



Article Sample-Pair Envelope Diamond Autoencoder Ensemble Algorithm for Chronic Disease Recognition

Yi Zhang ^{1,2,3,†}, Jie Ma ^{4,†}, Xiaolin Qin ^{1,2,*}, Yongming Li ^{4,*} and Zuwei Zhang ³

- ² University of Chinese Academy of Sciences, Beijing 100049, China
- ³ Academy of Chips Technology, China Electronics Technology Group, Chongqing 401332, China
- ⁴ School of Microelectronics and Communication Engineering, Chongqing University,
 - Chongqing 400044, China
- * Correspondence: qinxl2001@126.com (X.Q.); yongmingli@cqu.edu.cn (Y.L.)
- + These authors contributed equally to this work.

Abstract: Chronic diseases are severe and life-threatening, and their accurate early diagnosis is difficult. Machine-learning-based processes of data collected from the human body using wearable sensors are a valid method currently usable for diagnosis. However, it is difficult for wearable sensor systems to obtain high-quality and large amounts of data to meet the demands of diagnostic accuracy. Furthermore, existing feature-learning methods do not deal with this problem well. To address the above issues, a sample-pair envelope diamond autoencoder ensemble algorithm (SP_DFsaeLA) is proposed. The proposed algorithm has four main components. Firstly, sample-pair envelope manifold neighborhood concatenation mechanism (SP_EMNCM) is designed to find pairs of samples that are close to each other in a manifold neighborhood. Secondly, the feature-embedding stacked sparse autoencoder (FESSAE) is designed to extend features. Thirdly, a staged feature reduction mechanism is designed to reduce redundancy in the extended features. Fourthly, the sample-pair-based model and single-sample-based model are combined by weighted fusion. The proposed algorithm was experimentally validated on nine datasets and compared with the latest algorithm. The experimental results show that the algorithm is significantly better than existing representative algorithms and it achieves the highest improvement of 22.77%, 21.03%, 24.5%, 27.89%, and 10.65% on five criteria over the state-of-the-art methods.

Keywords: recognition of chronic disease; wearable sensor monitoring; sample-pair envelope concatenation; envelope learning; diamond-like feature learning; feature-embedded stacked sparse autoencoder; ensemble learning

1. Introduction

Chronic diseases have been the greatest threat to human life in recent decades and it is vital to diagnose and predict them before reducing the rate of fatality [1]. As chronic diseases are long-lasting and have slow progression, the lack of wearable sensor monitoring system is currently limiting disease diagnosis, thus affecting timely treatment of early patients. At present, wearable sensors can be used to continuously collect physiological signals from patients fast and then the obtained data can be processed by machine learning algorithms [2]. The structure of a machine-learning-based chronic disease recognition system is shown in Figure 1. Firstly, physiological data from the human body are collected by the wearable sensors. Then, they are stored in a chronic disease database by transmission. After that, machine learning algorithms are used to process and analyze data to provide diagnosis results back to the patient or hospital. For example, a continuous dataset of vital signs is obtained using optical and pressure sensors, followed by feature extraction and classification to obtain predictions for chronic diseases [3]. Data collected by a set of sensors are fused and then fed into a classifier model for early prediction of heart disease [4].



Citation: Zhang, Y.; Ma, J.; Qin, X.; Li, Y.; Zhang, Z. Sample-Pair Envelope Diamond Autoencoder Ensemble Algorithm for Chronic Disease Recognition. *Appl. Sci.* **2023**, *13*, 7322. https://doi.org/10.3390/app13127322

Academic Editor: Alexander N. Pisarchik

Received: 27 April 2023 Revised: 13 June 2023 Accepted: 15 June 2023 Published: 20 June 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 Chengdu Institute of Computer Application, Chinese Academy of Sciences, Chengdu 610041, China

Currently, many wearable medical sensor systems enable data collection but do not provide high-quality data and timely medical diagnosis. It is well known that feature learning in machine learning has a large impact on recognition results. Therefore, the improvement of feature learning in machine learning algorithms for chronic diseases is important. It is a major motivation of this paper.



Figure 1. The structure of chronic disease recognition system based on machine learning.

Machine learning technology has enabled new tools for recognition of chronic diseases because of its efficiency in data processing. Research on machine learning methods for chronic disease recognition is focused on the following two areas: feature learning and classifier design.

Feature learning is particularly important. Feature learning algorithms mainly include feature selection methods and feature extraction methods. Feature selection is the procedure of selecting the best features from all the features available to distinguish between classes. Feature extraction converts features from a higher dimensional to a lower dimensional space. Some techniques used for the diagnosis of chronic diseases include linear discriminant analysis (LDA), generalized discriminant analysis (GDA), principal component analysis (PCA), and so on.

Classifiers can discover hidden patterns in existing human physiological data. Common classifiers [5–10] include the radial basis function network (RBF Network), decision tree (DT), naive Bayesian (NB), logistic regression (LR), functional tree (FT), logistic model tree (LMT), support vector machine (SVM), k-nearest neighbor (KNN), multi-layer perceptron (MLP), random forest (RF), and recurrent neural network (RNN). Ekanayake et al. [11] selected high-quality features and used 11 machine learning methods for model training: KNN, support vector classifier (SVC) with a linear kernel, LR, decision tree classifier, SVC with radial basis function (RBF) kernel, Gaussian NB, RF, a classical neural network, extra trees classifier, Adaboost classifier, and extreme gradient boosting (XGB) classifier.

Motivation: These techniques developed by these studies improve the performance for chronic disease detection. In particular, feature learning can provide high-quality features, thus reducing classifier complexity and improving recognition performance. However, the existing feature learning algorithms did not take into account the characteristics of datasets on chronic diseases well, which typically are (1) small or medium-sized features and samples, usually with no more than 35 features and no more than 1500 samples according to the relevant literature; (2) complex correlation between features and class label (disease status); and (3) a requirement for high recognition accuracy. Deep learning has powerful feature extraction capabilities but cannot obtain sufficiently high quality of features in the condition of small samples. Existing non-parameter or low-parameter feature reduction algorithms, such as PCA and LDA, can considerably reduce the number of features in small samples. However, they use the original features and do not work well when the

original features have low quality. Therefore, it is important and challenging to obtain a high-quality feature set in the case of small or medium-sized features and samples.

To address the above issues, this paper proposes a solution–sample-pair envelope diamond autoencoder ensemble algorithm (SP_DFsaeLA). Initially, the sample-pair envelope manifold neighborhood concatenation mechanism (SP_EMNCM) is designed to capture the intrinsic structure of manifold neighboring samples. Then, the feature embedding stacked sparse autoencoder (FESSAE) is designed to extend features. To eliminate redundancy between the expanded features, a nonparametric staged feature reduction mechanism is designed. The first stage uses the L1 regularized feature reduction algorithm and the second stage uses an improved manifold dimensionality reduction algorithm to further reduce the number of features. This mechanism combining SP_EMNCM, feature expansion, and feature reduction is called the sample-pair diamond-like feature learning mechanism. Figure 2a illustrates the flow of non-parameter or low-parameter feature reduction methods, such as PCA, LDA, LPP, and Relief. Figure 2b illustrates the flow of the deep feature learning method. The proposed sample-pair diamond-like feature learning mechanism is illustrated in Figure 2c.



Figure 2. Proposed method and traditional feature-based learning methods: (**a**) non-parameter or low-parameter feature reduction; (**b**) deep feature learning; (**c**) sample-pair diamond-like feature learning mechanism.

The main contributions of this article are as follows:

1. A diamond-like feature learning mechanism is proposed for the first time in this paper, realizing feature expansion and reduction with a diamond topology. The traditional feature learning algorithm has low adaptivity, and the deep learning algorithm is dependent on a large number of samples. Both cannot meet the requirement for

high accuracy from chronic disease recognition with low samples. The proposed diamond-like feature learning mechanism includes feature expansion mechanism and feature reduction mechanism. It first expands features to enhance the representation capability, then reduces features to enhance the generalization capability, thereby reducing the requirement for a large number of samples. In short, this mechanism combines deep learning and traditional feature reduction methods' advantages together and can better adapt to the data characteristics of chronic diseases and the requirement for high recognition accuracy than the deep learning and traditional feature reduction methods.

- 2. Existing feature learning methods only focus on the original sample itself and ignore the neighborhood relationship of the sample. This leads to the samples' features being easily affected by noise because the features are reduced and the relative relationship (e.g., manifold structure) is possibly being broken easily. The SP_EMNCM proposed in this paper is used to mine the sample nearest manifold neighbor structure information to form an envelope-like structure, thus enriching the feature information and improving its representational capacity.
- 3. Existing deep stack autoencoders learn poor-quality features at small sample sizes and have limited feature complementary fusion performance due to insufficient complementarity. The FESSAE is designed as a lightweight deep network for feature extraction of chronic sensing data. The FESSAE network introduces original features into the training process and structure of the network to improve the complementary nature between the higher quality feature and original feature, thus achieving high-quality deep features in small and medium sample sizes.
- 4. Existing feature expansion methods do not consider both between-sample structural information and between-feature complementary information in the expansion process, which leads to the introduction of a large number of ineffective features. The proposed feature expansion mechanism includes SP_EMNCM and FESSAE. It considers both the manifold neighbor structure information between samples and the feature complementarity, thereby improving original features richer and better.
- 5. Existing feature reduction methods do not adequately select and extract features for chronic disease recognition. The proposed staged feature reduction mechanism in this paper first selects the expanded features based on L1 regularization, then reduces the dimensionality of the most important features based on manifold learning, thus making the features more compact without losing useful information.

The following is the organization of the paper: Section 2 discuss some existing machine learning methods for chronic disease recognition. The proposed method is mainly described in Section 3. The experimental results are analyzed and reported in Section 4. Finally, in Section 5, the main contributions and possible limitations are discussed.

2. Related Work

In this section, we first briefly discuss some existing machine learning methods for chronic disease recognition and then discuss their advantages and disadvantages.

Feature learning refers to the process of automatically discovering and extracting meaningful representations or features from original data. Ahmed H et al. [12] used Relief and univariate feature selection to select great features from the dataset. Shrivas et al. [13] introduced a union-based feature selection technique for predicting chronic kidney disease. Chormunge et al. [14] realized a new relevance and cluster-based feature selection method in order to reduce the dimensionality issue in data mining tasks. Sawhney et al. [15] used penalization functions combined with the existing fitness function of the binary firefly method to reduce the feature set and improve cancer classification accuracy. Jayaraman et al. [16] have combined particle swarm optimization and gravitational cuckoo search algorithms for managing the features that exist in heart disease classification systems. Paul A K et al. [17] used weighted least squares to select effective attributes, and Rasitha [18] used LDA to classify hypothyroid disease. Mohamed et al. [19] used PCA to reduce the

5 of 29

dimensions of medical data of type 2 diabetes. Shahbazi et al. [20] used GDA to minimize the number of features in the feature space and the overlap of samples. Lu H et al. [21] developed a patient network and machine learning method that combines the attributes of the patient network with sample features. Taghizadeh E et al. [22] chose Analysis of Variance, mutual information, additional tree classifiers, logistic regression for feature selection, and then used PCA again for feature reduction. Khan A et al. [23] understood the pattern of complications of type 2 diabetes and then analyzed codes and their relationships were then analyzed to construct a comorbidity network.

For classifier design, Ge et al. [24] researched a new multi-label neural network to predict chronic diseases. El-baz et al. [25] proposed a combinative classifier based on the KNN classifier in the prediction module of a hybrid intelligent system for breast cancer tumor recognition. Polat [26] used the SVM, KNN, RF, and LDA methods to classify medical databases after attribute weighting. Cheruku et al. [27] introduced a new hybrid decision support to a bat optimization method and a system based on rough set theory. In this hybrid system, the redundant feature is effectively reduced by generating fuzzy rules. Maniruzzaman et al. [28] proposed a Gaussian-based classification model for diabetes and investigated the performance of a Gaussian process using three kernels. Alhassan and Zainon [29] presented a deep-belief network for heart disease diagnosis. Abdollahi J et al. [30] used 10 machine learning algorithms as the basic algorithms in a stack generalization algorithm to predict chronic diseases and implemented a hybrid metaalgorithm for prediction. Fatan M [31] used advanced fusion techniques, deep learning segmentation methods, and survival analysis to automatically segment tumor and predict survival outcome in head-and-neck squamous cell carcinoma cancer. Rezaeijo S M [32] used hierarchical clustering to improve the validity of mpMR image-based for prostate tumor classifications.

In Table 1, the advantages and disadvantages of the method proposed in this paper and previous methods are listed.

Method	Strengths	Weaknesses
Ahmed H et al. [12]	Predicting heart disease with real-time streaming data.	Low adaptivity.
Shrivas et al. [13]	High robustness and computational efficiency.	Only focus on the original sample itself and ignore the neighborhood relationship of the sample.
Chormunge et al. [14]	Reduce the dimensionality issue in data mining tasks.	Sensitive to large changes in data distribution.
Sawhney et al. [15]	Effectively reducing the number of features and saving computational overhead.	Limited improvement in classification accuracy.
Rasitha [18]	Effective in reducing feature dimensionality.	Not sensitive to non-Gaussian distributed samples.
Mohamed et al. [19]	Relatively simple and computationally efficient.	only original features are operated on. When the quality of the original feature is not high, it does not work well.
El-baz et al. [25]	High classification accuracy and interpretability.	Increasing computational complexity.
Maniruzzaman et al. [28]	Robust to noise and outliers.	In the case of small sample sizes, it is not possible to obtain sufficiently high feature quality.

Table 1. Strengths and weaknesses of the proposed method and previous methods.

Table 1	L. Cont.
---------	----------

Method	Strengths	Weaknesses
Alhassan and Zainon [29]	Utilizes advanced deep learning architecture for feature learning	High computational and training resource
SP_DFsaeLA (proposed)	Both the manifold neighbor structure information between samples and the feature complementarity are considered, leading to a richer features and better improvement on accuracy. More adapted to the data characteristics of wearable chronic diseases recognition.	Requires careful tuning of hyperparameters.

In general, these researches have improved the performance of chronic disease detection. In particular, feature learning can provide high-quality features to improve recognition performance. However, existing feature learning algorithms do not take into account the characteristics of chronic disease datasets well. So, this paper proposed a diamond-like feature learning mechanism, realizing feature expansion and reduction with a diamond topology. It first expands features to enhance the representation capability, then reduces features to enhance the generalization capability, thereby reducing the requirement for a large number of samples.

3. Materials and Methods

3.1. Problem Formulation

Suppose a chronic disease dataset $X = \{(x_i, y_i)\}_{i=1}^N$, where x_i is *i*th sample and y_i is its corresponding label. The sample vector can be expressed as $x_i = \{x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(M)}\}$, where *M* is the number of features. Through feature learning operator (method) $\phi(\cdot)$, x_i is transformed to $\hat{x}_i = \phi(x_i)$ by feature learning. Traditional feature learning algorithms reduce the dimensionality to obtain $\tilde{x}_i = \phi_{tr}(x_i)$ and then train the classifier *C* to obtain prediction results. The traditional non-parameter or low-parameter feature learning algorithms are poorly adaptive and their feature extraction capability is unsatisfactory when the original features are complex. Deep learning method $\phi_d(\cdot)$ is conducted to obtain reduced dimensional features $\tilde{x}_i = \phi_d(x_i)$. Deep learning has powerful feature extraction capabilities but suffers from a small-sample-size problem. To solve this problem above, the SP_DFsaeLA algorithm is proposed to find the optimal $\phi(\cdot)$, thereby improving subsequent classification accuracy.

3.2. Proposed Algorithm's Framework

In this section, we introduce the proposed algorithm. Sample-pair diamond stacked sparse autoencoder ensemble learning algorithm (SP_DFsaeLA) is developed to better recognize chronic diseases according to the characteristics of chronic disease data. This algorithm consists of four parts, as illustrated in Figure 3. First, the sample-pair envelope manifold neighborhood concatenation mechanism (SP_EMNCM) is designed by searching the manifold nearest samples in manifold neighborhood and generating sample pairs. Second, the feature embedding stacked sparse autoencoder (FESSAE) is designed to extend features. Third, a staged feature reduction mechanism is designed to remove the feature redundancy. It includes L1 regularization and weighted locality preserving discriminant projection (L1_wLPPD). This mechanism for achieving feature expansion followed by reduction is called the diamond-like feature learning mechanism (DFLM), as show in Figure 4. Fourth, a sample-pair-based model is constructed for SP_EMNCM and DFLM

and a single-sample-based model is constructed for single sample and DELM. The samplepair-based model and single-sample-based model are combined by weighted fusion.



Figure 3. Proposed algorithm's framework.



Figure 4. Diamond-like feature learning mechanism (DFLM).

The main symbols used in this paper are listed with their meanings in Table 2.

Table 2. Symbol description.

Symbol	Meaning
$X \in R^{N*M}$	Input data. <i>N</i> and <i>M</i> are sample size and feature size.
Ŷ	Sample label.
$\mathbf{v} \subset \mathbf{p}N*2M$	Generated sample pairs. <i>N</i> and 2 <i>M</i> are the sample size and
$A_{pair} \in K$	feature size, respectively.
W_1	Encoder's weight.
W2	Decoder's weight.
<i>b</i> ₁ , <i>b</i> ₂	Bias vector of the autoencoder.
d	Number of hidden units.
$H \in R^{N*d}$	Hidden feature.
$d^{(k)}$	kth autoencoder number of hidden layer units.
β	Coefficient for the sparsity regularization term.
λ	Coefficient for the L2 weight regularization term.
ρ	Sparse parameter.
â	All training samples' average activation value on the <i>j</i> -th
	hidden neuron
$\hat{X} \in \mathcal{R}^{N*\hat{M}}$	Data after feature expansion.
S^{ϕ}_B	Inter-class variance matrix.
$S^{oldsymbol{\phi}}_W$	Intra-class variance matrix.

3.3. Sample-Pair Envelope Manifold Neighborhood Concatenation Mechanism (SP_EMNCM)

SP_EMNCM is designed to catch the inherent structure of similar manifold neighboring samples in the sample envelope space. It is known that geodesic distances can reveal similar relationships between samples located on nonlinear manifolds. Therefore, we introduce geodesic distance as the shortest distance into sample envelope space. The geodesic distance between the target sample and all samples is calculated, where the sample with the smallest geodesic distance is called the manifold nearest sample. Then, the original samples are concatenated with the manifold nearest samples to generate sample pairs.

Given a multi-dimensional sample envelope space \mathbb{R}^{N*M} , containing N sample vectors $\{x_1, x_2, ..., x_i, ..., x_n\}$, each sample vector can be expressed as $x_i = \{x_i^{(1)}, x_i^{(2)}, ..., x_i^{(m)}\}$. $y(x_i)$ is the class label of x_i . $d_e(x_i, x_j)$ denotes the Euclidean distance between x_i and x_j . The process of searching for the manifold nearest sample is as follows:

- 1. Divide the sample by class label. If $y(x_i) = y(x_j)$, x_i and x_j are divided into the same sample set.
- 2. Determine the neighborhood relationships. Compute distance matrix D_l from neighborhoods about each sample using Euclidean distance in the same class sample set.
- 3. Compute shortest distances. If the nearest neighbor graph has edge (x_i, x_j) , the shortest distance is $d_e(x_i, x_j)$. Otherwise, $d_g(x_i, x_j) = +\infty$. The shortest distance is defined as:

$$d_g(x_i, x_j) = \begin{cases} d_e(x_i, x_j) & \text{If } x_i \text{ and } x_j \text{ are neighors} \\ min\{d_g(x_i, x_j), d_g(x_i, x_k) + d_g(x_k, x_j)\} & \text{others} \end{cases}$$

Then, obtaining the shortest distance matrix $D_g = [d_g(x_i, x_j)]_{N \times N}$. The sample with the smallest distance value in each row of D_g is retained as the manifold nearest sample.

After the manifold nearest samples are searched, original samples X_i and the manifold nearest samples X_j are concatenated into sample pairs $Xpair = [X_i, X_j]$. The sample pair generation process is shown in the following pseudo code of Algorithm 1 and Figure 5.

Algorithm 1: Sample-pair envelope manifold neighborhood concatenation mechanism

```
Input: Dataset D = (X, Y)
```

```
1: Data set D classified by category according to sample labels
```

- **6:** Select the manifold nearest sample *xj* to *xi*
- 7: Concatenate samples *xi* and *xj* to generate sample pair *xpair*

```
Output: D_{pair} = (X_{pair}, Y_{pair})
```

SP_EMNCM is different from existing methods in that the latter are based on original samples and do not consider structural information between samples. SP_EMNCM combines the closest samples and extracts information between two samples by expanding the features. Figure 6 illustrates the differences between the single-sample-based model and the proposed sample-pair-based model. Seen from Figure 6, the sample-pair-based model not only considers the original samples, but also the manifold neighborhood relationship between the samples.

^{2:} For i = 1 : N

^{3:} Divide the sample by class label

^{4:} Calculate shortest distances between this sample and all other samples of its class

^{5:} Sort all shortest distances thus obtained

^{8:} End For



Figure 5. Sample-pair envelope manifold neighborhood concatenation mechanism (SP_EMNCM).



Figure 6. Sample-pair-based model and single-sample-based model.

3.4. Diamond-like Feature Learning Mechanism (DFLM)

3.4.1. FESSAE-Based Feature-Expansion Mechanism

In order to extend the features, an improved FESSAE was designed. FESSAE is a lightweight deep network.

FESSAE is improved compared to a stacked sparse autoencoder (SSAE). The FESSAE model is shown in Figure 7.

The key element of the FESSAE is the embedding combination between two adjacent autoencoders. Let $H^{(k)} = \begin{bmatrix} h_1^{(k)}, h_2^{(k)}, \dots, h_N^{(k)} \end{bmatrix}$ denote the *k*-th hidden layer's output matrix and $X^{(O)} \in \mathbb{R}^{N*M}$ the sample input to the first layer (original input) of the FESSAE. Firstly, the output of the previous hidden layer $H^{(k-1)}$ is added to the original input $X^{(o)}$ to obtain combined feature as follows:

$$E^{(k)} = \left[(X^{(o)})^T; H^{(k-1)} \right]$$
(1)

Then, the combined feature $E^{(k)}$ is transformed as follows:

$$L(E^{(k)}) = G^T E^{(k)}$$
⁽²⁾

where *G* is the appropriate sparse transformation matrix. The objective function of the feature-embedded unit is defined as follows to filter partially redundant features:

$$\max_{G} tr(G^{T}E^{(k)}(E^{(k)})^{T}G)$$

$$s.t. \quad \sum G_{ii} = d$$
(3)

where *d* is the number of high-quality features retained in feature extraction. After processing by the feature-embedded unit, the *k*-th autoencoder's input data are $L(E^{(k)})$.



Figure 7. FESSAE model.

+

The output data of the hidden layer are divided into two groups, and the ratio of the two groups of features is kept consistent with the ratio of the two types of data of the encoder for that layer. The hidden-layer output features are expressed as $H = [H_{\Gamma 1}, H_{\Gamma 2}]$. The addition group sparsity constraint is represented as follows:

$$\psi(H) = \sum_{g=1}^{2} ||H_{\Gamma_g}^{(k)}||_1 \tag{4}$$

After introducing embedding elements and sparse constraints in the structure during training, the objective function of the *k*-th autoencoder (k > 1) of the FESSAE is expressed as follows:

$$argmin_{\overline{N}} \sum_{i=1}^{N} ||L(E^{(k)}) - L'(E^{(k)})||^{2} \\ -\lambda(||W_{k1}||_{2} + ||W_{k2}||_{2}) + \beta(\sum_{j=1}^{d^{(k)}} KL(\rho ||\hat{\rho}_{j}) + \sum_{g=1}^{2} ||H_{\Gamma_{g}}^{(k)}||_{1})$$
(5)

where $L'(E^{(k)})$ is the *k*-th autoencoder's output. λ and β denote the regularization coefficient and sparsity coefficient, respectively. ρ denotes the sparse parameter. The numerical value increases monotonously as the difference between ρ and ρ_i^{\wedge} increases.

The proposed FESSAE is outlined in Algorithm 2.

Algorithm	2.	FFSSA	F-based	feature	expansion	algorithm
Algonum	∠.	LOSA	E-Daseu	reature	expansion	aigonnum

Input: Samples *X*^(*o*)

1: Set parameters: λ , β , ρ , $d^{(k)}$, number of iterations.

- 2: Pretraining:
- **3:** Train the first layer of FESSAE and extract the hidden layer's output $H^{(1)}$
- 4: For k = 2:K
- **5:** Calculate transformation matrix
- **6:** Calculate the output of the feature-embedded unit $L(E^{(k)})$ by Equations (1)–(2)
- **7:** Train *k*-th layer with objective function in Equation (5)
- 8: Extract output $H^{(k)}$ of *k*-th hidden layer
- 9: End For

10: End Pretraining

- 11: Stack hidden layers and softmax layer
- 12: Fine-tune entire network
- 13: Get the output of the last hidden layer as a deep features
- **Output:** Deep features

3.4.2. L1_wLPPD-Based Feature Reduction Mechanism

The feature-reduction mechanism consists of two stages: L1 regularization and weighted locality preserving discriminant projection.

The first stage used is L1 regularization, a common feature selection method, which obtains a sparse feature vector. The optimization objective function is as follows:

$$\operatorname{argmin}_{\theta} \sum_{i=1}^{N} (y_i - \sum_{m=1}^{\hat{M}} \theta_m x_{im}^{\wedge})^2 + \alpha \sum_{m=1}^{\hat{M}} |\theta_m|$$
(6)

where the first term is original error term and the second is L1 regularization term. α denotes regularization factor. θ_m is *m*-th feature's regression coefficient. Equation (6) is optimized using the proximal gradient-descent method, and each gradient-descent iteration is:

$$\theta^{(k+1)} = \operatorname{argmin}_{\theta} \frac{C}{2} ||\theta - z||_2^2 + \alpha ||\theta||_1$$
(7)

where $z = \theta^{(k)} - C^{-1}\nabla f(\theta^{(k)})$ and $f(\theta^{(k)}) = \sum_{i=1}^{N} (y_i - (\theta^{(k)})^T \hat{x}_i)^2$. *C* is a constant greater than zero, and the components of $\theta = [\theta_1, \theta_2, ..., \theta_m, ..., \theta_{(\hat{M})}]$ are independent of each other. The solution of the L1 regular expression obtained using the iterative soft-threshold function can be expressed as follows:

$$\theta^{(k+1)} = soft_{\alpha C^{-1}}(\theta^{(k)} - C^{-1}\nabla f(\theta^{(k)})) = sign(\theta^{(k)})(|\theta^{(k)}| - \frac{\alpha}{C})$$
(8)

where $sign(\cdot)$ is a symbolic function. The corresponding non-zero component features in $\theta^{(k+1)}$ are selected as the feature subset, that is, features selected by L1 regularization.

The second-stage feature reduction is based on a weighted locality-preserving discriminant projection (wLPPD). \tilde{X} denotes the samples after the first stage of reduction, p is the total number of classes. After sampling, the total number of samples is N_s and the number of samples for the p-th class is N_{sp} . The inter-class variance of the N_s nearest-neighbor

samples of the local sample center of the sample set X is S_B^{\emptyset} , defined as:

$$S_B^{\phi} = \sum_{p=1}^{P} \Delta_B^p \left(\Delta_B^p \right)^T \tag{9}$$

where Δ_B^p is the difference between the local sample center of the *p*-th class and the local sample center, that is, $\Delta_B^p = (1/N_{sp})\sum_{i=1}^{N_{sp}} \tilde{x}_i^{(p)} - (1/N_s)\sum_{i=1}^{N_s} \tilde{x}_i$. The intra-class variance

$$S_W^{\phi} = \sum_{p=1}^{P} \sum_{i=1}^{N_{sp}} \Delta_i^p (\Delta_i^p)^T$$
(10)

where Δ_i^p is the difference between the *i*-th sample of class *p* and the local sample class center of class *p*, that is, $\Delta_i^p = \widetilde{x}_i^{(p)} - (1/N_{sp}) \sum_{i=1}^{N_{sp}} \widetilde{x}_i^{(p)}$.

The regularization term is as follows:

$$J(A) = \sum_{i=1}^{N} \sum_{j=1}^{N} \left\| A^{T} \widetilde{x}_{i} - A^{T} \widetilde{x}_{j} \right\|^{2} W_{ij}$$

= $Tr(\sum_{i=1}^{N} \sum_{j=1}^{N} (2A^{T} \widetilde{x}_{i} \widetilde{x}_{i}^{T} A - 2A^{T} \widetilde{x}_{i} \widetilde{x}_{j}^{T} A) W_{ij})$
= $Tr(\sum_{i=1}^{N} A^{T} \widetilde{x}_{i} D_{ii} \widetilde{x}_{i}^{T} A - \sum_{i=1}^{N} \sum_{j=1}^{N} W_{ij} A^{T} \widetilde{x}_{i} \widetilde{x}_{j}^{T} A)$ (11)

T

where $D_{ii} = \sum_{i=1}^{N} W_{ij}$ is the diagonal matrix. *W* is the affinity matrix. Equation (11) also can be written as $Tr(A^T X (D - W) X^T A)$. Setting L = D - W, the locality preservation regularization term can be expressed as:

$$J(A) = Tr(A^T \widetilde{X} L \widetilde{X}^T A)$$
(12)

Using Equations (9)–(12), the proposed wLPPD can be expressed as:

$$\min_{A} Tr\left(A^{T} S_{W}^{\phi} A\right)$$
s.t.Tr $\left(A^{T} S_{B}^{\phi} A\right) - \gamma J(A) = \kappa I$
(13)

where κ is a constant and γ represents the Lagrange penalty factor. Through adding the Lagrange multiplier η , Equation (13) is rewritten as:

$$L(A,\eta) = Tr\left(A^T S_W^{\phi} A\right) - \eta \left(Tr\left(A^T S_B^{\phi} A\right) - \gamma J(A) - \kappa I\right)$$
(14)

Taking the partial derivative of Equation (14) and setting $\frac{\partial L(A,\eta)}{\partial A} = 0$, the result is:

$$\frac{S_W^{\varphi}A}{(S_B^{\varphi} - \gamma \widetilde{X}L\widetilde{X}^T)} = \eta A \tag{15}$$

After obtaining the projection matrix A, we take the top l eigenvectors corresponding to the biggest eigenvalues of A to obtain projection matrix A_l . Using the LPPD in each subspace separately, we can obtain T projection matrices: $A_{I}^{1}, A_{I}^{2}, \ldots, A_{I}^{T}$. The final projection matrix A_{l}^{E} is obtained through the weighted integration of A_{l}^{t} in each subspace projection transformation.

In addition, an ensemble learning method is adopted for the fusion mechanism. In particular, the sample sampling rate is set to δ_S and the sampling rate of the features to δ_F , and the mixed-feature dataset is sampled according to the bagging strategy q times, forming *q* subsets. The proposed L1_wLPPD staged feature-reduction mechanism is then summarized in Algorithm 3.

Algorithm 3: L1_wLPPD-based feature-reduction algorithm
Input: Data after feature expansion
1: First-stage feature reduction
2: Hybrid feature selection for \hat{X} based on Equations (6)–(8)
3: End first-stage feature reduction
4: Second-stage feature reduction
5: $\delta_S = 0.7, \delta_F = 0.5.$
6: Sample <i>q</i> times to form <i>q</i> subsets.
7: Train <i>q</i> -th SVM:
8: For i = 1:T
9: Choose ns train samples randomly
10: Calculate scatter matrixes S_B^{\varnothing} and S_W^{\varnothing} using Equations (9)–(10)
11: Calculate diagonal matrix D and the Laplacian matrix L

- **12:** Solve mapping matrix **A** using Equation (15)
- 13: End For
- 14: Obtain final mapping matrix A_i^E
- **15:** Map *q*-th subset to train SVM.
- 16: End second-stage feature reduction
- 17: Obtain the ultimate class label
- **Output:** Predicted result

4. Experimental Results and Analysis

Three sets of experiments are carried out to verify the effectiveness of the algorithm in this paper. The first experiment is based on the ablation method to verify the effectiveness of the SP_EMNCM and DFLM in the algorithm and to verify the innovation points of the algorithm. The second experiment compares the proposed algorithm with existing representative feature learning algorithms, representative autoencoders, and the state-ofthe-art algorithms to verify the proposed methods. The third experiment analyzes the effects of some important parameters including coefficient for the L2 weight regularization, sparsity regularization coefficient and sparse parameter, the type of classifier, and the number of classifiers.

4.1. Experimental Conditions

The proposed algorithm's performance is tested on several relevant datasets. Two publicly available chronic disease datasets (Pima Indians Diabetes and Statlog Heart Data Set) are selected for the experiments; they include data on cardiovascular diseases and diabetes, which are the two major chronic diseases [33,34]. The basic information on the datasets is presented in Table 3. The Statlog Heart Data Set and Pima Indians Diabetes are representative datasets for diabetes and heart disease, respectively. In addition, Table 3 includes some other chronic diseases, including Parkinson's and Alzheimer's.

We set three AEs in the proposed FESSAE. The number of hidden units (neurons in hidden layer) is selected by considering dimensional range of the feature vectors for different datasets, and a grid search is used to find the best values. The adjustable parameters in the proposed FESSAE include coefficient for regularization and sparsity parameters. The relevant parameters are listed in Table 4, including three parameters of tuned FESSAE objective function and the number of FESSAE hidden layer's neural units for each dataset.

In the experiments, the K-fold cross-validation technique was employed to evaluate the performance of the SP_DFsaeLA. The programing tool is R2018b MATLAB.

Dataset	Instances	Attributes	Class	Relevant Paper
Statlog Heart Data Set (Heart)	270	13	2	Reference [35]
Pima Indians Diabetes Data Set (PID)	768	8	2	Reference [36]
Parkinson Speech Dataset (PD)	1040	26	2	Reference [37]
Alzheimer's disease (AD)	90	32	3	Reference [38]
Breast Cancer Wisconsin Original (Wisconsin)	683	9	2	Reference [39]
Maxlettle Parkinson Dataset (Maxlettle)	195	22	2	Reference [40]
Statlog Vehicle Silhouettes (Vehicle)	846	18	4	Reference [41]
Breast Cancer Wisconsin Diagnostic(WDBC)	569	30	2	Reference [42]
lung cancer	32	56	3	Reference [43]

Table 3. Basic information of datasets used in study.

Table 4. Parameter information.

Parameter	Parameter Meaning	Parameter Value
λ	Coefficient for the L2 weight regularization term.	$\begin{array}{c} 1 \times 10^{-4}, 1 \times 10^{-3}, \\ 1 \times 10^{-2} \end{array}$
β	Coefficient for the sparsity regularization term.	1, 2, 3, 4, 5, 6
ρ	Sparse parameter.	[0.02, 0.1]
PID hidden units	Number of FESSAE hidden-layer units in PID dataset	120-40-16
Maxlettle hidden units	Number of FESSAE hidden-layer units in Maxlettle dataset	160-80-42
Heart hidden units	Number of FESSAE hidden-layer units in Heart dataset	100-60-24
PD hidden units	Number of FESSAE hidden-layer units in PD dataset	200-100-48
AD hidden units	Number of FESSAE hidden-layer units in AD dataset	240-120-60
Wisconsin hidden units	Number of FESSAE hidden-layer units in Wisconsin dataset	120-40-16
Vehicle hidden units	Number of FESSAE hidden-layer units in Vehicle dataset	120-60-32
WDBC hidden units	Number of FESSAE hidden-layer units in WDBC dataset	200-110-60
lung cancer hidden units	Number of FESSAE hidden-layer units in lung cancer dataset	424-210-106

4.2. Evaluation Criteria

Accuracy (Acc), sensitivity (Sens), precision (Prec), F1_score, and specificity (Spec), which are calculated from the confusion matrix, are chosen as metrics to assess the effectiveness of the model. The confusion matrix or error matrix is a useful tool for visualizing the classifier's whole performance. These datasets used in this study are multi-classified and binary, with a 2 \times 2 confusion matrix for binary and n \times n confusion matrix for multi-classification. The Acc, Sen, Pre, Spec, and F1_score of each category can be expressed as follows:

$$Acc = \frac{TP + TN}{TP + FP + FN + TN}$$
(16)

$$Prec = \frac{TP}{TP + FP} \tag{17}$$

$$Sens = \frac{TP}{TP + FN} \tag{18}$$

$$Spec = \frac{TN}{TN + FP} \tag{19}$$

$$F1_{score} = \frac{2(Prec \times Sens)}{Prec + Sens}$$
(20)

4.3. Ablation Experiments

4.3.1. Effectiveness Analysis of SP_EMNCM

In order to validate the efficacy of SP_EMNCM, we conducted a comparative analysis between single-sample-based model and the sample-pair-based model. The two models are illustrated in Figure 6. To eliminate the influence of other factors, such as different feature learning methods, the two methods are compared directly on the SVM and RF classifiers. Within Tables 5–10, the designation "SS" refers to a single-sample-based model, while "PS" refers to sample-pair-based model. The results are compared in Tables 5 and 6. Tables 7 and 8 compare the results obtained with the LDA. Results with the FESSAE are compared in Tables 9 and 10.

Table 5. Comparison of sample-pair-based and single-sample-based methods on SVM.

Dataset	Method	Acc (%)	Prec (%)	Sens (%)	Spec (%)	F1_Score (%)
DID	SS	74.61	77.09	70.5	78.72	73.44
PID	PS	74.29	79.30	79.48	79.11	79.32
Mailaula	SS	80.25	83.56	76.89	83.56	79.56
Maxiettie	PS	81.24	84.37	79.33	83.33	81.02
Usant	SS	82.92	87.24	77.50	88.33	81.94
Heart	PS	91.67	94.19	89.17	94.17	91.54
רות	SS	65.58	64.99	67.31	63.84	66.02
PD	PS	68.94	69.84	67.31	70.58	68.29
	SS	46.67	48.52	46.67	73.33	47.55
AD	PS	58.89	65.78	58.89	79.44	62.07
Missonsin	SS	96.86	95.54	98.32	95.40	96.91
WISCONSIN	PS	97.70	98.31	97.07	98.32	97.67
Vahiala	SS	82.16	82.23	82.15	94.04	82.19
venicie	PS	84.65	84.76	84.65	94.90	84.71
WDBC	SS	96.94	98.61	95.32	98.59	96.87
WDDC	PS	99.06	100	98.11	100	99.04
lung	SS	52.38	41.67	33.33	63.33	37.04
cancer	PS	91.75	87.78	91.11	93.67	89.33

Table 6. Comparison of sample-pair-based and single-sample-based methods on RF.

Dataset	Method	Acc (%)	Prec (%)	Sens (%)	Spec (%)	F1_Score (%)
DUD	SS	79.47	80.52	77.98	80.98	79.09
PID	PS	84.88	83.90	86.56	83.22	85.14
	SS	87.68	88.01	89.33	86.00	88.08
Maxlettle	PS	92.68	96.36	89.11	96.00	91.87
TT /	SS	87.50	90.57	84.17	90.83	87.17
Heart	PS	95.42	97.53	93.33	97.50	95.31

Method	Acc (%)	Prec (%)	Sens (%)	Spec (%)	F1_Score
55	81 35	81 22	91 73	80.96	81.41
PS	87.69	87.02	89.04	86.35	87.83
SS	50.00	52.04	50.00	75.00	50.87
PS	74.44	75.50	74.44	74.96	74.96
SS	98.53	97.95	99.15	97.91	98.54
PS	99.16	98.77	99.58	98.74	99.17
SS	86.68	86.49	86.68	95.57	86.59

86.68

90.82

96.71

98.12

38.89

91.11

96.95

96.67

98.57

75.00

93.67

Table 6. Cont.

Dataset

PD

AD

Wisconsin

Vehicle

WDBC

lung

cancer

PS

SS

PS

SS

PS

Table 7. Comparison of sample pair and single samples after LDA on SVM.

86.68

90.83

96.70

98.35

66.67

91.74

Dataset	Method	Acc (%)	Prec (%)	Sens (%)	Spec (%)	F1_Score (%)
NID	SS	74.98	77.87	70.16	79.83	73.67
PID	PS	80.21	80.08	80.58	79.85	80.30
Martin	SS	86.46	92.78	79.11	93.78	85.32
Maxiettie	PS	97.00	98.18	96.00	98.00	96.83
TTeent	SS	86.67	90.78	81.67	91.67	85.94
Heart	PS	94.17	95.12	93.33	95.00	94.13
	SS	67.21	66.63	68.84	65.58	67.65
PD	PS	69.81	70.78	68.88	71.73	69.09
	SS	62.22	64.68	62.22	81.11	63.23
AD	PS	74.44	75.49	74.44	87.22	74.95
TA7.	SS	97.28	96.34	98.32	96.22	97.31
vvisconsin	PS	98.54	97.96	99.16	97.91	98.54
Valatala	SS	83.54	83.50	83.53	94.51	83.51
venicie	PS	88.18	88.54	88.17	96.06	88.36
MIDDC	SS	98.56	99.52	97.65	99.52	98.58
WDBC	PS	99.29	100	98.58	100	99.28
lung	SS	57.14	61.11	61.11	78.33	61.11
cancer	PS	93.81	94.44	95.00	97.00	94.66

86.49

90.79

96.84

98.67

50.00

87.77

Fable 8. Comparison	of sample	pair and	l single samp	oles after	r LDA on RF

Dataset	Method	Acc (%)	Prec (%)	Sens (%)	Spec (%)	F1_Score (%)
DID	SS	81.52	82.54	80.22	82.84	82.24.
PID	PS	84.32	82.68	87.30	81.33	84.85
	SS	89.57	91.57	89.11	89.78	89.44
Maxlettle	PS	89.67	94.18	85.33	94.00	89.15
TT /	SS	89.17	92.96	85.00	93.33	88.71
Heart	PS	93.75	95.75	91.67	95.83	93.61
	SS	80.19	81.42	78.26	82.12	79.80
PD	PS	85.09	85.21	85.19	85.01	85.08
	SS	62.22	66.24	62.22	81.11	64.12
AD	PS	55.56	58.60	55.56	77.78	56.94
147	SS	98.32	97.56	99.16	97.48	98.34
vvisconsin	PS	98.95	97.98	100	97.91	98.97
X7.1.1.1	SS	89.19	89.46	89.20	96.39	89.33
venicle	PS	92.57	92.70	92.57	97.53	92.63

90.80

96.76

98.38

43.75

89.33

Dataset	Method	Acc (%)	Prec (%)	Sens (%)	Spec (%)	F1_Score (%)
	SS	98.59	99.53	97.64	99.52	98.57
WDBC	PS	98.59	99.52	97.65	99.52	98.57
lung	SS	83.33	83.33	88.89	91.67	86.02
cancer	PS	94.28	93.33	96.67	96.67	94.92

Table 8. Comparison of sample pair and single samples after LDA on RF.

 Table 9. Comparison of sample pairs and single samples after FESSAE on SVM.

Dataset	Method	Acc (%)	Prec (%)	Sens (%)	Spec (%)	F1_Score (%)
PID	SS	77.44	77.30	79.15	75.77	77.81
	PS	83.80	88.60	79.14	88.48	83.03
Mailaula	SS	99.00	100	98.00	100	98.95
Maxiettle	PS	98.89	100	97.78	100	98.82
TTerest	SS	91.25	91.64	90.83	91.67	91.16
Heart	PS	97.08	97.59	96.67	97.50	97.07
PD	SS	75.00	73.26	79.33	70.67	76.07
	PS	81.49	80.92	82.69	80.29	81.69
	SS	71.11	74.96	71.11	85.56	72.95
AD	PS	75.56	78.70	75.56	87.78	77.07
TA7.	SS	99.16	98.38	100	98.32	99.18
wisconsin	PS	99.58	99.18	100	99.16	99.58
X7.1.1.1.	SS	80.77	82.34	80.78	93.57	81.53
venicie	PS	84.53	85.56	84.54	97.82	85.05
MDDC	SS	97.17	97.80	96.71	97.62	97.18
WDBC	PS	99.76	99.53	100	99.53	99.76
lung	SS	85.71	83.33	91.67	91.67	87.30
cancer	PS	97.14	96.67	98.33	98.33	97.46

 Table 10. Comparison of sample pairs and single samples after FESSAE on RF.

Dataset	Method	Acc (%)	Prec (%)	Sens (%)	Spec (%)	F1_Score (%)
DID	SS	77.44	77.30	79.15	75.77	77.81
PID	PS	83.80	88.60	79.14	88.48	83.03
	SS	92.61	97.50	87.56	97.78	91.89
Maxlettle	PS	96.78	100	93.33	100	96.32
	SS	91.25	91.64	90.83	91.67	91.16
Heart	PS	97.08	97.59	96.67	97.50	97.07
22	SS	75.00	73.26	79.33	70.67	76.07
PD	PS	81.49	80.92	82.69	80.29	81.69
	SS	58.89	60.82	58.89	79.44	59.78
AD	PS	62.22	64.08	62.22	81.11	63.04
TA <i>T</i> :	SS	99.16	98.38	100	98.32	99.18
Wisconsin	PS	99.58	99.17	100	99.16	99.58
37.1.1.1.	SS	80.77	82.34	80.78	93.57	81.54
venicie	PS	84.53	85.57	84.54	94.82	85.05
MIDDC	SS	97.41	97.79	97.18	97.62	97.42
WDBC	PS	99.76	99.53	100	99.53	99.76
lung	SS	50.00	50.00	38.89	75.00	43.75
cancer	PS	87.62	87.78	91.11	93.67	89.33

As seen from Tables 5 and 6, all the sample-pair-based methods performed better than the single-sample-based method for these datasets on the SVM and RF classifiers. This indicates that the SP_EMNCM is more effective than the single-sample-based methods.

As seen from Tables 7 and 8, most of the sample-pair-based methods are better than the single-sample-based results in these databases with the LDA on the SVM and RF classifiers. Thus, in general, SP_EMNCM is effective for manifold feature reduction.

As seen from Tables 9 and 10, most of the sample pair results in the dataset are better than single sample results on the SVM and RF classifiers. Thus, in general, SP_EMNCM is effective under deep learning.

4.3.2. Effectiveness Analysis of DFLM

Verification of the performance of the different steps in DFLM is conducted by ablation experiments. First, the sample pairs are input to the classification model as original features (OFs) input. Second, the samples are processed with the FESSAE and the OFs converted into deep features (DFs). Third, OFs and DFs are combined to construct new features (CF). The CF is processed by the L1_wLPPD to become the reduced features (CF and DFLM). Different feature sets are used with the same classifier and dataset. A comparison of the results is shown in Table 11.

Dataset	Method	Acc (%)	Prec (%)	Sens (%)	Spec (%)	F1_Score (%)
	OFs	79.29	79.30	79.48	79.11	79.32
DUD	DFs	83.80	88.60	79.14	88.48	83.03
PID	CF	82.31	86.97	77.66	86.97	81.38
	CF and DFLM	84.54	88.50	80.99	88.11	84.18
	OFs	91.67	94.19	89.17	94.17	91.54
TTt	DFs	97.08	97.59	96.67	97.50	97.07
Heart	CF	96.25	97.51	95.00	97.50	96.13
	CF and DFLM	96.67	97.59	95.83	97.50	96.65
	OFs	68.94	69.84	67.31	70.58	68.29
רות	DFs	81.49	80.92	82.69	80.29	81.69
PD	CF	78.73	81.36	75.96	81.49	77.53
	CF and DFLM	79.93	80.05	79.81	80.05	79.60
	OFs	58.89	65.78	58.89	79.44	62.07
	DFs	75.56	78.70	75.56	87.78	77.07
AD	CF	68.89	69.97	68.89	84.44	69.41
	CF and DFLM	72.22	74.55	72.22	86.11	73.36
	OFs	97.70	98.31	97.07	98.32	97.67
TA7	DFs	99.58	99.18	100	99.16	99.58
vvisconsin	CF	98.95	98.76	99.16	98.74	98.95
	CF and DFLM	99.37	98.77	100	98.74	99.58
	OFs	84.65	84.76	84.65	94.90	84.71
\$7.1 * 1	DFs	84.53	85.56	84.54	97.82	85.05
venicle	CF	85.54	86.37	85.55	95.18	85.95
	CF and DFLM	87.80	88.73	87.82	95.93	88.27
	OFs	99.06	100	98.11	100	99.04
WDBC	DFs	99.76	99.53	100	99.53	99.76
WDBC	CF	98.82	99.09	98.58	99.05	98.82
	CF and DFLM	99.53	100	99.53	100	99.53
	OFs	81.24	84.37	79.33	83.33	81.02
Maulaula	DFs	98.89	100	97.78	100	98.82
Maxiettie	CF	97.84	100	95.56	100	97.65
	CF and DFLM	97.89	100	95.56	100	97.50
	OFs	91.75	87.78	91.11	93.67	89.33
lung concor	DFs	97.14	96.67	98.33	98.33	97.46
lung cancer	CF	87.62	87.78	91.11	93.67	89.40
	CF and DFLM	95.24	91.67	96.67	95.83	93.84

Table 11. Result of principal stages of the proposed method.

Table 11 shows that in some cases the CF and DFLM has better results. This result indicates that the proposed diamond-like feature learning mechanism (DFLM) is effective. The accuracy of DFs also performs well. The results show that FESSAE is effective, and it can learn as many high-quality features as possible and expand features well. The fact that

the CF and DFLM performs better than the CF means that the proposed feature reduction mechanism is effective. The accuracy of the CF does not improve as much as that of the OFs. We hypothesize that the reason is that the simple combination of the OFs and DFs leads to high redundancy. Therefore, it needs to consider feature reduction. Different datasets have different results, due to characteristics of data and disease. For example, most metrics of PID, Wisconsin and Heart datasets have higher results in CF and DFLM but, for the AD, PD, lung cancer, and Maxlettle datasets, the metrics have higher results in DFs in most cases. We hypothesize that this is due to the former's sample size being medium-sized with dimensionality of 200 to 768, but the number of features is small. While the latter's features are medium-scale with dimension of 20 to 35, the number of samples is large or small. The difference between the number of samples and features possibly is one of the reasons for the different performance of the metrics in different datasets and may also be related to the internal structure of the data.

To visualize the DFLM, the features extracted at the different stages are shown in Figure 8. Here, four datasets (PID, Heart, WDBC, and Wisconsin) are used.



Figure 8. Variation in the number of features in DFLM: (a) PID, (b) Heart, (c) WDBC, and (d) Wisconsin databases.

As shown in Figure 8, the change in the number of features tends to increase first and then decrease as the stages of the proposed method progress. The distribution of number of features in the stages appear as a diamond topology. The number of features increased after combining the OFs and DFs. After the first L1_wLPPD stage, the expanded features are reduced. The second L1_wLPPD stage further reduces the number of features. Finally, high-quality features are obtained for small or medium size of samples and features.

4.4. Algorithm Comparison

4.4.1. Comparison with Typical Feature Learning Algorithms

In order to evaluate the performance of the SP_DFsaeLA, its performance is compared to those of feature learning algorithms that are representative. These included both feature

selection methods, such as least absolute shrinkage and selection operator (LASSO) [44] and Relief [45], as well as feature extraction algorithms, such as PCA [46], LDA [47], and locality preserving projection (LPP) [48]. The SVM is used as the base classifier, since it is a commonly used classifier. The results are listed in Table 12.

Dataset	Performance Indices	OFs (%)	PCA (%)	LDA (%)	LPP (%)	Relief (%)	LASSO (%)	SP_DFsaeLA (Proposed) (%)
	Acc	79.29	80.03	80.21	80.03	78.17	77.59	84.54
	Prec	79.30	79.82	80.09	79.38	78.30	77.76	88.50
PID	Sens	79.48	80.59	80.58	81.33	78.34	77.59	80.99
	Spec	79.11	79.48	79.85	78.74	77.99	77.60	88.11
	F1_score	79.32	80.13	80.30	80.25	78.24	77.67	84.18
	Acc	81.24	91.73	97.00	95.84	90.62	84.52	97.89
	Prec	84.37	91.33	94.85	96.37	88.32	91.86	100
Maxlettle	Sens	79.33	93.78	100	95.78	93.78	77.33	95.56
	Spec	83.33	89.78	94.00	96.00	87.56	91.78	100
	F1_score	81.02	91.91	97.23	95.87	90.92	82.65	97.50
	Acc	91.67	95.42	94.17	94.58	93.33	88.75	96.67
	Prec	94.19	95.93	95.12	94.36	94.85	89.54	97.59
Heart	Sens	89.17	95.00	93.33	95.00	91.67	88.33	95.83
110011	Spec	94.17	95.83	95.00	94.17	95.00	89.17	97.50
	F1_score	91.54	95.39	94.12	94.61	93.19	88.76	96.65
	Acc	99.06	99 29	99 29	99.29	99 29	98 11	99 53
	Prec	100	100	100	100	99.53	99.05	100
WDBC	Sens	98.11	98.58	98.58	98.57	99.05	97.17	99.53
WDDC	Spec	100	100	100	100	99.52	99.06	100
	F1_score	99.04	99.28	99.28	99.27	99.28	98.09	99.53
	Acc	68 94	71.25	69.90	70 58	71.25	67 21	79 93
	Prec	69.84	71.29	70.80	70.30	71.23	66.67	80.05
רוק	Sens	67 31	70.19	68.27	69.42	70.38	68.65	79.81
ТD	Spec	70.58	72.31	71.54	71 73	72.12	65.77	80.05
	F1 score	68.29	70.88	69.26	70.12	70.86	67.49	79.60
		EQ 90	71 11	72.22	66.67	68.80	62.22	72.00
	Acc Proc	50.09 65 78	71.11	72.22 80.00	66.82	00.09 72.70	62.22	72.22
	Free	59.70 59.90	74.01	80.00 72.22	66.67	73.70	60.30	74.34
AD	Space	70.44	71.11	72.22 86.11	82.22	84.44	02.22 81.11	96 11
	F1 score	79.44 62.07	72 00	75 91	66 75	04.44 71 19	64.07	73 36
	I'I_SCOLE	02.07	72.00	75.91	00.75	71.19	04.07	75.50
	Acc	84.65	87.55	87.15	87.44	86.79	82.77	87.80
	Prec	84.76	87.61	87.56	87.46	87.20	83.20	88.73
Vehicle	Sens	84.65	87.55	87.17	87.44	86.78	82.77	87.82
	Spec	94.89	95.86	95.05	95.14	95.60	94.25	95.93
	F1_score	84.71	87.58	87.36	87.45	86.99	82.99	88.27
	Acc	97.70	98.53	98.54	98.54	98.11	97.70	99.37
	Prec	98.31	97.55	98.34	97.94	97.58	97.52	98.77
Wisconsin	Sens	97.06	99.57	98.74	99.16	98.74	97.91	100
	Spec	98.32	97.49	98.32	97.91	97.49	97.49	98.74
	F1_score	97.67	98.55	98.53	98.54	98.13	97.70	99.38
	Acc	91.75	91.75	94.28	84.65	91.42	88.57	95.24
1	Prec	87.78	87.78	94.44	84.76	92.22	90.00	91.67
lung	Sens	91.11	91.11	96.11	84.65	92.22	90.00	96.67
can-cer	Spec	93.67	93.67	97.00	94.89	95.67	94.33	95.83
	F1_score	89.33	89.33	95.24	84.71	92.22	90.00	93.84

 Table 12. Comparison of typical feature learning algorithms.

The results in Table 12 show that the SP_DFsaeLA compared with the other featurelearning algorithms has superior performance, can obtain better quality features than other feature learning algorithms, and can considerably improve the classification accuracy. The first possible reason is that the compared methods are based on the original features, but the proposed algorithm expands the original features, thereby obtaining higher-quality features efficiently. The second possible reason is the high effectiveness of the two-stage feature-reduction mechanism.

4.4.2. Comparison with Representative Stacked Autoencoders

To verify the accuracy of FESSAE, we experimentally compared its performance to that of some representative stacked autoencoders that are representative, including the stacked autoencoder (SAE) [49], stacked sparse autoencoder (SSAE) [50], stacked denoising autoencoder (SDAE) [51], stacked pruning sparse autoencoder (SPSAE) [52], and SSAE combined with LASSO (SSAE and LASSO) [53]. The performance of FESSAE without sparse term is verified, and the method is referred to as ESAE. There are some main reasons why the autoencoder, so representative autoencoders are compared for fairness; (2) the autoencoder is a kind of deep network, which is more suitable for datasets with a small number of samples and features compared with other kinds of deep learning methods.

Since the number of the samples and features is small or medium, the number of hidden layers and neurons cannot be large. For a fair comparison, the SP_EMNCM is not considered and all parameters are set to the same parameters for all the methods. The observed accuracy is presented in Table 13. Three datasets (PD, AD, and Vehicle) are used.

Table 13. Classification accuracy of different deep autoencoder classifiers.

Dataset	SAE [49] (%)	SSAE [50] (%)	SDAE [51] (%)	SPSAE [52] (%)	SSAE and LASSO [53] (%)	ESAE (%)	Proposed FESSAE (%)
PD	64.15	66.48	66.48	66.22	65.87	60.29	75.67
AD	57.67	61.67	59.58	61.78	57.66	60.94	71.11
Vehicle	67.30	70.00	72.00	74.76	80.06	74.89	80.77

As seen from Table 13, the FESSAE achieved the highest accuracy, showing that the proposed FESSAE is effective. Table 13 shows that the sparse constraints can substantially enhance the classifier accuracy. For all datasets, the proposed FESSAE achieved the best classification accuracy.

4.4.3. Comparison with Recent Chronic Disease Detection Algorithms

To further verify the accuracy of SP_DFsaeLA algorithm, its performance is compared to those of state-of-the-art algorithms proposed in recent years for classifying chronic diseases, including those by Hasan et al. [54], Wang et al. [55], Guia et al. [56], Hasan et al. [57], Lu H et al. [21], and Taghizadeh E et al. [22], Abdollahi J et al. [30]. For a more visual representation of the comparative algorithm information, we describe the state-of-the-art (SOTA) studies with their pros and cons in Table 14. Tables 15 and 16 show that all the models perform better in either positive or negative chronic disease prediction; the two cross-validation methods (CV, including k-fold and Holdout) are used for fair comparison. As seen from the table, the SP_DFsaeLA is better than the others in precision or F1_score or both. The cases are similar under fivefold and Hold out CV.

Literature	Method	Pros	Cons
Hasan M K et al. [54]	A framework for diabetes prediction is proposed in which data standardization, feature selection, and different Machine Learning classifiers.	High robustness.	Requires significant time to search for the best combination.
Wang Q et al. [55]	An effective prediction algorithm for diabetes is proposed, which include Naïve Bayes method, adaptive synthetic sampling method and random forest.	Relatively simple and computationally efficient.	Low universality.
Guia J et al. [56]	Six classification algorithms were used to predict chronic kidney disease, including support vector classifier, decision trees, random forest, Gaussian Naive Bayes, Multilayer Perceptron, and K-Nearest neighbors.	Saving computational overhead.	Classification performance is dependent on tuning parameters.
Hasan K A et al. [57]	Using logistic regression, analysis of variance and to identify by using multiple supervised machine learning algorithms.	Ability to select the most significant risk factors associated with diabetes.	Limited improvement in classification accuracy.
Lu H et al. [21]	A patient network and machine learning approach is developed to combine the attributes of the patient network with the sample features and eight machine learning models are used to predict disease.	Discover the potential characteristics of the patient.	Classification performance is too dependent on the dataset.
Taghizadeh E et al. [22]	Three groups of machine learning algorithms were employed: four feature selection procedures are employed and compared to select the most valuable feature; Principal Component Analysis, 13 classification algorithms accompanied with automated hyperparameter tuning.	Effectively reducing the number of features.	The sample size was relatively large.
Abdollahi J et al. [30]	Selection of 10 machine learning algorithms as the basic algorithms in a stack generalization algorithm to predict chronic diseases and implementation of a hybrid meta-algorithm for prediction.	High classification accuracy and interpretability.	Only original features are operated on. When the quality of the original feature is not high, it does not work well.

Table 14. SOTA studies with their pros and cons.

Table 15. Performance comparison of proposed method with state-of-the-art chronic disease detection algorithms.

Dataset	Performance Indices	Literature [54] (5-Fold)	Literature [55] (5-Fold)	Literature [56] (Holdout)	Literature [57] (5-Fold)	Proposed Method (5-Fold)	Proposed Method (Holdout)
	Acc (%)	88.8	87.1	80.04	75.71	84.54	83.18
	Sens (%)	78.9	85.4	73.51	-	80.99	74.07
PID	Spec (%)	93.4	-	-	-	88.11	92.45
	Prec (%)	84.2	80.6	84.93	-	88.50	90.91
	F1_score (%)	-	82.9	78.80	-	84.18	81.63

Dataset	Performance Indices	Literature [54] (5-Fold)	Literature [55] (5-Fold)	Literature [56] (Holdout)	Literature [57] (5-Fold)	Proposed Method (5-Fold)	Proposed Method (Holdout)
	Acc (%)	97.3	95.6	97.65	95.61	99.76	98.82
	Sens (%)	83.0	95.1	95.35	-	99.53	100
WDBC	Spec (%)	95.7	-	-	-	100	98.82
	Prec (%)	87.8	96.0	100	-	100	97.67
	F1_score (%)	-	95.5	97.62	-	99.76	98.82
	Acc (%)	98.1	97.1	96.84	96.79	99.37	98.95
	Sens (%)	53.3	97.1	97.87	-	100	100
Wisconsin	Spec (%)	99.5	-	-	-	98.74	97.87
	Prec (%)	86.7	97.2	95.83	-	98.77	97.96
	F1_score (%)	-	97.1	96.84	-	99.38	98.97
	Acc (%)	73.9	86.7	85.42	82.96	96.67	95.83
	Sens (%)	74.8	81.7	75.00	-	95.83	100
Heart	Spec (%)	73.0	-	-	-	97.50	91.67
	Prec (%)	69.7	90.8	94.74	-	97.59	92.31
	F1_score (%)	-	86.0	83.72	-	96.65	96.00

Table 15. Cont.

Table 16. Comparison of proposed method with state-of-the-art chronic disease detection algorithms.

Dataset	Performance Indices	Literature [21] (5-Fold)	Literature [30] (5-Fold)	Literature [22] (5-Fold)	Proposed Method (5-Fold)
	Acc (%)	79.38	81.13	80.19	84.54
	Sens (%)	87.50	83.02	79.25	80.99
PID	Spec (%)	71.25	79.25	81.13	88.11
	Prec (%)	75.27	80.00	80.77	88.50
	F1_score (%)	80.92	81.48	80.00	84.18
	Acc (%)	96.85	98.82	95.29	99.76
	Sens (%)	96.83	100	92.86	99.53
WDBC	Spec (%)	96.88	97.67	97.67	100
	Prec (%)	96.83	97.67	97.50	100
	F1_score (%)	96.83	98.82	95.12	99.76
	Acc (%)	94.41	97.89	95.79	99.37
	Sens (%)	94.37	100	95.74	100
Wisconsin	Spec (%)	94.44	95.74	95.83	98.74
	Prec (%)	94.37	96.00	95.74	98.77
	F1_score (%)	94.37	97.86	96.00	99.38
	Acc (%)	83.33	95.83	85.42	96.67
	Sens (%)	91.67	100	83.33	95.83
Heart	Spec (%)	75.00	91.67	87.50	97.50
	Prec (%)	78.57	92.31	86.96	97.59
	F1_score (%)	84.62	96.00	85.00	96.65

4.5. Parameter Analysis

4.5.1. Analysis of FESSAE Parameters

The FESSAE model is a critical part of the SP_DFsaeLA. Therefore, it is critical to assess the impact of different FESSAE parameter settings on overall performance. First, the effect of the FESSAE's sparsity parameter is discussed. The sparsity parameter, ranging from 0.02 to 0.1, is chosen based on previous research.

As seen from Figure 9, different sparse parameters give different results. The sparse parameter had an apparent impact for the FESSAE algorithm's accuracy. The optimal sparse parameter is different for different datasets, and there is no fixed selection criterion.



The datasets used have different numbers of samples, feature dimensions, and numbers of categories, and the optimal sparse parameter may be related to these factors.

Figure 9. Effect of sparse proportion on FESSAE with (**a**) PID; (**b**) Heart; (**c**) WDBC; and (**d**) Wisconsin datasets.

The effects of the λ in the range of 10^{-5} to 10^{-3} and β in the range of 1 to 6 on the performance of the proposed FESSAE are analyzed together. Figure 10 shows that, when the λ is fixed, the β had relatively little effect on the network. When the β is fixed, the closer the λ is to 0, the more robust is the model.





Figure 10. Effect of penalty term coefficients on FESSAE performance: (**a**) single sample of PID; (**b**) sample pair of PID; (**c**) single sample of Heart; (**d**) sample pair of Heart.

4.5.2. Analysis of Classifier Type

The impact of different classifiers (extreme learning machines (ELM), SVM, and RF) on the proposed algorithm's performance is experimentally studied. The results are listed in Table 17.

Table 17. Classification accurac	y of	proposed	l a	lgorithm	with	different	classifiers
----------------------------------	------	----------	-----	----------	------	-----------	-------------

Dataset —	SVM (%)				RF (%)		ELM (%)			
	Acc	Spec	Sens	Acc	Spec	Sens	Acc	Spec	Sens	
PID	84.54	88.11	80.99	83.61	85.85	81.38	80.83	92.54	69.14	
Maxlettle	100	100	100	98.89	100	97.78	99.00	98.00	100	
Heart	96.67	97.50	95.83	96.67	97.50	95.83	96.67	97.50	95.83	
PD	81.73	81.35	82.12	81.44	78.27	84.62	80.67	80.00	81.35	
AD	71.11	85.56	71.11	61.11	80.56	61.11	67.78	83.89	67.78	
Wisconsin	99.37	98.74	100	99.16	98.32	100	99.16	99.15	99.15	
Vehicle	87.80	95.93	87.82	87.06	95.67	87.06	87.55	95.16	87.55	
WDBC	99.76	100	99.53	97.41	97.63	97.16	98.82	98.58	99.06	

Table 17 shows that, for five of the eight datasets, the highest accuracy, sensitivity, and specificity are achieved with the SVM. For the AD dataset, classification accuracy obtained using SVM is 71.11%, which is 10 and 3.33% higher than those with RF and ELM, respectively, and the sensitivity is 71.11%, which is 10 and 3.33% higher than those with RF and ELM. In addition, the specificity is 85.56%, which is 5% and 1.67% better than those with RF and ELM, and it had better stability, since the SVM is most accurate in most cases. Overall, the SVM outperformed the RF and ELM.

4.5.3. Analysis of the Number of Classifiers

To verify whether subclassification has an effect on the performance, we designed experiments with different numbers of sub-classifiers and the results are shown in Figure 11. From Figure 11 it is clear that the accuracy of the different numbers of sub-classifiers did not differ significantly.



Figure 11. Effect of the number of classifiers on accuracy.

5. Discussion and Conclusions

In recent years, chronic diseases have become a serious threat to human health. Currently, a practical approach is to first use wearable sensors to collect data from the human body and then process the data using machine learning. Machine learning algorithm is an effective tool for the analysis of sensing data of chronic disease. The machine learning method mainly includes two parts: feature learning and classifier, where feature learning is very important. Therefore, it is important and challenging to study highly efficient feature learning methods for sensor monitoring of chronic disease. Traditional feature learning methods are restricted from original features and cannot construct high-quality features, whereas deep feature learning methods suffer from the small-sample problems.

To overcome these limitations, this paper proposed a solution—sample-pair envelope diamond stacked sparse autoencoder ensemble learning algorithm (SP_DFsaeLA). First, the sample-pair envelope manifold neighborhood concatenation mechanism (SP_EMNCM) is designed by searching the manifold nearest samples in manifold neighborhood and generating sample pairs. Second, the feature embedding stacked sparse autoencoder (FESSAE) is designed to extend features. Third, a staged feature reduction mechanism is designed to reduce extended feature redundancy. The first stage uses the L1 regularized feature-reduction algorithm and the second stage uses an improved manifold dimensionality reduction algorithm to further reduce features. Fourth, the sample-pair-based model and single-sample-based model are combined by weighted fusion.

In the experimental section, three sets of experiments are organized to verify the proposed method's effectiveness to validate the effectiveness of the innovation points and to research the algorithm's parameters. The results show that the SP_EMNCM is valid apparently. Compared with feature extraction methods, such as LDA, the accuracy of SP_DFsaeLA's classification recognition is higher because the chronic disease samples are often random. However, LDA is not sensitive to non-Gaussian-distributed samples. Compared with feature selection algorithms, including Relief, the advantages of SP_DFsaeLA are its greater learnability and lower tolerance to faults. Relief gives higher weights to all features that are correlated with the class, so the limitation of Relief is in that it does not effectively remove redundant features. Compared with deep learning methods (SAEs), the accuracy of SP_DFsaeLA is improved by up to 20.5%. However, SAE only considers the feature extraction without considering the intrinsic structural information between samples.

As mentioned previously, effective feature learning of sensor data for chronic diseases is important and challenging, and there are some contributions for solving the problem. Firstly, a diamond-like feature learning mechanism is proposed in this paper. This mechanism combines deep learning and traditional feature-reduction methods' advantages and can better adapt to the data characteristics of chronic diseases and the requirement for high recognition accuracy than the deep learning and traditional feature-reduction methods. Secondly, this paper proposes a sample-pair envelope manifold neighborhood concatenation mechanism (SP_EMNCM), which has the advantage of enriching the feature information. Thirdly, this paper designs lightweight FESSAE, which has the advantage of improving the complementarity of original features with deep features and achieving high-quality deep features in small or medium sample sizes. Fourth, a feature expansion mechanism is proposed by combining SP_EMNCM and FESSAE, which has the advantage of improving the original features richer and better by considering the sample structure relationship and feature complementary relationship. Fifth, a two-stage feature reduction mechanism for L1_wLPPD is proposed, and the advantages of this method are making the features more compact without losing high-quality features.

As far as we know, the proposed method is designed for feature learning of chronic diseases' sensor data, and no similar published reports have been found. In addition, the SP_DFsaeLA is a framework approach, which is more inclusive of concrete algorithms. The proposed method can produce different variations by using different feature learning methods and different classifiers. Therefore, the SP_DFsaeLA has good generalization. There are limited types of the feature learning algorithms, but they are representative and some of them had been used for sensing of chronic diseases. Compared with other deep learning algorithms, the autoencoder has lightweight structure and parameters. Therefore, it is discussed in this paper as a representative deep neural network. That is a major reason why only autoencoder rather than other deep learning methods is involved.

Although the proposed method is effective, it has limitations for datasets with a large number of samples and features. In future work, other types of deep neural networks can be considered for further verification and improvement. In addition, the proposed algorithm can be validated on more datasets and embedded in portable systems for practical diagnosis of chronic diseases.

Author Contributions: Conceptualization, Y.L.; methodology, X.Q. and Y.L.; supervision, X.Q. and Z.Z.; software, formal analysis, writing—original draft preparation, Y.Z.; data acquisition, Z.Z. and J.M.; investigation and visualization, J.M. writing—reviewing and editing, Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: No new data were created.

Acknowledgments: We thank the support of NSFC (No. U21A20448, 61771080); this work was supported by Sichuan Science and Technology Program (No. 2019ZDZX0006) and the Talents by Sichuan provincial Party Committee Organization Department, and Science and Technology Service Network Initiative (KFJ-STS-QYZD-2021-21-001).

Conflicts of Interest: The authors declare no conflict of interest pertaining to this work.

References

- Alhassan, A.M.; Zainon, W.M.N.W. Review of feature selection, dimensionality reduction and classification for chronic disease diagnosis. *IEEE Access* 2021, 9, 87310–87317. [CrossRef]
- Yin, H.; Jha, N.K. A health decision support system for disease diagnosis based on wearable medical sensors and machine learning ensembles. *IEEE Trans. Multi Scale Comput. Syst.* 2017, 3, 228–241. [CrossRef]
- Wu, J.; Chang, L.; Yu, G. Effective data decision-making and transmission system based on mobile health for chronic disease management in the elderly. *IEEE Syst. J.* 2020, 15, 5537–5548. [CrossRef]
- 4. Muzammal, M.; Talat, R.; Sodhro, A.H.; Pirbhulal, S. A multi-sensor data fusion enabled ensemble approach for medical data from body sensor networks. *Inf. Fusion* **2020**, *53*, 155–164. [CrossRef]
- Abreu, P.H.; Santos, M.S.; Abreu, M.H.; Andrade, B.; Silva, D.C. Predicting breast cancer recurrence using machine learning techniques: A systematic review. ACM Comput. Surv. 2016, 49, 1–40. [CrossRef]

- Gunarathne, W.; Perera, K.; Kahandawaarachchi, K. Performance evaluation on machine learning classification techniques for disease classification and forecasting through data analytics for chronic kidney disease (CKD). In Proceedings of the 2017 IEEE 17th International Conference on Bioinformatics and Bioengineering, Washington, DC, USA, 23–25 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 291–296.
- Yildirim, P. Chronic kidney disease prediction on imbalanced data by multilayer perceptron: Chronic kidney disease prediction. In Proceedings of the 41st IEEE Computer Software and Applications Conference, Turin, Italy, 4–8 July 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 193–198.
- 8. Zou, Q.; Qu, K.; Luo, Y.; Yin, D.; Ju, Y.; Tang, H. Predicting diabetes mellitus with machine learning techniques. *Front. Genet.* **2018**, *9*, 515. [CrossRef]
- Rubini, L.J.; Eswaran, D.P. Generating comparative analysis of early stage prediction of chronic kidney disease. *Int. J. Mod. Eng. Res.* 2015, 5, 49–55.
- 10. Sinha, P.; Sinha, P. Comparative study of chronic kidney disease prediction using KNN and SVM. *Int. J. Eng. Res. Technol.* **2015**, *4*, 608–612.
- Ekanayake, I.U.; Herath, D. Chronic kidney disease prediction using machine learning methods. In Proceedings of the 2020 Moratuwa Engineering Research Conference, Moratuwa, Sri Lanka, 28–30 July 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 260–265.
- 12. Ahmed, H.; Younis, E.M.; Hendawi, A.; Ali, A.A. Heart disease identification from patients' social posts, machine learning solution on spark. *Future Gener. Comput. Syst.* 2020, 111, 714–722. [CrossRef]
- 13. Shrivas, A.K.; Sahu, S.K.; Hota, H.S. Classification of chronic kidney disease with proposed union based feature selection technique. *Soc. Sci. Res. Netw. Electron. J.* **2018**, *4*, 26–27. [CrossRef]
- Chormunge, S.; Jena, S. Correlation based feature selection with clustering for high dimensional data. J. Electr. Syst. Inf. Technol. 2018, 5, 542–549. [CrossRef]
- Sawhney, R.; Mathur, P.; Shankar, R. A firefly algorithm based wrapper-penalty feature selection method for cancer diagnosis. In Proceedings of the 18th International Conference on Computational Science and Its Applications, Melbourne, VIC, Australia, 2–5 July 2018; Springer: Berlin/Heidelberg, Germany, 2018; pp. 438–449.
- 16. Jayaraman, V.; Sultana, H.P. Artificial gravitational cuckoo search algorithm along with particle bee optimized associative memory neural network for feature selection in heart disease classification. *J. Ambient. Intell. Humaniz. Comput.* **2019**, 1–10. [CrossRef]
- Paul, A.K.; Shill, P.C.; Rabin, M.R.I.; Murase, K. Adaptive weighted fuzzy rule-based system for the risk level assessment of heart disease. *Appl. Intell.* 2018, 48, 1739–1756. [CrossRef]
- 18. Rasitha, G. Predicting thyroid disease using linear discriminant analysis (LDA) data mining technique. *Int. J. Mod. Trends Eng. Res.* **2016**, *4*, 4–6. [CrossRef]
- 19. Mohamed, E.I.; Linder, R.; Perriello, G.; Di Daniele, N.; Pöppl, S.J.; De Lorenzo, A. Predicting type 2 diabetes using an electronic nose-based artificial neural network analysis. *Diabetes Nutr. Metab.* **2002**, *15*, 215–221. [PubMed]
- Shahbazi, F.; Asl, B.M. Generalized discriminant analysis for congestive heart failure risk assessment based on long-term heart rate variability. *Comput. Methods Programs Biomed.* 2015, 122, 191–198. [CrossRef]
- 21. Lu, H.; Uddin, S.; Hajati, F.; Moni, M.A.; Khushi, M. A patient network-based machine learning model for disease prediction: The case of type 2 diabetes mellitus. *Appl. Intell.* 2022, 52, 2411–2422. [CrossRef]
- 22. Taghizadeh, E.; Heydarheydari, S.; Saberi, A.; JafarpoorNesheli, S.; Rezaeijo, S.M. Breast cancer prediction with transcriptome profiling using feature selection and machine learning methods. *BMC Bioinform.* **2022**, *23*, 410. [CrossRef]
- 23. Khan, A.; Uddin, S.; Srinivasan, U. Comorbidity network for chronic disease: A novel approach to understand type 2 diabetes progression. *Int. J. Med. Inform.* **2018**, *115*, 1–9. [CrossRef]
- Ge, R.; Zhang, R.; Wu, Q.; Wang, P. Prediction of chronic diseases with multi-label neural network. *IEEE Access* 2020, 127, 24–25. [CrossRef]
- El-Baz, A.H. Hybrid intelligent system-based rough set and ensemble classifier for breast cancer diagnosis. *Neural Comput. Appl.* 2015, 26, 437–446. [CrossRef]
- 26. Polat, K. Similarity-based attribute weighting methods via clustering algorithms in the classification of imbalanced medical datasets. *Neural Comput. Appl.* **2018**, *30*, 987–1013. [CrossRef]
- 27. Cheruku, R.; Edla, D.R.; Kuppili, V.; Dharavath, R. RST-BatMiner: A fuzzy rule miner integrating rough set feature selection and bat optimization for detection of diabetes disease. *Appl. Soft Comput.* **2017**, *67*, 764–780. [CrossRef]
- Maniruzzaman; Kumar, N.; Abedin, M.; Islam, S.; Suri, H.S.; El-Baz, A.S.; Suri, J.S. Comparative approaches for classification of diabetes mellitus data: Machine learning paradigm. *Comput. Methods Programs Biomed.* 2017, 152, 23–34. [CrossRef]
- 29. Alhassan, A.M.; Wan Zainon, W.M.N. Taylor bird swarm algorithm based on deep belief network for heart disease diagnosis. *Appl. Sci.* **2020**, *10*, 6626. [CrossRef]
- 30. Abdollahi, J.; Nouri-Moghaddam, B.; Ghazanfari, M. Deep Neural Network Based Ensemble learning Algorithms for the healthcare system (diagnosis of chronic diseases). *arXiv* 2021, arXiv:2103.08182.
- Fatan, M.; Hosseinzadeh, M.; Askari, D.; Sheikhi, H.; Rezaeijo, S.M.; Salmanpour, M.R. Fusion-based head and neck tumor segmentation and survival prediction using robust deep learning techniques and advanced hybrid machine learning systems. In Head and Neck Tumor Segmentation and Outcome Prediction: Second Challenge, Proceedings of the HECKTOR 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, 27 September 2021; Springer International Publishing: Cham, Switzerland, 2022; pp. 211–223.

- 32. Rezaeijo, S.M.; Hashemi, B.; Mofid, B.; Bakhshandeh, M.; Mahdavi, A.; Hashemi, M.S. The feasibility of a dose painting procedure to treat prostate cancer based on mpMR images and hierarchical clustering. *Radiat. Oncol.* **2021**, *16*, 182. [CrossRef] [PubMed]
- 33. Hegde, S.; Mundada, M.R. Early prediction of chronic disease using an efficient machine learning algorithm through adaptive probabilistic divergence based feature selection approach. *Int. J. Pervasive Comput. Commun.* **2020**, 17, 20–36. [CrossRef]
- 34. Simon, D.S.; Fraser Paul, J. Kidney disease in the global burden of disease study 2017. Nat. Rev. Nephrol. 2019, 15, 193.
- Lichman, M. UCI Machine Learning Repository. 2013. Available online: http://archive.ics.uci.edu/ml (accessed on 1 January 2022).
- Smith, J.W.; Everhart, J.E.; Dickson, W.C.; Knowler, W.C.; Johannes, R.S. Using the ADAP learning algorithm to forecast the onset of diabetes mellitus. In Proceedings of the Symposium on Computer Applications in Medical Care, Washington, DC, USA, 6–9 November 1988; Volume 10, pp. 261–265.
- Sakar, B.E.; Isenkul, M.E.; Sakar, C.O.; Sertbas, A.; Gurgen, F.; Delil, S.; Apaydin, H.; Kursun, O. Collection and analysis of a Parkinson speech dataset with multiple types of sound recordings. *IEEE J. Biomed. Health Inform.* 2013, 17, 828–834. [CrossRef] [PubMed]
- Tan, X.; Liu, Y.; Li, Y.; Wang, P.; Zeng, X.; Yan, F.; Li, X. Localized instance fusion of MRI data of Alzheimer's disease for classification based on instance transfer ensemble learning. *Biomed. Eng. Online* 2018, *17*, 49. [CrossRef]
- 39. Mangasarian, O.L.; Wolberg, S. Breast cancer diagnosis and prognosis via linear programming. *Oper. Res.* **1995**, *43*, 570–577. [CrossRef]
- Little, M.; McSharry, P.E.; Hunter, E.J.; Spielman, J.; Ramig, L.O. Suitability of dysphonia measurements for telemonitoring of Parkinson's disease. *IEEE Trans. Bio-Med. Eng.* 2009, 56, 1015. [CrossRef]
- 41. Asuncion, A. UCI Machine Learning Repository, University of California, Irvine, School of Information and Computer Sciences. 2007. Available online: http://www.ics.uci.edu/~mlearn/MLRepository.html (accessed on 1 January 2022).
- 42. Merz, C.J. UCI Repository of Machine Learning Databases. 1998. Available online: http://archive.ics.uci.edu/ (accessed on 5 January 2022).
- 43. Beer, D.G.; Kardia, S.L.; Huang, C.C.; Giordano, T.J.; Levin, A.M.; Misek, D.E.; Lin, L.; Chen, G.; Gharib, T.G.; Thomas, D.G.; et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.* **2002**, *8*, 816–824. [CrossRef]
- 44. Yamada, M.; Jitkrittum, W.; Sigal, L.; Xing, E.P.; Sugiyama, M. High-dimensional feature selection by feature-wise kernelized lasso. *Neural Comput.* **2014**, *26*, 185–207. [CrossRef] [PubMed]
- 45. Sun, Y.; Lou, X.; Bao, B. A novel relief feature selection algorithm based on mean-variance model. *J. Inf. Comput. Sci.* **2011**, *8*, 3921–3929.
- 46. Wold, S.; Esbensen, K.; Geladi, P. Principal component analysis. Chemom. Intell. Lab. Syst. 1987, 2, 37–52. [CrossRef]
- 47. Li, C.H.; Kuo, B.C.; Lin, C.T. LDA-based clustering algorithm and its application to an unsupervised feature extraction. *IEEE Trans. Fuzzy Syst.* **2010**, *19*, 152–163. [CrossRef]
- 48. He, X.; Niyogi, P. Locality preserving projections. Adv. Neural Inf. Process. Syst. 2003, 16, 153–160.
- 49. Hinton, G.E.; Salakhutdinov, R.R. Reducing the dimensionality of data with neural networks. *Science* 2006, 313, 504–507. [CrossRef] [PubMed]
- 50. Goswami, G.; Vatsa, M.; Singh, R. Face verification via learned representation on feature-rich video frames. *IEEE Trans. Inf. Secur.* 2017, 12, 1686–1698. [CrossRef]
- 51. Görgel, P.; Simsek, A. Face recognition via deep stacked denoising sparse autoencoders (DSDSA). *Appl. Math. Comput.* **2019**, 355, 325–342. [CrossRef]
- 52. Kampffmeyer, M.; Løkse, S.; Bianchi, F.M.; Jenssen, R.; Livi, L. The deep kernelized autoencoder. *Appl. Soft Comput.* 2018, 71, 816–825. [CrossRef]
- Zhu, H.; Cheng, J.; Zhang, C.; Wu, J.; Shao, X. Stacked pruning sparse denoising autoencoder based intelligent fault diagnosis of rolling bearings. *Appl. Soft Comput.* 2020, 88, 106060. [CrossRef]
- 54. Hasan, M.K.; Alam, M.A.; Das, D.; Hossain, E.; Hasan, M. Diabetes prediction using ensembling of different machine learning classifiers. *IEEE Access* 2020, *8*, 76516–76531. [CrossRef]
- 55. Wang, Q.; Cao, W.; Guo, J.; Ren, J.; Cheng, Y.; Davis, D.N. DMP_MI: An effective diabetes mellitus classification algorithm on imbalanced data with missing values. *IEEE Access* **2019**, *7*, 102232–102238. [CrossRef]
- De Guia, J.D.; Concepcion, R.S.; Bandala, A.A.; Dadios, E.P. Performance comparison of classification algorithms for diagnosing chronic kidney disease. In Proceedings of the 2019 IEEE 11th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management, Laoag, Philippines, 29 November–1 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–7.
- Hasan, K.A.; Hasan, M.A.M. Prediction of clinical risk factors of diabetes using multiple machine learning techniques resolving class imbalance. In Proceedings of the 2020 23rd International Conference on Computer and Information Technology, Dhaka, Bangladesh, 19–21 December 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.

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.