A Deep-Learning-Based Oil-Well-Testing Stage Interpretation Model Integrating Multi-Feature Extraction Methods

Xin Feng 1, Qiang Feng 2, Shaohui Li 3, Xingwei Hou 1 and Shugui Liu 1, *

1 State Key Laboratory of Precision Measuring Technology and Instruments, Tianjin University, Tianjin 300072, China; fengxin182@tju.edu.cn (X.F.); houxingwei@tju.edu.cn (X.H.)
2 CNPC Bohai Drilling Engineering Company Ltd., Tianjin 300457, China; qfengtj@163.com
3 Tianjin Research Institute of Water Transport Engineering, Tianjin 300000, China; shlisfw@hotmail.com
* Correspondence: sgliu@tju.edu.cn

Received: 9 March 2020; Accepted: 9 April 2020; Published: 20 April 2020

Abstract: The interpretation of well-testing data is a key means of decision-making support for oil and gas field development. However, conventional processing methods have many problems, such as the stochastic nature of the data, feature redundancies, the randomness of the initial weights or thresholds, and fluctuations in the generalization ability with slight changes in the network parameters. These result in a poor ability to characterize data features and a low generalization ability of the interpretation models. We propose a new integrated well-testing interpretation model based on a multi-feature extraction method and deep mutual information classifiers (MFE-DMIC). This model can avoid the low model classification accuracy caused by the simple training structures, lacking of redundancy elimination, and the non-optimal classifier configuration parameters. First, we obtained the initial features according to four classical feature extraction methods. Then, we eliminated feature redundancies using a deep belief network and united the maximum information coefficient method to achieve feature purification. Finally, we calculated the interpretation results using a hybrid particle swarm optimization–support vector machine classification system. We used 572 well-testing field samples, including five working stages, for model training and testing. The results show that the MFE-DMIC model had the highest total stage classification accuracy of 98.18% as well as the least of features (nine) compared with the classical feature extraction and classification methods and their combinations. The proposed model can reduce the efforts of oil analysts and allow accurate labeling of samples to be predicted.

Keywords: data interpretation; well testing; integrated interpretation model; deep belief network

1. Introduction

During oil and gas exploitation, the use of an efficient stage interpretation scheme for well-testing data not only guides staff toward revising the production flow but can also provide an important means to manage reservoirs scientifically [1]. Due to temporary operation flow adjustments and the potential for uncertain events to occur at strata and borehole locations [2], well-testing data have the characteristics of being both stochastic and non-scheduled. Even when the data are from the same operation stage but different testing wells, a large difference exists in the length and curve shapes, which poses significant challenges when interpreting the data [3]. These differences directly contribute to an inaccurate matching relationship between the testing data being played back and the operation stage. This mismatch is primarily attributed to the poor information-characterizing ability of the features extracted from the data and to the low generalized interpretation ability of the models [1].
In recent years, time and frequency-domain methods—such as the gradient descent [4], time-series shapelets [5], wavelet packet transform (WPT) [6], and the empirical mode decomposition (EMD) [7]—have been shown to be advantageous for enhancing the information-characterizing ability of data features. Firoozabadi et al. [4] built a well-prediction model based on a pressure drop and achieved a prediction accuracy of 86%. Ahmadi et al. [5] proposed a well-testing identification model for the pressure transient test data using the concept of shapelets. The total predicting accuracy for the noisy data was 91.8%. Xu et al. [6] extracted features using a three-level WPT and used a neural network to predict the corner wear of a high-speed steel drill bit. The best drill-wear prediction accuracy was 0.9157. Zheng et al. [7] extracted the features of the real-time mud signals based on the intrinsic mode functions through an ensemble EMD. The relevance index was as high as 0.8929. However, the features extracted by the methods mentioned above are single-type features, and the number of samples used in previous studies was too small. Furthermore, in complicated working conditions, the small number of shallow computing units and the simple training structures of these methods create difficulties in the direct assessment of whether the features used are reasonable and effective. Therefore, to improve the information-characterizing ability of these features, it is necessary to perform a multiangle analysis of data characteristics.

To improve model generalizability, researchers have been studying this problem, but no universal scheme has emerged [8]. The artificial neural network (ANN) [9–11], support vector machine (SVM) [12–14], and deep learning [15,16] have been reported as modern mathematical methods to conduct the overall and regional analyses on data with multiple recognition stages, obtaining good generalization ability. Ahmadi et al. [11] developed a well-productivity prediction model of a horizontal well based on the ANN linked to the particle swarm optimization (PSO) tool. The attained low-average absolute error in percentage was lower than 0.82%. Xu et al. [17] proposed an oil–water flow pattern recognition method based on the support vector classification in a vertical well. The identification accuracy obtained using the proposed method was 97.95%. Although these studies obtained good interpretation results, they lacked the ability to optimize the model while maintaining training accuracy, and the features to be identified were obvious or predictable. Meanwhile, the time–frequency features extracted by the current research method or its combination method can describe the between-class differences of the studied samples at different stages. However, due to the lack of learning and selection of distinguishable features within or between classes, the feature extraction structure is not universal. For complex measurement environments, a model with a strong generalized interpretation ability is urgently needed. We observed that the deep belief network (DBN) [16,18] had high prediction and classification performance in intrusion detection [19], transformer fault diagnosis [20], image processing [21], and speech recognition [22]. Based on these advantages, this paper proposes a multi-feature extraction method and deep mutual information classifiers (MFE-DMIC) model with a strong generalized interpretation ability.

The purpose of the proposed model is to improve the interpretation performance of well-testing data and to obtain satisfactory classification of each stage. Figure 1 shows the framework of the proposed model. To improve the information-characterizing ability of the data features, we first performed multi-feature extraction (MFE) using four classical methods to extract the initial diverse features from the accumulation of 30 million pressure sampling points from 572 field well test samples. Powerful database support is lacking in conventional reservoir dynamic flow analysis. Next, we employed the deep belief network (DBN) to learn, excavate, and reconstruct features in depth. The maximum information coefficient (MIC) algorithm combines the gridding of two-dimensional space with statistical calculation of the probability distribution of each piece of information. It is suitable for the correlation analysis of linear and nonlinear data and has low computational complexity and high robustness. Then, we used the MIC [23] to remove data redundancy. These two processes not only retain the time–frequency characteristics that can represent potential changes in the data, but they also obtain high-level feature expressions with high distinguishing abilities. To ensure the efficient operation of each processing unit and enhance the generalized interpretation ability of the
model, we analyzed the effects of the feature redundancies, model scale, and parameter values on the training accuracy and used the multi-support vector machine (multi-SVM) system optimized by the particle swarm optimization (PSO) to classify the features. In this way, the feature learning and automatic tagging for the unlabeled data were completed.

Figure 1. Multi-feature extraction method and deep mutual information classifiers (MFE-DMIC) algorithm framework. MIC: maximum information coefficient.

2. Methods

2.1. Multi-Feature Extraction

A complete well-testing process has five stages: lowering the oil string, the waiting stage, well opening, well closing, and pulling up the oil string. Although the procedure of well testing is scheduled and planned, affected by the on-site working conditions, the running time of each stage is uncertain, and its configuration cannot be predicted, designed, or added to the program before lowering the electronic pressure gauge to the oil well. Some operating platforms need to adjust the frequency and time of well opening and closing according to the surface pressure test results. In addition, the underground energy and pressure recovery capacity of different test wells are different, which will lead to different waiting times for each testing stage. As shown in Figure 2a,b, even when the data are from the well opening stage but two different testing wells, large differences in the position, length, and occurrence of the peak and inflection point on the timeline exist. Therefore, it is impossible to design an accurate operation schedule in advance to provide an operation guide and executive standard for each test stage.

Figure 2. Diagram of oil test data analysis.
Figure 2 also shows the similarity of the distinguishable features between different stages. It shows that a stepped pressure rise will be caused by lowering the oil string, pressure gauge problems, and a low sampling frequency. The step pressure drop will occur when the oil string is pulled up and in variable-rate drawdown tests. Continuous peaks will appear due to screen plugging, multiple pressure-relieving operations during well closing, and mechanical vibrations. Furthermore, regarding the failure of well opening and closing, the packer loss of seal during the closing operation and the geological hazards will result in burrs similar to the curve peaks. The above analysis shows that it is not enough to analyze the data from the perspective of the time domain. The time–frequency domain analysis method has the ability to carry out multi-resolution data feature extraction, which can characterize the global and local features better using the above similarities and uncertainties. So, we propose the use of the multi-feature extraction (MFE) method to detect and distinguish changes in the production operation, noise interference, and formation blockage. We obtained five types of features, including the wavelet packet decomposition-approximate entropy feature (WPD-AE), the empirical mode decomposition-approximate entropy feature (EMD-AE), the fast Fourier transform (FFT) coefficient feature, the gradient feature, and the gradient extreme value feature.

The $n^\text{th}$ group data are defined as $s(n_0, N, x, y)$ with data length $N$, where $y$ is the amplitude of the data, the index number is $x \in [0, N - 1]$, and $\text{SD}$ denotes the standard deviation of $s(n_0, N, x, y)$. The dimension and tolerance threshold in EMD are donated as $m$ and $r$, respectively.

WPD and EMD are two of the best methods for the nonlinear analysis of data, as they are conducive to the analysis of different types of stepped rises/falls in the pressure from the perspectives of the step width, curve rise/fall slop, and the time length of each stage. By using WPD, we extracted the WPD coefficients $\{d^N_i(n_0)\}$, where $il \in [1, 2^c]$, and $C_i$ is the number of decomposed layers. By using EMD, we obtained multiple intrinsic mode function (IMF) components $\{c^N_i(n_0)\}$ and a residual $r(n_0)$, where $nl$ is the number of IMF components. Furthermore, we separately obtained the WPD-AE feature $a(n_0) = \text{ApEn}_m^N(n_0, d_i)$ and the EMD-AE feature $\beta(n_0) = \text{ApEn}_m^N(n_0, c_i)$ by calculating the approximate entropy (AE) of $\{d^N_i(n_0)\}$ and $\{c^N_i(n_0)\}$.

FFT can detect the density of the peak group and the sampling frequency. We obtained the frequency distribution of data in the frequency domain by calculating the top $k_0$ FFT coefficients

$$X(k_0, n_0) = \sum_{nx=1}^{N} s(n_0, N, nx) \exp[-j2k_0\pi(nx - 1)/N]$$

(1)

Production experience [24] shows that gradient features and peak values can not only accurately predict the motion trend of the well-testing data but they can also judge the differences between peaks and burrs. Considering that global analysis is more advantageous for extracting well-testing stage features than transient analysis, we used linear regression (LR) to extract the features defined as follows, including three-interval regression parameters $\text{reg}_3(n_0) = \{r^0_{3,1}, r^0_{3,2}, r^0_{3,3}\}$ and multi-interval extreme parameters $\text{Reg}_{nd}(n_0) = [\min(r^0_{nd, jd}), \max(r^0_{nd, jd})]$, where $r^0_{nd, jd}$ is the gradient value of the $jd$th interval data after dividing the $n_0^\text{th}$ group data into $nd$ parts; it can be expressed as

$$r^0_{nd, jd} = \sum_{kd=0}^{(N/nd)-1} (x_i - \bar{x})(y_i - \bar{y}) / \sum_{kd=0}^{(N/nd)-1} (x_i - \bar{x})^2,$$

(2)

where $nd \in Z$, $nd > 3$, $jd \in [1, nd]$, $\bar{x} = \sum_{kd=0}^{(N/nd)-1} x_{kd} / (nd/N)$, and $\bar{y} = \sum_{kd=0}^{(N/nd)-1} y_{kd} / (nd/N)$ represent the mean of the index number and the magnitude of the $jd$ interval data, respectively; and $kd \in [0, N/nd - 1]$.

Finally, the initial feature vector is expressed as

$$\text{attr}(n_0) = [a(n_0), \beta(n_0), X(k_0, n_0), \text{reg}_3(n_0), \text{Reg}_{nd}(n_0)].$$

(3)

The detailed MFE method and its parameter configuration are presented in Algorithm 1.
can prevent the poor generalization ability caused by the randomness of the initial weights or thresholds in deep learning networks, the local optimal initial parameter values can be obtained \cite{25}.

The DBN used in this paper was composed of a multilayer restricted Boltzmann machine (RBM) and a back propagation (BP) neural network. RBM has the advantages of possessing a simple structure, convenient network combination, and flexible setting of the neuron number in each layer. As shown in Figure 3, by training the RBM networks layer by layer and stacking the trained RBM networks into deep learning networks, the local optimal initial parameter values can be obtained \cite{25}.

```
Algorithm 1. Method 1: MFE method

Input: s(n₀, N, x, y).

Pretreatment: Normalize N to 1000, y to [0,100].

Output: attr(n₀).

1: Set Cₐ = 3, nl = 3, m=2, r=0.2SD, kₐ = 20, nd = 10 and 100.
2: Decompose s(n₀, N, x, y) and get \{δₐₙ(n₀)\}.
3: Get \{cₐₙ(n₀)\}, r(n₀).
4: Get a(n₀), β(n₀).
5: Calculate X(k₀, n₀) using Equation (1).
6: Calculate Reg₃(n₀) and Regₐ₀(n₀).
7: Get attr(n₀).
```

Thus, as shown in Table 1, the MFE method can obtain a vector with 38 multiple features for the n₀th group data.

| Method          | WPD-AE | EMD-AE | FFT  | LR(1) | LR(2) | LR(2) |
|-----------------|--------|--------|------|-------|-------|-------|
| Feature Number  | α(n₀)  | β(n₀)  | X(n₀)| reg₃(n₀)| Regₐ₀(n₀)| Regₐ₀(n₀)|
| Number          | 8      | 3      | 20   | 3     | 2     | 2     |

2.2. DBN Feature Learning

The DBN used in this paper was composed of a multilayer restricted Boltzmann machine (RBM) and a back propagation (BP) neural network. RBM has the advantages of possessing a simple structure, convenient network combination, and flexible setting of the neuron number in each layer. As shown in Figure 3, by training the RBM networks layer by layer and stacking the trained RBM networks into deep learning networks, the local optimal initial parameter values can be obtained \cite{25}.

![Figure 3. Illustration of the restricted Boltzmann machine framework.](image-url)

The introduction of the back propagation (BP) network to fine-tune the DBN network parameters can prevent the poor generalization ability caused by the randomness of the initial weights or thresholds and non-global optimization of parameters. The feature’s learning steps were as follows:

Step 1: Input the MFE features into the DBN network and train each layer of the RBM network for 100 iterations in an unsupervised manner.

Step 2: Add the fine-tuned BP network after DBN and optimize the DBN network weights 100 times to decrease the difference between the highest layer output of the network and the tagged data.
2.2.1. RBM Training

Define $\theta = \{a, b, \omega\}$ as the parameters of the RBM model. Then, use the contrastive divergence (CD) method \cite{26} to obtain an optimal value of $\theta$. The RBM training process and its parameter configuration are shown in Algorithm 2.

**Algorithm 2.** Method 2: RBM training process

| Input: $s(n_0, N, x, y)$ |
|--------------------------|
| **Pretreatment:** Normalize $y$ to $[0, 1]$. Divide $s(n_0, N, x, y)$ into $M = 80$ parts, satisfying $S = \cup_{i=1}^{M} S_i, |S_i| = n_{\text{batch}}$. |
| **Output:** optimal $\theta$. |
| **Parameters:** training period $\sigma$, learning rate $\eta$, bias vectors $a$ and $b$, and weight matrix $\omega$. |

1: Set $\sigma = 50$ and $\eta = 0.1$; $a$, $b$, and $\omega$ are random values. The number of nodes in $v$ and $h$ are set as needed.

2: Foreach $M \in \{1, \ldots, 80\}$ Do

3: Obtain $\Delta w$, $\Delta a$, and $\Delta b$ using the CD method.

4: Update $\theta$.

5: End Foreach

6: End Foreach

7: Get the correction of $\theta$.

2.2.2. DBN Optimization

The corrected $\theta$ of each layer in the RBM network can guarantee the best mapping of the feature vectors of only that layer. Therefore, we used the gradient descent method-based BP network to obtain the global parameters $\theta_{\text{DBN}}$ of the DBN. We used the Polack–Ribiere conjugate gradients method \cite{27} to compute the search directions. The optimal function value was controlled within a reasonable range based on the Wolfe–Powell stopping criteria \cite{28} using polynomial approximations.

2.3. MIC Purification

We constructed a new network structure with the same weight $\theta_{\text{DBN}}$ as the training DBN model. Then, we completed the input features $fv$ mapping from the first layer to the higher layers again. Therefore, the highest layer output features $y_{h-\text{level}}(n_0)$ from the $n_0$th group data can be expressed as

$$y_{h-\text{level}}(n_0) = f(fv; \theta_{\text{DBN}}).$$

Considering that the problem of feature redundancy still exists in $y_{h-\text{level}}(n_0)$, we introduced the MIC to realize the feature purification. The main steps are as follows:

Step 1: Calculate the MIC value $M(y_{h-\text{level}}, n_0, I_{h-\text{level}})$ using the Minepy package, where $I_{h-\text{level}}$ denotes the index number of each feature in $y_{h-\text{level}}(n_0)$.

Step 2: Update the index sorting based on the MIC value from large to small and obtain the rank, which is expressed as

$$\text{Rank} = \left[ I_{h-\text{level}}^{\ast}, M^{-1}(y_{h-\text{level}}, n_0, I_{h-\text{level}}^{\ast}) \right].$$

(5)

where $I_{h-\text{level}}^{\ast}$ denotes the updated feature index number, and $M^{-1}(y_{h-\text{level}}, n_0, I_{h-\text{level}}^{\ast})$ denotes the corresponding MIC value.

Step 3: Calculate the SVM classification accuracy of the above rank, select the top $n$ features to obtain the highest accuracy, and define these feature as $y_{n_{\text{MIC}}(n_0)}$. 
2.4. PSO-SVM Classification

2.4.1. PSO

The LIBSVM software package developed by Lin Chih-Jen [29] offers several advantages, including convenient modification, and easy transplantation. Its SVM function only requires two parameters to be set: the penalty parameter \( c \), and kernel function parameter \( g \).

We introduced PSO to avoid the occurrence of overlearning and underfitting states and K-fold cross validation to obtain the best parameters \( c \) and \( g \), \( K = 10 \) [30]. The proposed PSO-SVM training model is shown in Figure 4. It is evident that the training set is combined with the radial basis kernel function to optimize \( c \) and \( g \). We used the testing set to validate the high classification accuracy of the model.

![Figure 4. The PSO-SVM training model. K-CV: K-fold cross validation.](image)

2.4.2. SVM Classification

After PSO, we used the LIBSVM with optimal \( c \) and \( g \) for the final feature classification. The predicted testing tag \( \text{Tag}^t \) is expressed as

\[
\text{Tag}^t = \arg\max \{ i_c | f(y_{\text{MIC}}^n(n_0), i_c) = j_c \}, \tag{6}
\]

where \( f(\cdot, i_c) (i_c = 1, 2, \ldots, \text{Dec}) \) denotes the decision function of the \( i_c \)-th classifier, \( \text{Dec} \) is the combination of \( j_c \) and 2, and \( j_c \) is the number of categories.

For a given test sample, we define the number of classifiers that classify \( y_{\text{MIC}}^n(n_0) \) as a category \( j_c \) as \( \text{num}_{j_c} \),

\[
\text{num}_{j_c} = \{i_c | f(y_{\text{MIC}}^n(n_0), i_c) = j_c \}. \tag{7}
\]

Then, the category with the largest \( \text{num}_{j_c} \) is the category to which the sample belongs.

2.5. Data Source

We collected the dataset used in this paper from the well-testing platform in Huabei Oilfield, China. The reservoirs have the feature of special lithology, structural fractures, and strong edge water. The sample containing the data of a complete well-testing operation was used as the data sample, and a sample containing the data of one working stage was used as the stage sample. From 2009 to 2018, based on the field data collected from 572 wells, we obtained a total of 4004 stage samples of oil testing data and the corresponding operation stages. We acquired all of the data using a downhole pressure storage gauge. As for the proportions of the training set, verification set, and testing set, we adopted the 6:2:2 mode [31]. Considering that this paper does not involve the selection of hidden layers of DBN, the combination of the verification set and testing set will not affect the training effect; the final ratio of the training set and verification set/testing set was 6:4. We randomly classified 2402 stage samples as the training set and 1602 stage samples as the verification set/testing set. Since each data sample contained two well opening and closing operations, the ratio of the sample number in each stage was 1:1:2:2:1. Figure 5 shows the relationship of different types of samples.
The purpose of this paper is to put forward an idea of using multi-methods cooperation to realize high-precision classification in the well-testing stage. It mainly includes the following four aspects:

1. Propose the MFE multi feature extraction method on the basis of FFT, WPD, EMD, and gradient, and determine which method can complete the initial extraction of well-testing data features to the maximum extent.

Figure 5 shows the process of data acquisition to data playback. First, we lowered the electronic pressure gauge to the depth of the oil well where the measurement was required. Then, surface staff conducted the operations following the well-testing procedure. After the measurement was completed, the staff took the gauge out of the well, downloaded the memory data from the gauge, and interpreted the well-testing data using the method proposed in this paper.

![Figure 6. Well-testing workflow from data acquisition to data playback.](image)

We set the tag value for each stage as shown in Table 2. Surface staff identified the work stage based on the returned tag value and completed the interpretation work.

| Table 2. Working stages and the corresponding stage tags. |
|--------------------------------------------------------|
| Stage Order | Working Stage         | Stage Tag |
|-------------|-----------------------|-----------|
| S1          | Lower the oil string  | 1         |
| S2          | Waiting stage         | 2         |
| S3          | Well opening          | 3         |
| S4          | Well closing          | 4         |
| S5          | Pull-up the oil string| 5         |

2.6. Proposed Well-Testing Stage Classification Scheme

The purpose of this paper is to put forward an idea of using multi-methods cooperation to realize high-precision classification in the well-testing stage. It mainly includes the following four aspects:

1. Propose the MFE multi feature extraction method on the basis of FFT, WPD, EMD, and gradient, and determine which method can complete the initial extraction of well-testing data features to the maximum extent.
2. Set the layer number of the deep learning network, determine the optimal number of neuron nodes, and introduce the BP feedback network to optimize the parameters of the DBN network, so as to ensure the efficient configuration of the deep learning network.

3. Use MIC to analyze the priority of feature elements in the feature vector and eliminate redundant features.

4. Optimize SVM classifier parameters.

We set up an integrated well-testing interpretation model named MFE-DMIC. The model workflow is shown in Figure 7. The standard well testing operating procedure was used as a template, and each data sample was divided into five stages and stage labels were matched, forming a training set. The multi-dimensional feature attr extracted by the MFE method was learned by the parameter-optimized (including network weight ω and number of nodes R) DBN network. The output reconstructed feature was \( y_{\text{h-level}} \). The MIC method output the feature \( y_{\text{MIC}} \) after removing redundancies and reducing the feature dimension. Finally, the SVM optimized by the PSO could be used successfully for classification.

![Diagram](image)

**Figure 7.** MFE-DMIC workflow.

In Figure 7, we set up three hidden layers, acting on feature learning, feature reconstruction, and feature dimension reduction. The input layer was composed of approximately 10 times as many neurons as the number of feature elements, and thus, the DBN had a strong learning ability and information processing capacity. The number of neurons in each of the three hidden layers \( R_{\text{hidd}} \) declined by an equal ratio from input to output: 400, 200, and 100. The number of neurons in the first visible layer \( R_{\text{first}} \) was set to 38. Here, \( \text{Attr}_{\text{train}}(2402 \times 38) \) is defined as the feature training set, and \( \text{Attr}_{\text{test}}(1602 \times 38) \) as the testing set features. The number of neurons in the highest layer \( R_{\text{high}} \) was adjusted and set according to the simulation results of the training set. The main definitions are given in the Appendix A at the end of the paper (see Table A1). The simulation algorithm of the MFE-DMIC model is presented in Algorithm 3.
Algorithm 3. MFE-DMIC model

**Input:** \( s(n_{train}), \text{Tag}(n_{train}), s(n_{test}) \).

**Output:** \( \text{Tag}'(n_{test}) \).

**Model optimization and \( s(n_{train}) \) training:**
1: Set \( R_{midd} \).
2: **Foreach** \( n_{train} \in \{1, \ldots, 2402\} \) **Do**
3: Extract the initial features using proposed **Method 1**;
4: Get \( \alpha(n_{train}), \beta(n_{train}), X(n_{train}), \text{regs}(n_{train}), \text{Reg}_ad(n_{train}) \).
5: Get \( \text{attri}(n_{train}) \) by Equation (3);
6: **End Foreach**
7: Get \( \text{Attr}(2402 \times 38) = \{\text{attri}(n_{train})\} \).
8: Perform the RBM training using **Method 2**.
9: Optimize \( \theta_{DBN} \) using the BP network.
10: Get the optimal \( \theta_{high} \) by analyzing the impact of different \( \theta_{high} \) on classification accuracy.
11: Sort \( y_{h-level} \) by Equation (4).
12: Sort \( y_{h-level} \) by using the MIC and get \( y_{MIC}' \).
13: Get the optimal \( c \) and \( g \) based on \( y_{MIC}' \) and \( \text{Tag}(n_{train}) \).

**Interpretation of \( s(n_{test}) \):**
14: Set the optimal \( \theta_{DBN}, \theta_{high}, \text{SVM parameters} \) \( c \) and \( g \).
15: **Foreach** \( n_{test} \in \{1, \ldots, 1602\} \) **Do**
16: Perform steps 3–5 and get \( \text{attri}(n_{test}) \);
17: Perform steps 8 and 11–12;
18: Get \( \text{Tag}' \) by Equation (7);
19: **End Foreach**

3. Experimental Simulations and Results

We conducted all the training, testing, and validation experiments using MATLAB 2017a.

Figure 8 compares the interpretation efficiency of the DBN with a different number of highest-layer neurons. The BP network in the DBN optimization experiment had five layers in all. The number of fine-tuning BP optimization iterations for the weight \( \theta \) of the DBN was in the range 1–100.

![Figure 8. Trend of classification accuracy with different top-level features.](image-url)
Figure 8 shows that when $R_{\text{high}}$ was 30 and 50, and the number of weight optimization iterations was 30, and the classification accuracy stabilized quickly. When $R_{\text{high}}$ was 30, the DBN still achieved the highest system classification ability with fewer optimization iterations.

Table 3 shows the identification efficiency before and after the MIC processing with different numbers of highest layer neurons. To obtain self-validation classification accuracy, we derived the training set and the verification set from $\text{Attr}_{\text{train}}\ (2402 \times 38)$. We calculated the network output for the validation dataset using the sigmoid function. We obtained the self-verification classification accuracy by calculating the ratio of correct predictions to the total number of stage samples in verification set.

Table 3. Identification efficiency before and after the MIC processing at different numbers of high-layer neurons in DBN.

| Model Type | 20- | 30- | 40- | 50- | 60- |
|------------|-----|-----|-----|-----|-----|
| Number of highest-level nodes | 20 | 30 | 40 | 50 | 60 |
| Self-verification accuracy | 95.43% | 98.45% | 98.38% | 98.1% | 98.35% |
| Mean of loss | 0.2481 | 0.2133 | 0.2153 | 0.2157 | 0.2285 |
| Standard deviation of loss | 0.0154 | 0.0192 | 0.0189 | 0.0207 | 0.0221 |
| Number of MIC features | 10 | 9 | 25 | 35 | 27 |
| MIC classification accuracy | 95.44% | 98.19% | 98.13% | 98.13% | 97.63% |

In this paper, the tools-minepy toolbox was used to calculate the MIC values of each feature element in the DBN output feature vector under the condition of optimal meshing. According to the order of MIC values, we evaluated the SVM classification accuracy of the feature element subset of stage samples in the testing set.

We obtained the MIC classification accuracy using $y^9_{\text{MIC}}$. The input training features were $\text{Attr}_{\text{train}}\ (2402 \times 38)$, and the testing features were $\text{Attr}_{\text{test}}\ (1602 \times 38)$. Because the $y^9_{\text{MIC}}$ with the maximum self-verification classification rate was selected as the purified features, the SVM classification achieved the highest theoretical classification efficiency. Table 3 shows that the DBN with 30 highest-level neurons not only obtained the highest self-verification classification rate, but also showed good classification characteristics for its corresponding feature $y^9_{\text{MIC}}$.

Figure 9 shows the optimal feature ranking when $R_{\text{high}}$ was 30 or 50. The horizontal axis represents the feature priority ranking obtained by the MIC method. The vertical axis represents the classification accuracy obtained by the SVM classifier after the current feature was combined with all the features that were prioritized.

![Signature Sequence (a) $R_{\text{high}} = 30$](image1)

![Signature Sequence (b) $R_{\text{high}} = 50$](image2)

Figure 9. Optimal feature ranking when $R_{\text{high}}$ was 30 or 50.

Figure 9 shows that as the number of features increased, the classification accuracy of the system gradually increased and tended to stabilize. In addition, the features appearing after the classification
rate stabilized had little influence on the classification and interpretation of the system. These features were classified as redundant information that needed to be eliminated by the MIC operation. According to the results in Table 3, when the number of DBN highest-layer neurons was 30, the number of new features was the smallest—equal to 9—and had the best redundancy elimination effect.

Figure 10 shows the loss-value changing trend with the perplexity at different $R_{high}$ values. The horizontal axis represents the perplexity of the visualized t-distributed stochastic neighbor embedding (t-SNE) algorithm, and the vertical axis represents the loss value; it is evident that the loss value decreased as the perplexity increased.

![Figure 10. Loss curves at a different $R_{high}$ and perplexity values.](image)

In the PSO-based [11] parameter optimization experiment, the population number was 20, and the termination algebra was 200. Figure 11 shows the fitness curve of the parameters. In Figure 9, both the optimal fitness value and the average fitness value are kept above 93%, which illustrates that the particles in the optimization algorithm always maintained the best optimization ability. Finally, the optimum values of the SVM parameters were $c = 12.126$ and $g = 137.18$.

![Figure 11. Fitness (accuracy) curve of the PSO based parameter optimization.](image)

As shown in Figure 12, we employed the grid search method to provide statistical information regarding the classification accuracy near the optimal parameters. For $c$ and $g$, the range was $[2^{-10}, 2^{10}]$. 
and the steps size was 0.1. When c and g were optimal, the SVM achieved the maximum K-CV classification accuracy of 99.71%.

![3D view of classification accuracy distribution](image1)

![Contour view of classification accuracy distribution](image2)

**Figure 12.** The distribution of the classification accuracy at different parameters (3D view and contour view).

Figure 13 shows the stage tag corresponding to the predicted output and the actual standard tag. All training samples used in these methods were randomly selected. The rows correspond to the predicted class (Output Class) and the columns correspond to the true class (Target Class). The diagonal cells correspond to observations that were correctly classified.

| Output Class | S1 | S2 | S3 | S4 | S5 |
|--------------|----|----|----|----|----|
| S1           | 216| 0  | 0  | 0  | 0  |
| S2           | 19 | 198| 0  | 0  | 0  |
| S3           | 0  | 0  | 474| 0  | 0  |
| S4           | 1  | 14 | 447| 0  | 0  |
| S5           | 2  | 0  | 4  | 0  | 227|

DBN-MIC-SVM

| Output Class | S1 | S2 | S3 | S4 | S5 |
|--------------|----|----|----|----|----|
| S1           | 221| 0  | 0  | 0  | 0  |
| S2           | 15 | 198| 0  | 0  | 0  |
| S3           | 0  | 0  | 480| 0  | 0  |
| S4           | 1  | 12 | 447| 0  | 0  |
| S5           | 1  | 0  | 0  | 0  | 227|

DBN-MIC-PSO-SVM

**Figure 13.** Confusion matrices comparison of the SVM classification algorithm, before and after the PSO.

This paper introduced the t-SNE visualization method to achieve the visualization of all data samples by minimizing the standardized Euclidean distance between the original data and the reconstructed data. Figure 14 shows the t-SNE visualization classification effect of different dimensional features at perplexity = 100. In Figure 14, the serial numbers 1–5 correspond to the five well-testing stages.

We obtained the feature $Attr_{(400x38)}$ presented in Figure 12a using the MFE method. Because of the lack of deep feature extraction, many mutual inclusions and overlaps existed. The features corresponding to Figure 14b,c took the output features $y_{\text{level}}$ of the DBN as input. When $R_{\text{high}}$ was 50, some points belonging to Stage 3 were separated by Stage 2 and Stage 4. When $R_{\text{high}}$ was 30, there was no interruption and isolation, and the points belonging to the same stage were relatively aggregated. Figure 14d presents the features $y_{\text{MIC}}$ extracted by MIC. Compared with Figure 14c, their aggregation degree within the class was similar, but $y_{\text{MIC}}$ had a larger between-class distance, which could be more conducive to the identification of the well-testing working stage.
In order to compare the classification performance of this method with the classical feature extraction methods, we analyzed the classical feature extraction methods. Here, the classical methods generally refer to the methods mentioned above and their combination methods. Table 4 shows the SVM classification accuracy of the classical feature extraction methods used individually. Table 5 summarizes the SVM classification accuracy of different feature extraction and optimization algorithms for the five stages to be identified during well testing. The total classification accuracy was equal to the ratio of the number of successful classification samples to the total number of samples, and the average classification accuracy was equal to the average classification accuracy in each stage. In Table 5, A-B represents the integration algorithm integrating A and B methods.

Table 4. SVM classification accuracy using the classical feature extraction methods in MFE.

| Accuracy | S1   | S2   | S3   | S4   | S5   | Total          |
|----------|------|------|------|------|------|----------------|
| FFT      | 66.34| 44.64| 14.96| 33.97| 84.62| 41.511(665/1602) |
| WPD-AE [6] | 30.77| 47.77| 66.88| 19.44| 38.89| 41.573(666/1602) |
| EMD-AE [7] | 72.12| 36.61| 22.44| 8.12 | 44.44| 29.900(479/1602) |
| reg3     | 28.85| 88.39| 53.63| 48.29| 47.86| 52.871(847/1602) |
| Reg10    | 90.38| 12.50| 72.86| 53.42| 36.75| 55.743(893/1602) |
| Reg100   | 19.23| 87.50| 61.54| 61.97| 50.85| 58.240(933/1602) |

Table 5. SVM classification accuracy using different feature extraction and optimization algorithms.

| Accuracy (%) | S1   | S2   | S3   | S4   | S5   | Average/Total |
|--------------|------|------|------|------|------|---------------|
| MFE          | 60.92| 86.87| 91.67| 93.06| 89.40| 84.38/91.573  |
| DBN30        | 85.57| 93.43| 94.92| 99.56| 99.12| 94.52/95.256  |
| DBN30-MIC    | 90.76| 100  | 96.34| 100  | 100  | 97.42/97.503  |
| DBN30-PSO    | 87.39| 100  | 97.15| 100  | 100  | 96.91/97.253  |
| MFE-DMIC     | 92.86| 100  | 97.56| 100  | 100  | 98.08/98.190  |

4. Discussion

Based on the accumulation of 30 million pressure sampling points from 572 field well test samples, numerous simulation results are presented to reflect the feature representation and generalization.
abilities of the proposed method. Powerful database support is lacking in conventional reservoir
dynamic flow analysis.

We proposed the MFE method based on characteristics of well-testing data by integrating
four widely accepted feature methods: FFT, WPD, EMD, and gradient. As shown in Table 4,
the information-characterizing ability varied among the methods. The FFT, EMD-AE, and Reg10
features better characterized Stage 1 and Stage 5. Reg3 and Reg100 had performed well in the
recognition of Stage 2 and Stage 4. WPD-AE and Reg10 obtained relatively high levels of accuracy
for Stage 3. Even when we used the SVM classifier with the optimal parameters, the maximum
classification efficiency of a single feature was only 58.2397%, which was too low to satisfy the
requirement of practical production. Therefore, the classical methods had significant limitations in
well-testing data interpretation and the ability to identify the working stage. The MFE, with added
computing units, was able to extract more types of features compared with a single-feature extraction
method, so it could obtain a better classification effect; the total identification accuracy was 91.57%.
Although the error at Stage 1 was too large to meet the production requirements, it is still a feasible
method for obtaining the variation and characteristics of the data in the time–frequency domain.

As for obtaining the deep expression of multidimensional features, we first used a DBN supervisory
network to study the time–frequency characteristics obtained by the MFE method. To the best of our
knowledge, it was the first time that feature deep learning with a DBN was performed. We tried to
determine the DBN structure by seeking the configuration parameters corresponding to the highest
classification accuracy. It was a self-comparison process. Finally, \( R_{\text{high}} \) was determined to be 30 by
comparing a set of results on the classification accuracy, loss-value, and s-SNE distribution, which were
obtained by setting different \( R_{\text{high}} \) values in the simulation, with theoretical and practical tags.
This process is a new method for the deep learning and reconstruction of oil test data characteristics.
Being an important processing unit of the MFE-DMIC, DBN retains the time–frequency characteristics
without losing high-level feature expressions with a high distinguishing ability. Table 5 shows that the
DBN network with the highest layer neuron number of 30 combined with any method can achieve
perfect classification results in the S1, S4, and S5.

Studies on the feature selection of well-testing data are limited. Researchers usually enter the
extracted features directly into the classification system and seldom purify them from an MIC [32]
perspective. Thus, based on the priority ordering of the feature vectors under the SVM classification,
we demonstrated that sorting and screening with the MIC is an efficient method for identifying the
correlations between features, eliminating redundant features, and reducing the number of feature
dimensions (reducing the number of dimensions of the MFE from 38 to 9). Compared with DBN30,
the DBN30-MIC achieved maximum classification accuracy more quickly (Figure 9) and had the
best redundancy elimination effect (Table 3). Additionally, the comparison between DBN30 and
DBN30-PSO in Table 5 shows that it is feasible to use the PSO optimized parameters \( c \) and \( g \) for SVM
classification. Although the comparison results before and after the optimization of SVM parameters
only showed a 1% improvement, a high classification accuracy of 98.19% was obtained step-by-step by
making full use of the advantages of each algorithm. From another point of view, we tried to introduce
DBN into the analysis of well test data, and the results were satisfactory. This is because of the second
learning of MFE features by the DBN network, whereby all integration methods using the DBN can
extract features with a high representation ability.

We also find that the classical methods have different classification effects when dealing with data
in different stages. In Figure 13, the PSO-optimized decision-making system significantly improved the
efficiency of stage classification, except for S1 (the stage of lowering the oil string). Similarly, in Table 5,
PSO has the worst ability to classify stage 1. The possible reason might be that the string-lowering
stage was affected by external factors such as mechanical friction and noise interference, causing the
data with similar characteristics to other stages to appear at this stage. These similar features need to
be considered from the point of view of feature extraction and purification, which cannot be effectively
distinguished only by optimizing the system. As for MIC, PSO is better than MIC in improving the
classification accuracy of stage 3 (see Table 5). This means that after the DBN learning, the features of stage 3 have been fully excavated, and the main factor affecting the successful classification of stage 3 data is the work efficiency of the classifier. Therefore, it can be concluded that the MIC is helpful for S1 stage classification and the PSO is helpful for S3 classification.

With the use of the above-mentioned classical methods, the accuracy of the model classification gradually improved. Thus, the proposed MFE-DMIC method could achieve a performance superior to other classical methods and combination methods. The data processing route from multidimensional feature depth extraction to model optimization was reasonable, reliable, and suitable for autonomous feature learning and accurate labeling of the samples to be predicted.

In future work, three aspects need to be studied in-depth. First, although the MFE method can effectively extract the features of well-testing data, it is not the only method suitable for that purpose. We should try to use more feasible methods to extract different kinds of features to characterize the data information to the greatest extent possible. Second, this study only optimized the number of neurons at the highest DBN layer and thus lacked global optimization. At present, although there is no scientific and universal method to determine the number of neurons in each hidden layer of the DBN network, inappropriate setting will also affect the efficiency of deep learning [33]. Hence, the influence of a DBN structure on the classification accuracy should be analyzed. The number of hidden layers and the number of neurons in each layer should be adjusted and set according to the simulation results of the training samples. Third, the results shown in Figure 13 indicate that the classification results of the string lowering stage should be improved. In future research, we will improve the identification of similarity data and the fault tolerance of the decision models to reduce the misinterpretation of well-testing stages.

5. Conclusions

This paper extracted the deep features from well-testing data to obtain high classification accuracy. We proposed a new MFE-DMIC model to solve the two problems of current interpretation methods, that is, the poor information-characterizing ability of features and the low generalized interpretation ability of the model. First, we extracted multiple features using the MFE. Then, we conducted feature learning, reconstitution, and MIC purification using the DBN-MIC method. Lastly, we conducted classification using the PSO-optimized multi-SVM classification method. The proposed model achieved an average single-stage classification of 98.08% and a total accuracy of 98.19%, which indicated excellent classification performance in the well-testing process with five working stages. This model can be used to reduce the effort of oil analysts and to lower the error rate of data interpretation, thus providing a reliable scientific method for both actual production and academic research on data that is stochastic and non-scheduled.

Author Contributions: All authors contributed equally to this paper, where the first and corresponding authors were responsible for the conceptualization, the first author also for methodology, validation, writing, and other authors mainly for writing. 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 (51375338) and Research on key measurement standards and Traceability Technology of water transport engineering, National key research and development plan (2018YFP021220).

Acknowledgments: The authors would like to thank the editor and the reviewers who provided many helpful comments and thereby contributed to the final manuscript.

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

Table A1. Main definitions of the paper.

| Symbol | Quantity | Symbol | Quantity |
|--------|----------|--------|----------|
| $n_0$   | Group number of samples | $R_{\text{MFE}}(n_0)$ | Multi-interval extreme parameters |
| $s(n_0, \bullet)$ | Data in $n_0$th group | nd | Partitions the $n_0$th group data are divided into |
| $N$ | Length of $s(n_0, \bullet)$ | attr($n_0$) | MFE feature of $n_0$th group data |
| $x$ | Index number of $s(n_0, \bullet)$ | $y_{\text{p-tecd}}$ | Features output from $R_{\text{high}}$ |
| $y$ | Amplitude of the data | $\Theta$ | Parameters of the RBM |
| $\{d^N_{nl}(n_0)\}$ | WPD coefficients of $n_0$th group data | $\Theta_{\text{DBN}}$ | Optimum DBN weights |
| $C_l$ | Number of WPD decomposed layers | $M(\bullet)$ | The function to calculate MIC value |
| $\{\phi^N_{nl}(n_0)\}$ | IMF components of $n_0$th group data | $y_{\text{MIC}}$ | Features after MIC purification |
| $nl$ | Number of IMF components | $\text{Attr}^{\text{train}}$ | Training set features |
| $X(k_0, n_0)$ | FFT coefficient of $s(n_0, \bullet)$ | $\text{Attr}^{\text{test}}$ | Testing set features |
| $k_s$ | Number of FFT coefficients | $R_{\text{first}}$ | Number of neurons in DBN first layer |
| $\text{ApEn}_{N_{s}}$ | Approximate entropy (AE) | $R_{\text{hid}}$ | Number of neurons in DBN hidden layers |
| $a(n_0)$ | WPD-AE of $s(n_0, \bullet)$ | $R_{\text{high}}$ | Number of neurons in DBN highest layer |
| $\beta(n_0)$ | EMD-AE of $s(n_0, \bullet)$ | $\text{Tag}(n_0)$ | Stage tag corresponding to $s(n_0, \bullet)$ |
| $\alpha_{\text{reg}}(n_0)$ | Three-interval regression parameters | $\text{Tag}'$ | Predicted testing tag |

References

1. Rahmanifard, H.; Plaksina, T. Application of artificial intelligence techniques in the petroleum industry: A review. Artif. Intell. Rev. 2019, 52, 2295–2318. [CrossRef]
2. Arnaout, A.; O’Leary, P.; Esmael, B.; Thonhauser, G. Distributed recognition system for drilling events detection and classification. Int. J. Intel. Syst. 2014, 11, 25–39. [CrossRef]
3. Bhattacharyya, B.; Solomatine, D.P. Machine learning in soil classification. Neural Netw. 2006, 19, 186–195. [CrossRef] [PubMed]
4. Firoozabadi, H.M.; Rahimzade, K.; Pourafshary, P.; Edalat, M. Analysis of production logging data to develop a model to predict pressure drop in perforated gas condensate wells. Petrol. Sci. Technol. 2011, 29, 1722–1732. [CrossRef]
5. Ahmadi, R.; Aminshahidy, B.; Shahrabi, J. Well-testing model identification using time-series shapelets. J. Petrol. Sci. Eng. 2016, 149, 292–305. [CrossRef]
6. Xu, J.; Yamada, K.; Seikiya, K.; Tanka, R.; Yamane, Y. Effect of different features to drill-wear prediction with back propagation neural network. Petrol. Explor. Dev. 2012, 39, 798–801. [CrossRef]
7. Zheng, Y.; Sun, X.; Chen, J.; Yue, J. Extracting pulse signals in measurement while drilling using optimum denoising methods based on the ensemble empirical mode decomposition. Petrol. Explor. Dev. 2012, 39, 798–801. [CrossRef]
8. Aguirre, L.A.; Teixeira, B.O.S.; Barbosa, B.H.G.; Teixeira, A.F.; Campos, M.C.M.M.; Mendes, E.M.A.M. Development of soft sensors for permanent downhole gauges in deepwater oil wells. Control Eng. Pract. 2017, 65, 83–99. [CrossRef]
9. Zadkarami, M.; Shahbazian, M.; Salahshoor, K. Pipeline leakage detection and isolation: An integrated approach of statistical and wavelet feature extraction with multi-layer perceptron neural network (MLPNN). J. Loss Prev. Proc. 2016, 43, 479–487. [CrossRef]
10. Wilamowski, B.M.; Kaynak, O. Oil well diagnosis by sensing terminal characteristics of the induction motor. IEEE Transl. Ind. Electron. 2000, 47, 1100–1107. [CrossRef]
11. Ahmadi, M.A.; Soleiman, R.; Lee, M.; Kashiwao, T.; Bahadori, A. Determination of oil well production performance using artificial neural network (ANN) linked to the particle swarm optimization (PSO) tool. Petroleum 2015, 1, 118–132. [CrossRef]
12. Wang, F.; Lin, W.; Liu, Z.; Wu, S.; Qiu, X. Pipeline leak detection by using time-domain statistical features. IEEE Sens. J. 2017, 17, 6431–6442. [CrossRef]
13. Wang, C.; Zhang, Y.; Song, J.; Liu, Q.; Dong, H. A novel optimized svm algorithm based on pso with saturation and mixed time-delays for classification of oil pipeline leak detection. J. Syst. Sci. Syst. Eng. 2019, 7, 75–88. [CrossRef]
14. Kumar, A.; Ramkumar, J.; Verma, N.K.; Dixit, S. Detection and classification for faults in drilling process using vibration analysis. In Proceedings of the Prognostics & Health Management, Cheney, WA, USA, 22–25 June 2014; IEEE: Piscataway, NJ, USA, 2014.
15. Zhang, X.; Zhang, H.; Guo, J.; Zhu, L. Auto measurement while drilling mud pulse signal recognition based on deep neural network. *J. Petrol. Sci. Eng.* 2018, 167, 37–43. [CrossRef]
16. Kim, J.K.; Han, Y.S.; Lee, J.S. Particle swarm optimization-deep belief network-based rare class prediction model for highly class imbalance problem. *Concurr. Comput. Pract. Exp.* 2017, 29, e4128. [CrossRef]
17. Xu, L.; Chen, J.; Cao, Z.; Zhang, W.; Xie, R.; Liu, X.; Hu, J. Identification of oil–water flow patterns in a vertical well using a dual-ring conductance probe array. *IEEE Transl. Instrum. Meas.* 2016, 65, 1249–1258. [CrossRef]
18. Zhang, Y.; Li, P.; Wang, X. Intrusion detection for IoT based on improved genetic algorithm and deep belief network. *IEEE Access* 2019, 7, 31711–31722. [CrossRef]
19. Hatcher, W.G.; Yu, W. A survey of deep learning: Platforms, applications and emerging research trends. *IEEE Access* 2018, 6, 24411–24432. [CrossRef]
20. Dai, J.; Song, H.; Sheng, G.; Jiang, X. Dissolved gas analysis of insulating oil for power transformer fault diagnosis with deep belief network. *IEEE Transl. Dielect. Electr. Insul.* 2017, 24, 2828–2835. [CrossRef]
21. Hinton, G.E. Training products of experts by minimizing contrastive divergence. *Neural Comput.* 2002, 14, 1771–1800. [CrossRef] [PubMed]
22. Sastry, S.P.; Shontz, S.M. Performance characterization of nonlinear optimization methods for mesh quality improvement. *Eng. Comput.* 2012, 28, 269–286. [CrossRef]
23. Ahmad, A.; Zabidin, S. Modification of nonlinear conjugate gradient method with weak wolfe-powell line search. In *Abstract and Applied Analysis*; Hindawi: London, UK, 2017; pp. 1–6.
24. Hsu, C.W.; Lin, C.J. A comparison of methods for multiclass support vector machines. *IEEE Transl. Neural Netw.* 2002, 13, 415–425.
25. The Concept and Partition Principle of Training Set, Verification Set and Testing Set (Chinese). Available online: https://www.cnblogs.com/hello-aij/p/11099824.html (accessed on 27 June 2019).
26. James, G.; Witten, D.; Hastie, T.; Tibshirani, R. *An Introduction to Statistical Learning*; Springer: New York, NY, USA, 2013.
27. Reshef, D.N.; Reshef, Y.A.; Finucane, H.K.; Grossman, S.R.; McVean, G.; Turnbaugh, P.J.; Sabeti, P.C. Detecting novel associations in large data sets. *Science* 2011, 334, 1518–1524. [CrossRef]
28. Hinton, G. A practical guide to training restricted Boltzmann machines. *Momentum* 2010, 9, 926–947.