

Article

Inversion of Rayleigh Wave Dispersion Curve Extracting from Ambient Noise Based on DNN Architecture

Qingsheng Meng^{1,2,3,*} , Yuhong Chen¹, Fei Sha^{4,*}  and Tao Liu¹

¹ College of Environmental Science and Engineering, Ocean University of China, Qingdao 266100, China; ltmilan@ouc.edu.cn (T.L.)

² Key Laboratory of Marine Environment Science and Ecology, Ministry of Education, Qingdao 266100, China

³ Key Laboratory of Marine Environmental Geological Engineering, Qingdao 266100, China

⁴ College of Engineering, Ocean University of China, Qingdao 266100, China

* Correspondence: qingsheng@ouc.edu.cn (Q.M.); shafei@ouc.edu.cn (F.S.);

Tel.: +138-5327-4653 (Q.M.); +187-6401-5021 (F.S.)

Abstract: The inversion of the Rayleigh wave dispersion curve is a crucial step in obtaining the shear wave velocity (V_S) of near-surface structures. Due to the characteristics of being ill-posed and nonlinear, the existing inversion methods presented low efficiency and ambiguity. To address these challenges, we describe a six-layer deep neural network algorithm for the inversion of 1D V_S from dispersion curves of the fundamental mode Rayleigh surface waves. Our method encompasses several key advancements: (1) we use a finer layer to construct the 1-D V_S model of the subsurface, which can describe a more complex near-surface geology structure; (2) considering the ergodicity and orderliness of strata evolution, the constrained Markov Chain was employed to reconstruct the complex velocity model; (3) we build a practical and complete dispersion curve inversion process. Our model tested the performance using a random synthetic dataset and the influence of different factors, including the number of training samples, learning rate, and the selection of optimal artificial neural network architecture. Finally, the field test dispersion data were used to further verify the method's effectiveness. Our synthetic dataset proved the diversity and rationality of the random V_S model. The results of training and predicting showed higher accuracy and could speed the inversion process (only ~15 s), and we proved the important effect of different factors. The outcomes derived from the application of this technique to the measured dispersion data in the Yellow River Delta exhibit a strong correlation with the outcomes obtained from the integration of the very fast simulated annealing method and the downhill simplex method, as well as the statistically derived shear wave velocity data of the sedimentary layers in the Yellow River Delta. From a long-term perspective, our method can provide an alternative for deriving V_S models for complex near-surface structures.

Keywords: Rayleigh wave; dispersion curve; inversion; deep learning; deep neural networks



Citation: Meng, Q.; Chen, Y.; Sha, F.; Liu, T. Inversion of Rayleigh Wave Dispersion Curve Extracting from Ambient Noise Based on DNN Architecture. *Appl. Sci.* **2023**, *13*, 10194. <https://doi.org/10.3390/app131810194>

Academic Editor: Michel Darmon

Received: 18 July 2023

Revised: 3 September 2023

Accepted: 9 September 2023

Published: 11 September 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Rayleigh wave exploration based on ambient noise is an important method for investigating near-surface shear wave velocity (V_S) structures. It is widely used in fundamental geological surveys [1,2], geotechnical engineering [3], geological hazard early warning [4,5], hydrogeological surveys [6], and other fields. In particular, it has incomparable advantages for urban subsurface surveying and measuring the V_S of weakly consolidated soils in offshore areas because other geophysical methods are easily susceptible to noise [7] or do not allow for borehole testing to be carried out.

Achieving the subsurface S-wave velocity profile by inverting the Rayleigh wave dispersion curve is a key step in surface wave analysis [8]. However, similar to most geophysical inversion problems, dispersion curve inversion has the characteristics of high nonlinearity, multiparameter, and multi-extremum [9], and the accuracy of inversion results is affected by the two main aspects. One is the choice of the initial model; in

general, the inversion problem exists relative to the forward problem and may generate different inversion results for the same physical process. Another is the dimension of the inverted parameters. The phase velocity of Rayleigh waves depends on the compression wave velocity (V_p), S-wave velocity (V_s), density (ρ), and layer thickness (h) of strata. However, the S-wave velocity has a deterministic effect on the surface wave dispersion, so dispersion curve inversion usually refers to inverting the S-wave velocity and layer thickness (h). In recent research, the local optimization inversion method assumed that the subsurface is subdivided into a reasonable and fixed number of layers, and the number and thickness of each layer could be defined by the user [10]. In fact, this method transforms the problem into a single parameter dispersion curve inversion (only for V_s). In case of simple stratigraphic structures, such a method can provide a reliable estimation of the S-wave velocity. However, due to the influence of natural and human factors on the near surface, the seismic wave velocities usually change abruptly with depth or local structural reversals, such as the subsurface cavities, weak interlayers, etc. In such a case, it is difficult to give a reasonable estimation of V_s by simply increasing the number of layers. Moreover, accurate depths of layers are critical for many applications. A promising approach is to invert V_s and h simultaneously. However, an increase in unknown parameters may cause problems. Cox et al. [11] proposed a method to perform multiple inversions with a global search approach, in which the layer thickness (h) is parameterized with a “stratified ratio”, and its value can be determined via user experience. Subsequently, many global optimization algorithms, such as the Monte Carlo method, the genetic algorithms, and the simulated annealing, have appeared in many studies [12–14]. The global optimization algorithms have been widely adopted because they do not rely on the initial model and do not require calculating directional derivatives. However, due to the larger amount of computation, low efficiency, and easy premature convergence, the application of the global optimization method is limited.

Artificial intelligence (AI) and cloud computing technology have entered many areas, such as national defense and military [15], economic construction [16], and social development [17], impacting our world on many levels. AI is a general field that encompasses machine learning and deep learning (DL) but also includes many more approaches that do not involve any learning. With machine learning, we input data as well as the answers expected from the data, and the output comes from the rules. These rules could then be applied to new data to produce original answers [18]. The core of deep learning is called the neural network model, a process of distributed storage of information and parallel collaborative processing of a nonlinear system, and it can implement complex nonlinear mapping. From this perspective, it is analogous to the geophysical inversion problem. Therefore, in recent years, geophysicists have gradually introduced it into the fields of seismic data processing and interpretation [19], lithology identification of logging data [20], and magnetotelluric inversion [21]. Caylak et al. [22] proposed a multilayer perceptron neural network for the inversion of Rayleigh wave data, but they did not describe how to select the training dataset. Cao et al. [23] obtained the Bayesian posterior distribution of the S-wave velocity at each layer of the model with a mixture density network (MDN), but the number of layers and velocity range of each layer must be known. They also did not explain how to obtain the prior models to train the MDN. Hu et al. [24] utilized convolutional neural networks (CNN) to derive 1D S-wave velocity from surface wave dispersion and pointed out that CNN can solve more complex nonlinear inversion problems compared with densely connected neural networks. However, the training process of CNN is slower, and the training dataset is derived from the general geophysical inversion method. Moreover, the neural network structure must be recreated when processing new data, which makes it difficult to promote in the application.

The successful applications mentioned above demonstrate that DL can invert Rayleigh wave dispersion curves into shear wave velocity models. Current research aims to improve the accuracy of inversion by exploring different neural network models and generating more representative training data. However, most existing studies focus on deep geological

structures, with inversion of layer thicknesses ranging from tens of kilometers to even more. In contrast, this study focuses on the more complex near-surface geological structures. The main difference between our research and previous neural network inversion methods lies in the construction of a more refined, layered shear wave velocity model, incorporating a Markov chain (MC) with ergodic properties in the random process. As a result, the synthesized training data are more representative.

In this study, we first established the method to construct a V_S model based on the constrained MC, and then the forward calculation of Rayleigh wave dispersion was employed to obtain the corresponding phase velocities. All these data provide the training dataset required by deep learning. Next, a multilayer deep neural network (DNN) architecture was constructed to implement the training process. We discussed the DNN performance in relation to loss function and accuracy, as well as the influence of the number of neural network layers and parameter adjustment on the inversion accuracy using a synthetic dataset. Further, we tested the DNN with field data collected in the Yellow River Delta and compared the result with the very fast, simulated annealing method and the downhill simplex method (VFSA-DHSM) approach and statistical data in the Yellow River Delta. Our purpose was to realize an easy and applicable method for Rayleigh wave dispersion curve inversion.

2. Materials and Methods

2.1. Generate Training Dataset

The S-wave velocity and its corresponding dispersion curve dataset constitute the training sample data pair for DNN. The diversity and sufficiency of the sample dataset determine the performance of deep learning. Sample data can be obtained via logging experiments, laboratory tests after sampling, and field Rayleigh wave measurements. However, these methods are time-consuming and expensive. An alternative approach is forward modeling by providing geological models and then calculating the corresponding dispersion data. In this method, establishing a large number of representative geological models will be very important.

2.1.1. Generate Velocity Models

Academic Yang Wencai, a well-known geophysicist in China, pointed out that there is no simple repetition in the history of the earth's evolution in the past 4600 million years, and it shows the ergodicity of all states [25]. Just as "There are no two identical leaves in the world", a variety of phenomena in the natural system confirms the validity of the ergodic hypothesis. As with the surface of the earth, the velocity structure of the near surface should also have ergodicity. In addition, the natural system has the inertia of maintaining its characteristic shape in evolution, making the strata show orderliness with the change in time and space. That is to say, the velocities will rise with the increase in depth due to the self-weight (inertia) of the strata. The variation in velocities may be shown in three cases. One is when the velocity increases approximately linearly with the depth; the second is when the velocity drops sharply at a certain depth, such as when there exists a cavity or weak interlayer in the subsurface; the third is when the velocity rises sharply at a certain depth, such as the existence of boulders. The strata with an oscillating or linearly decreasing velocity are rarely seen. According to the general rule of sediment deposit, the first case is a high-probability event, while the other two are low-probability events.

Compared with the giant earth system, the velocity structure of the near surface has a higher complexity. In the evolution process, any type of velocity structure may appear in different areas or at different times. So, the near-surface velocities possess ergodicity and completely satisfy the random process theory. According to the probability and mathematical statistics theory, the Markov Chain (MC) is a method used to describe stochastic processes for discrete-time index datasets and state spaces with Markov properties [26]. MC has ergodicity, which makes a good combination between MC and strata evolution. If we regard the variation in velocities as a time series, the depth of each layer is regarded

as equivalent to time, so the S-wave velocity is the state space of the corresponding time. Therefore, the MC of the random variation process of the velocity structure can be expressed as $\{v(t), t \in T\}$, and its Markov property can be expressed as

$$P\{v(t_n) = j | v(t_{n-1}) = i_n, v(t_{n-1}) = i_{n-1}, \dots, v(t_1) = i_1\} = P\{v(t_n) = j | v(t_{n-1}) = i_{n-1}\} = P_{ij} \tag{1}$$

In Equation (1), t_n is the time series, and $t_1 < t_2 < \dots < t_n \in T$, $v_n \in V$ is the corresponding state space variable; $P_{ij} (> 0)$ is the one-step transition probability for all states i, j and n . Equation (1) indicates that the probability of state j at time t_n is determined by state i at time t_{n-1} . Therefore, the velocity of the current layer only depends on the velocity of the previous layer and has no relation to the velocity of other layers. The orderliness of the velocity structure shows that the velocity increasing linearly with depth is a high-probability event, while other cases are low-probability events. Thus, we can define the MC transition probability of the velocity variation.

In our study, the basic steps of constructing a near-surface velocity model according to the MC theory are listed as follows:

Step 1: Generate a model with a random thickness. The total thickness of the model is limited to 0.05 km because we focus on solving shallow engineering geological problems. Here, we set the model layer number to 20. Then, we use a random function to generate 20 random numbers between 0 and 1, and finally, they are normalized and multiplied by the total thickness to obtain the thickness of each layer.

Step 2: Generate random S-wave velocity for the first layer. Because the surface layer is usually quaternary sedimentary, the velocity is generally low; the S-wave velocity may not be appropriate. So, we constrain the velocity range according to the geological conditions. The range in this study is $V_{min} = 0.15 \text{ Km/s}$, $V_{max} = 0.3 \text{ Km/s}$, which is more consistent with the velocity characteristics of the near surface.

Step 3: Generate random S-wave velocity for the remaining layers. We define the probability $\alpha = 0.8$ for the velocity increasing linearly with depth and the probability $\alpha = 0.1$ for other two cases. Then, we randomly generate a random number x between 0 and 1. Next, if $x \leq 0.8$, the following formula is used to calculate $V_{s,m+1}$:

$$V_{s,m+1} = V_{s,m} + \lambda_{m+1,m} \cdot V_{s,m} \tag{2}$$

In Equation (2), $\lambda_{m+1,m}$ is the random floating number between $(\lambda_{min}, \lambda_{max})$, which determines the increasing or decreasing rate. In this study, we defined $\lambda_{min} = 0.01$ and $\lambda_{max} = 0.35$.

If $0.8 < x \leq 0.9$, $V_{s,m+1}$ is taken as the random number between $(V_{s,m} + \lambda_{max} \cdot V_{s,m}, \min(V_{s,m} + 0.3, 1.0))$.

If $0.9 < x \leq 1.0$, $V_{s,m+1}$ is taken as the random number between $(\max(V_{s,m} - 0.3, 0.1), \lambda_{min} \cdot V_{s,m})$.

Step 4: Calculate the compressional wave velocity (V_p) and density (ρ). Although their influence on the Rayleigh wave velocity is much smaller than that of V_s , accurate values are also important for calculating dispersion data. Much previous research provided the empirical relationships among V_p , V_s and ρ [27,28], which apply to different lithology and depth ranges. In our study, we use the approximate relationship given by Tang et al. to calculate the V_p .

$$V_p = V_s / \left[0.5684 \times (Z/0.2)^{0.163} \right] \tag{3}$$

Here, Z is the depth (Km) and $Z = \sum_{i=1}^m h_i$; $h_i (i = 1, 2, \dots, m)$ is the thickness of each layer.

Equation (3) is suitable for sedimentary strata when the depth is less than 0.2 Km.

For the density ρ , we use the empirical relationship given by Ludwig et al. [29].

$$\rho = 1.6612V_p - 0.4721V_p^2 + 0.0671V_p^3 - 0.0043V_p^4 + 0.000106V_p^5 \quad (4)$$

Here, the unit of ρ is g/cm^3 .

The above method can generate a complex geological model with any layer number, but in our study, we randomly generated a 20-layer model and then interpolated it into a 100-layer smooth velocity model with a spacing of 0.5 m. The method of using the smooth velocity model is conducive to describing complex geological structures and is closer to the velocity variation characteristics of the near surface.

2.1.2. Calculate Dispersion Data

Next, we generated the fundamental mode Rayleigh wave phase velocity with a period range of 0.08 to 0.48 s with the Computer Programs in Seismology (CPS) software package [30]. The corresponding frequency range was 2.0 Hz–2.5 Hz. The reason for choosing this range was to keep consistent with measured data obtained in our previous research using ambient noise technology.

In our study, we calculated 22,527 sample data pairs via MC and forward calculation. Ten groups of sample pairs were randomly selected to test the rationality (Figure 1). From Figure 1, we can see three typical velocity structures. Hence, this method could provide sufficient and diverse training data for the following research.

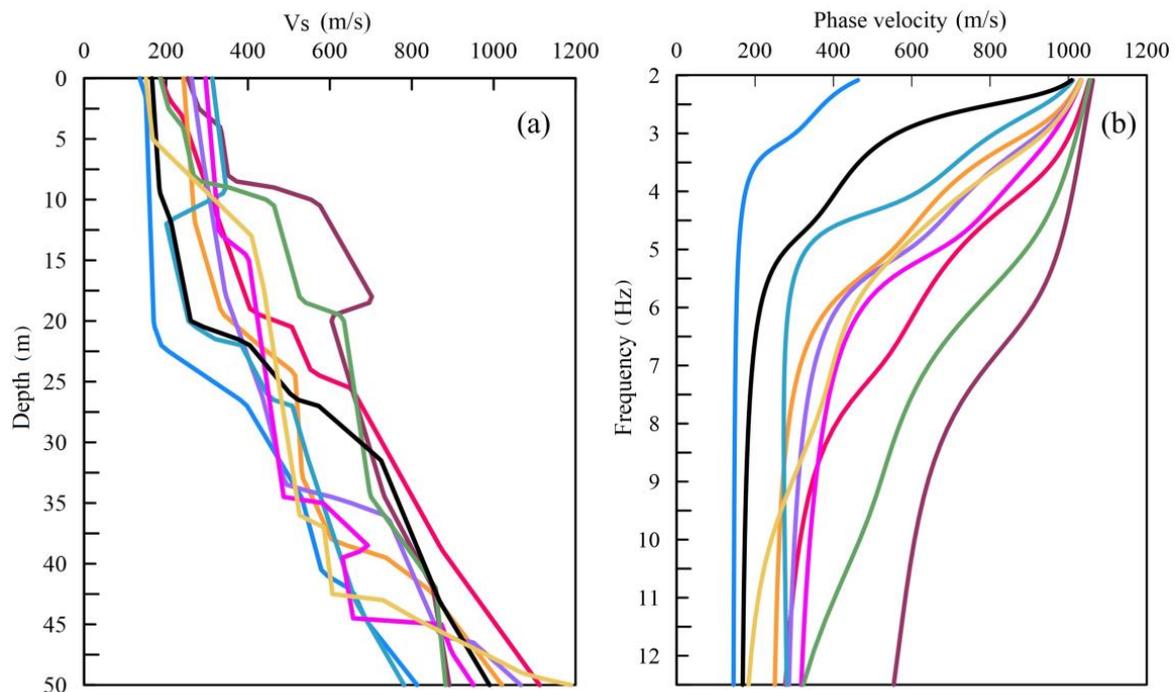


Figure 1. Theoretical layered shear wave velocity models and corresponding dispersion curves generated for training the algorithm. (a) Ten distinct velocity models were considered, where the inflection points of each curve correspond to the interfaces between layers. Within each layer, the velocities exhibit a linear variation with depth. (b) The dispersion curves corresponding to the ten velocity models.

2.2. DNN Architecture

We chose a DNN model with multiple hidden layers. It is also known as a multilayer perceptron and is an artificial neural network with a relatively simple connection. Its advantage is that it is easy to construct, suitable for complex discrete data series, and performs better than CNN, temporal recurrent neural networks (RNN), and other network

structures that are a subset of DNN. The DNNs usually have multiple hidden layers; adding hidden layers can better separate the characteristics of data and can thus better describe the nonlinear characteristics.

The DNN model introduced in this paper contains one input layer, four hidden layers, and one output layer (Figure 2). To invert the 1D S-wave velocity models from the dispersion curve, we took the frequency-dependent phase velocities as the input, composed of 101 uniformly sampled data with periods ranging from 0.08 to 0.48 s. Before training, we first scaled the original data via normalization so that all values were within the range [0, 1]. The outputs were the corresponding S-wave velocity model. According to our experience, the output dimensions of the first to fifth layers were set to 1600, 1200, 800, 200, and 101, respectively. The sixth layer was the output layer with dimension 101 due to the total thickness (50 m) being discretized into 101 layers with a spacing of 0.5 m. The Rectified Linear Unit (ReLU) activation function was chosen at each layer (except for the output layer) to avoid the problem of vanishing gradients. To reduce the complexity and instability of the model during the learning process, enhance the anti-disturbance ability, and thus avoid the danger of overfitting, the L2 regularization function was employed for the weights of each layer, and the L1 regularization function was used for the outputs of each layer.

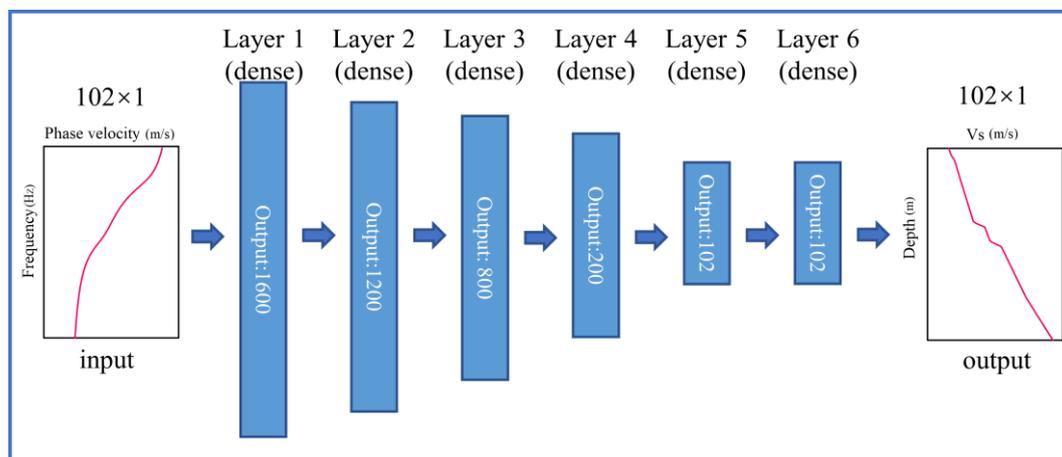


Figure 2. DNN architecture.

Using deep learning to invert the 1D V_S model is a regression problem. A large number of neural network loss functions can be used to account for the difference between predicted and true values. The most common nonlinear functions are the mean absolute error (MAE), mean squared error (MSE), and mean squared logarithmic error (MSLE). We chose the MSE loss function because it is more sensitive to outliers than MAE and is a commonly used loss function for regression problems. Corresponding to the loss function, the Adam optimizer is used to update the weights of the network. The evaluation metrics used in regression problems are different from those used in the classification problem, so the concept of accuracy does not apply to regression problems. To control the accuracy of the network, we defined the relative error as

$$e = \frac{1}{n} \sum_{i=1}^n \frac{\Delta v}{v} \quad (5)$$

Here, n is the number of layer depth nodes (101), Δv is the difference between true S-wave and predicted velocities, and v is the true S-wave velocities. From Equation (5), we define accuracy as

$$A = 1 - e \quad (6)$$

2.3. Training and Predicting

In our study, we wrote the Python (version 3.9) code to implement the calculation. Using the constructed sample data pairs, we randomly split 70% of the whole dataset as the training dataset and 30% as the validation dataset. The validation dataset is independent of the training dataset and is only used to guide parameter adjustment and avoid overfitting. At the same time, we separately calculated the 2000 synthetic dispersion curves as the test dataset. The corresponding S-wave velocity models are known and can be used to compare the results before and after inversion. In training, we used a personal computer with an Intel (R) i7-10700TCPU core, 64 GB memory, and no GPU acceleration. The learning rate was set to 0.01, the maximum number of epochs was set to 200 to ensure the convergence of the training process, and other parameters were set to the system's default value.

To avoid overfitting, we also calculated the loss function of each epoch using the accuracy expression defined in Equation (6) to calculate the accuracy of the validation dataset and use this data to evaluate the performance of the DNN.

Figure 3 shows the training loss and validation loss curves after 200 epochs, and Figure 4 shows the corresponding accuracy curve of the validation dataset. Figures 3 and 4 show that the loss function of each iteration is decreasing, and it tends to be stable after 110 epochs. The training loss and the validation loss keep changing in synchronization, and the accuracy of the validation dataset also shows the same rule after 110 epochs. The loss function and accuracy eventually stabilize without showing overfitting.

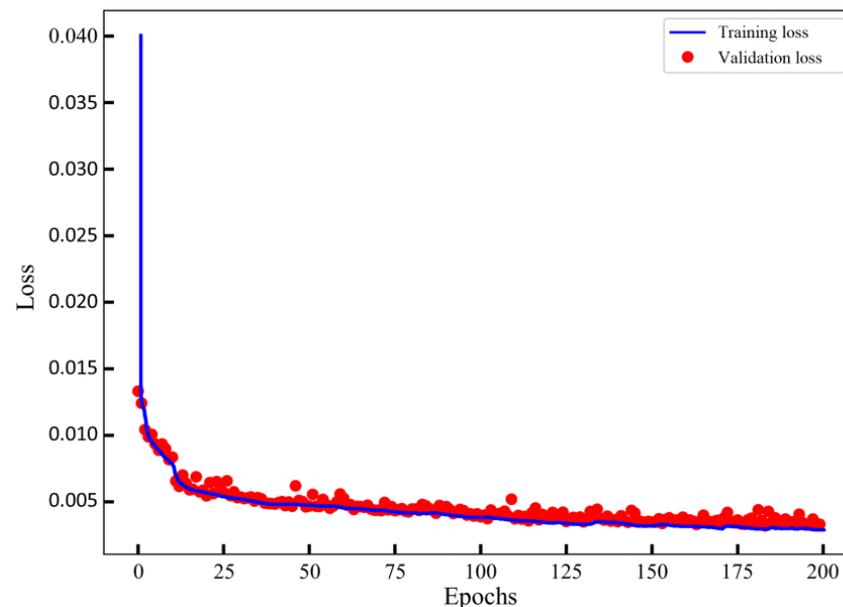


Figure 3. The loss function value of the training set and validation set.

Then, we inverted the 2000 test datasets. Figure 5 reveals the inversion results of the randomly selected dispersion curves. The figure shows that the inversion results of S-wave velocities agree with the theoretical velocity models. To better illustrate the effectiveness of the method, we also calculated the relative error distribution map of the test dataset. Figure 6a shows the relative error distribution of all data points for all samples, and Figure 6b shows the relative error distribution of all samples. Figure 6 indicates that 70% of data points have an error of less than 11.5% among all data points, and 70% of the samples have an error of less than 11.2% for all test samples. That is, the accuracy of DL inversion is close to 90%.

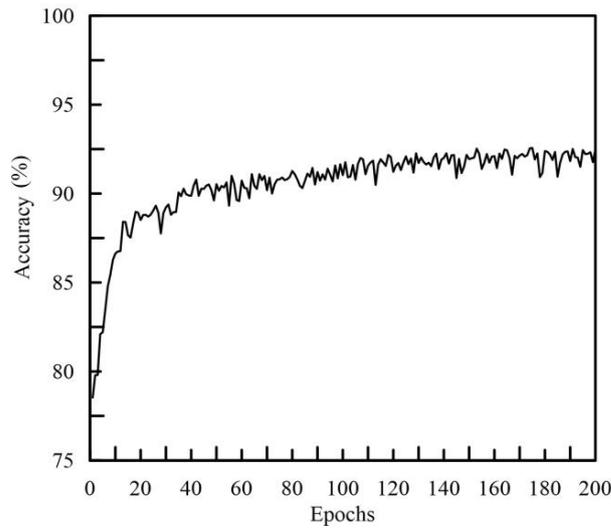


Figure 4. Accuracy of validation dataset.

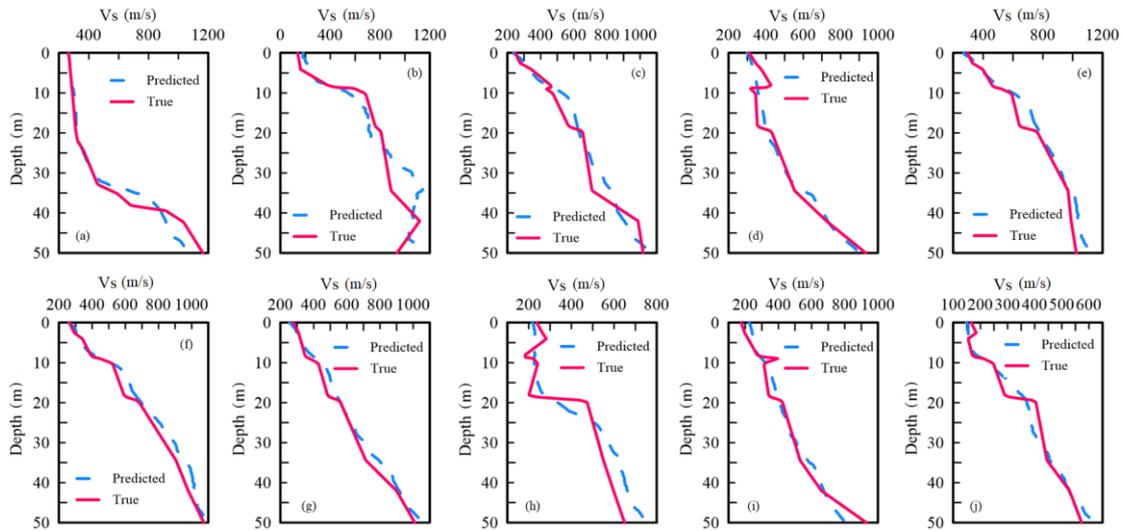


Figure 5. The inversion results of the randomly selected dispersion curves from the 2000 test samples. Among them, subfigures (a,e–g) exhibit shear wave velocities linearly increasing with depth in the subsurface layers, (d,h,j) denote a sharp decrease in shear wave velocity at a certain depth in the subsurface, and (b,c,i) stand for a sharp increase in shear wave velocity at a certain depth in the subsurface.

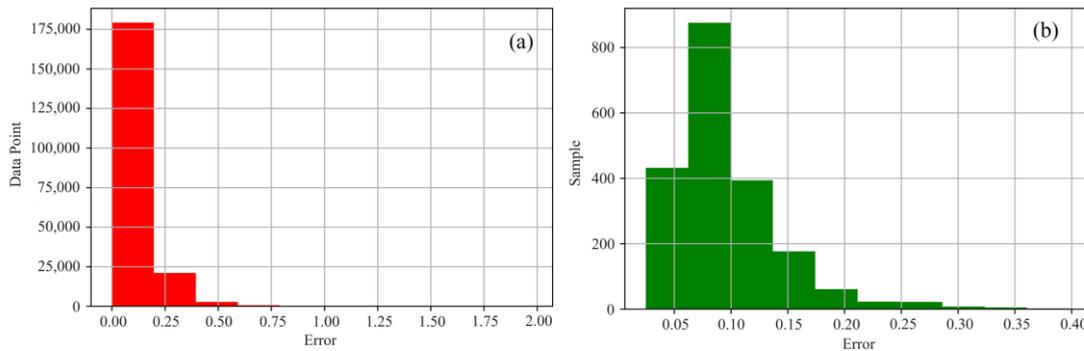


Figure 6. Distribution of the relative errors for the test dataset. (a) Relative errors for all data points, and (b) relative errors for all samples.

2.4. Parameters Test

Few studies have provided detailed information about the influence of network structure and parameter tuning on network performance. These issues, however, are crucial to the accuracy of dispersion curve inverting. So, we calculated the relative errors of the number of training samples, learning rates, and different network layers to discover their relationships.

2.4.1. The Number of Training Samples

As we know, insufficient training datasets will lead to poor approximation. Will the model's performance continue to improve as the training data volume increases? According to Hestness et al. [31], the performance will improve when the size of the dataset is increased. Conversely, Joulin et al. [32] found that the model performance decreased with a higher dataset volume. In this section, we calculated the relative error of nine different training sample numbers for the validation dataset. Two parallel experiments were conducted for each sample size in order to determine the average error, as shown in Figure 7. The figure shows that the error does not continuously decrease with an increase in the number of training samples. However, when the number of samples reaches 5600, the error reaches its lowest value (8.9%).

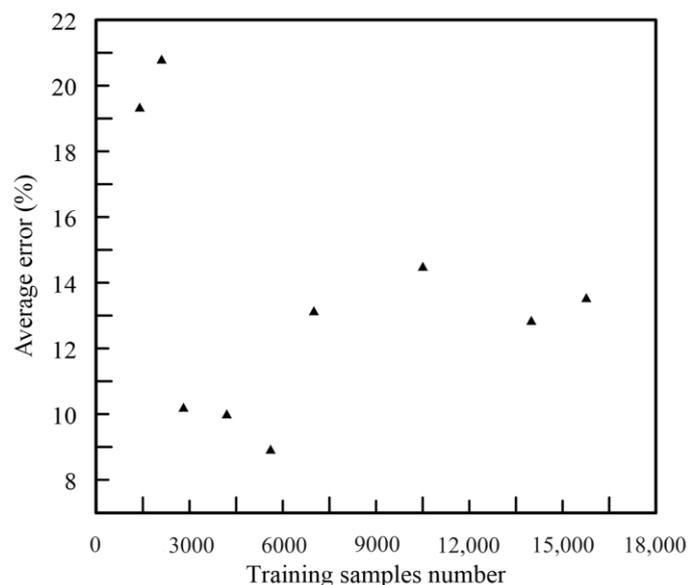


Figure 7. Average error for different numbers of training samples.

2.4.2. Learning Rate

In contrast to the parameters automatically updated through training (e.g., weights in neurons), hyperparameters are defined as manually set parameters with fixed values prior to the start of the learning process. The proper configuration of hyperparameters is crucial for achieving optimal performance in DNN. One of the most important hyperparameters for deep learning network models is the learning rate, typically between 0.0 and 1.0. The learning rate governs the speed of convergence of the model. The smaller the learning rate, the slower the gradient descent and the longer the convergence time. However, the higher the learning rate, the larger the update step size of the weight, and it will be easier to skip the optimal solution. We chose nine different learning rates, set up two parallel sets of experiments, and calculated the corresponding errors to test the effect of learning rates on DL performance. From Figure 8, we conclude that the learning rate impacts the performance of the DL network. There is no simple linear relationship between the learning rate and the relative error, but the model performs best when the learning rate is 0.01. The relative error at this time is 8.89%.

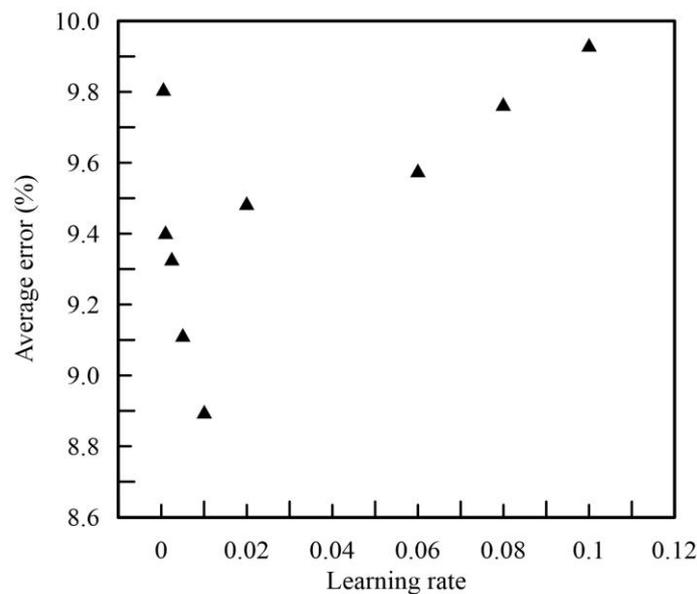


Figure 8. Average error of different learning rates.

2.4.3. The Number of Network Layers

The series of layers between input and output is referred to as hidden layers. Having more than one hidden layer allows a neural network to approximate the nonlinear relationship between input and output data. However, having too many hidden layers may lead to overfitting, where the model performs poorly when faced with new data. Currently, there are no general rules available for selecting the number of layers in research. In this study, we employed seven network models with varying numbers of layers, ranging from three to nine, and calculated their corresponding errors. As shown in Figure 9, the optimal number of layers for training the theoretical synthetic dataset was six, with an average relative error of 8.9%. This indicates that the problem of Rayleigh wave dispersion curve inversion exhibits highly nonlinear characteristics, and shallow neural network models are ineffective in learning the features effectively.

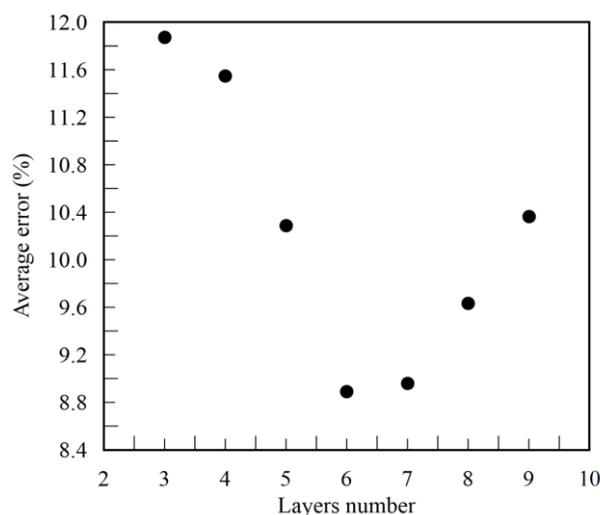


Figure 9. Average error of different DL network layers.

3. Data Application Example in Yellow River Delta

Considering the good performance of synthetic data, we applied our approach to real dispersion data to demonstrate how it works well in inversion. In October 2021, we employed the centerless circular array (CCA) for observation (Figure 10) and collected the

ambient noise data at an offshore beach in the Yellow River Delta [33]. The ambient noise survey was a completely new method; its measurement was relatively straightforward, and the process did not need an active source [34,35]. This method can overcome a number of obstacles that make traditional geophysical technologies difficult or impossible to implement in some special areas, such as weakly consolidated soils offshore where it is difficult to maintain the quality of the borehole and hence impossible to conduct in-hole S-wave velocity testing [5].

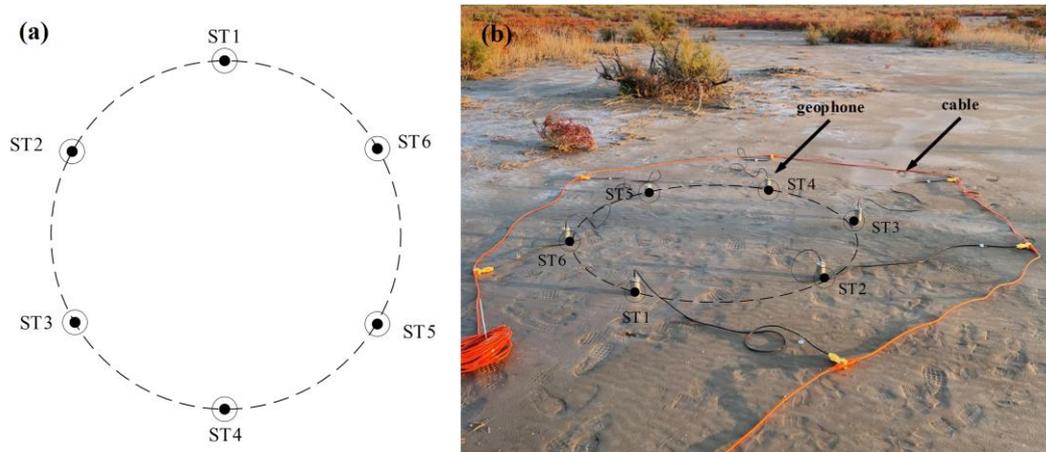


Figure 10. Map of the centerless circular array (CCA). (a) Schematic of CCA; (b) field photograph.

In testing, we set up two stations, S1 and S2, located 1.2 km apart. Each station utilized six identical vertical component velocity detectors (PS-2 Hz) referred to as ST1, ST2, ST3, ST4, ST5, and ST6. These six detectors were evenly distributed on a circular array centered at each station, with a radius of one meter. We employed this CCA array to collect random noise data and subsequently processed the obtained raw data. Through resampling and calculating the coefficients of principal component analysis, we obtained the dispersion curves for stations S1 and S2 stations [33] (black line in Figure 11).

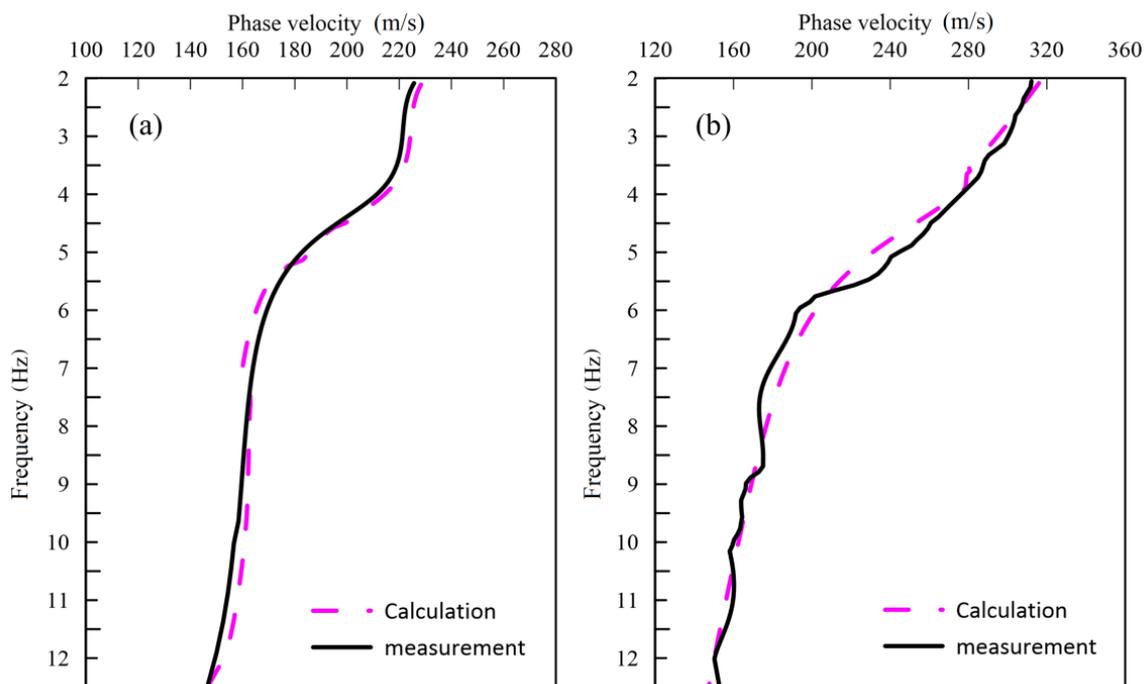


Figure 11. Comparison of observed and predicted dispersion curves. (a) S1 station; (b) S2 station.

Since the output depth range needs to vary with the periodic range of the input Rayleigh wave dispersion velocities, to invert these dispersed data using DNN, the dispersion data must be preprocessed first. The frequency range of the measured dispersion curve was 2.08–12.5 Hz (0.08–0.48 s), but the spacing between each dispersion point was not uniform. For this reason, we used an interpolation algorithm to resample these data into 101 points at equal intervals to ensure consistency with our network structure; these data would be used for S-wave velocity inversion.

In part 1.1, we generated lots of random velocity models. However, the S-wave velocity range of these velocity models was between 300 and 1200 m/s, while the velocity range of the testing area in the Yellow River Delta was 80–350 m/s above a 100 m depth. The weights trained with the synthetic dispersion data could not accurately predict S-wave velocities because DL did not learn the features of the measured dispersion curve. To this end, we recalculated the velocity models with a range of 80 m/s to 350 m/s and the corresponding dispersion data over 8000 pairs. After training these datasets (~15 s), we obtained the weights, and then the 1D V_S models (blue dashed line in Figure 12) of the two stations were inverted.

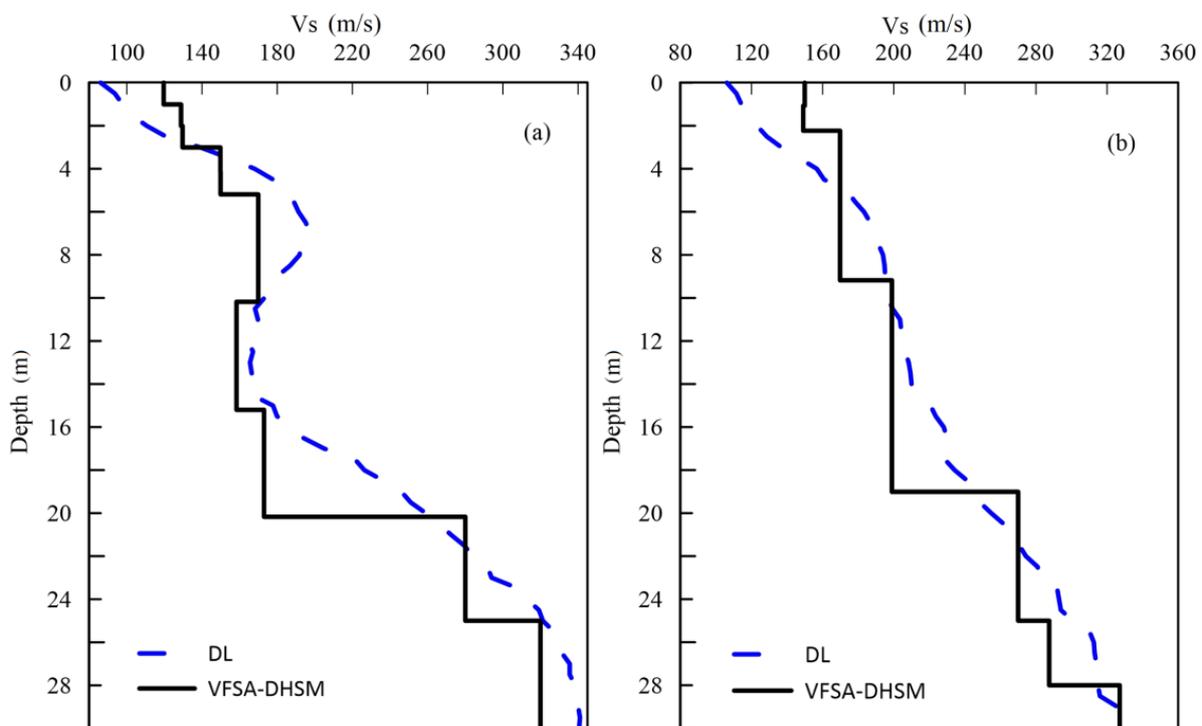


Figure 12. Comparison of inverted V_S model from DL and VFSA-DHSM method for observed dispersion data. (a) S1 station; (b) S2 station.

In order to compare the inversion results, we also used the VFSA-DHSM method [36] to invert the dispersion curve. The results are presented in Figure 12 (black solid line). As seen from the figure, DL results are in good agreement with that of the VFSA-DHSM. In addition, Liu et al. [37] gave statistics on the S-wave velocity in the Yellow River Delta area. They listed the S-wave velocity at every 10 m interval within a depth of 100 m; unfortunately, no more detailed logging information is available. Here, we tabulated the data and our inversion results in Table 1. Again, we found consistency among them and further confirmed the validity of our DNN method.

Table 1. Comparison of V_S between general statistics and DL inversion (Table 1).

Depth Range (m)	DL Inversion (m/s)		Liu 2015 (m/s) [37]
	S1	S2	
0–10	80–200	110–200	120–170
11–20	150–250	200–250	150–230
21–30	260–340	250–320	200–300

4. Conclusions and Discussions

4.1. Conclusions

This study investigated the DNN neural network to obtain 1D S-wave velocities of the near surface using the frequency-dependent Rayleigh wave dispersion curves. Based on the results, the following conclusions can be drawn:

1. Based on the ergodicity and orderliness of strata evolution and the constrained MC theory, we can construct innumerable rational velocity models that can effectively describe the more complex near-surface conditions. This approach could offer several benefits in terms of generating a wide range of training datasets that are necessary for effectively training DNN;
2. The effectiveness of the proposed DNN was first tested using a synthetic dataset. The training and validation loss curves and the accuracy of the validation dataset show that the performance of DNN tends to be stable after 110 epochs, and the accuracy of inversion results reaches ~90%;
3. Calculation of the relative errors for different network layers shows that the errors do not decrease all the way and will reach a minimum at a certain number of samples. The relative errors of different learning rates also have no simple linear relationship; the model performed best when the learning rate was 0.01 in our study, and the best number of network layers was six in our DNN architecture. This may be attributed to the highly nonlinear properties of the Rayleigh wave dispersion problem;
4. To apply DNN to real dispersion curves extracted from ambient noise in the Yellow River Delta, we rebuilt the 8000 pairs training dataset, and the training process took ~15 s. The results showed good consistency with that of the VFSA-DHSM inversion and previous statistics of the S-wave velocities, which may help to provide an alternative for deriving S-wave velocity models for complex near-surface structures.

4.2. Discussions

DNN provides a promising alternative method for inversion. It performs inversion by learning and extracting features from a large training dataset, allowing for fast mapping of dispersion data to shear wave models with good training. However, improving the accuracy and generalization ability of DNN inversion requires more complex and suitable neural network models, as well as more realistic and comprehensive sample datasets.

In this study, we investigated three parameters that significantly impact the convergence of the training process: the number of training samples, learning rate, and number of hidden layers. However, it should be noted that the complexity of the neural network model may lead to differences in parameter settings for different models. The optimal parameters obtained in this study are not fixed rules. In many cases, for other specific data sequences, it is necessary to determine these optimal parameters through repeated testing.

Furthermore, the randomness in the selection of velocity models for DNN training datasets can affect the learning process. In our research, we only use synthetic data to validate the network performance. However, the adequacy and diversity of the samples are still worth considering, which is a topic we must explore in the future.

Author Contributions: Conceptualization and methodology, Q.M. and Y.C.; formal analysis, Q.M. and Y.C.; data curation, Q.M. and Y.C.; writing—original draft preparation, Q.M. and Y.C.; writing—review and editing, F.S.; visualization, Y.C.; project administration, Q.M.; funding acquisition, Q.M. and T.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (Nos. 42272327) and the Social and Livelihood Project of Shandong Province (2021, 202131001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: The authors would like to thank Yang Li and Wenjing Wang, who participated in recording the ambient noise in the field. We also appreciate two anonymous reviewers for their valuable comments to improve the article.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Berg, E.M.; Lin, F.; Allam, A.; Schulte-Pelkum, V.; Ward, K.M.; Shen, W. Shear Velocity Model of Alaska Via Joint Inversion of Rayleigh Wave Ellipticity, Phase Velocities, and Receiver Functions Across the Alaska Transportable Array. *J. Geophys. Res. Solid Earth* **2020**, *125*, e2019JB018582. [[CrossRef](#)]
- Xu, P.F.; Li, S.H.; Ling, S.Q.; Guo, H.L.; Tian, B.Q. Application of SPAC method to estimate the crustal S-wave velocity structure. *Chin. J. Geophys.* **2013**, *56*, 3846–3854.
- Rahman, M.Z.; Kamal, A.S.M.M.; Siddiqua, S. Near-Surface Shear Wave Velocity Estimation and V_s30 Mapping for Dhaka City, Bangladesh. *Nat. Hazards* **2018**, *92*, 1687–1715. [[CrossRef](#)]
- Li, W.; Chen, Y.; Liu, F.; Yang, H.F.; Liu, J.L.; Fu, B.H. Chain-Style Landslide Hazardous Process: Constraints From Seismic Signals Analysis of the 2017 Xinmo Landslide, SW China. *J. Geophys. Res. Solid Earth* **2019**, *124*, 2025–2037. [[CrossRef](#)]
- Zhou, X.; Wang, W.J.; Li, Y.; Meng, Q.S. Research and application of microtremor in the geological disaster. *Period. Ocean. Univ. China* **2021**, *51*, 58–64. [[CrossRef](#)]
- Serdjukov, A.; Yablokov, A.; Chernyshov, G.; Azarov, A. The Surface Waves-Based Seismic Exploration of Soil and Ground Water. *IOP Conf. Ser. Earth Environ. Sci.* **2017**, *53*, 012010. [[CrossRef](#)]
- Zhang, W.; Lv, Y.; Liang, D.H.; Wu, J.Q.; Liu, W.; Zhu, C.Q. Application of active and passive-sourced seismic surface wave exploration to the detecting of shallow overburden karst area. *Geol. Bull. China* **2022**, *41*, 416–424.
- Socco, L.V.; Foti, S.; Boiero, D. Surface-Wave Analysis for Building near-Surface Velocity Models—Established Approaches and New Perspectives. *Geophysics* **2010**, *75*, 75A83–75A102. [[CrossRef](#)]
- Yu, D.K.; Song, X.H.; Jiang, D.W.; Zhang, X.Q.; Zhao, S.; Zhao, P.Q.; Cai, W.; Yuan, S.C. Improvement of Artificial Bee Colony and its application in Rayleigh wave inversion. *Chin. J. Geophys.* **2018**, *61*, 1482–1495.
- Lai, C.G. Surface Waves in Dissipative Media: Forward and Inverse Modelling. In *Surface Waves in Geomechanics: Direct and Inverse Modelling for Soils and Rocks*; Lai, C.G., Wilmański, K., Eds.; Springer: Vienna, Austria, 2005; pp. 73–163. [[CrossRef](#)]
- Cox, B.R.; Teague, D.P. Erratum: Layering Ratios: A Systematic Approach to the Inversion of Surface Wave Data in the Absence of A-Priori Information. *Geophys. J. Int.* **2017**, *211*, 378. [[CrossRef](#)]
- Dal Moro, G.; Pipan, M.; Gabrielli, P. Rayleigh Wave Dispersion Curve Inversion via Genetic Algorithms and Marginal Posterior Probability Density Estimation. *J. Appl. Geophys.* **2007**, *61*, 39–55. [[CrossRef](#)]
- Pei, D.H.; Louie, J.N.; Pullammanappallil, S.K. Application of Simulated Annealing Inversion on High-Frequency Fundamental-Mode Rayleigh Wave Dispersion Curves. *Geophysics* **2007**, *72*, R77–R85. [[CrossRef](#)]
- Socco, L.V.; Boiero, D. Improved Monte Carlo Inversion of Surface Wave Data. *Geophys. Prospect.* **2008**, *56*, 357–371. [[CrossRef](#)]
- Hadlington, L.; Binder, J.; Gardner, S.; Karanika-Murray, M.; Knight, S. The Use of Artificial Intelligence in a Military Context: Development of the Attitudes toward AI in Defense (AAID) Scale. *Front. Psychol.* **2023**, *14*, 1164810. [[CrossRef](#)]
- Rahmani, A.M.; Rezazadeh, B.; Haghparast, M.; Chang, W.-C.; Ting, S.G. Applications of Artificial Intelligence in the Economy, Including Applications in Stock Trading, Market Analysis, and Risk Management. *IEEE Access* **2023**, *11*, 80769–80793. [[CrossRef](#)]
- Jiang, Y.; Han, L.; Gao, Y. Artificial Intelligence-Enabled Smart City Construction. *J. Supercomput.* **2022**, *78*, 19501–19521. [[CrossRef](#)]
- Chollet, F. *Deep Learning with Python*; Manning Publications Co.: Shelter Island, NY, USA, 2018; ISBN 978-1-61729-443-3.
- Dai, H.; MacBeth, C. Automatic Picking of Seismic Arrivals in Local Earthquake Data Using an Artificial Neural Network. *Geophys. J. Int.* **1995**, *120*, 758–774. [[CrossRef](#)]
- Zhang, G.; Wang, Z.; Chen, Y. Deep Learning for Seismic Lithology Prediction. *Geophys. J. Int.* **2018**, *215*, 1368–1387. [[CrossRef](#)]
- Spichak, V.; Popova, I. Artificial Neural Network Inversion of Magnetotelluric Data in Terms of Three-Dimensional Earth Macroparameters. *Geophys. J. Int.* **2000**, *142*, 15–26. [[CrossRef](#)]
- Çaylak, Ç.; Kaftan, İ. Determination of Near-Surface Structures from Multi-Channel Surface Wave Data Using Multi-Layer Perceptron Neural Network (MLPNN) Algorithm. *Acta Geophys.* **2014**, *62*, 1310–1327. [[CrossRef](#)]
- Cao, R.; Earp, S.; de Ridder, S.A.L.; Curtis, A.; Galetti, E. Near-Real-Time near-Surface 3D Seismic Velocity and Uncertainty Models by Wavefield Gradiometry and Neural Network Inversion of Ambient Seismic Noise. *Geophysics* **2019**, *85*, KS13–KS27. [[CrossRef](#)]

24. Hu, J.; Qiu, H.; Zhang, H.; Ben-Zion, Y. Using Deep Learning to Derive Shear-Wave Velocity Models from Surface-Wave Dispersion Data. *Seismol. Res. Lett.* **2020**, *91*, 1738–1751. [[CrossRef](#)]
25. Yang, W.C. On Complexity and Ergodicity. *Sci. Technol. Rev.* **2008**, *3*, 1.
26. Ching, W.-K.; Ng, M.K. Markov Chains: Models, Algorithms and Applications. In *Markov Chains: Models, Algorithms and Applications*; International Series in Operations Research & Management Science; Springer: Boston, MA, USA, 2006; pp. 87–109. [[CrossRef](#)]
27. Brocher, T.M. Empirical Relations between Elastic Wavespeeds and Density in the Earth's Crust. *Bull. Seismol. Soc. Am.* **2005**, *95*, 2081–2092. [[CrossRef](#)]
28. Tang, Y.X.; Xiang, X.M.; Sun, J.; Zhang, Y.S. A Generic Shear Wave Velocity Profiling Model for Use in Ground Motion Simulation. *Geosciences* **2020**, *10*, 408. [[CrossRef](#)]
29. Ludwig, W.J.; Nafe, J.E.; Drake, C.L. *Seismic Refraction, the Sea*; Wiley-Interscience: New York, NY, USA, 1970; pp. 53–84.
30. Herrmann, R.B. Computer Programs in Seismology: An Evolving Tool for Instruction and Research. *Seismol. Res. Lett.* **2013**, *84*, 1081–1088. [[CrossRef](#)]
31. Hestness, J.; Narang, S.; Ardalani, N.; Damos, G.; Jun, H.; Kianinejad, H.; Patwary, M.M.A.; Yang, Y.; Zhou, Y. Deep Learning Scaling Is Predictable, Empirically. *arXiv* **2017**. [[CrossRef](#)]
32. Joulin, A.; van der Maaten, L.; Jabri, A.; Vasilache, N. Learning Visual Features from Large Weakly Supervised Data. *arXiv* **2015**. [[CrossRef](#)]
33. Meng, Q.S.; Li, Y.; Wang, W.J.; Chen, Y.H. Experimental study on key factors of the ambient noise CCA prospecting method. *Period. Ocean. Univ. China* **2023**, *53*, 134–141. [[CrossRef](#)]
34. Cho, I.; Tada, T.; Shinozaki, Y. A New Method to Determine Phase Velocities of Rayleigh Waves from Microseisms. *Geophysics* **2004**, *69*, 1535–1551. [[CrossRef](#)]
35. Cho, I.; Tada, T.; Shinozaki, Y. Centerless Circular Array Method: Inferring Phase Velocities of Rayleigh Waves in Broad Wavelength Ranges Using Microtremor Records. *J. Geophys. Res.* **2006**, *111*, B09315. [[CrossRef](#)]
36. Yokoi, T.; Black, K.; Nakagawa, H.; Suzuki, H.; Saito, H. Exploration at Yoshino General Park in Joso City, Ibaraki Pref., Japan Where the Ground Failure by Liquefaction Took Place Due to the 2011 off the Pacific Coast of Tohoku Earthquake. *BUTSURI-TANSA(Geophys. Explor.)* **2013**, *66*, 13–24. [[CrossRef](#)]
37. Liu, F.X.; Liu, L.; Liu, J.L. Characteristic Analysis of Shear Wave Velocity of the Yellow River Delta. *China Eng. Consult.* **2015**, *10*, 90–93.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.