimate any function within a restricted range of continuous data points.

The initial stage in the training process of an Artificial Neural Network (ANN) involves establishing its structural framework. The aforementioned structure consists of two primary components. The first aspect pertains to the characteristics of the network, encompassing its distinct functionalities, the quantity of neurons employed, and the presence or absence of hidden layers. The subsequent aspect pertains to the learning principles it adheres to. These regulations aid in determining factors such as rates, weights, and biases. Majdi and Beiki (2010) additionally provide guidance on the selection of parameters throughout the training process to enhance the performance of the network. Equation (2.3) displays the output of the neural network, while Equation (2.4) represents the function employed in this study.

$$Y_j = f\left(\sum_{i=1}^{n} w_{ij} x_{ij} + b_j\right) \quad (2.3)$$

$$(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}} \quad (2.4)$$

The variables W and b denote the weights as well as the bias of the neural network, correspondingly. Figure 2.2 is the representation of artificial neural network methodology.

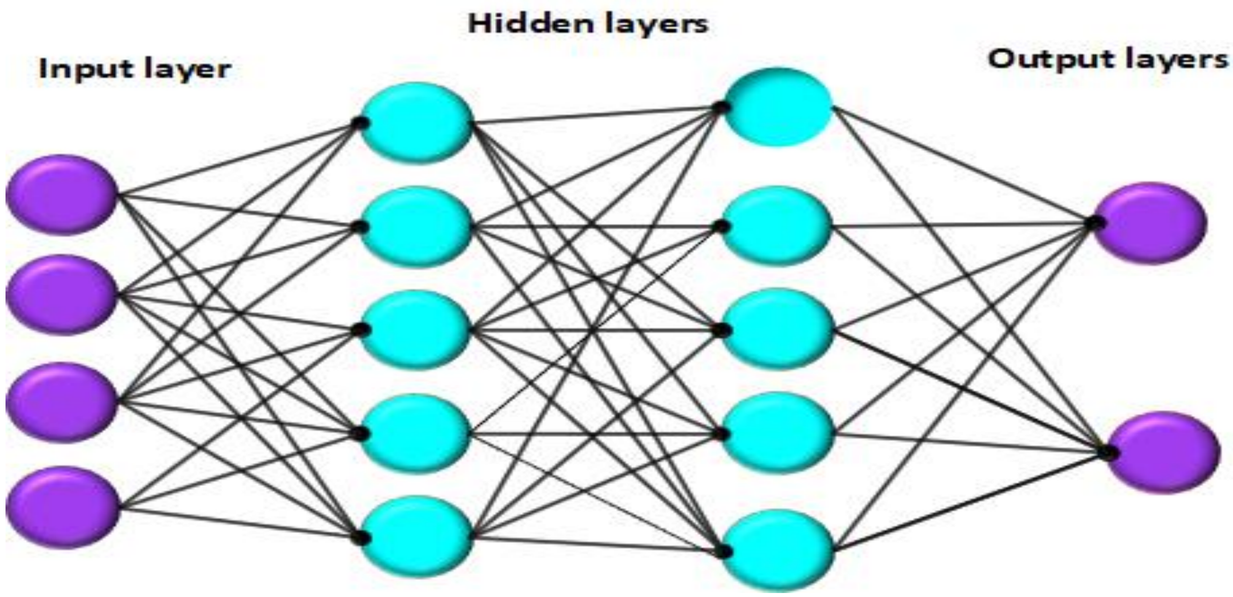


Figure 2.3: Artificial Neutral Network Structure

2.4. ANFIS for Predicting Vs

The Adaptive Neuro-Fuzzy Inference System (ANFIS) is a new type of computer model that takes the best parts of the Fuzzy Inference System (FIS) and the Artificial Neural Networks (ANN). The integration of fuzzy inference systems was initially proposed by Jang (1993) as a means to address the limitations identified in previous research (Gholizadeh and Sabzi, 2017). There exist 3 primary categories of FIS, namely Tsukamoto fuzzy models, Sugeno fuzzy models, and Mamdani fuzzy models. The distinguishing factors of those models lie in their distinct fuzzy rules and the methodologies employed for the aggregation and defuzzification processes, wherein a fuzzy collection is transformed into a crisp numerical value. The Sugeno fuzzy inference method was selected for this research.

The performance of ANN is intricately linked towards quantity and quality of the training data at their disposal. According to Kong and Kosko (1992), the reliability of findings obtained from artificial neural networks (ANNs) may be compromised when there is a limited amount of training data available. In instances of this nature, the integration of fuzzy logic with artificial neural networks (ANN) has been shown to enhance the productivity of the neural network and yield more precise outcomes (Nava and Taylor, 1996).

Upon examining the architecture of Adaptive Neuro-Fuzzy Inference Systems (ANFIS), it becomes evident that there exist two distinct categories of parameters that might undergo training: the antecedent parameters and the consequent parameters. The system employs a least squares technique to determine the subsequent parameters. In contrast, the antecedent parameters are modified through the utilization of a gradient descent technique. The Sugeno-type Adaptive Neuro-Fuzzy Inference System (ANFIS) configuration involves the allocation of distinct layers for the purpose of accomplishing certain tasks, hence facilitating the enhancement of the model. Learning process in ANFIS consists of two main components: a forward process and a backward process. In the initial phase, the antecedent parameters remain unchanged, while the least square method is employed to refine the consequent variables. Within context of the given scenario, the gradient descent technique is employed to iteratively modify the preceding parameters associated with the membership functions of the input variables. Subsequently, the system determines the output by calculating the weight average of subsequent varibles. In the event that the output is incorrect, the antecedent parameters are adjusted using the back-propagation technique, as discussed by Beale et al. (1992) and Guan et al. (2008).

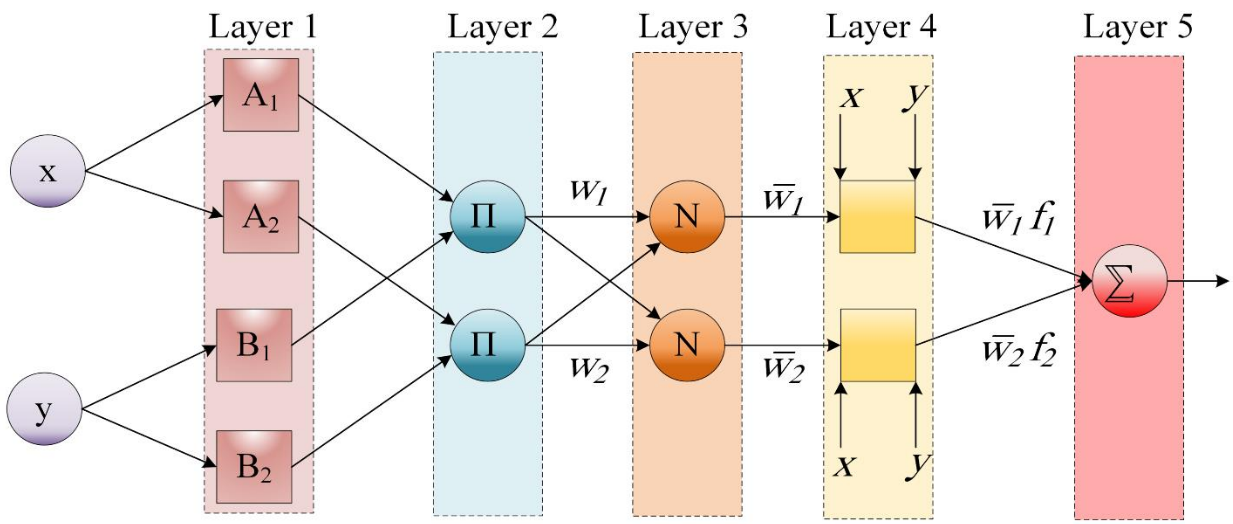


Figure 2.4: the structure of an Adaptive Neuro-Fuzzy Inference System (ANFIS).

This diagram elucidates the structure and foundational principles of the ANFIS structrure. As outlined in this figure and as expounded by Kayadelen (2011), the computational process of ANFIS is segmented into five core components.

1. A layer for membership functions that handles both input and output.
2. The Rule layer, where every node is fixed. The outcome of each node in this layer is derived from the input values of its preceding layer.
3. A layer dedicated to normalization.
4. In the adaptive layer, each node functions as a linear entity. The adjustment of the constants for these functions is achieved by a combination of back-propagation and least squares approximation methods.
5. The final layer is the output layer, where outcomes are aggregated from the outputs of the nodes in the prior layer.

During the training process of the Adaptive Neuro-Fuzzy Inference System (ANFIS), the nonlinear parameters extracted from the training data are utilized in initial layers, which corresponds towards fuzzy functions. Simultaneously, the linear variables in fourth layer are established in order to gain the desired output (Jang et al., 1997).

2.5. MGGP for Predicting Vs

The fundamental concepts behind the genetic algorithm were initially presented in the early 1960s by Goldberg and Holland (1988). There are 2 primary factors that distinguish GA from other entities: There are two key aspects to consider in this context: firstly, the outcome is contingent upon chance, and secondly, the analysis encompasses a set of potential answers rather of focusing solely on a single one. The algorithm exhibits proficiency in amalgamating diverse solutions to identify the optimal one, leveraging the individual qualities of each solution (Sivanandam & Deepa, 2008). The Genetic Programming System for the recognition of Physical Systems was made to make accurate models for making predictions. The process involves doing a search inside a dataset in order to identify the most optimal connections between input variables and output variables. In order to increase the simplicity, robustness, and accuracy of standard bio-inspired genetic programming (GP) algorithms, GPTIPS employs the utilization of symbolic multigene regression (SMGR) as proposed by Searson (2015).

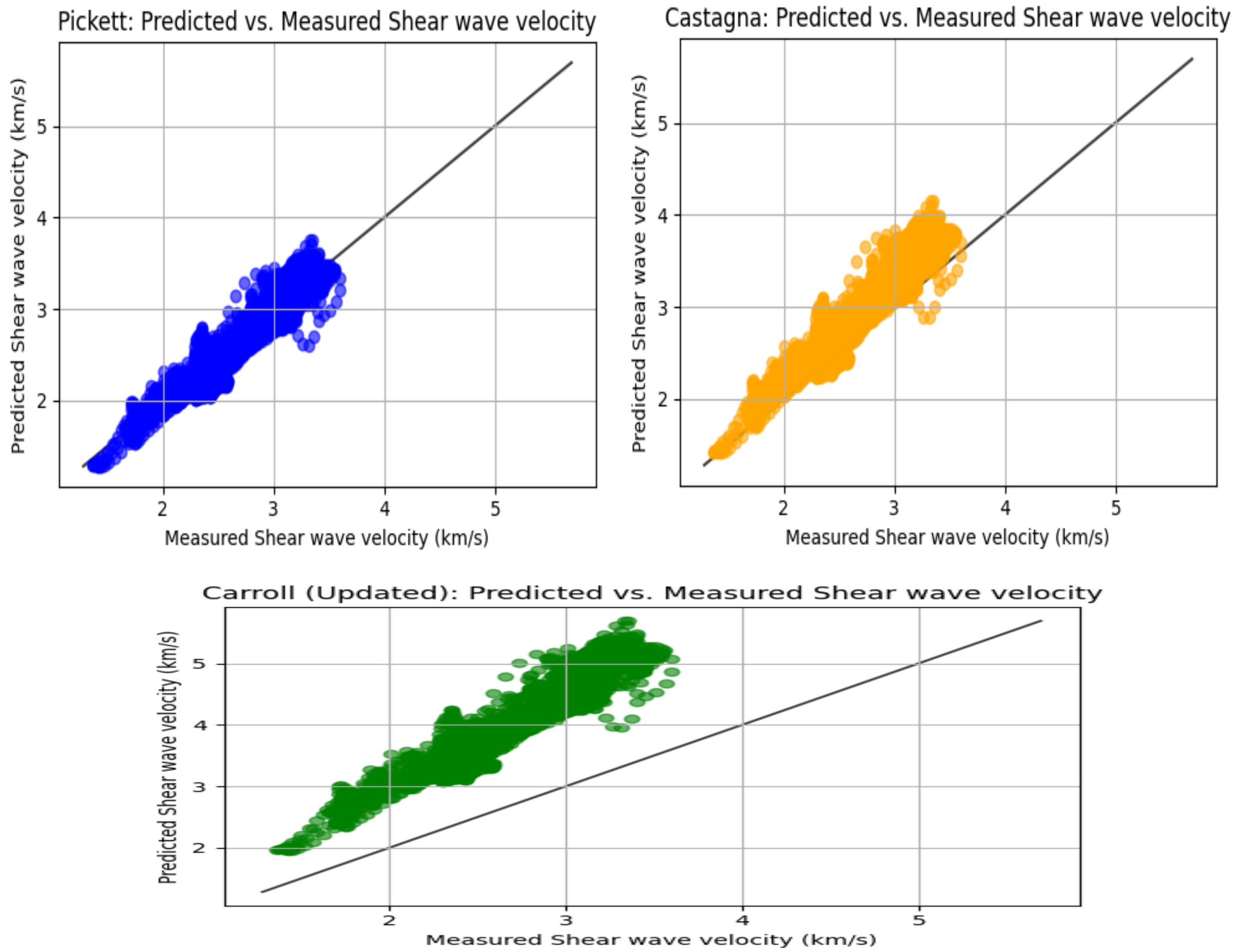


Figure 2.5: the shear wave velocity was determined using the utilization of empirical correlations, specifically the (a) Pickett, (b) Castagna and (c) Carroll methods

Genetic Programming (GP) is a variant of the genetic algorithm that incorporates the ability to modify the chromosomes (Koza & Koza, 1992). The primary distinction between Genetic Programming and Genetic Algorithms is in their respective approaches to data representation. GA often employs fixed-length binary strings as the preferred representation method, whereas GP evaluations exhibit a greater diversity in their data representation strategies. Gene Expression Programming (GEP) is a relatively recent variant of Genetic Programming (GP) as introduced by Ferreira in 2001. The neuro-genetic system employing a Genetic Algorithm (GA) commences by initializing a set of chromosomes in a random manner. As the progression unfolds, operators from the preceding generation exert an impact on the subsequent generation. The evaluation of a chromosome's suitability for progression to the subsequent stage is based on its function and sub-function. Throughout the progression of the process, the number of units that successfully pass on to subsequent generations gradually diminishes, culminating in the survival of a solitary chromosome that exhibits the most optimal match. The Multi-Gene Genetic Programming (MGGP) model represents a sophisticated variant of Genetic Programming (GP), which combines conventional regression techniques with GP methodologies. The distinguishing characteristic of MGGP is in its capacity to assess data across numerous dimensions and its rapid rate of learning.

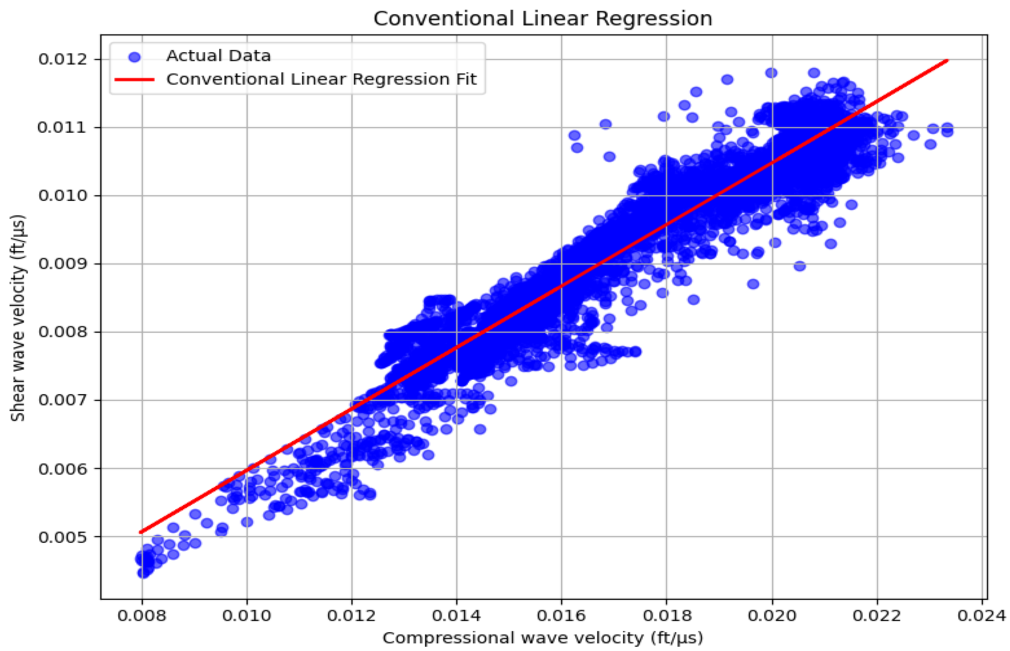


Figure 2.6 the shear wave velocity was determined through the application of conventional linear regression.

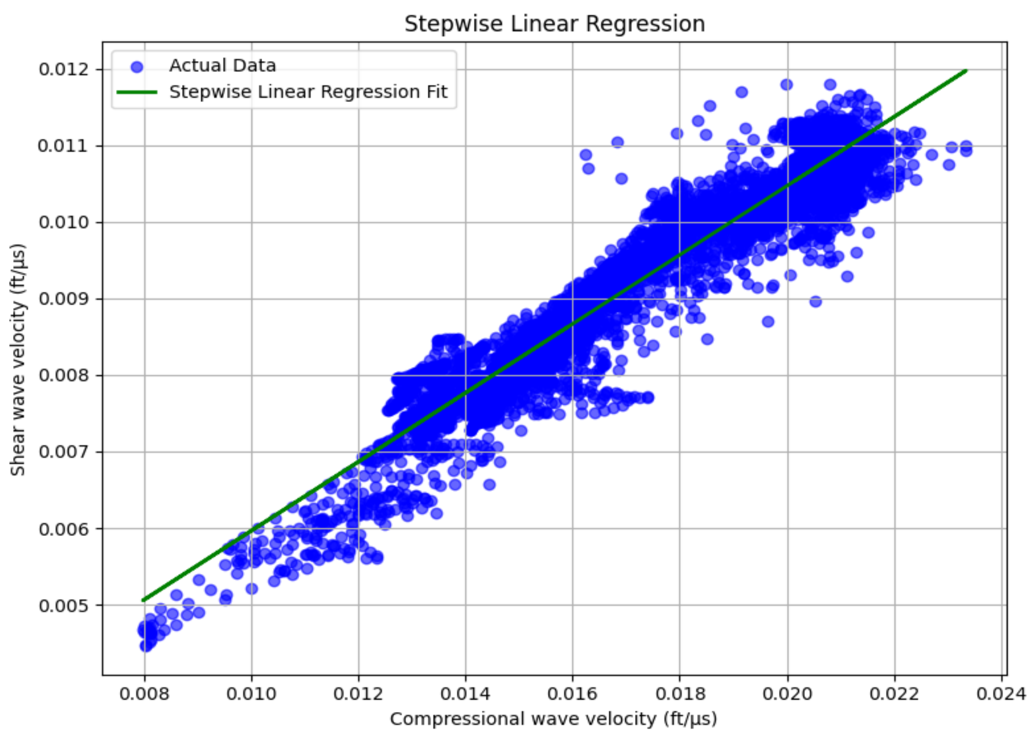


Figure 2.7 the shear wave velocity was determined through the application of a statistical technique known as Stepwise Linear Regression

According to Farmani et al. (2019), the outcomes derived from the MGGP model hold greater significance compared to those obtained using the conventional GP model.

In the context of Genetic Programming (GP), the initial solutions consist of trees that are generated randomly. The objective of investigating the evolutionary patterns of these is to develop a precise method that aligns with the provided input data. In contrast, GPTIPS adopts a distinct methodology. Instead of commencing with individual trees, the approach initiates with clusters of multi-tree solutions, referred to as Multi-Gene Genetic Programming (MGGPs). Consider these Multi-Granularity Graph Partitions (MGGPs) as collections of arrays that consist of numerous trees. The efficacy of each proposed solution is subsequently assessed. Based on this assessment, a specific portion of the population is chosen as the parental group. The process of selecting the candidate is conducted by a probability Pareto tournament. Upon being chosen, these proposed solutions undergo a process of evolutionary modifications, such as mutations and crossings, in order to generate the subsequent generation. The aforementioned process persists until a predetermined condition specified by the user is satisfied (Searson, 2015). According to Gandomi and Atefi (2020), GPTIPS provides users with the ability to determine the stopping criteria for the algorithm. This includes options such as terminating the algorithm when an optimal model fit is attained, when a maximum allowable runtime is exceeded, or after a predetermined number of generations has been finished. The present study utilizes the MGGP approach to forecast Vs data based on well log data. Figure 2.8 provides a visual representation of the structural framework employed in the MGGP approach.

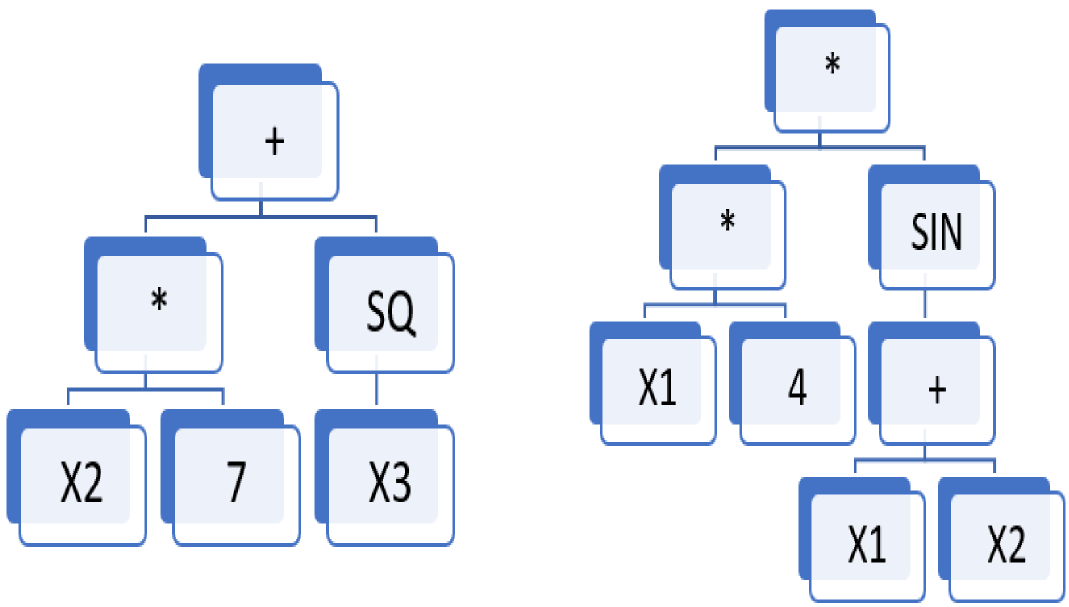


Figure 2.8 Schematic of MGGP structure

3. Methodology

3.1Collecting Data

This study delves into intricate details of an oil reservoir situated in the south-western region of Iran. A notable characteristic of this reservoir is its high gas oil ratio (GOR), a phenomenon previously explored in a study conducted by Izadpanahi et al. (2020). Two primary geological configurations define the reservoir's structure: the Ilam and Sarvak formations.

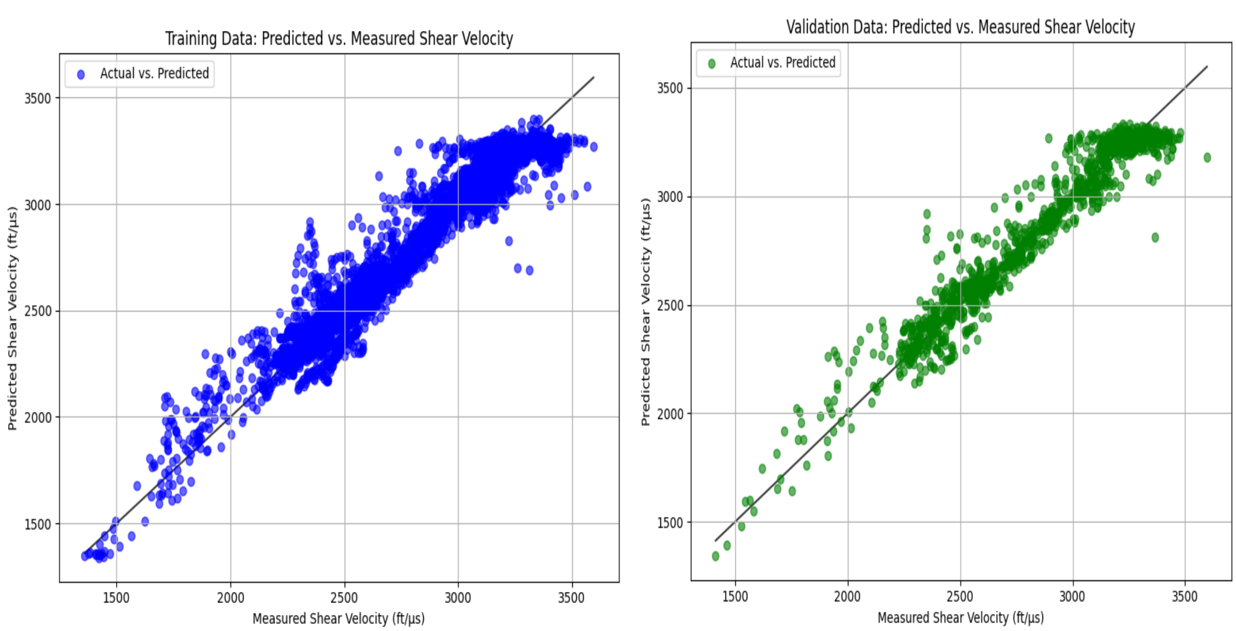
This field has an exceptionally large reservoir fractured and is predominantly made up of carbonate materials. The reservoir's geometry is intriguing; it boasts an asymmetrical shape oriented in an east-west direction. The gradients of its northern and southern slopes are 22 and 14 degrees, respectively, which might influence the flow dynamics and reservoir management strategies. The average elevation of the area earmarked for research stands at 1500 meters, which could have implications for drilling and extraction techniques.

The Ilam Formation, one of the two primary formations in the reservoir, has a unique lithological composition. It is primarily made up of limestone, but what makes it distinct is the presence of micrite, a fine-grained carbonate. The thickness of the Ilam Formation varies across the reservoir. In the northeastern, western, and southwestern regions, there's a decline in thickness. In contrast, the northern section sees an increase. This variation in thickness ranges between 36 meters at its thinnest to 51 meters at its thickest.

To predict Vs., or shear wave velocity, within the formation, a selection of logs was employed. These logs serve as tools to understand the formation's properties and characteristics at various depths. The chosen logs for this endeavor included depth logs, which provide information about the depth at which measurements are taken; effective porosity logs; Vp or compressional wave velocity logs; gamma ray logs in both their natural and spectral forms, which can indicate the presence of certain minerals or substances; neutron logs , which can provide data on hydrogen content; density logs, offering insights into the rock's density at various depths; and caliper logs, which measure the borehole's size and shape. Each of these logs play a crucial role in painting a comprehensive picture of the Sarvak Formation’s internal structure and potential.

Hence, for the construction of models, a total of 8192 sets of data and 7 logs were utilized. Table 2.2 presents comprehensive data utilized in the current investigation. Figure 2.1 additionally illustrates the logs used as input for the model.

Artificial Intelligent models stand out in the realm of computational modeling due to their simplicity, effectiveness, and efficiency in predicting output variables. Unlike traditional deterministic models that rely heavily on understanding the underlying physics and mathematical relationships, intelligent models lean on pattern recognition, data-driven insights, and learning algorithms. This makes them particularly adept at handling complex datasets and providing quick predictions.



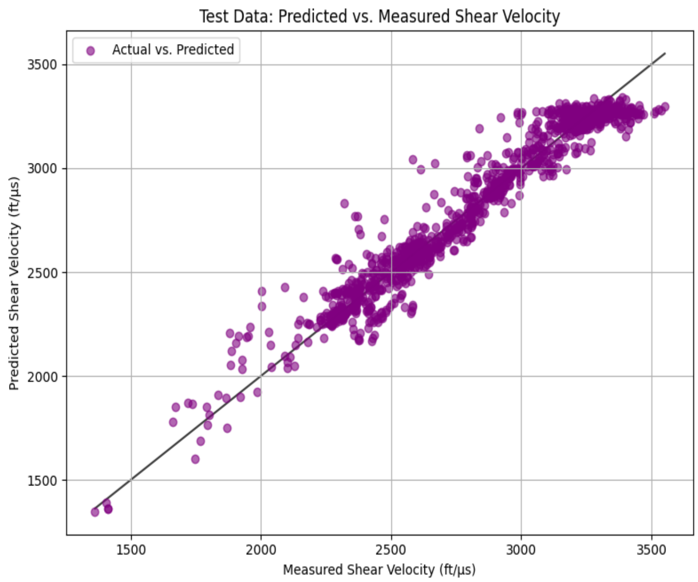


Figure 3.1: MLP-ANN Vs measurement (a) Train, (b) validate (c) test.

In a recent study, the task of estimating shear wave velocity was undertaken using a variety of empirical correlation methods. These methods were sourced from previous publications and encompassed five distinct approaches: regression, artificial neural networks (ANN), adaptive neuro-fuzzy inference systems (ANFIS), multi-gene genetic programming (MGGP), and one more that seems to be missing from the provided list. A visual representation of the methodologies and inputs employed in this research can be found in Figure 2.2. To provide a comprehensive understanding of each method, the literature was extensively reviewed. This review yielded insights into the advantages and disadvantages associated with each of the aforementioned strategies, and these findings have been systematically compiled and presented in Table 2.3.

In order to accommodate the extensive range of input parameters and the substantial volume of input data, the model has been designed with three hidden layers, each consisting of 25 neurons. In this approach, 80% of the dataset was allocated for training purposes, while 5% was reserved for validation and the remaining 15% was utilized for testing.

In the methodology adopted for this study of ANFIS the data was divided in a manner that 80% of it was earmarked for the training phase, while the remaining 20% was allocated for testing. This division was based on the dispersion of the data, ensuring a representative split that would allow for robust model training and subsequent validation.

For the implementation of the MGGP method, the dataset was divided in a specific manner: 80% was allocated for the training phase, a smaller 5% was set aside for the validation phase, and the remaining 15% was used for the testing phase. This distribution ensures that the model is not only trained on a significant portion of the data but also validated and tested to ensure its generalizability and robustness.

3.3 **Statistical Analysis**

The accuracy and trustworthiness of the models are evaluated using statistical methods. The statistical evaluation parameters used in this work are formulated mathematically and reported as follows (Mohamadi-Baghmolaei et al., 2021):

Determination coefficient:

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(VS_i^{Pred} - VS_i^{Exp})^2}{\sum_{i=1}^{N}(VS_i^{Pred} - average(VS_i^{Exp}))^2} \quad (3.1)$$

Average relative error:

$$Average\ \% = \frac{100}{N}\sum_{i=1}^{N}\left(\frac{Vs_i^{Pred} - Vs_i^{exp}}{Vs_i^{exp}}\right) \quad (3.2)$$

Average absolute relative error

$$Average\ Absolute\ Error\ \% = \frac{100}{N}\sum_{i=1}^{N}\left(\left|\frac{Vs_i^{Pred} - Vs_i^{exp}}{Vs_i^{exp}}\right|\right) \quad (3.3)$$

Root Mean square error:

$$RMSE = \left(\frac{\sum_{i=1}^{N}\left(VS_i^{Pred} - VS_i^{Exp}\right)^2}{N}\right)^{1/2} \quad (3.4)$$

Mean Squared Error:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(Vs_i^{Pred} - Vs_i^{exp})^2 \quad (3.5)$$

3.3.1 Examination of Sensitivity

The main purpose of sensitivity analysis is to ascertain the impact of various factors on the final decision. The aforementioned analysis can be conducted in order to identify the parameters that exert a substantial influence on the target value. In order to ascertain the degree of relevance, researchers may utilize the equation proposed by Hosseinzadeh and Hemmati-Sarapardeh (2014) as well as the equation put out by Naghizadeh et al. (2022).

$$r(I_k . V_s) = \frac{\sum_{i=1}^{n}(I_{k.i} - \bar{I}_k)(V_{s.i} - \bar{V}_{s.i})}{\sqrt{\sum_{i=1}^{n}(I_{k.i} - \bar{I}_k)^2 \sum_{i=1}^{n}(V_{s.i} - \bar{V}_{s.i})^2}} \quad (3.6)$$

Where r is the relevancy factor, $I_{k.i}$ is input variable, $\bar{I}_k$ is mean of the input variable, $V_{s.i}$ is predicted Vs and $\bar{V}_{s.i}$ is mean of predicted Vs. A numerical value between -1 and +1, the relevance factor indicates how strongly the given input is related to the desired output. (Hosseinzadeh and Hemmati-Sarapardeh, 2014).

4. Results and Discussion

4.1. Performance Assessment of Empirical Models

Theoretical framework of this research, a multitude of empirical relationships have been provided by experts to predict Vs. However, a recurring challenge with many of these models is their lack of holistic understanding. Specifically, they frequently fall short of accurately representing the complex interplay between reservoir rock qualities and the one-of-a-kind conditions present in every reservoir. This limitation can lead to models that, while theoretically sound, may not be universally applicable or accurate in diverse settings.

From the data scrutinized in this study, it becomes clear that only a fraction of these correlations—specifically three—yield results that can be deemed satisfactory. The detailed outcomes of the shear wave predictions, as derived from each model, are visually represented in Figure 2.5. For clarity and consistency, it's crucial to note that while formulating these empirical correlations, the unit of measurement used throughout is the Vs, and it's denoted in kilometers per second (km/s).

Furthermore, an analysis of the errors associated with each prediction method has been tabulated in Table 4.1. Among the various techniques employed, the picket approach emerges as the most reliable, registering the lowest error rate. This suggests that, for this specific field, the Shear wave velocity data maintains a steady linear relationship with the compressional wave velocity. This relationship is further characterized by a consistent coefficient value of 1.9. Such findings underscore the importance of selecting the right empirical model based on the specific characteristics of the reservoir in question.

Table 4.1 Evaluation of Empirical Model Errors.

Variables	Pickett	Carroll	Castagna
AARE %	6.745	126.8	14.4
ARE %	3.99	127.2764	-13.281
MSE	5.95×10^{4}	1.35×10^{1}	0.0019

RMSE	2.97×10^{-4}	6.73×10^{-2}	9.52×10^{-4}
R^2	0.88031	0.0995	0.7426

4.2. Regression Performance

In the scope of this research, two distinct regression methodologies were employed: the conventional regression and the stepwise regression. The conventional regression method is a straightforward approach where the relationship between the input variables is defined, and the best-suited coefficient for each input is determined. This method primarily focuses on the direct relationship between the inputs and the output.

On the other hand, the stepwise regression method offers a more nuanced approach. Not only does it define the relationship between inputs, but it also takes into account the interactions among these inputs. Leveraging statistical techniques, this method refines the prediction model by eliminating terms that exert minimal influence on the prediction of the shear wave velocity. This ensures that the model is not only more streamlined but potentially more accurate as it focuses on the most impactful variables. The results derived from both these regression methods are visually represented in Figure 2.6 and Figure 2.7. For a more granular understanding, Table 4.2 enumerates the coefficients linked with each variable for both the conventional and stepwise regression models. It's noteworthy that during the stepwise regression analysis, variables that were deemed to have negligible effects on the outcome were automatically excluded from the model. This methodological refinement, which emphasizes the synergy of variables, ensures a more precise prediction model.

Table 4.2 the coefficients of each variable derived from both ordinary and stepwise regression methods.

Input		**Ordinary Regression**		**Stepwise Regression**	
Elements	**Variable**	**Variables**	**Coefficient**	**Variables**	**Coefficient**
Depth	X_1	(Intercept)	-0.00072	(Intercept)	0.027808
Vp	X_2	X_1	1.451×10^{6}	X_1	-3.80×10^{-6}
GR	X_3	X_2	0.251166	X_2	0.464578
SGR	X_4	X_3	-2.94×10^{-5}	X_3	0.000429
Neutron	X_5	X_4	-5.40×10^{-6}	X_4	-0.00030
Caliper	X_6	X_5	-0.00561	X_5	0.018653
Eff. Poro	X_7	X_6	5.52×10^{-5}	X_6	-0.00209
Rhob	X_8	X_7	0.000156	X_8	-0.00579
		X_8	0.000724	X_1: X_2	-0.00020
				X_1: X_4	6.18×10^{-8}
				X_1: X_5	-9.04×10^{-6}
				X_1: X_6	5.74×10^{-7}
				X_2: X_3	4.84×10^{-3}
				X_2: X_4	-6.28×10^{-3}
				X_2: X_5	-1.51×10^{-1}
				X_2: X_6	-0.0046
				X_2: X_8	0.155382
				X_3: X_4	5.21×10^{-7}
				X_3: X_5	-0.00034
				X_3: X_8	-0.0002
				X_4: X_5	7.02×10^{-5}

				X_4: X_8	0.000126
				X_5: X_6	8.27×10^{-4}
				X_5: X_8	-0.00279
				X_6: X_8	0.000280

To gauge the efficacy of both methods, the inaccuracies associated with each are documented in Table 4.3. A perusal of the data in Table 7 reveals a clear distinction in performance. The stepwise regression method outshines the conventional approach, showcasing enhanced capability and accuracy in determining Vs. This underscores the importance of choosing an appropriate regression technique, especially when dealing with complex datasets where interactions between variables can significantly influence outcomes.

Table 4.3 Regression Error Analysis.

Variables	Stepwise Regression	Linear Regression
AARE %	3.661	4.356
ARE %	0.2161	0.3177
MSE	1.35×10^{-7}	1.80×10^{-7}
RMSE	3.62×10^{-4}	4.25×10^{-4}
R^2	0.9784315	0.9571591

4.3. Performance of Multilayer Perceptron Artificial Neural Network (MLP-ANN)

This approach involves an endeavor to determine the most favorable combination of hidden layers and neurons through a process of experimentation and refinement. Table 4.4 details the specifics of the deployed network.

Table 4.4 Performance of Multilayer Perceptron Artificial Neural Network (MLP-ANN).

Characteristics	Number
Data	arbitrary
Training Function	Levenberg-Marquardt
Efficiency	Mean Squared Error
Hidden Layers	3
Neurons	25
Max Number of Epochs	700
Mu	1×10^{-11}
Gradient	9.59×10^{-11}
Performance	9.02×10^{-9}

Figure 4.4 shows the MLP-ANN method's prediction performance for Vs. For the training, validation, and

testing phases, the coefficient of determination (R^2) values are 0.9983, 0.9911, and 0.9998, respectively. This particular model exhibits a notable capability in accurately estimating the shear wave velocity throughout the entire process including training, validation, and testing. Furthermore, Table 4.5 presents the error values associated with each stage.

Table4.5 Analysis of Errors in Multilayer Perceptron Artificial Neural Networks (MLP-ANN).

Variables	Train	Validation	Test
AARE %	1.16	1.78	1.79
ARE %	0.069	0.076	− 0.1376
Mean Squared Error	1.16×10^{-8}	4.20×10^{-8}	4.26×10^{-8}
RMSE	5.80×10^{-9}	2.10×10^{-8}	2.13×10^{-8}
R^2	0.9972	0.9901	0.9898

4.4. ANFIS Performance

The parameters integral to this method were fine-tuned through a trial-and-error process. This iterative approach ensures that the model is optimized for the best possible performance on the given dataset. The specifics of the Adaptive Neuro-Fuzzy Inference System (ANFIS) model developed through this process are detailed in Table 4.6.

Table 4.6 ANFIS characteristics.

The parameters of the ANFIS	Dtails/Value
FS	Sugeno-type
Starting FIS for learning	Sub-clustering
Membership Function	Gaussian
Output Membership Function	Linear
Cluster's Area of Effect	0.202
Fuzzy Rules	3
Starting Size	0.010
Rate of decreasing size	0.90
Rate of increasing size	1.11

A crucial metric to evaluate the performance of such models is the R2 value, which indicates the proportion of the variance in the dependent variable that is predictable from the independent variables. For the training phase, the R2 value was an impressive 0.9729, suggesting that the model could explain approximately 97.29% of the variance in the training data. For the testing phase, the R2 value was 0.8015, indicating that about 80.15% of the variance in the test data was explained by the model. These results can be visualized in Figure 4.1.

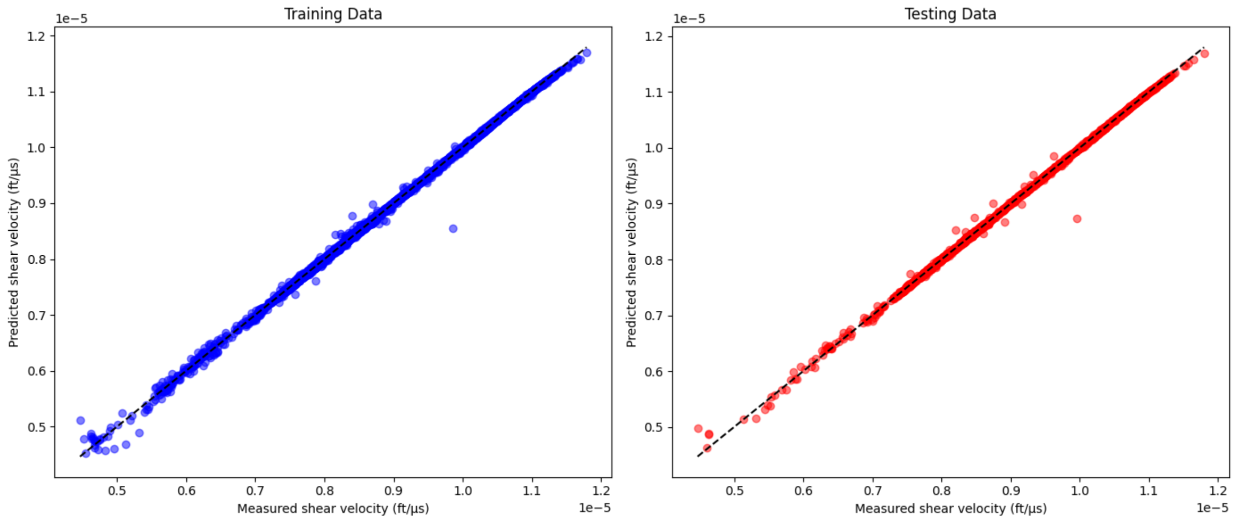


Figure 4.1 Measured Vs using ANFIS, (Blue) Training (Red) Testing.

To provide a comprehensive understanding of the model's accuracy, the error values associated with this method have been tabulated in Table 4.7. These values offer insights into the discrepancies between the model's predictions and the actual observed values, thereby providing a measure of the model's reliability and precision.

Table 4.7The ANFIS Error Analysis.

ANFIS	Training	Testing
AARE%	3.050	4.60
ARE%	0.1618	1.4781
Mean Squared Error	1.102×10^{-7}	1.16×10^{-7}
RMSE	5.613×10^{-8}	5.32×10^{-8}
R^2	0.9769	0.8115

4.5. MGGP Performance

In the application of the Multi-Gene Genetic Programming (MGGP) method, the results yielded were somewhat less robust compared to other methodologies previously discussed. However, it's essential to note that the MGGP method possesses certain inherent characteristics that render its performance appealing, despite the seemingly weaker outcomes. These unique features of MGGP, which will be elaborated upon shortly, can sometimes outweigh the benefits of marginally better predictive accuracy offered by other methods.

The parameters pivotal to the MGGP method were not predetermined. Instead, they were ascertained through a process of trial and error. This iterative approach ensures that the model is fine-tuned to the nuances of the dataset, optimizing its performance. While the initial results might suggest that MGGP's performance is not on par with other methods, it's crucial to delve deeper into the unique advantages MGGP offers. These could range from better interpretability, adaptability to different data structures, or even efficiency in terms of computational resources. Such benefits might make MGGP a preferred choice in specific scenarios or for particular applications, even if its raw predictive power is slightly inferior to other methods.

Table 4.8 provides a detailed overview of the characteristics and parameters associated with the developed model using the Multi-Gene Genetic Programming (MGGP) method. Two critical parameters for genetic programming, population and generation, were chosen. The population size, which represents the number of individual solutions in each generation, was set at 500. The number of generations, indicating the number of times the population would evolve, was determined to be 100. These values are indicative of the model's search breadth and depth in the solution space, respectively.

Table 4.8 MGGP characteristics.

Description	Numeric Values
Size of the population	510
Number of Gene	9
Size of Tournament	210
Generation size	100
Depth (maximum)	10
Operators	Times, R divide, Square, Exp, Sqrt, Log, Abs

The outcomes derived from the MGGP method are visually represented in Figure 4.2. A key metric to gauge the performance of the model across different phases is the R2 value. This coefficient of determination provides insights into how well the model's predictions match the actual data. For the training phase, the R2 value stood at 0.9691, suggesting that the model could explain approximately 96.91% of the variance in the training data. In the validation phase, the R2 value was 0.9135, indicating a 91.35% explanation of the variance. Lastly, for the testing phase, the R2 value was 0.8166, denoting that the model could account for about 81.66% of the variance in the test data.

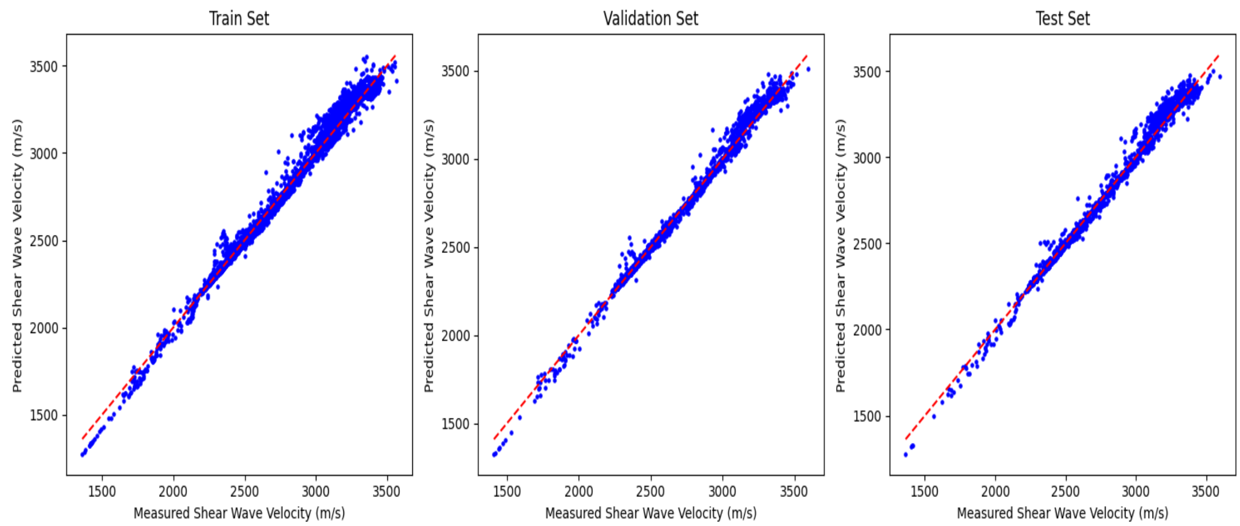


Figure 4.2: Measured Vs using MGGP a) Training Data b) Validation Data c) Test Data.

To offer a more comprehensive understanding of the model's accuracy and reliability, the error values associated with the MGGP method across all phases have been tabulated in Table 4.9. These error metrics provide a quantitative measure of the discrepancies between the model's predictions and the actual observed values, giving a clearer picture of the model's overall performance.

Table 4.9 Examining MGGP's Error

MGGP	Training Data	Validation Data	Testing Data
AARE%	3.49	4.30	4.25
ARE%	0.2015	-0.3836	1.0536
Mean Squared Error	1.36×10^{-7}	1.19×10^{-7}	1×10^{-7}
RMSE	6.80×10^{-8}	5.96×10^{-8}	5.02×10^{-8}
R^2	0.9691	0.9135	0.8166

4.5.1. Shear Wave Velocity Correlation

One of the standout features of the Multi-Gene Genetic Programming (MGGP) method is its ability to generate explicit relationships or correlations for the target variable, in this case, Vs (shear wave velocity), based on various input parameters. This capability is particularly valuable as it offers a clear, interpretable equation that can be used for predictions, as opposed to some other methods that might act as "black boxes."

From the MGGP method, a specific correlation was derived. Notably, this correlation achieved an R2 value of 0.9691 for its predictions. This coefficient of determination is indicative of the model's accuracy, suggesting that the correlation can explain approximately 96.91% of the variance in the data. Impressively, this R2 value surpasses those of many empirical correlations, highlighting the efficacy of the MGGP-derived relationship.

For a more detailed understanding and to facilitate the application of this correlation, the constants and coefficients integral to the equation have been systematically documented in Table 4.10. These constants, when plugged into the correlation, allow for accurate predictions of Vs based on the given input parameters. This combination of interpretability and high predictive accuracy makes the MGGP method a compelling choice for such applications.

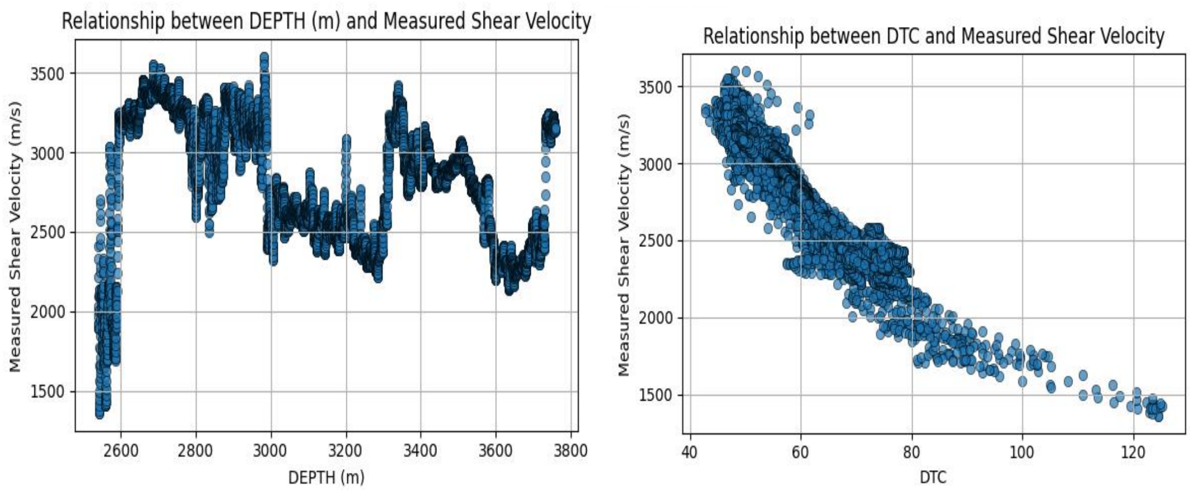
Table 4.10 Constants proposed for correlating Vs

Constant	Values	Parameter	Elements
A	1.617×10^{-5}	X_1	Depth
B	0.899	X_2	Vp
C	8.335	X_3	Gamma ray
D	0.002021	X_4	SGR
E	0.3327	X_5	Neutron
F	26.03	X_6	Caliper
G	9.4`5	X_7	Eff. Poro
H	2.120×10^{-5}	X_8	Rhob
I	14.22	X_1	Depth

4.5.2. Evaluation of sensitivity

To discern the importance of various input variables on the prediction of shear wave velocity (Vs), a parametric sensitivity analysis was conducted using the Multi-Gene Genetic Programming (MGGP) method. By plotting the estimated Vs against each input parameter, each variable's effect on the final result can be calculated. The R2 value, or the coefficient of determination, serves as a reliable metric in this context. A higher R2 value indicates a stronger influence of the input parameter on the output, making it a more significant predictor.

The detailed significance of each input variable, in relation to the prediction of Vs, has been tabulated in Table 4.11. For a more visual representation, Figure 4.3 graphically depicts the parameters and their respective impacts on the estimation of Vs.



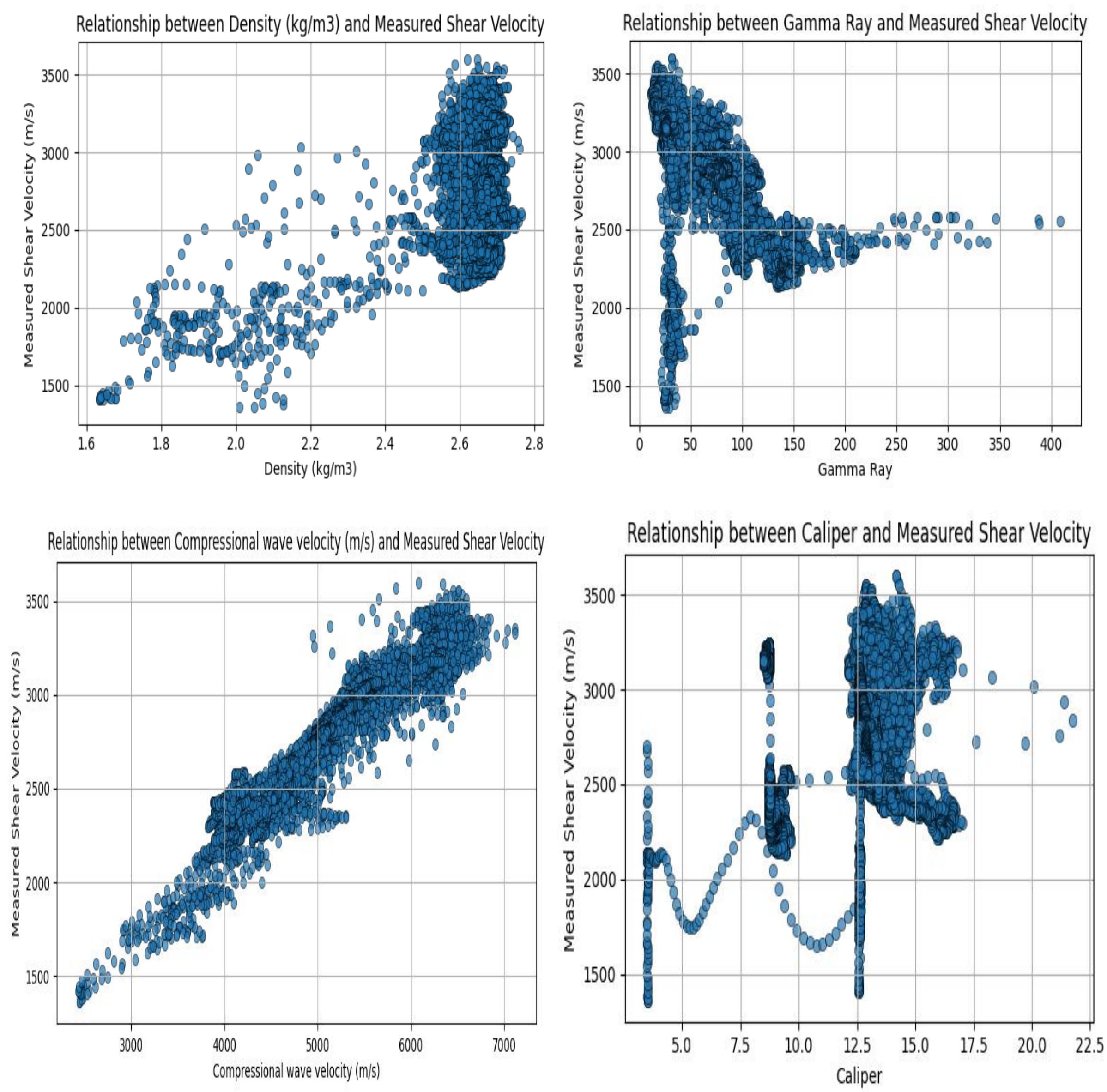


Figure 4.3 A plot of the observed Vs against the driving variables

Table 4.11the influence of input Parameters on Vs

Parameters	R^2
Depth	0.820
Vc	0.9501
GR	0.421
SGR	0.3753
Neutron	0.8621
Caliper	0.0987
Eff. Poro	0.5824
Rhob	0.6731

From the evaluation, it becomes evident that the VP stands out as most influential parameter on Vs. Following Vp, the parameters in decreasing order of significance are the neutron log, depth, density log, and effective porosity. The remaining parameters, while still having an effect, exhibit lesser importance in estimating the Vs.
Interestingly, results derived from the MGGP method align closely with the findings from the sensitivity analysis conducted using the Multi-Layer Perceptron Artificial Neural Network (MLP-ANN) method. This consistency across two different methodologies reinforces the reliability of the identified significant parameters and their respective influences on the estimation of Vs.

6. Evaluation of Error

In the methodology section of this research, it was highlighted that five distinct error models were employed to evaluate the performance of the various methods discussed in the paper. These error models serve as metrics to gauge the accuracy and reliability of each method in predicting the shear wave velocity. From the results derived for the test data, the Multi-Layer Perceptron Artificial Neural Network (MLP-ANN) model emerges as the most reliable method. This is evidenced by its impressive R2 value of 0.9898. Such a high coefficient of determination indicates that the MLP-ANN model can explain approximately 98.98% of the variance in the test data, which is a testament to its predictive prowess. Furthermore, other statistical error metrics also corroborate the high accuracy of the MLP-ANN in predicting the shear wave velocity, making it a standout performer among the methods evaluated.
For a more detailed breakdown of the errors associated with each method during the test phase, one can refer to Table 4.12. This table provides a comprehensive view of the discrepancies between the predicted and actual values for each method, allowing for a comparative analysis of their performances.

Table 4.12 Constants of postulated correlation for Vs.

For Testing	MLP-ANN	ANFIS	MGGP
AARE%	1.79	4.60	4.34
ARE%	-0.1396	1.4621	1.0446
Mean Squared Error	4.16×10^{-8}	1.06×10^{-7}	1×10^{-7}
RMSE	2.23×10^{-8}	5.31×10^{-8}	5.02×10^{-8}
R^2	0.998	0.8015	0.8166

Figure 4.4 offers a visual representation of the outcomes of each technique, plotted against depth of well. This graphical depiction provides insights into how each method's predictions vary with depth. Upon examining this figure, it becomes evident that the MLP-ANN technique consistently delivers most accurate predictions with comparison to the other models. Its predictions closely align with the actual values across various depths, further solidifying its position as the most reliable method for predicting vs. in this study.

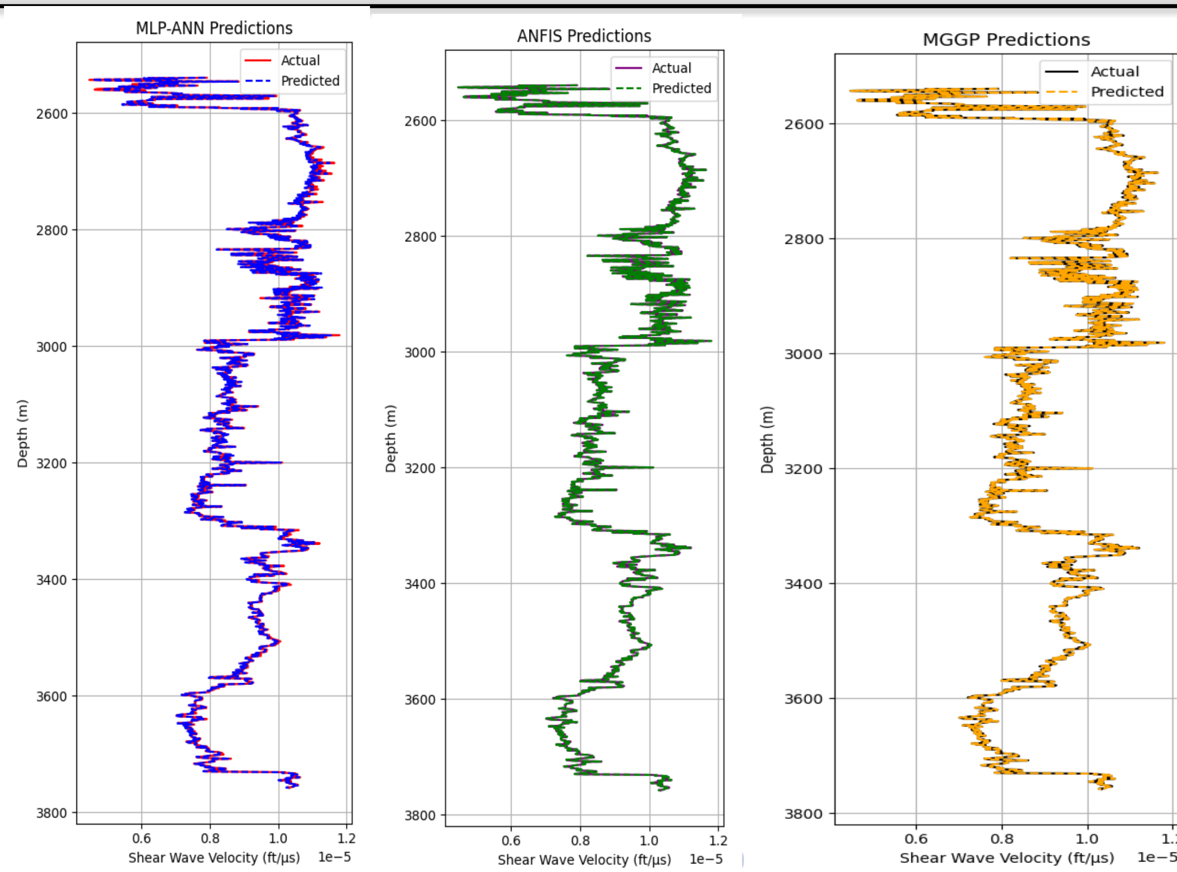


Figure 4.4: Actual and predicted Vs versus water depth utilizing (a) MLP-ANN (b) ANFIS

5. Conclusion

The estimation of shear wave velocity using ML methods and petro physical logs is the focus of this study. Three powerful machine learning techniques were chosen after an exhaustive literature analysis was conducted. The shear wave velocity in a carbonate oil reserve in Iran was then predicted using these techniques. Several significant findings emerged from the research:

Analyzing Correlational Effectiveness: The Pickett correlation performed better than the Carrol and Castagna correlations in this inquiry than any of the other correlations considered. R2 values for Pickett were 0.8812, for Castagna they were 0.7436, and for Carroll they were 0.0996.

Superiority of Stepwise Regression: The stepwise linear regression method demonstrated enhanced performance when compared to the conventional linear approach.

Optimization of Machine Learning Parameters: The parameters for the machine learning methods were fine-tuned through an iterative trial and error process. This involved multiple runs for each method, adjusting parameter values to achieve optimal results.

Dominance of MLP-ANN: When compared to the other approaches, MLP-ANN proved to be the most accurate at estimating future shear wave velocities. R2 values of 0.9983, 0.9906, and 0.9998 were obtained throughout the training, validation, and testing phases, respectively, suggesting a high degree of congruence between the model's predictions and the actual field data.

Performance of ANFIS: The ANFIS model, in comparison to MLP-ANN and MGGP, lagged in its performance during the test phase, with R2 values of 0.9729 for training and 0.8015 for testing.

Results from sensitivity studies utilizing both MLP-ANN and MGGP agreed on the order of importance of input factors on the outcome. After considering the neutron log, depth, density log, and effective porosity, it was found that the compressional wave velocity had the greatest impact on the shear wave velocity. MGGP's Correlation Ability: The study leveraged MGGP's capability to establish a correlation between input and output parameters. A specific correlation was derived, boasting an R2 value of 0.9691, which surpassed the performance of empirical correlations.

6. REFERENCE

Adjei, F., Afare, P., Osae, T., 2020. Predicting of shear wave velocity using Artificial Neural Networks arXiv preprint arXiv: 2011.12109.

Ahmad, M., Amin, K., Ali, A., & Ahmad, R. W. (2025). A Comparative Evaluation of Poisson, Negative Binomial, and Zero-Inflated Models for Count Data. *world*, 3(8).

Akhundi, H., Ghafoori, M., Lashkaripour, G.-R., 2014. Prediction of shear wave velocity using artificial neural network technique, multiple regression and petrophysical data: a case study in Asmari reservoir (SW Iran). Open J. Geol. 4, 303–313. http://www.scirp.org/journal/PaperInformation.aspx?PaperID=47965.

Anemangely, M., Ramezanzadeh, A., Amiri, H., Hoseinpour, S.-A., 2019. Machine learning technique for the prediction of shear wave velocity using petrophysical logs. J. Petrol. Sci. Eng. 174, 306–327.

Asoodeh, M., Bagheripour, P., 2012. Prediction of compressional, shear, and stoneley wave velocities from conventional well log data using a committee machine with intelligent systems. Rock Mech. Rock Eng. 45 (1), 45–63.

Asoodeh, M., Bagheripour, P., 2013. Neuro-fuzzy reaping of shear wave velocity correlations derived by hybrid genetic algorithm-pattern search technique. Cent. Eur. J. Geosci. 5 (2), 272–284.

Beale, M.H., Hagan, M.T., Demuth, H.B., 1992. Neural Network Toolbox User's Guide, vol. 103. The MathWorks Inc.

Behnia, D., Ahangari, K., Moeinossadat, S.R., 2017. Modeling of shear wave velocity in limestone by soft computing methods. International Journal of Mining Science and Technology 27 (3), 423–430.

Brocher, T.M., 2005. Empirical relations between elastic wavespeeds and density in the Earth's crust. Bull. Seismol. Soc. Am. 95 (6), 2081–2092.

Carroll, R.D., 1969. The determination of the acoustic parameters of volcanic rocks from compressional velocity measurements. In: International Journal of Rock Mechanics and Mining Sciences & Geomechanics Abstracts. Elsevier.

Castagna, J.P., Backus, M.M., 1993. Offset-dependent Reflectivity—Theory and Practice of AVO Analysis. Society of Exploration Geophysicists.

Du, Q., Yasin, Q., Ismail, A., Sohail, G.M., 2019. Combining classification and regression for improving shear wave velocity estimation from well logs data. J. Petrol. Sci. Eng. 182, 106260.

Eskandari, H., Rezaee, M., Mohammadnia, M., 2004. Application of multiple regression and artificial neural network techniques to predict shear wave velocity from wireline log data for a carbonate reservoir South-West Iran. CSEG recorder 42, 48.

Eskandari, H., Rezaie, M., Javaherian, A., Mohammad, N.M., 2003. Shear Wave Velocity Estimation Utilizing Wireline Logs for a Carbonate Reservoir (south-west IranEzebialu, K., Ubituogwale, M., Odegua, E., Idehen, A., 2020. Field development planning based on static, dynamic and geomechanical modelling of X field, Niger delta, Nigeria. In: SPE Nigeria Annual International Conference and Exhibition. Society of Petroleum Engineers. Farmani, Z., Azin, R., Mohamadi-Baghmolaei, M., Fatehi, R., Escrochi, M., 2019. Experimental and theoretical study of gas/oil relative permeability. Comput. Geosci. 23 (3), 567–581.

Feng, R., Grana, D., Balling, N., 2021a. Imputation of missing well log data by random forest and its uncertainty analysis. Comput. Geosci. 152, 104763.

Feng, R., Grana, D., Balling, N., 2021b. Variational inference in Bayesian neural network for well-log prediction. Geophysics 86 (3), M91–M99.

Ferreira, C., 2001. Gene Expression Programming: a New Adaptive Algorithm for Solving Problems arXiv preprint cs/0102027.

Figueiredo, B., Tsang, C.-F., Rutqvist, J., Bensabat, J., Niemi, A., 2015. Coupled hydromechanical processes and fault reactivation induced by CO2 injection in a threelayer storage formation. International Journal of Greenhouse Gas Control 39, 432–448.

Finisha, B., Haris, A., Ronoatmojo, I., 2018. Geomechanical modeling of reservoir rock using 2D seismic inversion: its application to wellbore stability in the onshore of Northwest Java Basin, Indonesia. In: AIP Conference Proceedings. AIP Publishing LLC.

Gandomi, A.H., Atefi, E., 2020. Software review: the GPTIPS platform. Genet. Program. Evolvable Mach. 21 (1), 273–280.

Samina. Amin/artificial-neural-network-ann-f541b0807e39, https://medium.com/

Gholizadeh, F., Sabzi, F., 2017. Prediction of CO2 sorption in poly (ionic liquid) s using ANN-GC and ANFIS-GC models. International Journal of Greenhouse Gas Control 63, 95–106.

Goldberg, D.E., Holland, J.H., 1988. Genetic algorithms and machine learning.

Goodman, R.E., 1989. Introduction to Rock Mechanics. Wiley, New York.

Grandi, S., Rao, R.V., Toksoz, M.N., 2002. Geomechanical Modeling of In-Situ Stresses around a Borehole, Massachusetts Institute of Technology. Earth Resources Laboratory.

Guan, J., Zurada, J., Levitan, A., 2008. An adaptive neuro-fuzzy inference system-based approach to real estate property assessment. J. R. Estate Res. 30 (4), 395–422.

Ayman Mutahar Al Rassas 2021 Optimized ANFIS Model Using Aquila Optimizer for Oil Production Forecasting School of Petroleum Engineering, , 10.3390/pr9071194

Hadi, F., Nygaard, R., 2018. Shear Wave Prediction in Carbonate Reservoirs: Can Artificial Neural Network Outperform Regression Analysis? 52nd US Rock Mechanics/Geomechanics Symposium. American Rock Mechanics Association.

Haykin, S., 1999. Neural Networks a Comprehensive Introduction. Prentice Hall, New Jersey.

Heidari, M., Khanlari, G., Momeni, A., 2010. Prediction of elastic modulus of intact rocks using artificial neural networks and non-linear regression methods. Australian Journal of Basic and Applied Sciences 4 (12), 5869–5879.

Hertz, J., Krogh, A., Palmer, R.G., 1991. Introduction to the Theory of Neural Computation. Addison-Wesley, Redwood City, CA.

Hosseinzadeh, M., Hemmati-Sarapardeh, A., 2014. Toward a predictive model for estimating viscosity of ternary mixtures containing ionic liquids. J. Mol. Liq. 200, 340–348.

Izadpanahi, A., Azin, R., Osfouri, S., Malakooti, R., 2020. Optimization of Two Simultaneous Water and Gas Injection Scenarios in a High GOR Iranian Oil Field. 82nd EAGE Annual Conference & Exhibition. European Association of Geoscientists & EngineersJang, J.-S.R., Sun, C.-T., Mizutani, E., 1997. Neuro-fuzzy and soft computing-a computational approach to learning and machine intelligence [Book Review]. IEEE Trans. Automat. Control 42 (10), 1482–1484.

Kayadelen, C., 2011. Soil liquefaction modeling by genetic expression programming and neuro-fuzzy. Expert Syst. Appl. 38 (4), 4080–4087.

Khatibi, S., Aghajanpour, A., 2020. Machine learning: a useful tool in geomechanical studies, a case study from an offshore gas field. Energies 13 (14), 3528.

Kong, S.-G., Kosko, B., 1992. Adaptive fuzzy systems for backing up a truck-and-trailer. IEEE Trans. Neural Network. 3 (2), 211–223.

Koza, J.R., Koza, J.R., 1992. Genetic Programming: on the Programming of Computers by Means of Natural Selection. MIT press.

Maghsoudian, A., Izadpanahi, A., Esfandiarian, A., 2021. Deterministic smart tools to predict recovery factor performance of saline water injection in carbonated reservoirs. In: 82nd EAGE Annual Conference & Exhibition. European Association of Geoscientists & Engineers.

Majdi, A., Beiki, M., 2010. Evolving neural network using a genetic algorithm for predicting the deformation modulus of rock masses. Int. J. Rock Mech. Min. Sci. 47 (2), 246–253.

KHAN, R., SHAH, A. M., & KHAN, H. U. (2025). Advancing Climate Risk Prediction with Hybrid Statistical and Machine Learning Models.

Mehrgini, B., Izadi, H., Memarian, H., 2019. Shear wave velocity prediction using Elman artificial neural network. Carbonates Evaporites 34 (4), 1281–1291.

Mohamadi-Baghmolaei, M., Azin, R., Osfouri, S., Zendehboudi, S., Hajizadeh, A., Izadpanahi, A., 2021. Mass transfer during transient condensate vaporization: experimental and modeling study. J. Mol. Liq. 325, 114022.

Khan, R., Shah, A. M., Ijaz, A., & Sumeer, A. (2025). Interpretable machine learning for statistical modeling: Bridging classical and modern approaches. *International Journal of Social Sciences Bulletin*, *3*(8), 43-50.

Naghizadeh, A., Larestani, A., Amar, M.N., Hemmati-Sarapardeh, A., 2022. Predicting viscosity of CO2–N2 gaseous mixtures using advanced intelligent schemes. J. Petrol. Sci. Eng. 208, 109359.

Khan, R. EFFECT OF OUTLIERS ON CLASSICAL VS. ROBUST REGRESSION TECHNIQUES.

Najibi, A.R., Ghafoori, M., Lashkaripour, G.R., Asef, M.R., 2017. Reservoir geomechanical modeling: in-situ stress, pore pressure, and mud design. J. Petrol. Sci. Eng. 151, 31–39.

Muhammad Ahmad, Khan, R. ., Ahmad, R. W. ., Wahab, F. ., & Nizamani, S. . (2025). Quantifying the Impact of Dot Balls on Winning Probability in T20 Cricket. *ACADEMIA International Journal for Social Sciences*, *4*(3), 4865-4885.

Nava, P., Taylor, J., 1996. The optimization of neural network performance through incorporation of fuzzy theory. In: Proceedings of the Eleventh International Conference on Systems Engineering.

Nourafkan, A., Kadkhodaie-Ilkhchi, A., 2015. Shear wave velocity estimation from conventional well log data by using a hybrid ant colony–fuzzy inference system: a case study from Cheshmeh–Khosh oilfield. J. Petrol. Sci. Eng. 127, 459–468.

Hanif, M. A. ., Abdul Wadood, Ahmad, R. W. ., Shah, S. A. ., & Khan, R. . (2025). Real-Time Anomaly Detection in IoT Sensor Data Using Statistical and Machine Learning Methods. *ACADEMIA International Journal for Social Sciences*, *4*(3), 5203-5227.

Olayiwola, T., Sanuade, O.A., 2020. A Data-Driven Approach to Predict Compressional and Shear Wave Velocities in Reservoir Rocks. Petroleum.

Ahmad, M., Rehman, A. A., Khan, R. ., & Bibi, H. . (2025). Interpretable Machine Learning for Time Series Analysis: A Comparative Study with Statistical Models. *ACADEMIA International Journal for Social Sciences*, *4*(3), 4001-4009.

Roidar khan, Shehzad khan, Naseemullah, Aasim Ullah, & Atif khan. (2025). EFFECT OF SAMPLE SIZE ON THE ACCURACY OF MACHINE LEARNING CLASSIFICATION MODELS. *Spectrum of Engineering Sciences*, *3*(7), 826–834.

Pickett, G.R., 1963. Acoustic character logs and their applications in formation evaluation. J. Petrol. Technol. 15 (6), 659–667.

Sumeer, A., Ullah, F., Khan, S., Khan, R., & Khan, W. (2025). Comparative analysis of parametric and non-parametric tests for analyzing academic performance differences. *Policy Research Journal*, *3*(8), 55-62.

Plumb, R., Edwards, S., Pidcock, G., Lee, D., Stacey, B., 2000. The Mechanical Earth Model Concept and its Application to High-Risk Well Construction Projects. IADC/ SPE Drilling Conference, OnePetro.

Ahmad, M., Qamar, H., Rehman, A. A., & Khan, R. (2025). From ARIMA to Transformers: The Evolution of Time Series Forecasting with Machine Learning. *Journal of Asian Development Studies*, *14*(3), 219-233.

Khan, R., Khan, A., Muhammad, I., & Khan, F. (2025). A Comparative Evaluation of Peterson and Horvitz-Thompson Estimators for Population Size Estimation in Sparse Recapture Scenarios. *Journal of Asian Development Studies*, *14*(2), 1518-1527.

Plumb, R.A., Hooyman, P., Veeningen, D., Dutta, N., Ritchie, G., Bennaceur, K., 2004. A new geomechanics process reduces operational risk from exploration to production. In: Gulf Rocks 2004, the 6th North America Rock Mechanics Symposium (NARMS). American Rock Mechanics Association.

Ahmad, M., Khan, S., Ahmad, R. W., & Rehman, A. A. (2025). COMPARATIVE ANALYSIS OF STATISTICAL AND MACHINE LEARNING MODELS FOR GOLD PRICE PREDICTION. *Journal of Media Horizons*, 6(4), 50-65.Rajabi, M., Bohloli, B., Ahangar, E.G., 2010. Intelligent approaches for prediction of compressional, shear and Stoneley wave velocities from conventional well log data: a case study from the Sarvak carbonate reservoir in the Abadan Plain (Southwestern Iran). Comput. Geosci. 36 (5), 647–664.

Ramjohn, R., Gan, T., Sarfare, M., 2018. 3D Geomechanical Modeling for Wellbore Stability Analysis: Starfish, ECMA, Trinidad and Tobago. SPE Trinidad and Tobago Section Energy Resources Conference. Society of Petroleum Engineers.

Ranjbar, N., Kuenzel, C., 2017. Cenospheres: a review. Fuel 207, 1–12.

Reyes, A., Huang, J.J., Yuan, Z., Temple, P., Liang, Q., Morana, J., Hashemian, Y., 2020. Cloud-Based Coherent Wellbore Stability, Improving Trajectory Design for Unconventional Well Construction Planning. IADC/SPE International Drilling Conference and Exhibition, Society of Petroleum Engineers.

Rezaee, M.R., Ilkhchi, A.K., Barabadi, A., 2007. Prediction of shear wave velocity from petrophysical data utilizing intelligent systems: an example from a sandstone reservoir of Carnarvon Basin, Australia. J. Petrol. Sci. Eng. 55 (3–4), 201–212.

Searson, D.P., 2015. Gptips 2: an open-source software platform for symbolic data mining. In: Handbook of Genetic Programming Applications. Springer, pp. 551–573.

Shukla, D., Solanki, C.H., 2020. Estimated empirical correlations between shear wave velocity and SPT-N value for indore City using NLR and ANN. Indian Geotech. J. 1–17.

Sivanandam, S., Deepa, S., 2008. Genetic Algorithms. Introduction to Genetic Algorithms. Springer, pp. 15–37.

Suppachoknirun, T., Tutuncu, A.N., 2017. Hydraulic fracturing and production optimization in eagle ford shale using coupled geomechanics and fluid flow model. Rock Mech. Rock Eng. 50 (12), 3361–3378.

Wang, P., Peng, S., 2019. On a new method of estimating shear wave velocity from conventional well logs. J. Petrol. Sci. Eng. 180, 105–123.

Wood, D.A., 2020. Bakken stratigraphic and type well log learning network exploited to predict and data mine shear wave acoustic velocity. J. Appl. Geophys. 173, 103936.

Zhang, F., Damjanac, B., Maxwell, S., 2019. Investigating hydraulic fracturing complexity in naturally fractured rock masses using fully coupled multiscale numerical modeling. Rock Mech. Rock Eng. 52 (12), 5137–5160.

Zhang, Y., Zhong, H.-R., Wu, Z.-Y., Zhou, H., Ma, Q.-Y., 2020. Improvement of petrophysical workflow for shear wave velocity prediction based on machine learning methods for complex carbonate reservoirs. J. Petrol. Sci. Eng. 192, 107234.

Zoback, M.D., 2010. Reservoir Geomechanics. Cambridge University Press.

Zoveidavianpoor, M., Samsuri, A., Shadizadeh, S.R., 2013. Adaptive neuro fuzzy inference system for compressional wave velocity prediction in a carbonate reservoir. J. Appl. Geophys. 89, 96–107