



A Comparative Study of Discrete Direction vs. Continuous Distance-Based Cost Function in Energy Classification

Mehdi Khashei¹ · Sepideh Etemadi¹ · Negar Bakhtiarvand¹

Received: 7 June 2022 / Accepted: 16 April 2023
© The Author(s) 2023

Abstract

Over the past decades, different classification approaches with different characteristics have been developed to achieve more efficient and accurate results. Although the loss function used in the training procedure is a significant influential factor in the performance of classification models, it has been less considered. In general, in previous research, two main categories of continuous and semi-continuous distance-based loss functions are often applied to estimate the unknown parameters of classification models. Among these, continuous distance-based cost functions are among the most commonly used and most popular loss functions in diverse statistical and intelligent classifiers. In particular, the fundamental of this category of the loss functions is based on the continuous reduction of the distance between the fitted and actual values with the aim of improving the performance of the model. However, since the goal function of classification models belongs to the class of discrete ones, the application of learning procedures based on a continuous distance-based function is not coordinated with the nature of these problems. Consequently, it is theoretically illogical and practically at least inefficient. Accordingly, in order to fill this research gap, the discrete direction-based loss function in the form of mixed-integer programming is proposed in the training procedure of statistical, shallow/deep intelligent classifiers. In this paper, the impact of the loss function type on the classification rate of the classifiers in the energy domain is investigated. For this purpose, the logistic regression (LR), multilayer perceptron (MLP), and deep multilayer perceptron (DMLP), which are respectively among the most widely used statistical, shallow intelligent, and deep learning classifiers, are exemplarily chosen. Numerical outcomes from 13 benchmark energy datasets show that, in all benchmarks, the performances of the discrete direction learning-based classifiers, i.e., discrete learning-based logistic regression (DILR), discrete learning-based multilayer perceptron (DIMLP), and discrete learning-based deep multilayer perceptron (DIDMLP), is higher than its conventional versions. In addition, the proposed DILR, DIMLP, and DIDMLP models can on average yield an 89.88%, 94.53%, and 96.02% classification rate, which indicate a 6.78%, 5.90%, and 4.69% improvement from the classic versions, which only produce an 84.17%, 89.26%, and 91.72% classification rate. Consequently, the discrete direction-based learning methodology can be a more suitable, effective, and valuable alternative for training processes in statistical and shallow/deep intelligent classification models.

Keywords Classification processes · Discrete and continuous learning algorithms · Classification rate · Logistic regression (LR) · Multilayer perceptrons (MLPs) · Deep multilayer perceptrons (DMLPs) · Energy applications

Abbreviations

ANN	Artificial neural network	DIDMLP	Discrete learning-based deep multilayer perceptron
BN	Bayesian network	DILR	Discrete learning-based logistic regression
CNN	Convolutional neural networks	DIMLP	Discrete learning-based multilayer perceptron
DT	Decision tree	XGBoost	Extreme gradient boost
DMLP	Deep multilayer perceptron	ELM	Extreme learning machine
		FN	False negative
		FP	False positive
		LGBM	Light gradient boosting machine
		KNN	K-nearest neighbor
		GA	Genetic algorithm
		MLP	Multilayer perceptron

✉ Mehdi Khashei
Khashei@cc.iut.ac.ir

¹ Department of Industrial and Systems Engineering, Isfahan University of Technology (IUT), Isfahan 84156-83111, Iran

SVM	Support vector machine
NB	Naïve Bayes
RF	Random forest
LDA	Linear discriminant analysis
LR	Logistic regression
TN	True negative
TP	True positive

1 Introduction

Classifiers in the energy sector play a fundamental role in the decisions' quality made by resource managers, policy-makers, and planners. A variety of classification methods with different characteristics such as certainty (Fuzzy and/or Crisp), type (Statistical and/or Intelligent), complexity (Deep and/or Shallow), linearity (Linear and/or Nonlinear), structure (Single and/or Hybrid), cost function (Distance and/or Direction) etc., have been developed for classification. These models have been applied in various applications of the energy sector such as system stability, network efficiency, wave energy, solar energy, electrical energy, gas turbine, and consumption management. Some well-known statistical approaches in this field include Linear Discriminant Analysis (LDA), K-Nearest Neighbor (KNN), Logistic Regression (LR), Naïve Bayes (NB), and Bayesian Network (BN). Support Vector Machine (SVM), Multilayer Perceptron (MLP), Artificial Neural Network (ANN), Decision Tree (DT), Random Forest (RF), Light Gradient Boosting Machine (LGBM), Extreme Learning Machine (ELM), Extreme Gradient Boost (XGBoost) are among the most well-known intelligent classifiers. Deep Multilayer Perceptrons (DMLP) and Convolutional Neural Networks (CNN) are also two deep intelligent classifiers that are applied in the energy domain more than other deep classifiers.

Logistic regression, multilayer perceptron, and deep multilayer perceptron are common linear single statistical and nonlinear single shallow/deep intelligent classifiers that are extensively applied in modeling and data mining [1]. The distance and direction are also among the most popular and widely used cost/loss functions in the different classification approaches [2]. Musbah et al. [3] have compared and evaluated the performance of some different classifiers in order to determine the most accurate ones to specify the energy source for supplying the demand. These classifiers include the Gaussian Naïve Bayes, K-Nearest Neighbor (KNN), Decision Tree (DT), and Random Forest (RF). Empirical results of this study indicates that the DT classifier can outperform other classification methods. Song et al. [4] have assessed and compared the accuracy of KNN, DT, DA, and SVM classifiers to classify energy consumers. Their results show that the KNN can yield more accurate results than other classifiers. Chen et al. [5] have evaluated the performance of RF, DT, and Extreme Gradient Boost (XGBoost) tree

methods in order to classify the energy requirements of rural and urban households. In this study, the XGBoost yields the best results. Banihashemi et al. [6] have applied the DT method for classifying the level of energy consumption of buildings. Wang et al. [7] have designed the CNN classifier to forecast solar irradiance based on weather classification. The final goal is photovoltaic power prediction by using weather classification. Empirical results of this study indicates that the proposed model outperforms MLP, KNN, and SVM methods. Liu et al. [8] have classified solar radiation zones using the SVM. The parameters of the model are determined by the Genetic Algorithm (GA) optimization.

Yan et al. [9] have developed a Bayes classification method to classify household appliances. Their results show that their developed classifier can yield desired accuracy. Wang et al. [10] have established a multi-feature KNN model to specify occupancy distribution for controlling the ventilation, heating, and air-conditioning systems in buildings. Jiang and Yao [11] have established a C-Support Vector classifier for modeling personal thermal sensation. The findings of this study can be used for correctly setting the conditioning system. Shao et al. [12] have predicted electricity prices in Canada and New York markets by using a Bayesian ELM classifier. The presented model had higher performance than other considered classifiers. Bai et al. [13] have classified the Chinese climate for building energy efficiency. Protásio et al. [14] have classified Eucalyptus clones for combustion and energy purposes. Sabia et al. [15] have classified the energy performances of wastewater treatment plants. Patnaik et al. [16] have diagnosed microgrid faults in the XGBoost classifier, which is equipped with preprocessing and feature extraction procedures. Radhakrishnan et al. [17] have detected disturbances in the photovoltaic power network by applying hybrid structures of NB, J48 DT, and LR models. They indicate that their presented model can yield better accuracy than individual classifiers. Eskandari et al. [18] have also classified the fault of photovoltaic systems based on the hybrid structure of KNN, SVM, and NB methods. They also indicate that their proposed classifier can yield more accurate results than its components. Li et al. [19] have designed a hybrid structure of the ELM, XGBoost, and the LGBM for detecting the Intrusion of cyber-physical energy systems. Bi et al. [20] have designed an ANN classifier to diagnose faults in wind turbine generators.

In addition to the type of classifier, that significantly affects the obtained performance, the type of cost/loss function is another factor that may meaningfully affect the classification rate. Although several researchers have used regular cost/loss functions such as mean squared error (MSE), sum squared error (SSE), mean absolute error (MAE), root mean squared error (RMSE), etc. However, some others have developed new or compared existing cost/loss functions in order to find the desired ones for classification purposes. In general, cost/loss functions that are used in different classifiers are categorized

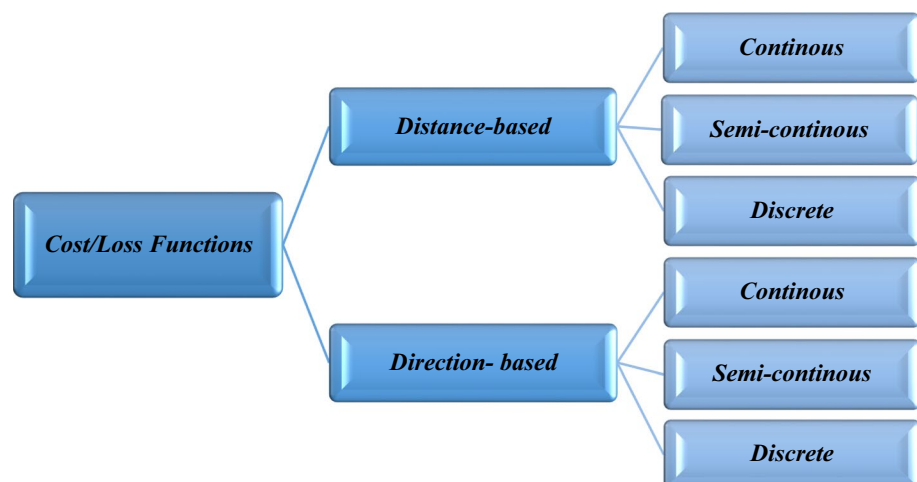
into two main categories (1) Distance-based and (2) Direction-based cost/loss functions. In distance-based cost/loss functions, the distance, or difference of actual and predicted values; or their probability is considered as cost function; while, in direction-based cost/loss functions, the matching of actual and predicted values, or their probability is considered as cost function. Moreover, cost/loss functions can be classified into three classes, including (1) Continuous, (2) Semi-continuous, and (3) Discrete cost/loss functions based on the type of function. The outputs of a continuous-based cost/loss function can continuously take the whole of values in its relevant range. While, in the semi-continuous cost/loss functions, the output can continuously change only in some parts of its relevant range. Similarly, in the discrete cost/loss functions, the output can only take some discrete values. In this way, the cost/loss functions can be generally categorized into six main categories as shown in Fig. 1.

Zhang et al. [21] have proposed tangent loss, as a continuous distance-based loss function, that calculates the tangent of differences between actual and predicted values. Their results demonstrate the higher performance of the proposed tangent loss-based deep neural networks than cross-entropy-based classifiers in natural language processing and computer vision tasks. Hazarika and Gupta [22] have developed a new ϵ -insensitive Huber loss function, that belongs to the semi-continuous distance-based loss functions, for dealing with noise in datasets. The results show that the random vector functional link with the proposed loss function outperforms SVM and ELM in biomedical datasets. Ozyildirim and Kiran [23] have studied the relationships between loss functions and the accuracy of classifiers. Their results indicate that, in certain circumstances, the MLP based on the squared hinge loss function is superior to the classifier with cross-entropy. Torre et al. [24] have presented a weighted kappa loss function for deep learning classifiers that fall in the category of the continuous direction-based loss function. Based on obtained results, the superiority of kappa-based

classifiers to the logarithmic-based classifier is confirmed. Liang and Zhang [25] have designed a support vector machine based on the quantile function for classification purposes of uncertain data. Their results show the higher performance of the classifier with ϵ -insensitive pinball loss function than the hinge loss function for classifying real-world and artificial datasets. In other research, the SVM is applied with semi-continuous direction-based loss functions, such as hinge loss [26], ramp loss [27], and truncated pinball loss [28], rescaled hinge loss [29] in the learning process.

Literature indicates that despite the fact that discrete direction-based cost/loss functions have more consistency with goal function of classification; the continuous distance-based cost/loss functions are often used. Therefore, presenting a classification methodology that uses the discrete distance-based loss function for the training procedure and optimizing the model's parameters, to match the training procedure with the goals of classification, is an efficient and reasonable approach. It is also a superior approach to conventional classification approaches for achieving maximum classification rates. Accordingly, by replacing the usual learning process with the discrete learning-based procedure in the logistic regression, multi-layer perceptron, and deep multilayer perceptron classifiers as the most widely used and popular statistical and shallow/deep intelligent models, the new versions of continuous learning-based LR, MLP, and DMLP are developed. Also, the evaluation of the efficiency and superiority of the discrete learning-based logistic regression, discrete learning-based multilayer perceptron, discrete learning-based deep multilayer perceptron to a continuous learning-based LR, MLP, and DMLP classifiers in various domains of energy sector based on 13 widely used benchmark datasets in the UCI database are demonstrated based on the classification rate.

Fig. 1 Cost/loss functions classification chart



Accordingly, the primary goal of this study is to conduct a comparative study of the classification rate of discrete learning-based statistical and shallow/deep intelligent models against continuous learning-based versions that are applied in the energy decision-making sector. In other words, the main purpose of this paper is to highlight the significance of the direction learning-based procedure in the classification modeling, on the quality of decisions made in the energy sector with statistical and intelligent decision support systems. The rest of the paper is designed in the following style. In Sect. 2, the multilayer perceptron based on the conception of the discrete direction-based learning process is mathematically formulated as an example. In Sect. 3, the selected datasets for evaluating the discrete learning-based models are described and the evaluation metrics of the models are introduced. In addition, the analysis and comparison of the discrete direction learning-based models' performances and the conventional models are presented in Sect. 4. In Sect. 5, some discussions and future research gaps are suggested. Lastly, the conclusions are presented in Sect. 6.

2 The Discrete Direction Learning-Based Multilayer Perceptron (DIMLP)

The multilayer perceptron is a common intelligent classification method that is extensively used in modeling and data mining. The main reasons for the popularity of the MLP are basically related to its desired accuracy which comes from some unique features, including self-adaptive data-driven procedures, flexible nonlinear modeling, and general approximation. Despite all advantages reported for the MLP, it also has some disadvantages and limitations that may reduce its performance and accuracy. Noteworthy endeavor has been made in the literature to improve the accuracy and classification rate of multilayer perceptrons. However, in all of them, a common continuous distance-based loss function is applied in their training procedures. In this way, in this paper, a novel multilayer perceptron has been established based on a discrete learning-based algorithm.

Commonly, an m -variable MLP binary classifier, comprising the target variable $Y \in \{-1, +1\}$ and m features, $X_1, X_2, \dots, X_m \in \mathfrak{R}$ can be generally shown as follows [30]:

$$Y_t = f\left(\beta_0 + \sum_{j=1}^p \beta_j \cdot g\left(\beta_{0j} + \sum_{i=1}^m \beta_{ij} \cdot X_{t,i}\right)\right) + u_t \quad t = 1, 2, 3, \dots, N \tag{1}$$

where, β_j, β_{ij} are connection weights of the network, g and f are respectively the hidden and output layer activation functions, p and m are respectively the number of input and hidden nodes, N is the sample size, and u_t is the stochastic disturbance term. A popular training method for estimating

the unknown weights and biases of MLP is according to a gradient optimization in which a continuous distance-based loss function (sum squared of misclassification) is minimized [31]. Regardless of the fact that this loss function is one of the widely used and popular and in the learning process of classifiers in various applications. The inconsistency between the nature of the loss function, which is continuous based on distance, and the purpose of classification models, which is discrete, is unreasonable and inefficient.

The idea of the training procedure based on continuous-distance is such that at each stage of training, the fitted values are continuously approaching the actual values, which is in coordination with the goals of causal models and time series that have continuous outputs. Therefore, using this type of learning process in causal and time series models is efficient and rational due to improving the performance of the model. In the case of classification models and time series classification, the output of this type of learning process should be reported in a discrete form. Therefore, after the end of the learning process, the predicted values convert to discrete variables. However, if the direction of the training and discretization process is aligned, the use of such procedures is effectual and efficient. Otherwise, if the direction of the training procedure and discretization is non-aligned, the impact of the training procedure is eliminated, and therefore, by imposing the computational cost, this type of training procedure is unreasonable and quite ineffective. Generally, developing accurate classifiers in the light of a continuous-distance learning-based model may be totally unsuitable and quite inefficient.

Therefore, in this paper, to achieve a more reasonable training process for an intelligent classifier, a discrete direction-based learning methodology is presented and implemented on the multilayer perceptrons. The principal notion of the presented training methodology is to maximize a discrete matching function of the predicted and actual values instead of minimizing a continuous distance-based loss function. Based on this, the procedure of estimating the unknown parameters of the MLP in the discrete direction-based learning method can be shown as follows:

$$Max \sum_{t=1}^N Match(y_t, \hat{y}_t) \tag{2}$$

where, the $Match(y_t, \hat{y}_t)$ is the matching function of the actual ($y_t \in \{-1, +1\}$) and the predicted ($\hat{y}_t \in \mathfrak{R}$) at time t , which is shown in the binary form as follows:

$$Match(y_t, \hat{y}_t) = \begin{cases} +1 & \text{if } (y_t)(\hat{y}_t^s) \geq 0 \\ -1 & \text{if } (y_t)(\hat{y}_t^s) < 0 \end{cases} \tag{3}$$

where, the $\hat{y}_t^s = Std.(\hat{y}_t)$ is the standardized value of the \hat{y}_t at time t which compute as follows:

$$\hat{y}_t^s = \frac{\hat{y}_t - \bar{\hat{y}}}{\hat{y}^{Max} - \hat{y}^{Min}} \tag{4}$$

where, the \hat{y}^{Min} , \hat{y}^{Max} , and $\bar{\hat{y}}$ are the Minimum, maximum, and Mean of \hat{y}_t . Now, Eq. (2) can be rewritten by the *Sign* function as follows:

$$Max \sum_{t=1}^N Sign[(y_t) \cdot (\hat{y}_t^s)] \tag{5}$$

Or in a more simplified as:

$$Max \sum_{t=1}^N Sign(y_t) \cdot Sign(\hat{y}_t^s) \tag{6}$$

In this manner, we have that:

$$Max \sum_{t=1}^N (y_t) \cdot Sign \left(Std. \left(f \left(\hat{\beta}_0 + \sum_{j=1}^p \hat{\beta}_j \cdot g \left(\hat{\beta}_{0j} + \sum_{i=1}^m \hat{\beta}_{ij} \cdot X_{i,t} \right) \right) \right) \right) \\ = \sum_{t=1}^N (y_t) \cdot Sign \left(Std. \left(f \left(\sum_{j=0}^p \hat{\beta}_j \cdot g \left(\sum_{i=0}^m \hat{\beta}_{ij} \cdot X_{i,t} \right) \right) \right) \right) \tag{7}$$

Lastly, using the sigmoid (*Sigm.*) and linear function as hidden and output transfer functions, respectively, Eq. (7) can be converted to a mixed-integer programming form as Eq. (8). In which, ϵ and M are a very small and a very large very number, respectively:

$$Max \sum_{t=1}^N Sign(y_t) \cdot Sign(\hat{y}_t^s) = \sum_{t=1}^N (y_t) \cdot Sign \left(Std. \left(\sum_{j=0}^p \hat{\beta}_j \cdot Sigm \left(\sum_{i=0}^m \hat{\beta}_{ij} \cdot X_{i,t} \right) \right) \right) \\ \left. \begin{aligned} &Std. \left(\sum_{j=0}^p \hat{\beta}_j \cdot Sigm \left(\sum_{i=0}^m \hat{\beta}_{ij} \cdot X_{i,t} \right) \right) \geq \epsilon - M(1 - a_t) \quad t = 1, 2, 3, \dots, N \\ &Std. \left(\sum_{j=0}^p \hat{\beta}_j \cdot Sigm \left(\sum_{i=0}^m \hat{\beta}_{ij} \cdot X_{i,t} \right) \right) \leq -\epsilon + M(1 - b_t) \quad t = 1, 2, 3, \dots, N \\ &Std. \left(\sum_{j=0}^p \hat{\beta}_j \cdot Sigm \left(\sum_{i=0}^m \hat{\beta}_{ij} \cdot X_{i,t} \right) \right) \geq -M(1 - c_t) \quad t = 1, 2, 3, \dots, N \\ &Std. \left(\sum_{j=0}^p \hat{\beta}_j \cdot Sigm \left(\sum_{i=0}^m \hat{\beta}_{ij} \cdot X_{i,t} \right) \right) \leq M(1 - c_t) \quad t = 1, 2, 3, \dots, N \\ &a_t + b_t + c_t = 1 \quad t = 1, 2, 3, \dots, N \\ &Sign(\hat{y}_t^s) = a_t - b_t \quad t = 1, 2, 3, \dots, N \\ &a, b, c_t \in \{0, 1\} \quad t = 1, 2, 3, \dots, N \\ &\hat{\beta}_j \text{ free of sign} \end{aligned} \right\} S.T. \tag{8}$$

The learning process of the MLP or determining the optimal weights and biases of the model is typically conducted by using backpropagation and gradient descent techniques. The backpropagation learning algorithm can change the weights of the multilayer perceptron automatically and produces and generate output for a set of input features. The gradient decent technique is a measure that determines the alteration in error when there is a change in weights and it minimizes a cost (loss) function (sum squared of classification errors). To design an MLP model, the weights are initialized by applying a random value weight assignment. Then, the backpropagation algorithm and gradient descent optimization are applied to regulate the connection weights of the MLP. The weights are adjusted in such a manner that the cost function and loss in error value are minimized. Generally, it can be said that the procedure of training and estimating unknown weights of the MLP classifier is conducted by minimizing a continuous distance-based loss function (sum squared of misclassification). Despite the widespread use of this approach, there is a critical issue concerning the characteristic of classification itself and the training procedure. The intelligent classifiers have a continuous distance-based loss function, which is in conflict with the discrete nature of classification. Generally, utilizing a continuous loss function for the classification issue which has an actually discrete target function is irrational or at least quite inefficient. The purpose of this study is to propose a discrete learning-based approach for the learning process of the MLP model to obtain a more reasonable and efficient classification result. The fundamental notion of the proposed approach is maximizing a discrete matching function of the predicted and actual values instead of minimizing the sum of errors squares as a continuous distance-based loss function.

Finally, some of the most theoretical/practical advantages and disadvantages of the proposed DIMLP classifier are presented. In general, several features have been presented in the literature for theoretical/practical evaluating/ comparing different classification approaches. Some of the most important of them, from the theoretical point of view, involve convergence, speed of convergence, the universality of modeling, and modeling of uncertainty [32]. Due to the fact that the main difference between the proposed DIMLP model rather than the conventional MLP model is only its cost function. Thus, the proposed model, similar to its conventional version, are universal approximator, but cannot model uncertain patterns. However, since the proposed cost function is formulated as mixed-integer programming; thus, its speed of convergence is generally lesser than the conventional MLP, while both converge. In a similar fashion, some of the most important practical features, based on the degree of importance, are accuracy, computational time and cost,

interpretability, ease of use and implementation [33]. Since the proposed cost function is more matched by the goal function of classification; it is expected that the proposed DIMLP model can yield more accurate results than its conventional version. Nevertheless, the computational time and cost of the proposed DIMLP model, due to its discrete variables, is higher than the conventional MLP classifier. Therefore, the most serious disadvantage of the proposed model is its computational time and cost. Moreover, the proposed model, similar to its conventional version and other intelligent classifiers, has low interpretability. Thus, the second disadvantage of the proposed model is its low interpretability. Analogously, the designing process and procedure of determining the desired architecture of the proposed model, similar to conventional MLP, is a problematical task. Thus, the third disadvantage of the proposed DIMLP model is that its use and implementation are not simple.

3 Data Description and Evaluation Metrics

In this section, a brief explanation of the used datasets in this paper is provided. In addition, criteria applied for evaluating the performance of the presented DILR, DIMLP, and DIDMLP in comparison with its classic versions are presented. In this research, the classification rate, as presented in Eq. (9) is mainly considered to compare and assess the performance of the presented classifiers with the conventional ones. Moreover, the F1-score, precision, and recall are also considered to assess the performance of classifiers. The formulation of these criteria is given in Eq. (10) to Eq. (12), respectively [34]. In this research, 13 benchmark datasets regarding the classification category from the UCI website have been selected [35]. These case studies consist of simulated or real examples in various energy applications such as System stability, System simulation, Network efficiency, Wave energy, Solar energy, Consumption management, Gas turbine, and Electrical energy from 1989 to 2019 are considered to comprehensively evaluate the direction-based classifier. The number of attributes in the models varied from 3 to 81 and the instance size variation of these datasets is 167 to 71,999 data points. These examples have several kinds of explanatory variables, containing two main categories of single and mixed explanatory variables. The subcategories of single features include integer, real, and categorical. In addition, three sub-categories of mixed explanatory variables, covering 2–1) integer, real, 2–2) categorical, real, and 2–3) categorical, integer. More detailed information about these datasets, such as the number of attributes, the attributes' characteristic, and the instances' size have been summarized in Table 1.

Classification Rate

$$= \frac{\text{True Negative} + \text{True Positive}}{\text{False Negative} + \text{True Negative} + \text{False Positive} + \text{True Positive}} \quad (9)$$

$$\text{F1 Score} = \frac{2(\text{Recall} \times \text{precision})}{\text{Recall} + \text{precision}} \quad (10)$$

$$= \frac{2(\text{True Positive})}{(\text{False Negative} + \text{False Positive}) + 2(\text{True Positive})}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{False Negative} + \text{True Positive}} \quad (11)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{False Positive} + \text{True Positive}} \quad (12)$$

where true negative (TN) is the negative data that is correctly recognized as negative, true positive (TP) is positive data that is correctly identified as positive, false negative (FN) is the positive data that is misdiagnosed as negative, and false positive (FP) is the negative data that is misidentified as positive.

4 Empirical Result

In this section, the proposed DILR, DIMLP, and DIDMLP classifiers and all the analyses are presented in detail for one of these datasets, Auto MPG, as an example. This dataset contains seven attributes of 392 records, which are utilized to forecast car fuel consumption. The predictor attributes are cylinders (3, 4, 5, 6, 8), displacement, horsepower, weight, acceleration, model year (70–82), and origin (1–3) (respectively, X_1 to X_7) to classify cars as low consumption and high consumption. Statistical characteristics of sample are shown in Table 2. The plot of the attributes against each other based on the target variable is shown in Fig. 2. In this paper, 314 data points of the data, i.e., approximately 80% of dataset, are randomly chosen as training set, and the remaining 78 data points, i.e., approximately 20% of dataset, are regarded as the test set. Furthermore, the estimation procedure of classifiers is repeated 100 times to remove the effect of the random choice of the data. In this paper, all modeling of the DILR, DIMLP, DIDMLP, and classic LR, MLP, and DMLP classifiers are run in MATLAB and GAMS package software. The performance of the presented DILR, DIMLP, DIDMLP, and conventional LR, MLP, and DMLP classifiers, as well as the improvement percentage of the discrete direction learning-based models compared to the conventional versions, are reported in Table 3.

Empirical outcomes of this example demonstrate that the presented DILR, DIMLP, and DIDMLP models, by benefiting from the discrete direction-based learning approach, can

Table 1 The general characteristics of the selected benchmarks

No.	Name	Year	Sample size	Number of attributes	Attribute type	Domain of application
1	Electrical Grid Stability Simulated Data	2018	10,000	11	Real	Smart network (predicting system stability)
2	Optical Interconnection Network	2018	640	9	Integer, real	Electronic (network efficiency)
3	Solar Flare	1989	323	7	Categorical	Solar energy
4	Tamilnadu Electricity Board Hourly Readings	2013	45,781	3	Real	Consumption
5	Appliances Energy Prediction	2017	19,735	27	Real	Appliances
6	Auto MPG	1993	392	7	Real, categorical	Consumption management
7	Gas Turbine CO and NOx Emission	2019	7152	10	Real	Predicting hourly net energy of gas turbine
8	Combined Cycle Power Plant	2014	9568	4	Real	Electrical energy
9	Superconductivity Data	2018	21,263	81	Real	Electrical engineering
10	Servo	1993	167	4	Integer, categorical	System simulation
11	Yacht Hydrodynamics	2013	308	6	Real	Predicting the hydrodynamic performance of a yacht
12	Wave Energy Converters	2019	71,999	48	Real	Forecasting total production capacity
13	Energy Efficiency	2012	768	7	Integer, real	Evaluation of heating and cooling load

Table 2 Statistical characteristics of the selected dataset (Auto MPG)

Features	Mean	Standard deviation	Mode	Median	Minimum	Maximum	Skewness	Kurtosis
X_2	194.412	104.644	97	151	68	455	0.702	-0.778
X_3	104.469	38.491	150	93.500	46	230	1.087	0.697
X_4	2977.584	849.403	2130	2803.500	1613	5140	0.520	-0.809
X_5	15.541	2.759	14.500	15.500	8	24.800	0.292	0.444
X_6	75.980	3.684	73	76	70	82	0.020	-1.167

obtain a 94.64%, 97.85%, and 99.77% classification rate, respectively. While the conventional LR, MLP, and DMLP models, which applied the continuous distance-based loss function, can only achieve a 91.53%, 95.28%, and 98.56% classification rate, respectively. It is shown that the proposed DILR, DIMLP, and DIDMLP models can improve the performance of their classic ones by 3.40%, 2.70%, and 1.23%, respectively.

In addition, these improvements are not limited to classification rates and are also confirmed by other performance measurements. For example, the proposed DILR, DIMLP, and DIDMLP models can reach 94.39%, 97.74%, and 99.76% in the precision criterion, respectively, while the classic LR, MLP, and DMLP classifiers can only produce 91.17%, 95.05%, and 98.48%, respectively. In this way, the improvements of the proposed statistical, shallow, and deep classifiers rather than their traditional versions in the precision are, respectively, equal to 3.53%, 2.82%, and 1.29%. Furthermore, these improvements in the recall criterion and F1-score are equal to 3.03%, 2.42%, and 1.10%, and 3.28%, 2.62%, and 1.20%, respectively. These

results clearly illustrate that the superiority of the proposed classifiers is not dependent on the type of model, e.g., statistical, shallow intelligent/deep intelligent, and the performance indicators, e.g., precision, classification rate, F1-score, and recall. Accordingly, they can primarily indicate the importance and effectiveness of the consistency between discrete direction-based learning processes and classification cost function. Moreover, it can be seen that these performance measurements are overall consistent.

After that, to more comprehensively assess the performance of the presented DILR, DIMLP, and DIDMLP models, in addition to their classical versions, their classification rates are compared with some of the widely used and most popular classifiers with different characteristics. For this purpose, a total of 17 famous statistical and shallow/deep intelligent models in both single and hybrid forms are considered (Table 4). Numerical outcomes illustrate that the presented DILR, DIMLP, and DIDMLP models can on average improve 0.99%, 4.39%, and 6.44% of the classification rate of the considered classifiers. The presented DILR, as the single statistical classifier, outperforms all

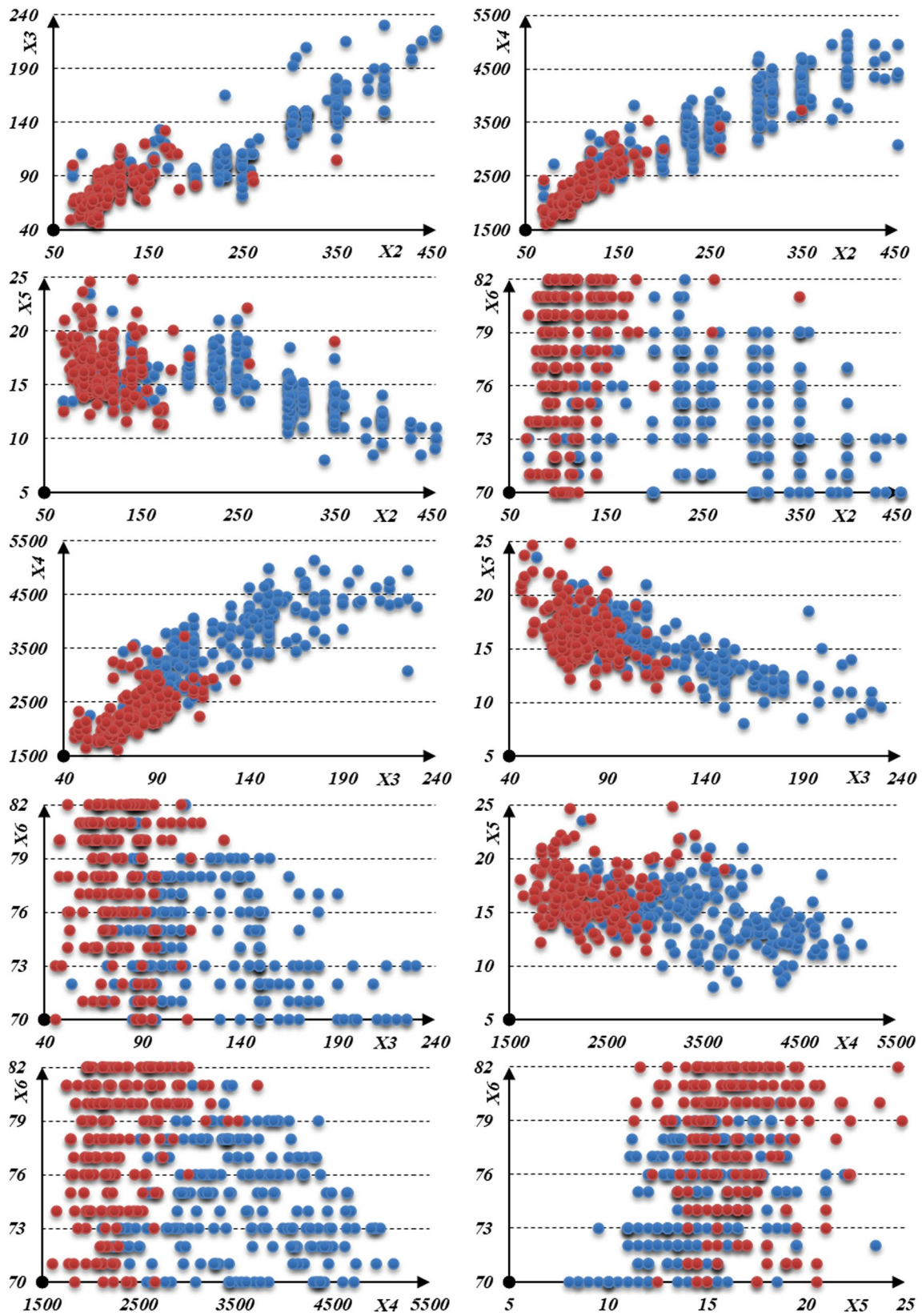


Fig. 2 The plot of attributes against each other according to their classes (blue: 0, red: 1)

Table 3 Classification rate of the discrete learning-based models and classic models (Auto MPG dataset)

Criteria	Model	Learning algorithms		Improvement
		Discrete learning-based version	Continuous learning-based version	
Classification Rate	LR	94.64%	91.53%	3.40%
	MLP	97.85%	95.28%	2.70%
	DMLP	99.77%	98.56%	1.23%
Precision	LR	94.39%	91.17%	3.53%
	MLP	97.74%	95.05%	2.82%
	DMLP	99.76%	98.48%	1.29%
Recall	LR	95.18%	92.38%	3.03%
	MLP	98.07%	95.75%	2.42%
	DMLP	99.79%	98.70%	1.10%
F1-score	LR	94.78%	91.77%	3.28%
	MLP	97.90%	95.40%	2.62%
	DMLP	99.78%	98.59%	1.20%

Bold values indicates the average and improvement values

single and hybrid statistical classifiers. It even can yield more classification rates than some shallow intelligent classifiers, such as decision tree, random forest, and probabilistic neural network. Consequently, the presented DILR model can on average improve 4.52% and 2.74% of the classification rate of single and hybrid classifiers, respectively.

Similarly, the presented DIMLP, as a single intelligent model, outperforms all statistical and single/hybrid shallow intelligent classifiers. The DIMLP even can yield more classification rate than the deep naive Bayes classifier, as a deep intelligent model. Empirical outcomes demonstrate that the presented DIMLP can on average improve the 7.37%, 3.75%, and 1.94% classification rates of statistical, as well as single and hybrid shallow intelligent models, respectively. The presented DIDMLP, as a single deep intelligent model, can also improve by 2.68%, 0.54%, and 0.94% the classification rate of the deep naive Bayes classifier, deep convolutional neural network, and principal component analysis-based deep genetic cascade of the SVM, respectively.

Finally, to remove the effects of data characteristics on the performance of the models, 12 other benchmark datasets were also used. The classification rates of the conventional DILR, DIMLP, and LR, MLP classifiers, as well as the improvement of the discrete direction learning-based models in comparison with the classic versions, are separately given for each datasets in Tables 5 and 6, respectively. According to the numerical outcomes, the presented DILR, DIMLP, and DIDMLP models have better performance than the continuous learning-based LR, MLP, and DMLP models in all fields of application and categories, including system stability, system simulation, network efficiency, wave energy, solar energy, consumption management, gas turbine, and electrical energy. Of course, it is not an unexpected outcome

and was a pre-expected consequence. Because, as mentioned previously, it can be generally indicated that the classification rate of the discrete direction learning-based models will not be worse than its continuous distance-based learning version, due to more consistency between goal and cost function.

These improvements are not constant in all categories and vary from each category to another. The results show that the lowest and highest improvement of DILR is in the fields of electrical energy (combined cycle power plant) and system simulation (servo) with a classification rate of 0.27% and 34.38%. Overall, the DILR can on average enhance the 6.78% classification rate of the classical LR. Finally, it must be also notated that the DILR model can achieve the perfect classification rate (i.e., 100%) in some case studies. The frequency of this matter is approximately equal to 23.08%, e.g., in 3 out of 13 cases, while it is never obtained by the LR in no cases and domains. These results first indicate that it is a regular event in the discrete learning algorithms, while in the traditional continuous ones it is completely an uncommon event. Second, it indicates that the concept of linear separability is totally changed in the discrete learning algorithms and completely differs from its traditional continuous version. In general, based on the empirical outcomes and consequences, the DILR model can be considered an appropriate alternative to classic classification models in different application fields to achieve a better classification rate.

Based on the reported performances in Table 6, the lowest and highest improvement of DIMLP is in the fields of wave energy (wave energy converters) and consumption management (Tamilnadu Electricity Board Hourly Readings) with a classification rate of 0.41% and 29.74%. Overall, the DIMLP can on average enhance the 5.90% classification rate

Table 4 Comparison of the presented DILR, DIMLP, and DIDMLP with other classifiers (Auto MPG dataset)

Model	Type/category (structure)	Classification rate	Improvement		
			DILR	DIMLP	DIDMLP
Probit regression (PR)	Single/statistical (N/A)	91.48%	3.49%	6.96%	9.06%
Linear discriminant analysis (LDA)	Single/statistical (N/A)	90.61%	4.48%	7.99%	10.11%
Quadratic discriminant analysis (QDA)	Single/statistical (N/A)	90.89%	4.16%	7.66%	9.77%
K-nearest neighbor (KNN)	Single/statistical (N/A)	91.22%	3.78%	7.27%	9.37%
Entropy difference approach (EDA)	Single/statistical (N/A)	88.52%	6.95%	10.54%	12.71%
Generalized logistic regression (GLR)	Hybrid/statistical (N/A)	92.13%	2.76%	6.21%	8.29%
Self-organizing map-based logistic regression (SOM-LR)	Hybrid/statistical (N/A)	92.48%	2.37%	5.81%	7.88%
Genetic algorithm-based K-nearest neighbor (GA-KNN)	Hybrid/statistical (N/A)	91.75%	3.18%	6.65%	8.74%
Decision tree (DT)	Single/intelligent (shallow)	93.76%	0.97%	4.36%	6.41%
Random forest (RF)	Single/intelligent (shallow)	94.33%	0.36%	3.73%	5.77%
Probabilistic neural network (PNN)	Single/intelligent (shallow)	93.29%	1.48%	4.89%	6.95%
Support vector machine (SVM)	Single/intelligent (shallow)	95.87%	-1.25%	2.07%	4.07%
Hybrid MLP-SVM	Hybrid/intelligent (shallow)	96.06%	-1.45%	1.86%	3.86%
Particle swarm optimization- support vector machine (PSO-SVM)	Hybrid/intelligent (shallow)	95.91%	-1.29%	2.02%	4.02%
Deep naive Bayes classifier	Single/intelligent (deep)	97.17%	-2.57%	0.70%	2.68%
Deep convolutional neural network	Single/intelligent (deep)	99.23%	-4.60%	-1.39%	0.54%
Principal component analysis-based deep genetic cascade of the SVM	Hybrid/intelligent (deep)	98.84%	-4.22%	-1.00%	0.94%
Average		93.74%	0.99%	4.39%	6.44%

Bold values indicates the average and improvement values

Table 5 Classification rate of the presented DILR and conventional LR classifiers for all datasets

No.	Dataset	Classification rate		Improvement
		DILR	Classic LR	
1	Electrical Grid Stability Simulated Data	87.38%	80.33%	8.77%
2	Optical Interconnection Network	83.44%	77.08%	8.24%
3	Solar Flare	98.83%	97.96%	0.89%
4	Tamilnadu Electricity Board Hourly Readings	50.02%	49.88%	0.29%
5	Appliances Energy Prediction	66.40%	63.48%	4.61%
6	Auto MPG	94.64%	91.53%	3.41%
7	Gas Turbine CO and NOx Emission	99.95%	99.35%	0.61%
8	Combined Cycle Power Plant	95.87%	95.61%	0.27%
9	Superconductivity Data	93.69%	89.62%	4.55%
10	Servo	98.20%	73.08%	34.38%
11	Yacht Hydrodynamics	100.00%	80.43%	24.32%
12	Wave Energy Converters	100.00%	98.88%	1.13%
13	Energy Efficiency	100.00%	97.00%	3.09%
	Average	89.88%	84.17%	6.78%

Bold values indicates the average and improvement values

Table 6 Classification rate of the presented DIMLP and conventional MLP classifiers for all datasets

No.	Dataset	Classification rate		Improvement
		DIMLP	MLP	
1	Electrical Grid Stability Simulated Data	92.41%	86.13%	7.29%
2	Optical Interconnection Network	89.75%	86.33%	3.96%
3	Solar Flare	99.61%	98.46%	1.17%
4	Tamilnadu Electricity Board Hourly Readings	73.55%	56.69%	29.74%
5	Appliances energy prediction	80.15%	71.29%	12.43%
6	Auto MPG	97.85%	95.28%	2.70%
7	Gas Turbine CO and NOx Emission	100.00%	99.55%	0.45%
8	Combined Cycle Power Plant	98.39%	97.73%	0.68%
9	Superconductivity Data	98.09%	94.21%	4.12%
10	Servo	99.14%	86.69%	14.36%
11	Yacht Hydrodynamics	100.00%	89.86%	11.28%
12	Wave Energy Converters	100.00%	99.59%	0.41%
13	Energy efficiency	100.00%	98.63%	1.39%
	Average	94.53%	89.26%	5.90%

Bold values indicates the average and improvement values

of the classical MLP. Finally, it must be also notated that the DIMLP can achieve the perfect classification rate (i.e., 100%) in some case studies. The frequency of this matter is approximately equal to 30.77%, e.g., in 4 out of 13 cases, while it is never obtained by the MLP in cases and domains.

Moreover, the average classification rate, precision, F1-scores, and recall of the presented DILR, DIMLP, and DIDMLP classifiers and the conventional LR, MLP, and DMLP models for the 13 benchmark datasets are reported in Table 7. Accordingly, the average classification rate of the DILR, DIMLP, and DIDMLP classifiers in the field of energy applications based on the 13 aforementioned datasets is 89.88%, 94.53%, and 96.02%, which is achieved at 6.78%, 5.90%, and 4.69% improvement than the 84.17%, 89.26%,

and 91.72% classification rate of the conventional versions, respectively. These performances and improvements are also plotted in Fig. 3.

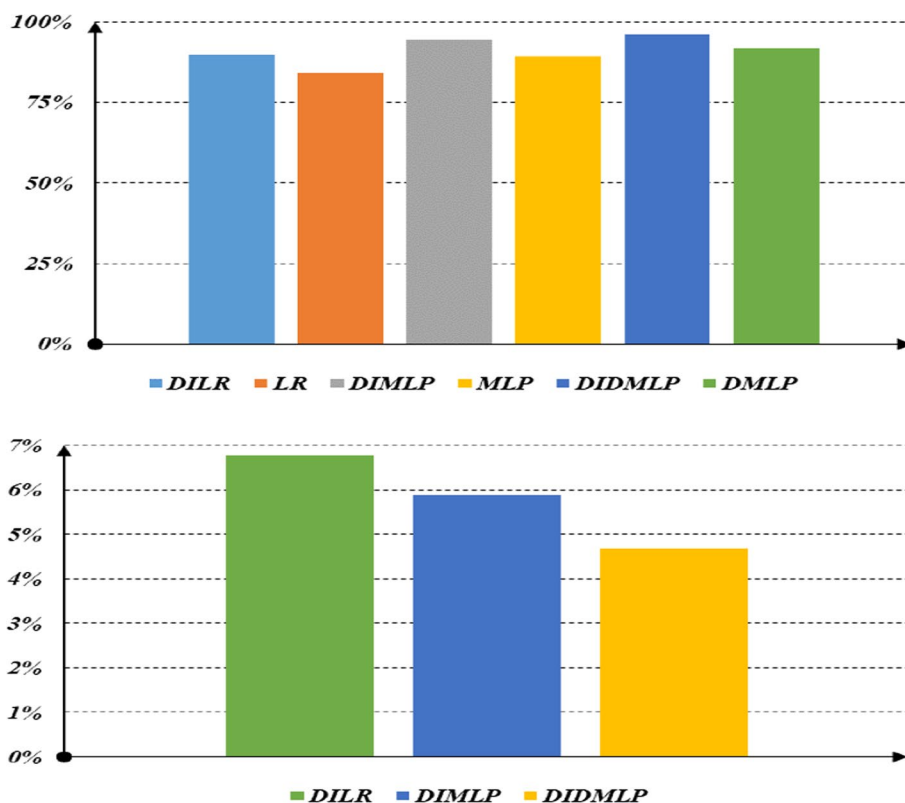
Furthermore, these improvements are not limited to the classification rate and are also repeated in other criteria. For example, the proposed DILR, DIMLP, and DIDMLP models can achieve an 88.76%, 93.84%, and 95.50% in the precision criterion, respectively, while the classic LR, MLP, and DMLP classifiers can only yield 82.72%, 88.09%, and 90.75%, respectively. In this way, the improvements of the proposed statistical, shallow, and deep classifiers rather than their traditional versions in the precision are, respectively, equal to 7.31%, 6.52%, and 5.23%. Besides, these improvements in the recall criterion and F1-score are equal to 5.48%,

Table 7 Classification rate of the presented discrete direction learning-based and conventional versions for all datasets

Criteria	Model	Learning algorithms		Improvement
		Discrete learning-based version	Continuous learning-based version	
Classification rate	LR	89.88%	84.17%	6.78%
	MLP	94.53%	89.26%	5.90%
	DMLP	96.02%	91.72%	4.69%
Precision	LR	88.76%	82.72%	7.31%
	MLP	93.84%	88.09%	6.52%
	DMLP	95.50%	90.75%	5.23%
Recall	LR	91.57%	86.81%	5.48%
	MLP	95.44%	91.05%	4.82%
	DMLP	96.68%	93.10%	3.85%
F1-score	LR	90.14%	84.71%	6.41%
	MLP	94.63%	89.55%	5.68%
	DMLP	96.09%	91.91%	4.54%

Bold values indicates the average and improvement values

Fig. 3 The classification rate and improvement of the DILR, DIMLP, and DIDMLP against conventional versions



4.82%, and 3.85%, and 6.41%, 5.68%, and 4.54%, respectively. These results show that the superiority of the proposed classifiers is not dependent on (1) the type of model, e.g., statistical, shallow intelligent/deep intelligent, (2) the performance indicators, e.g., precision, classification rate, F1-score, and recall, and (3) the data characteristics, such as example size, type of attributes, and number of attributes. Accordingly, they can indicate the importance and effectiveness of the consistency between discrete direction-based learning processes and classification cost function.

5 Discussions and Future Research Suggestions

This section discusses the accomplishments of this survey, followed by some recommendations for future research.

- This paper conducts a comparative study of the classification rate of discrete learning-based statistical and shallow/deep intelligent models against classic versions that are applied in the energy decision-making sector.
- This is the first research that evaluates the efficiency and classification accuracy of discrete learning-based statistical and discrete learning-based shallow/deep intelligent classification models against the most popular continuous

learning-based versions. The results have demonstrated that maximizing a discrete matching function of predicted and actual values in the training process of classifiers can reduce the misclassification rate.

- This study reveals that the consistency between loss function during the training procedure and the nature of classifier models in terms of discrete or continuous have a significant effect on the performance and generalizability of classifiers. However, this critical and effective matter has been overlooked in the training procedures of conventional statistical and shallow/deep intelligent models.
- The empirical results provide further support to the assumption that considering a continuous distance-based loss function for classification purposes is non-sequitur or at least insufficient.
- In this paper, the proposed approach has been implemented on the LR, MLP, and DMLP models as the most frequent and popular statistical and shallow/deep intelligent classifiers. However, the presented discrete direction-based learning approach is a completely general methodology that can be applied to other classes of classification models.

The following issues are some of the potential research gaps for future works:

- Conducting a comparative study for assessing the efficiency and accuracy of the proposed classifiers against statistical, shallow/deep intelligent classifiers with other types of loss functions.
- Developing an advanced version of the proposed classifiers using a combination of several loss functions in the categories of discrete, semi-continuous, and continuous.
- Implementing the discrete direction learning-based approach on other class of models including rule and tree classifiers, instance-based classifiers, Bayesian classifiers, other neural networks, and SVM.
- Examining the data properties such as dimension, noise, and outlier, missing attribute values on the performance of classifiers in the categories of linear/ nonlinear, statistical/intelligent, and shallow/deep models that are developed based on different loss functions.
- Developing the discrete direction learning-based method for other types of classification problems such as multi-class and multi-label issues.
- Investigating the effectiveness of the type of loss function as discrete, semi-continuous, and continuous on the accuracy of linear/nonlinear, crisp, fuzzy, statistical/intelligent, and shallow/deep classifiers in diverse fields of science.
- Assessing the generalizability of discrete direction learning-based classifiers in other scopes of science such as medicine, finance, environment, management, and engineering.

6 Conclusion

As data science and machine learning have developed over the years, most classification methods have become more complex to achieve results that are more accurate. Various techniques have been developed to enhance the accuracy of the statistical and intelligent classification models. These techniques include data preprocessing, feature selection, hybridization, and ensembling. Logically, the consistency between the goal function and the learning procedure is one of the most important factors influencing the achievement of the maximum classification rate, especially in the energy sector; nevertheless, it was ignored in all previously developed classifiers models. Accordingly, this paper conducts a comparative study of the classification rate of a discrete learning-based approach based on the most common statistical and shallow/deep intelligent models versus classic classifiers, which are widely used for decision-making in the energy sector. The empirical results demonstrated that although the classic classifiers yield a greater degree of accuracy at the price of complexity, however, the discrete learning-based approach is much superior to classic versions by utilizing more consistency between the discrete direction-based learning process and classification cost function. The

outcomes of the study support the claim that, in the light of this perspective, which is replacing a continuous learning-based loss function with a discrete learning-based loss function, the rate of classification in the energy applications has achieved higher accuracy in comparison to the classic statistical, shallow/deep intelligent versions.

Authors' Contributions All authors contributed equally in preparing this manuscript.

Funding No specific financial support was received to carry out this study.

Availability of Data and Material Data is available from the corresponding author on reasonable request.

Code Availability The code used in the study is available from the corresponding author on reasonable request.

Declarations

Conflict of Interest The authors declare that they have no competing interests.

Consent for Publication The authors declare that they consent to the publication of this study.

Consent to Participate The authors declare that they consent to participate in this study.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Khashei, M., Hamadani, A., Bijari, M.: A fuzzy intelligent approach to the classification problem in gene expression data analysis. *Knowl.-Based Syst.* **27**, 465–474 (2012). <https://doi.org/10.1016/j.knosys.2011.10.012>
2. Wang, Q., Ma, Y., Zhao, K., Tian, Y.: A comprehensive survey of loss functions in machine learning. *Data Sci, Ann* (2020). <https://doi.org/10.1007/s40745-020-00253-5>
3. Musbah, H., Aly, H.H., Little, T.A.: Energy management of hybrid energy system sources based on machine learning classification algorithms. *Electr. Power Syst. Res.* **199**, 107436 (2021). <https://doi.org/10.1016/j.epsr.2021.107436>
4. Song, K., Anderson, K., Lee, S.: An energy-cyber-physical system for personalized normative messaging interventions: identification and classification of behavioral reference groups. *Appl. Energy* **260**, 114237 (2020). <https://doi.org/10.1016/j.apenergy.2019.114237>

5. Chen, G., Zhu, Y., Wiedmann, T., Yao, L., Xu, L., Wang, Y.: Urban-rural disparities of household energy requirements and influence factors in China: classification tree models. *Appl. Energy* **250**, 1321–1335 (2019). <https://doi.org/10.1016/j.apenergy.2019.04.170>
6. Banihashemi, S., Ding, G., Wang, J.: Developing a hybrid model of prediction and classification algorithms for building energy consumption. *Energy Procedia* **110**, 371–376 (2017). <https://doi.org/10.1016/j.egypro.2017.03.155>
7. Wang, F., Zhang, Z., Liu, C., Yu, Y., Pang, S., Duić, N., Shafie-Khah, M., Catalão, J.P.: Generative adversarial networks and convolutional neural networks based weather classification model for day ahead short-term photovoltaic power forecasting. *Energy Convers. Manag.* **181**, 443–462 (2019). <https://doi.org/10.1016/j.enconman.2018.11.074>
8. Liu, Y., Zhou, Y., Wang, D., Wang, Y., Li, Y., Zhu, Y.: Classification of solar radiation zones and general models for estimating the daily global solar radiation on horizontal surfaces in China. *Energy Convers. Manag.* **154**, 168–179 (2017). <https://doi.org/10.1016/j.enconman.2017.10.043>
9. Yan, D., Jin, Y., Sun, H., Dong, B., Ye, Z., Li, Z., Yuan, Y.: Household appliance recognition through a Bayes classification model. *Sustain. Cities Soc.* **46**, 101393 (2019). <https://doi.org/10.1016/j.scs.2018.12.021>
10. Wang, W., Chen, J., Hong, T.: Modeling occupancy distribution in large spaces with multi-feature classification algorithm. *Build. Environ.* **137**, 108–117 (2018). <https://doi.org/10.1016/j.buildenv.2018.04.002>
11. Jiang, L., Yao, R.: Modelling personal thermal sensations using C-Support Vector Classification (C-SVC) algorithm. *Build. Environ.* **99**, 98–106 (2016). <https://doi.org/10.1016/j.buildenv.2016.01.022>
12. Shao, Z., Zheng, Q., Yang, S., Gao, F., Cheng, M., Zhang, Q., Liu, C.: Modeling and forecasting the electricity clearing price: A novel BELM based pattern classification framework and a comparative analytic study on multi-layer BELM and LSTM. *Energy Econ* **86**, 104648 (2020). <https://doi.org/10.1016/j.eneco.2019.104648>
13. Bai, L., Yang, L., Song, B., Liu, N.: A new approach to develop a climate classification for building energy efficiency addressing Chinese climate characteristics. *Energy* **195**, 116982 (2020). <https://doi.org/10.1016/j.energy.2020.116982>
14. de Paula Protásio, T., Scatolino, M.V., Lima, M.D.R., de Araújo, A.C.C., de Figueiredo, I.C.R., Bufalino, L., Hein, P.R.G., Trugilho, P.F.: Insights in quantitative indexes for better grouping and classification of Eucalyptus clones used in combustion and energy cogeneration processes in Brazil. *Biomass Bioenergy* **143**, 105835 (2020). <https://doi.org/10.1016/j.biombioe.2020.105835>
15. Sabia, G., Petta, L., Avolio, F., Caporossi, E.: Energy saving in wastewater treatment plants: a methodology based on common key performance indicators for the evaluation of plant energy performance, classification and benchmarking. *Energy Convers. Manag.* **220**, 113067 (2020). <https://doi.org/10.1016/j.enconman.2020.113067>
16. Patnaik, B., Mishra, M., Bansal, R.C., Jena, R.K.: MODWT-XGBoost based smart energy solution for fault detection and classification in a smart microgrid. *Appl. Energy* **285**, 116457 (2021). <https://doi.org/10.1016/j.apenergy.2021.116457>
17. Radhakrishnan, P., Ramaiyan, K., Vinayagam, A., Veerasamy, V.: A stacking ensemble classification model for detection and classification of power quality disturbances in PV integrated power network. *Measurement* **175**, 109025 (2021). <https://doi.org/10.1016/j.measurement.2021.109025>
18. Eskandari, A., Milimonfared, J., Aghaei, M.: Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on IV characteristics. *Sol. Energy* **211**, 354–365 (2020). <https://doi.org/10.1016/j.solener.2020.09.071>
19. Li, Y., Xue, W., Wu, T., Wang, H., Zhou, B., Aziz, S., He, Y.: Intrusion detection of cyber physical energy system based on multivariate ensemble classification. *Energy* **218**, 119505 (2021). <https://doi.org/10.1016/j.energy.2020.119505>
20. Bi, R., Zhou, C., Hepburn, D.M.: Detection and classification of faults in pitch-regulated wind turbine generators using normal behaviour models based on performance curves. *Renew. Energy* **105**, 674–688 (2017). <https://doi.org/10.1016/j.renene.2016.12.075>
21. Zhang, X., Lu, W., Pan, Y., Wu, H., Wang, R., Yu, R.: Empirical study on tangent loss function for classification with deep neural networks. *Comput. Electr. Eng.* **90**, 107000 (2021). <https://doi.org/10.1016/j.compeleceng.2021.107000>
22. Hazarika, B.B., Gupta, D.: Random vector functional link with ϵ -insensitive Huber loss function for biomedical data classification. *Comput Methods Program Biomed* **215**, 106622 (2022). <https://doi.org/10.1016/j.cmpb.2022.106622>
23. Ozyildirim, B.M., Kiran, M.: Levenberg–Marquardt multi-classification using hinge loss function. *Neural Netw.* **143**, 564–571 (2021). <https://doi.org/10.1016/j.neunet.2021.07.010>
24. de La Torre, J., Puig, D., Valls, A.: Weighted kappa loss function for multi-class classification of ordinal data in deep learning. *Pattern Recogn. Lett.* **105**, 144–154 (2018). <https://doi.org/10.1016/j.patrec.2017.05.018>
25. Liang, Z., Zhang, L.: Support vector machines with the ϵ -insensitive pinball loss function for uncertain data classification. *Neurocomputing* **457**, 117–127 (2021). <https://doi.org/10.1016/j.neucom.2021.06.044>
26. Deng, N., Tian, Y., Zhang, C.: Support vector machines: optimization based theory, algorithms, and extensions. CRC Press (2012)
27. Collobert, R., Sinz, F., Weston, J., Bottou, L.: Trading convexity for scalability. In: Proceedings of the 23rd international conference on machine learning, pp. 201–208 (2006). <https://doi.org/10.1145/1143844.1143870>
28. Shen, X., Niu, L., Qi, Z., Tian, Y.: Support vector machine classifier with truncated pinball loss. *Pattern Recogn.* **68**, 199–210 (2017). <https://doi.org/10.1016/j.patcog.2017.03.011>
29. Xu, G., Cao, Z., Hu, B.G., Principe, J.C.: Robust support vector machines based on the rescaled hinge loss function. *Pattern Recogn.* **63**, 139–148 (2017). <https://doi.org/10.1016/j.patcog.2016.09.045>
30. Zhang, G.P.: Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing* **50**, 159–175 (2003). [https://doi.org/10.1016/S0925-2312\(01\)00702-0](https://doi.org/10.1016/S0925-2312(01)00702-0)
31. Al-Andoli, M., Tan, Sh., Sim, K., Lim, Ch., Goh, P.: Parallel deep learning with a hybrid BP-PSO framework for feature extraction and malware classification. *Appl. Soft Comput.* **130**, 109756 (2022). <https://doi.org/10.1016/j.asoc.2022.109756>
32. Khashei, M.: Soft intelligent decision making. Ph.D. Thesis, Isfahan University of Technology, Department of Industrial and Systems Engineering (2012)
33. Hajirahimi, Z., Khashei, M.: An optimal hybrid bi-component series-parallel structure for time series forecasting. *IEEE Trans. Knowl. Data Eng.* (2023). <https://doi.org/10.1109/TKDE.2022.3231008>
34. Etemadi, S., Khashei, M.: Accuracy versus reliability-based modelling approaches for medical decision making. *Comput. Biol. Med.* **141**, 105138 (2022). <https://doi.org/10.1016/j.compbiomed.2021.105138>
35. Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [<http://archive.ics.uci.edu/ml/>]. Irvine, CA: University of California, School of Information and Computer Science.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.