

Article

A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min–Max Normalization

Mohammed Shantal, Zalinda Othman *  and Azuraliza Abu Bakar

Center for Artificial Intelligence Technology, Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi 43600, Selangor, Malaysia; moh.shantal@sebhau.edu.ly (M.S.); azuraliza@ukm.edu.my (A.A.B.)

* Correspondence: zalinda@ukm.edu.my

Abstract: In the realm of data analysis and machine learning, achieving an optimal balance of feature importance, known as feature weighting, plays a pivotal role, especially when considering the nuanced interplay between the symmetry of data distribution and the need to assign differential weights to individual features. Also, avoiding the dominance of large-scale traits is essential in data preparation. This step makes choosing an effective normalization approach one of the most challenging aspects of machine learning. In addition to normalization, feature weighting is another strategy to deal with the importance of the different features. One of the strategies to measure the dependency of features is the correlation coefficient. The correlation between features shows the relationship strength between the features. The integration of the normalization method with feature weighting in data transformation for classification has not been extensively studied. The goal is to improve the accuracy of classification methods by striking a balance between the normalization step and assigning greater importance to features with a strong relation to the class feature. To achieve this, we combine Min–Max normalization and weight the features by increasing their values based on their correlation coefficients with the class feature. This paper presents a proposed Correlation Coefficient with Min–Max Weighted (CCMMW) approach. The data being normalized depends on their correlation with the class feature. Logistic regression, support vector machine, k-nearest neighbor, neural network, and naive Bayesian classifiers were used to evaluate the proposed method. Twenty UCI Machine Learning Repository and Kaggle datasets with numerical values were also used in this study. The empirical results showed that the proposed CCMMW significantly improves the classification performance through support vector machine, logistic regression, and neural network classifiers in most datasets.

Keywords: data normalization; data standardization; feature weighting; correlation matrix; correlation coefficient; classification method; regression method



Citation: Shantal, M.; Othman, Z.; Bakar, A.A. A Novel Approach for Data Feature Weighting Using Correlation Coefficients and Min–Max Normalization. *Symmetry* **2023**, *15*, 2185. <https://doi.org/10.3390/sym15122185>

Academic Editors: Sunil Jha, Malgorzata Rataj, Xiaorui Zhang and Jian-Qiang Wang

Received: 16 October 2023
Revised: 27 November 2023
Accepted: 1 December 2023
Published: 11 December 2023



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

1. Introduction

Data preprocessing is one of the most crucial steps in machine learning. Using important and necessary methods to prepare raw data correctly will positively affect the output and improve a model's performance [1,2]. The data preparation stage includes many tasks, such as removing outliers and noise, integrating data from diverse sources, dealing with missing data, and transforming data into a scale suitable for analysis. This stage aims to eliminate the impact of systematic sources of variation as much as is feasible [3,4]. While data scaling is considered a treatment for domination issues in most cases, it will have a reverse impact when giving the same contribution to a feature that is irrelevant to the target feature. In this case, using feature weighting to select essential features is one of the solutions, and using it to balance a feature based on its importance is the other possible solution. In many data-mining algorithms, such as neural network algorithms, clustering, and distance-based algorithms, data rescaling is an essential stage prior to the training phase

to avoid misleading results and speed up the learning process [5–7]. To avoid numeric features dominating other numeric values, transforming all feature values into a standard range will give equal importance to all features [4,6]. Many research findings suggest that data normalization and standardization substantially impact classification and clustering results [8]. The effect of data normalization on the results of many machine learning (ML) methods has been studied, such as its effect on classification accuracy [9,10], neural network training [11,12], the clustering process [8], and outlier detection methods [13]. When data have been normalized, all features have an equal contribution to ML results, but it does not mean that these features are equally important. In some cases, data have many irrelevant and redundant features. Ref. [14] proposed the Adaptive Distinct Feature Normalization Strategy, which enhances the fusion of sparse topics and deep features by adaptively normalizing them separately. This results in a clearer feature description and decreases confusion in complex scenes. The presence of undesired features makes learning difficult and increases the feature's space. Generally, normalization gives all features the same contribution to a model. But removing any dominant features may reduce the model's performance. In this case, the behavior of data plays a vital role in the performance [15].

In the same context, feature weighting is one of the essential stages since it is used to adjust the feature values based on their contribution to the model and the result [16]. Many variable importance measurements are presented and categorized into sub-categories based on the techniques used. One of those techniques is the correlation coefficient, a parametric regression technique. Another different strategy is the nonparametric strategy which includes multiple techniques; one example is Random Forest [17]. Many researchers have used variable importance measurement strategies and applied them to enhance the classifier's performance, such as [18], naive Bayes text classifiers [19,20], the fuzzy clustering method, and feature weighting used for the neural network [21,22], with SVMs [23]. Also, feature weighting has been used as a feature selection strategy to know the influence of features on results and then exclude irrelevant, redundant features [24–26], as well as the information gain attribute [27].

In contrast, a correlation coefficient is a measurement tool used to look for relationship patterns between different characteristics [28]. Therefore, a bivariate study measures the degree of association between two variables and the direction of the relationship [29]. In this context, the correlation coefficient can be considered as a variable importance method amongst parametric regression techniques when used to indicate linear dependence [17]. Various types of correlation coefficient formulas have been presented depending on the data type (numerical, ordinal, or nominal) and the type of correlation (linear or non-linear relationship) [30]. Pearson's correlation coefficient is a measurement method to explore the linearity of a correlation [31] for numerical data only. Although the correlation coefficient has been used mainly in statistical analysis to discover relationships among variables, there are many other uses for correlation metrics, especially in data mining. For example, correlation coefficients have been used for feature selection [32–34], missing data imputation methods [35–37], and feature quality measurement to find the best splitting features and points of decision trees [38].

Normalization is a preprocessing step applied to data to give features equal importance and prevent the domination of a few features. The impact of normalization methods has been studied in many works [39–41]. The Min–Max normalization method is one of the best normalization methods that has been found to improve the performance of classifiers [42,43]. Data standardization has also produced better outcomes in neural network training, though the benefit decreases with increasing network and sample sizes [11]. Choosing the appropriate normalization method is essential, as it affects the performance of supervised learning algorithms [44]. Also, Ref. [40] examined the impact of four normalization techniques on forecasting. They found that standardization calculations have a high affectability, and that a cautious approach is needed to deal with the outcomes. The impact of normalization on the performance of different methods such as outlier detection, violent video classification, backpropagation neural networks, and classification performance has

been studied, and it was found that the performance of normalization depends on both the data and the technique used [13,45]. Due to the importance of normalization, [46] proposed a new normalization method to only deal with integer values. The authors of [47] proposed two normalization methods, a “new approach to Min-Max” and a “new approach to Decimal Scaling,” to reduce the impact of large feature values and represent them in a small range. These methods are based on the Min–Max and Decimal Scaling normalization techniques. The study performed well with the k-means clustering method, but only that method was used for evaluation.

At the same time, feature weighting is another best strategy to prepare data to improve a method’s accuracy. Feature weighting methods are strategies for correlation analysis and dimension reduction, where the weight represents the contribution of each feature dimension to classification [48]. Ref. [49] proposed feature weighting methods based on similarity calculations using k-means, fuzzy c-means, and mean shift clustering. The weight of each feature is obtained by calculating the difference between original data and cluster centers, dividing by the mean of differences, and multiplying by the feature value in the dataset.

Ref. [50] proposed a support vector machine (SVM) with an information gain approach to improve fraud detection accuracy in card transactions. The approach involves normalizing the data using Min–Max normalization and reducing features through information gain-based feature selection. Discretization is carried out before normalization. Ref. [51] proposed a two-stage strategy combining normalization and supervised feature weighting using the Pearson correlation coefficient (PCC) and Random Forest Feature Importance estimation. The first stage involves normalizing the data using standardization. In contrast, the second stage involves calculating feature weights through the PCC and Random Forest Feature Importance and multiplying each feature value by weight. Ref. [52] proposed a dynamic feature-weighting algorithm to address multi-label classification by minimizing an objective function to bring together samples with the same label and separate samples with different labels. The weight function is used to evaluate the method with a multi-label classifier. Ref. [53] proposed a Correlation-Based Hierarchical k-Nearest Neighbor Imputation (CoHiKNN) method that weights the distance based on the correlation between the features and the label. The correlation coefficient is used to weigh the distance and impute missing values using k-NN. The weighting strategy multiplies each difference between two points by the correlation coefficient value of the feature. Ref. [54] proposed a hybrid data-scaling method combining Min–Max and Cox-Box to improve fault classification performance in compressors. The Min–Max method rescales data to one range, while Cox-Box transforms non-linear distributions into normal ones.

However, although it is essential to normalize the data and give all features the same contribution to override the feature’s dominance over the results of the model, in many cases, the importance of variables varies from one feature to another. Also, normalization has some limitations, such as destroying the original structure of the dataset [50]. In addition, using the feature weight for selecting important features has some limitations, in that the selected features still have an equal contribution. Also, using a subset of the dataset can decrease the quality and results of the correlation coefficient due to low instance numbers, affecting the correct weight assigned to each feature [53].

In this work, we tried to balance the importance of normalization with the various importance of features and reduce feature domination by employing data standardization, as well as maximizing the feature contribution by maximizing the value based on its correlation coefficient. The maximization limit parameter is proposed to control the new range of feature values. We investigate the influence of normalization on the classification and regression methods using three data types in two phases. First, we implement the un-normalized data (raw data) and normalized data using one normalization method. In the second phase, the normalized data are used to weight the values depending on the association among the values with the class feature. This step calculates the correlation coefficient between the class feature and the rest of the features. This step increases the

contribution of each feature based on its relationship with the class. Each value will increase based on its correlation, where features with higher correlation will experience a greater increase compared to features with lower correlation. This weighting strategy significantly influences the model construction, directing more attention towards features exhibiting high correlation. The features with strong correlation will potentially increase the model's accuracy and achieve more power in classification algorithms.

2. Materials and Methods

A new weighting strategy is proposed using the correlation value between features for data representation. The new data representation values are used to increase the contribution of data. Each value will be maximized depending on the feature's actual value and Pearson correlation coefficient. Calculating the correlation between the class and the rest of the features is the initial step for weighting the feature values. The new maximum values will be calculated based on the actual and correlation coefficient values. First, all features will be normalized using Min–Max (0,1) to give the same contribution to all features, then increase the values using its correlation coefficient. The ensuing maximization step carefully calculates new maximum values by considering both the original feature values and their corresponding correlation coefficients. This dual-factor approach ensures that the augmented values strike a balance between the inherent importance of features and their contextual significance, enhancing the robustness of the data representation methodology. The strategy's dual consideration of raw feature values and correlation coefficients establishes a nuanced approach that not only maximizes the impact of each feature but also captures the intricacies of their relationships within the dataset.

As the Min–Max (0,1) normalization method represents all features in the same range (0–1), the proposed method will modify this concept, where each feature will have its own range. If we assume that f_1 is a feature with 0 correlation, all the values will represent the same value due to having no correlation with the class label. On the other hand, the features with a correlation of 1 with the class label will represent the new maximum value. The maximum value is determined based on the coefficient of change (C) and the correlation R -value. The relationship is direct between the coefficient of change and the maximum value, indicating that as the coefficient increases, the maximum value also increases, and vice versa. The maximum possible value for each feature corresponds to the correlation value of 1 (strong correlation) for that feature. In this scenario, strong correlation is the only case where new values can attain the maximum values.

Datasets and Experiments

The proposed method will calculate the correlation among features using the Pearson correlation coefficient (PCC) (Equation (1)) and Min–Max normalization (MMN) (Equation (2)). new_min and new_max denote the updated range of feature values, with new_min set to 0 and new_max set to 1 in the context of Min–Max scaling (0,1). The new weighted value is created using Equation (3).

$$r = \frac{\sum_{i=1}^m (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^m (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^m (y_i - \bar{y})^2}}. \quad (1)$$

$$MMN, x'_{i,n} = \frac{x_{i,n} - \min(x_i)}{\max(x_i) - \min(x_i)} (new_Max - new_Min) + new_Min. \quad (2)$$

$$\begin{aligned} \text{New weighted value, } v' &= \left(\frac{V_{i,j} - mi_j}{ma_j - mi_j} (new_max - new_min) + new_min \right) \\ &+ C \left(\frac{V_{i,j} - mi_j}{ma_j - mi_j} (new_max - new_min) + new_min \right) \cdot Corr(V_j, V_{target}). \end{aligned} \quad (3)$$

where the following are defined in Table 1:

Table 1. Definition of parameters in Equation (3).

v'	the new weighted value	mi_j	Minimum value of column j
$V_{i,j}$	each original value	ma_j	Maximum value of column j
C	the parameter to adjust the new range	new_min	New minimum value (0)
$Corr(V_j, V_{target})$	the correlation coefficient among column j and column label using PCC	new_max	New maximum value (1)

The following, Algorithm 1 explains the steps of the proposed method CCMMW.

Algorithm 1 CCMMW algorithm

Algorithm: Normalization + Weighting data using Min-Max01 and CC to improve the accuracy of classification methods
 Input: Un-Data: un-normalize data
 Output: CCoT Correlation Coefficient of Label feature, Normalized data (No-Data), Correlation Weighted
 Normalized Data (CCMMW-Data)
 CCoT = Calculate the correlation coefficient of each feature with the Label feature $corr(V_j, V_{Target})$ using the PCC method as in Equation (1)
 Calculate No-Data using MMN normalization method, Equation (2)
 For $i = 1$ to $n - 1$, where n is the number of features in the data
 Calculate CCMMW-Data using the proposed method as in Equation (3)

Assume a dataset (D) with six features (F1 to F6) where D is the normalized Min–Max (0,1) data. As a result of the MMN method, each small value of features is presented as 0, and the maximum value is presented as 1, as in the original normalized Min–Max 0,1 data in Table 2. Also, the correlation among the complete dataset is shown in Table 3, where each correlation coefficient (CC) value represents the correlation coefficient value between the feature and the class label feature. Table 2 shows three sets of values as the results of CCMMW. The first is CCMMW(1), where the table is produced with a C parameter value of 1. Also, CCMMW(5) is produced using the CCMMW method with a value of 5 for the C parameter. Finally, CCMMW(10) is the results of the CCMMW method using a value of 10 for the C parameter.

Table 2. CC values among features and the label feature using the proposed weight feature.

Original Normalized Min–Max (0,1) Data						CCMMW (1)						CCMMW (5)						CCMMW (10)					
F1	F2	F3	F4	F5	F6	F1	F2	F3	F4	F5	F6	F1	F2	F3	F4	F5	F6	F1	F2	F3	F4	F5	F6
0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.28	0.31	0.38	0.45	0.50	0.25	0.38	0.56	0.88	1.25	1.50	0.25	0.50	0.88	1.50	2.25	2.75
0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.55	0.63	0.75	0.90	1.00	0.50	0.75	1.13	1.75	2.50	3.00	0.50	1.00	1.75	3.00	4.50	5.50
0.75	0.75	0.75	0.75	0.75	0.75	0.75	0.83	0.94	1.13	1.35	1.50	0.75	1.13	1.69	2.63	3.75	4.50	0.75	1.50	2.63	4.50	6.75	8.25
1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.10	1.25	1.50	1.80	2.00	1.00	1.50	2.25	3.50	5.00	6.00	1.00	2.00	3.50	6.00	9.00	11.00

Table 3. CC among features and the label feature.

Feature	F1	F2	F3	F4	F5	F6
CC value	0.00	0.10	0.25	0.50	0.80	1.00

As shown in Figure 1, the algorithm aims to improve the accuracy of classification methods by applying normalization and correlation-based weighting techniques to the

input data. It starts by calculating the correlation coefficient between each feature and the target feature using the PCC method. Then, it normalizes the data using the Min–Max method to scale the values between 0 and 1. Next, it calculates the correlation-weighted normalized data for each feature, incorporating the correlation coefficients obtained. The algorithm outputs the correlation coefficient of the target feature, the normalized data, and the correlation-weighted normalized data. By normalizing the data and incorporating feature–target correlations, the algorithm aims to enhance the classification accuracy by giving more importance to relevant features in the classification process.

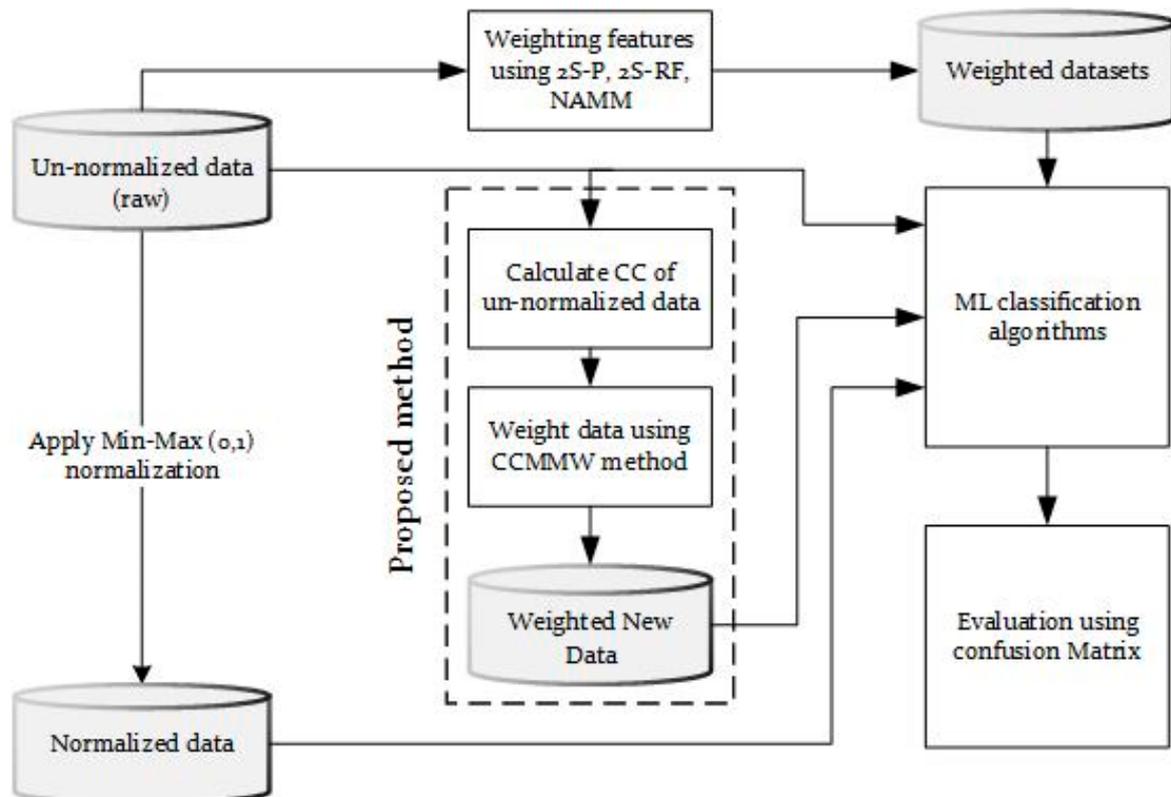


Figure 1. The general CCMMW methodology.

The general research methodology is presented in Figure 1; two strategies are used to transform the data prior to new data creation. One is transformed using existing and new approaches to normalization methods. The second uses CC values to weigh the features for representing the data, and then uses the new data to improve the accuracy of classification approaches. Then, the performance of the proposed method is compared using un-normalized and normalized data with different classifiers.

Experiments were conducted on twenty datasets comprising both discrete and continuous features to assess the proposed method’s consistency. These datasets were carefully selected from reputable sources, including the Machine Learning Repository UCI and Kaggle Repository databases. These databases are renowned for their diverse and well-curated datasets, making them valuable resources for empirical studies in this field. Table 4 provides an overview of the datasets. Also, various data sizes are considered, where various instances and features are used in the study.

In this experiment, to analyze the effect of weighting features based on its CC on the classification results, comparisons are made between the un-normalized (RAW) data, the normalized Min–Max (0,1) (MM) method, the Two-Stage Min–Max with Pearson (2S-P) method, the Two-Stage Min–Max with RF Feature Importance (2S-RF) method, the New Approach Min–Max Normalization method (NAMM), and the proposed CCMMW approach. All codes are written in Python 3.7. Jupyter (Anaconda 3). The SciKitLearn

Library was used to implement all methods in this study. Five classification methods were used to investigate the impact on the performance and the improvement of the proposed method. The classifiers are the logistic regression (LR), SVM, k-nearest neighbor (kNN), neural network (NN), and naive Bayes (NB) classifiers. Finally, all experimental results were performed on numerical datasets.

Table 4. Experiment datasets.

Datasets	Type of Data	# of Instances	# of Features	# Classes
Breast Cancer 1	real	569	30	2
QSAR	real	1055	41	2
Sonar	real	208	60	2
PARKINSON	real	195	22	2
Wine	Integer + real	6463	12	2
Monkey	Integer	556	6	2
German	real	1000	24	2
Musk	Integer	6598	166	2
liver	Integer + real	345	6	2
wholesale	Integer	440	8	2
Spam	real	4601	57	2
Heart s Cleveland	Integer + real	1190	11	2
Magic	real	19,020	10	2
Blood	Integer	748	4	2
Breast Cancer Coimbra	real	116	9	2
Vehicle	Integer	846	18	4
Bupa	Integer + real	345	6	2
Glass	real	214	9	6
Letter	Integer	20,000	16	26
Ecoli	real	336	7	8

In the evaluation stage, the effect of using the CC to weigh the normalized values was evaluated based on the type of algorithm that was applied to the data. For classification purposes, the result was evaluated based on the accuracy rate. Ten cross-validations and the experiment were repeated ten times, and the results were averaged for further analysis. For the k-NN classifier, 3 neighbors were considered as k values. In the kernel function of the SVM, RBF was used as the kernel parameter value, and all other settings used the default values.

3. Results

3.1. The Evaluation of Different Classifiers Based on the Best Result of CCMMW

The best results of CCMMW from 10 available values in the experiment results are shown in Figure 2. The ten results were due to the setting of the C parameter, which we used to adjust the new range of data. The impact of the various values of C is discussed in the further section of this discussion.

In the following results, the classification accuracy is presented. Bold values are the highest out of the six methods, and the CCMMW values are underlined when the value is the highest value after the raw values.

3.2. Logistic Regression (LR) Classifier

Table 5 shows that the CCMMW method obtained the highest accuracy in 12 out of the 20 datasets, and an extra four datasets if the raw data are excluded. It shows that the best accuracy improvement using the CCMMW + LR classifier compared with the MM method was 5.46%, obtained from the Letter dataset. The accuracy of MM was 71.46%, whereas CCMMW gave 76.92%. Although this case had the best raw data accuracy, the proposed method was still better than the other strategies. Also, the accuracy decreased in three cases out of twenty, whereas the worst was -1.24% .

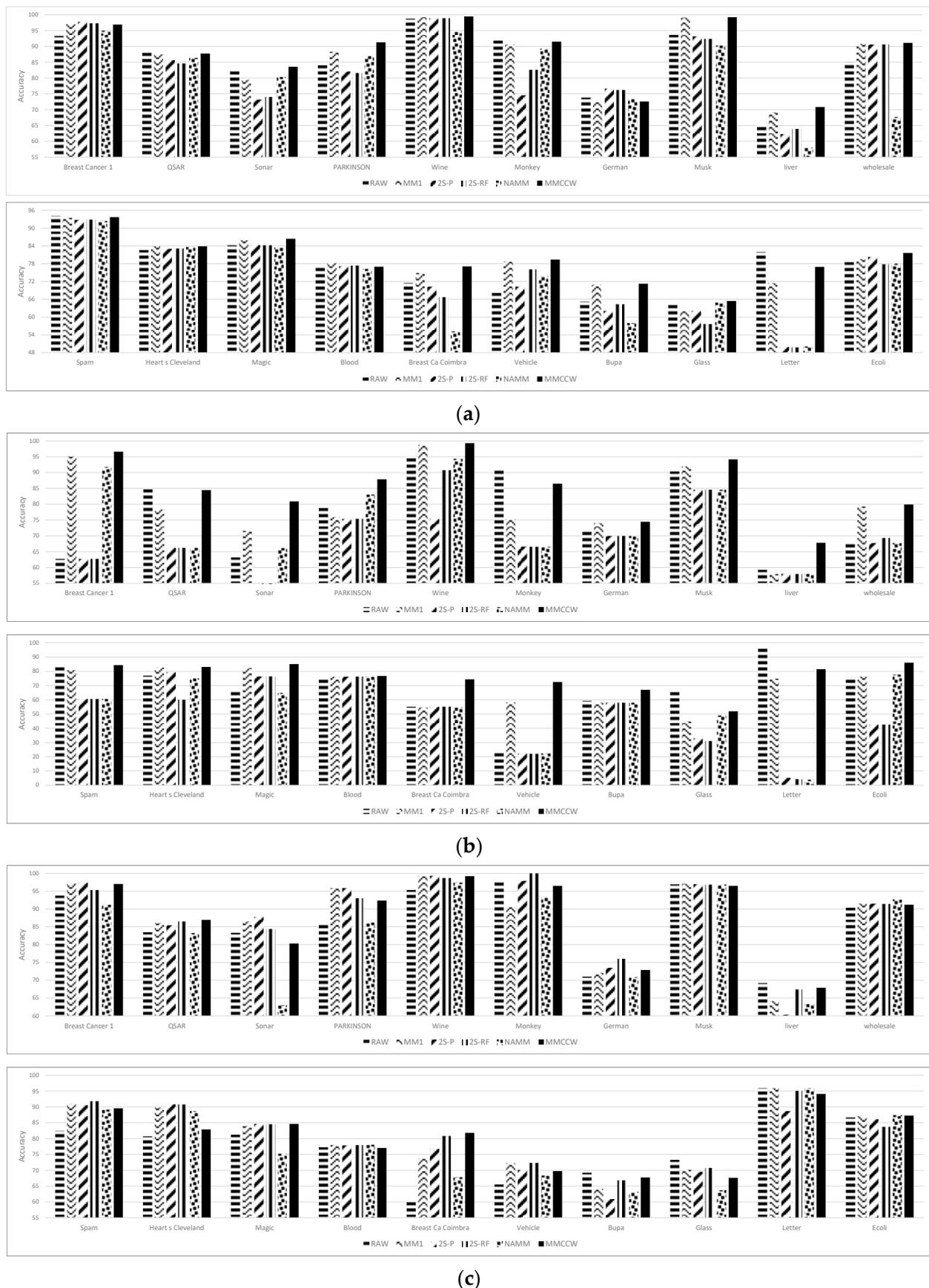


Figure 2. Cont.

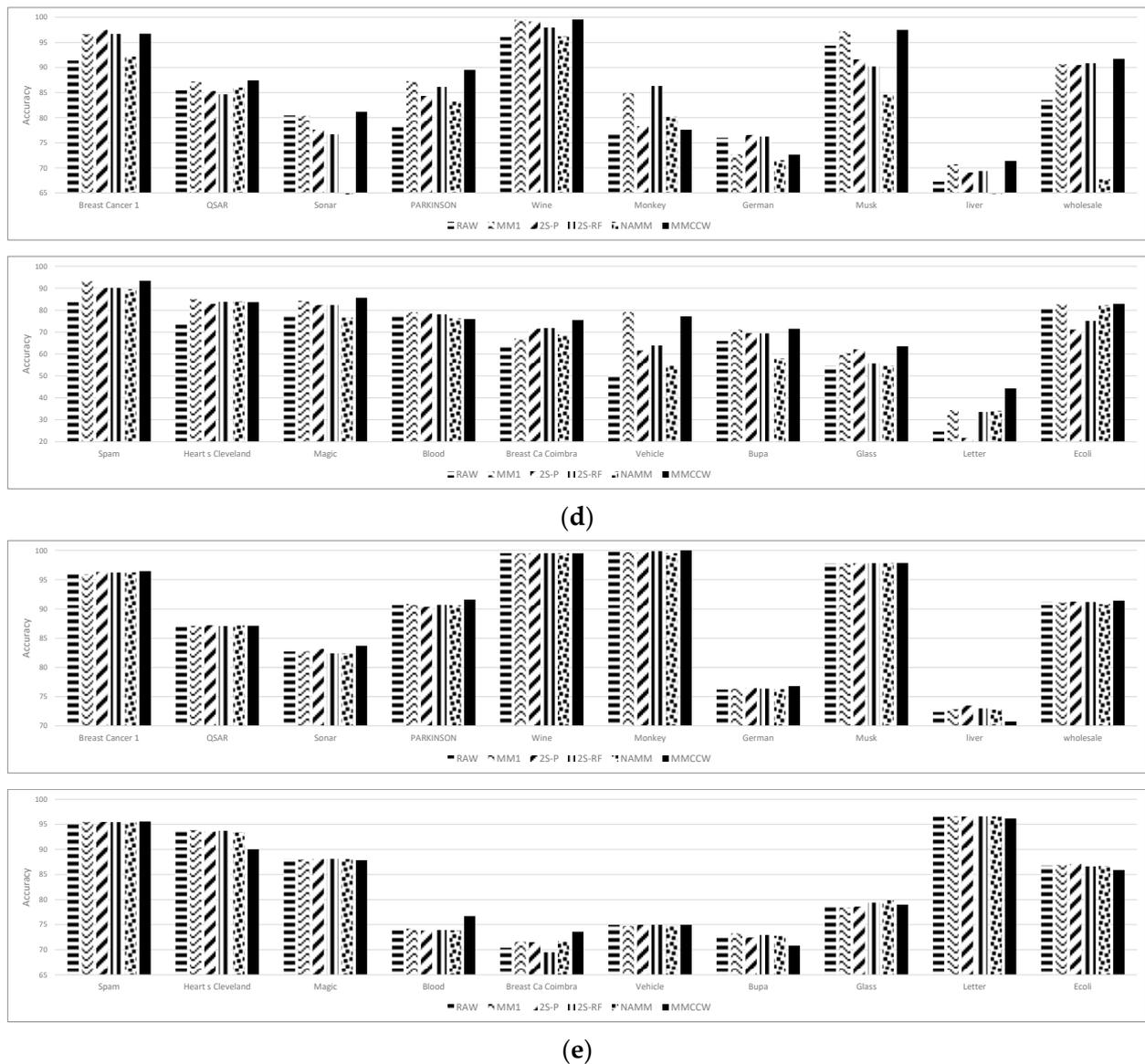


Figure 2. The CCMMW performance on different classifiers. (a) LR. (b) SVM. (c) k-NN. (d) NN. (e) NB.

As in Figure 2a, we find that, through a comparison between the proposed method and the MM normalization method, the proposed method excelled in most cases, as it excelled with 17 databases in this study, while MM excelled in only three cases, with an outperformance that did not exceed 1.24%. In addition, 2S-P outperformed in three cases while CCMMW outperformed in 17 cases, with a high difference in the Letter dataset, where 2S-P was 49.96% while CCMMW was 76.92%. Also, when comparing the results with 2S-RF, we find that, the same as with 2S-F, the proposed method outperformed 2S-RF with 17 datasets. In comparison, only 2S-RF outperforms in the Breast Cancer 1, German, and Blood datasets. The last compared method is the NAMM, where the proposed method performed well in all datasets except German, where the NAMM result was 73.53% while that of CCMMW was 72.56%. Only three differences between the NAMM and CCMMW had a less than 1% improvement, while all other results obtained were between 1.27 and 26.78% with Spam and Letter.

Table 5. Performance of CCMMW on logistic regression classifier.

Dataset	RAW	MM	2S-P	2S-RF	NAMM	CCMMW
Breast Cancer 1	93.29	97.05	97.73	97.29	94.97	96.89
QSAR	88.11	87.46	85.69	84.61	86.33	<u>87.75</u>
Sonar	82.16	79.42	73.19	74.02	80.32	83.56
PARKINSON	84.13	88.35	82.08	81.58	86.98	91.27
Wine	98.82	99.25	98.87	98.93	94.66	99.47
Monkey	91.88	90.64	74.55	82.59	89.34	<u>91.50</u>
German	73.82	72.35	76.66	76.20	73.53	72.56
Musk	94.33	99.12	93.18	92.38	90.26	99.27
liver	65.33	69.17	62.26	63.91	57.98	70.85
wholesale	84.98	90.75	90.59	90.61	67.73	91.14
Spam	94.10	93.47	92.72	92.86	92.41	<u>93.68</u>
Heart s Cleveland	82.58	83.86	82.99	83.13	83.65	<u>83.83</u>
Magic	84.22	85.87	84.28	84.21	83.48	86.40
Blood	76.88	78.22	77.16	77.38	76.21	76.98
Breast Ca Coimbra	71.49	74.80	70.28	66.74	55.19	77.05
Vehicle	68.14	78.68	70.24	76.08	73.85	79.41
Bupa	65.23	70.94	62.09	64.35	57.98	71.25
Glass	64.88	61.98	62.11	57.65	64.87	65.45
Letter	81.94	71.46	49.96	49.81	50.14	<u>76.92</u>
Ecoli	78.51	79.55	80.22	77.80	78.00	81.65

3.3. Support Vector Machine Classifier

Table 6 shows that CCMMW obtained the highest accuracy in 16 out of the 20 datasets, and an extra four datasets when raw data are excluded, when using model building based on weighting correlation features. When comparing CCMMW + SVM with the MM normalization method, as in Figure 2b, all the CCMMW results outperformed the MM results. The minimum increase was obtained with the German dataset, which was 0.38%, while the best result obtained was with Breast Ca Coimbra, which increased by 19.64%. Also, CCMMW outperformed 2S-P with all datasets. Some large differences were observed, such as 76.22%, 50.15%, 43.50%, and 27.49% with the Letter, Vehicle, Ecoli, and Sonar datasets, due to the bad performance of 2S-P with the SVM. The same was observed with 2S-RF, where some large differences in accuracy were obtained compared with CCMMW. Also, due to the bad performance of 2S-RF with the SVM, the difference reached 77.12% with the Letter dataset. The lower difference obtained with 2S-RF was 0.49%. Finally, compared with the NAMM, all CCMMW results outperformed the NAMM results. Overall, CCMMW+SVM outperformed the other data preparation methods, increasing the accuracy compared with the rest of the normalization and feature weighting methods. The significant results obtained from all datasets showed an improvement in accuracy. Only the raw data and CCMMW produced the best results with all datasets. The performance of CCMMW with the SVM increased significantly, such as 19.64%, 14.24%, 11.97%, and 11.18 with Breast Ca Coimbra, Vehicle, PARKINSON, and Monkey, respectively.

3.4. k-Nearest Neighbor (k-NN) Classifier

Table 7 shows that CCMMW obtained the highest accuracy in only two out of the 20 datasets, and an extra two datasets if the raw data are excluded. As shown in Table 7 and Figure 2c, the performance of CCMMW with the k-NN classifier compared with MM shows us that CCMMW was only improved with the QSAR, Liver, Breast Ca Coimbra, and Bupa datasets. The accuracy of other datasets obtained by CCMMW was less than the accuracy of MM. Also, compared to 2S-P and 2S-RF, CCMMW only outperformed six and eight datasets, respectively. The NAMM was the only method that CCMMW outperformed, whereas the proposed method outperformed in 14 out of 20 datasets.

Table 6. Performance of CCMMW on SVM classifier.

Dataset	RAW	MM	2S-P	2S-RF	NAMM	CCMMW
Breast Cancer 1	62.74	95.11	62.74	62.74	91.79	96.61
QSAR	85.41	78.34	66.25	66.25	66.25	<u>84.47</u>
Sonar	63.19	71.54	53.38	52.69	66.13	80.87
PARKINSON	78.86	75.89	75.38	75.39	83.07	87.86
Wine	95.03	98.81	75.35	90.77	94.37	99.30
Monkey	91.37	75.30	66.56	66.54	66.55	<u>86.48</u>
German	71.78	74.07	70.00	70.00	70.00	74.45
Musk	90.39	91.93	84.59	84.59	84.59	94.16
liver	59.28	57.97	57.98	57.95	58.00	67.80
wholesale	67.73	79.27	67.73	69.32	67.73	79.93
Spam	83.90	80.92	60.60	60.60	60.60	84.30
Heart s Cleveland	76.97	82.62	79.50	59.96	75.10	83.03
Magic	65.87	82.28	76.38	76.49	64.84	85.08
Blood	75.24	76.21	76.21	76.21	76.20	76.70
Breast Ca Coimbra	55.20	54.67	55.18	55.17	54.69	74.31
Vehicle	23.99	58.22	22.33	22.03	22.31	72.48
Bupa	59.34	57.98	57.95	57.97	57.97	67.05
Glass	65.75	44.93	32.88	31.06	48.97	<u>51.95</u>
Letter	97.52	74.82	5.20	4.30	3.85	<u>81.42</u>
Ecoli	75.84	76.22	42.55	42.56	77.71	86.05

Table 7. Performance of CCMMW on k-NN classifier.

Dataset	RAW	MM	2S-P	2S-RF	NAMM	CCMMW
Breast Cancer 1	93.75	97.24	97.38	95.36	91.20	97.02
QSAR	83.47	86.06	85.43	86.53	83.25	86.97
Sonar	83.34	86.53	87.83	84.40	63.02	80.28
PARKINSON	85.75	95.86	95.89	93.01	86.20	92.37
Wine	95.34	99.23	99.32	98.71	97.42	99.23
Monkey	97.59	90.61	97.88	100.00	93.27	96.50
German	71.00	72.12	73.48	76.05	70.87	72.89
Musk	96.92	97.08	96.97	96.84	96.87	96.48
liver	69.20	64.15	60.30	67.44	63.37	<u>67.87</u>
wholesale	90.39	91.55	91.50	91.43	92.70	<u>91.20</u>
Spam	82.47	90.88	90.55	91.83	89.20	89.57
Hearts Cleveland	80.74	89.79	90.86	90.78	88.71	82.89
Magic	81.23	83.99	84.67	84.56	75.24	84.64
Blood	77.23	78.00	77.85	77.92	78.06	77.05
Breast Ca Coimbra	59.97	73.64	76.78	80.84	67.82	81.82
Vehicle	65.50	72.26	70.12	72.32	68.31	69.73
Bupa	69.24	64.12	60.89	66.84	63.24	<u>67.75</u>
Glass	73.33	70.18	69.50	70.74	63.68	67.61
Letter	96.00	96.01	88.73	95.09	95.93	94.11
Ecoli	86.71	87.09	86.14	83.73	87.47	87.26

Overall, the two improvements of the CCMMW + k-NN classifier were with the Breast Ca Coimbra dataset, while the improvement with QSAR was 0.91%. Another two improvements compared with the other methods (except for the raw data) were the liver and Bupa datasets, which reached 3.72% and 3.63% improvement, respectively. As seen in Figure 2c, the performance of CCMMW with the k-NN method had no improvement in general, where the results obtained by CCMMW were the worst when comparing most numbers.

3.5. Neural Network (NN) Classifier

The first thing we figured out from Figure 2d is that the raw data did not produce any of the best values in all datasets. Also, Table 8 show that Improvements were obtained with

most of the datasets, where the accuracy increased with 15 datasets. The best improvement when using CCMMW with the NN classifier was 9.91% with the Letter dataset, while the second-best improvement was with Breast Ca Coimbra. The worst decrease with the NN classifier was 7.30%, where the accuracy decreased from 84.91% to 77.61%. Also, compared with the 2S-P method, the results show that CCMMW outperformed 2S-P, whereas, with 16 datasets, CCMMW produced higher accuracy than 2S-P. As with 2S-P, the comparison between 2S-RF and CCMMW favored the proposed method as it also outperformed when applied to most of the datasets. Although the NAMM outperformed CCMMW in three datasets (Monkey: 2.62% higher; Hearts Cleveland: 0.24% higher; and Blood: 0.32% higher), CCMMW performed as expected with the rest of the datasets. Some of the performance differences were high, such as 24% higher with Wholesale, 23.78% with Sonar, and 22.39% with Vehicle. Overall, we can conclude that CCMMW performed better when applied to data when using the NN classifier.

Table 8. Performance of CCMMW on neural network classifier.

Dataset	RAW	MM	2S-P	2S-RF	NAMM	CCMMW
Breast Cancer 1	91.39	96.70	97.50	96.73	92.20	96.77
QSAR	85.48	87.19	85.33	84.67	86.01	87.44
Sonar	80.45	80.39	77.64	76.68	57.38	81.16
PARKINSON	78.28	87.30	84.36	86.14	83.29	89.54
Wine	96.11	99.55	99.16	98.00	96.23	99.57
Monkey	77.13	84.91	78.32	86.33	80.23	77.61
German	76.02	72.73	76.58	76.26	71.52	72.62
Musk	94.36	97.26	91.64	90.23	84.59	97.50
liver	67.29	70.72	69.10	69.36	57.99	71.38
wholesale	83.55	90.66	90.50	90.86	67.73	91.73
Spam	83.64	93.18	90.16	90.16	89.51	93.40
Heart s Cleveland	73.61	85.08	82.89	83.77	83.90	83.66
Magic	77.93	84.25	82.35	82.32	76.60	85.65
Blood	78.20	79.05	78.34	78.14	76.27	75.95
Breast Ca Coimbra	64.01	67.11	71.56	71.83	68.85	75.40
Vehicle	50.64	79.14	61.62	63.87	54.78	77.17
Bupa	65.89	71.17	69.47	69.45	57.97	71.51
Glass	54.53	60.36	62.09	55.73	54.68	63.57
Letter	24.62	34.36	21.77	33.60	34.11	44.27
Ecoli	81.34	82.67	71.16	75.06	82.31	82.82

3.6. Naive Bayes Classifier

Table 9 shows that CCMMW obtained the highest accuracy in 10 out of the 20 datasets. Compared with the rest of the normalization and raw methods, as in Figure 2e, CCMMW produced high accuracy, outperforming others. Ten datasets were given high accuracy when using CCMMW with the NB classifier, while MM exceeded with only two datasets: Wine and Magic. The best accuracy increase for CCMMW with the NBC was 2.44% with Blood. Weighting features using 2S-P produced higher accuracy than CCMMW with eight datasets, while 2S-RF outperformed with eleven datasets. The same happened with the NAMM, which obtained the two best accuracies overall, 97.86% with the Musk dataset and 79.86% with the Glass dataset.

Table 9. Performance of CCMMW on naive Bayes classifier.

Dataset	RAW	MM	2S-P	2S-RF	NAMM	CCMMW
Breast Cancer 1	96.08	95.91	96.33	96.22	96.24	96.45
QSAR	87.00	87.13	87.20	87.05	87.17	87.13
Sonar	82.73	82.77	83.16	82.42	82.38	83.71
PARKINSON	91.06	90.85	90.38	90.70	90.63	91.58

Table 9. Cont.

Dataset	RAW	MM	2S-P	2S-RF	NAMM	CCMMW
Wine	99.52	99.50	99.50	99.52	99.50	99.51
Monkey	99.79	99.68	99.66	99.87	99.53	100.0
German	76.22	76.33	76.47	76.39	76.28	76.81
Musk	97.81	97.81	97.84	97.83	97.86	97.86
liver	72.36	72.88	73.46	72.99	72.80	70.76
wholesale	91.23	91.11	91.25	91.18	90.80	91.39
Spam	95.44	95.49	95.47	95.42	95.41	95.55
Heart s Cleveland	93.79	93.85	93.57	93.71	93.41	90.01
Magic	88.15	88.05	88.09	88.12	88.09	87.85
Blood	73.88	74.25	73.82	73.95	73.89	76.69
Breast Ca Coimbra	70.39	71.61	71.47	69.48	71.67	73.58
Vehicle	74.93	74.77	74.96	74.98	74.82	74.96
Bupa	72.42	73.23	72.47	72.91	72.77	70.81
Glass	78.78	78.40	78.61	79.39	79.86	78.98
Letter	96.62	96.60	96.57	96.61	96.60	96.18
Ecoli	86.79	86.91	87.13	86.63	86.73	85.89

4. Discussion

For an overall discussion, we calculated the average accuracy of the methods with the twenty datasets.

As shown in Figure 3, the average accuracy is presented as an indicator of the accuracy results on various datasets. Also, Table 10 shows the average accuracy in numbers. The Logistic regression outperformed other preparation methods, where it had the best results in 12 out of the 20 datasets, and an extra four results outperformed other methods, except for the raw data. In the second method, the SVM classifier shows a clear improvement in accuracy when using the CCMMW strategy. Out of the 20 datasets, 16 gave the best accuracy with the SVM. In comparison, an extra four datasets gave the best accuracy when raw data are excluded. For the k-NN classifier, the proposed method's accuracy was slightly lower than the others. For the NN classifier, the results still show a good improvement in the accuracy of CCMMW. In 15 out of 20 datasets, its accuracy outperformed the other methods. Finally, although the averaged results present the proposed method as the fourth-best result, this is in part due to a lower difference between the results, where its accuracy with 11 datasets outperforms all other methods.

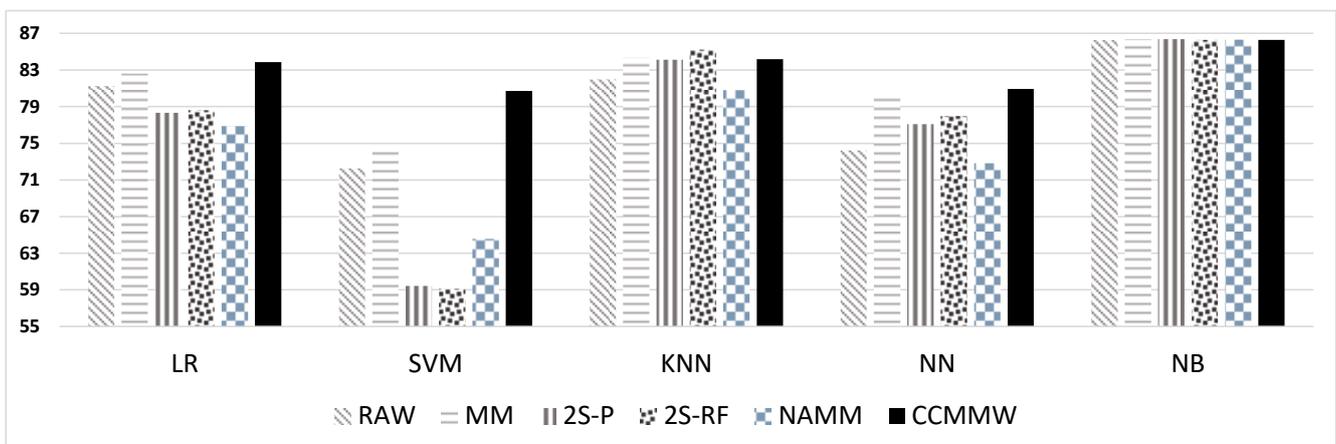


Figure 3. CCMMW performance: average accuracy (20 datasets) of five classification methods (classifiers).

Table 10. Average classification accuracy of data preparation method.

Method	RAW	MM	2S-P	2S-RF	NAMM	CCMMW
Logistic Regression	81.24	82.62	78.34	78.61	76.9	83.84
SVM	72.27	74.36	59.44	59.13	64.54	80.72
KNN	81.96	84.32	84.10	85.22	80.79	84.16
NN	74.22	80.19	77.08	77.96	72.81	80.94
NB	86.25	86.36	86.37	86.27	86.32	86.29

The Effect of C Parameter on the CCMMW Results

As we see in Figures 4a–e and 5a–e, the primary role of parameter C is to adjust the range of the data, since the relationship between C and the maximum value of the new range is direct, so the range increases as the value of C increases.

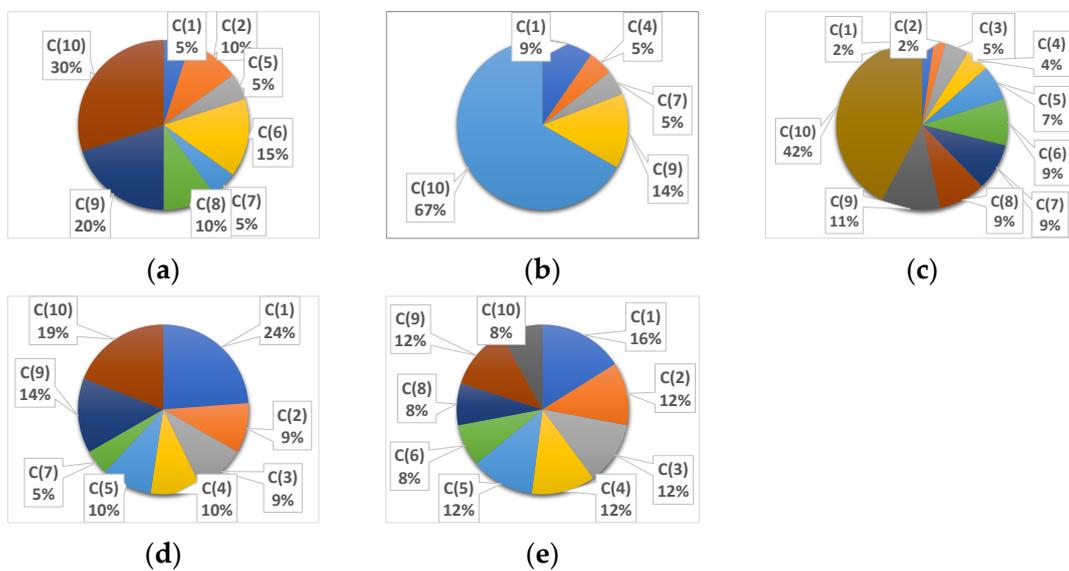


Figure 4. The value of C. (a) LR. (b) SVM. (c) k-NN. (d) NN. (e) NB.

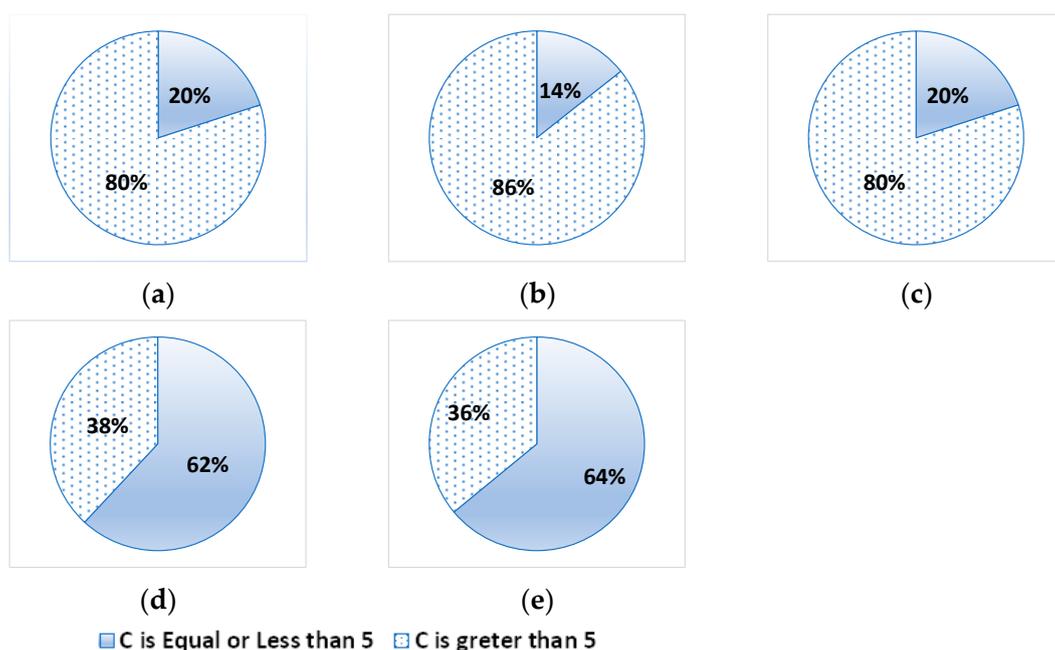


Figure 5. Value range of C. (a) LR. (b) SVM. (c) k-NN. (d) NN. (e) NB.

As we see in Figure 4a, the best results of CCMMW were obtained with a high C value of 10. Most of the best results were obtained with C values of more than 5, as in Figure 4a (around 80%). Only three of the best results were obtained with C values less than 5, meaning that increasing the range of values by increasing the C parameter values will positively impact the results with the LR classifier. Like LR, CCMMW with a high C parameter value gave good results with the SVM classifier. As shown in Figure 4b, most of the best results with the SVM were obtained specifically with a C value of 10, which is 67% of the result, as shown in Figure 5b. Only with the German and Wholesale datasets was the best accuracy result reached with a C value of 1. In some cases, multiple C values obtained the same result. All results of the SVM show that increasing the C parameter's value positively impacted the classifier's results, where all results showed increasing accuracy. Although the results of CCMMW with the k-NN classifier did not outperform other classifiers, where it is considered as having a weak impact on the results among the classifiers, the best results with CCMMW were obtained by setting a high C parameter value of 10, as seen in Figure 5c, where it represents 80% of the high results. All of the higher results were obtained with C values of 10, as seen in Figure 4c. In contrast to the previous methods, most of the good results of the proposed method were obtained by using a value of 5 or less for C, which is around 62% of the higher results, as seen in Figure 5d. Although some of the higher accuracies of CCMMW were obtained with C values greater than 5, the best results of the other methods with the NN were with C values less than 5. We can conclude that the impact of the C values on the NN were not clearly impactful, which means that adjusting the range towards higher possible values is not the best approach, as seen in Figure 4d. Figure 5e illustrates that higher accuracy was mostly achieved with small C values. For the naive Bayes (NB) model, the optimal result was obtained at $C = 1$, yielding 16%, whereas for C values of 2, 3, 4, and 5, the accuracy remained consistent at 12% each. This cumulative value amounts to 64% in total. According to the experimental results, data normalization has the potential to develop prediction models with the highest prediction accuracy. However, compared to the results of models with normalized and non-normalized data, the accuracy was improved based on many factors such as the used normalization method, the type of data, and the classification methods.

5. Conclusions

Our experiment discusses the impact of the maximization of the role of features based on their correlation. The impact of the combined data normalization and feature weight is apparent. Although normalization impacted the results positively by giving an equal contribution of features to avoid eliminating features with high range values, the weighted features outperform those methods by increasing the contribution of the features based on the dependency measurement of features, such as the CC of importance of feature measurement. In this study, we have presented a novel feature weighting with normalization method named the Correlation Coefficient Min–Max Weighted (CCMMW) method. The relation between the correlation in the normalization area has not received much attention in the literature. We used the CC values to give features with a strong association more contribution to the learning step in ML methods to improve the performance of those methods. We used the LR, SVM, k-NN, NN, and NB classification methods to evaluate the effect of CCMMW. The performance improvement of the SVM, NN, and LR classifiers was clear, as most results showed increased accuracy. Only the k-NN classifier's accuracy results were unsuitable, where the proposed method outperformed the Min–Max normalization method in 40% of the datasets. Still, other normalization and weighted methods outperform CCMMW in most cases. Also, adjusting the upper limit of the feature's maximum value plays the main role in reaching the best result of the proposed method. In future work, various normalization methods could be explored, as each method offers distinct advantages that may impact the outcomes of data transformation. Additionally, incorporating alternative weight measurement methods may enhance the significance of features in constructing classification models, potentially leading to improved accuracy.

Author Contributions: Conceptualization, methodology, and software: M.S.; validation: M.S., Z.O. and A.A.B.; formal analysis, investigation, resources, data curation, and writing—original draft preparation: M.S.; writing—review and editing: Z.O. and A.A.B.; supervision: Z.O. and A.A.B.; funding acquisition: Z.O. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded in part by the Malaysian Ministry of Higher Education under the Fundamental Research Grant Scheme (FRGS): FRGS/1/2021/ICT06/UKM/02/1; The Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia: FTM1 and Center for Artificial Intelligence Technology, Universiti Kebangsaan Malaysia: FTM3.

Data Availability Statement: The data presented in this study are openly available in Machine Learning Repository UCI and Kaggle Repository databases.

Acknowledgments: We would like to thank the Data Mining and Optimization Lab of the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, for the expert knowledge sharing.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

CCMMW	Correlation Coefficient with Min–Max Weighted
PCC	Pearson correlation coefficient
MMN	Min–Max normalization
CC	correlation coefficient
2S-P	Two-Stage Min–Max with Pearson
2S-RF	Two-Stage Min–Max with RF Feature Importance
NAMM	New Approach Min–Max
NB	naive Bayes
NN	neural network
LR	logistic regression
kNN	k-nearest neighbor

References

- Niño-Adan, I.; Manjarres, D.; Landa-Torres, I.; Portillo, E. Feature weighting methods: A review. *Expert Syst. Appl.* **2021**, *184*, 115424. [\[CrossRef\]](#)
- Han, T.; Xie, W.; Pei, Z. Semi-supervised adversarial discriminative learning approach for intelligent fault diagnosis of wind turbine. *Inf. Sci.* **2023**, *648*, 119496. [\[CrossRef\]](#)
- Muralidharan, K. A note on transformation, standardization and normalization. *Int. J. Oper. Quant. Manag.* **2010**, *IX*, 116–122.
- García, S.; Luengo, J.; Herrera, F. *Data Preprocessing in Data Mining*; Springer: Cham, Switzerland, 2015; Volume 72.
- Mohamad Mohsin, M.F.; Hamdan, A.R.; Abu Bakar, A. The Effect of Normalization for Real Value Negative Selection Algorithm. In *Soft Computing Applications and Intelligent Systems*; Noah, S.A., Abdullah, A., Arshad, H., Abu Bakar, A., Othman, Z.A., Sahran, S., Omar, N., Othman, Z., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 194–205.
- Han, J.; Kamber, M.; Pei, J. 3—Data Preprocessing. In *Data Mining*, 3rd ed.; Han, J., Kamber, M., Pei, J., Eds.; Morgan Kaufmann: Boston, MA, USA, 2012; pp. 83–124. [\[CrossRef\]](#)
- Cui, Y.; Wu, D.; Huang, J. Optimize TSK Fuzzy Systems for Classification Problems: Minibatch Gradient Descent With Uniform Regularization and Batch Normalization. *IEEE Trans. Fuzzy Syst.* **2020**, *28*, 3065–3075. [\[CrossRef\]](#)
- Trebuña, P.; Halčinová, J.; Fil’o, M.; Markovič, J. The importance of normalization and standardization in the process of clustering. In Proceedings of the 2014 IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMII), Herl’any, Slovakia, 23–25 January 2014; pp. 381–385.
- Adeyemo, A.; Wimmer, H.; Powell, L.M. Effects of Normalization Techniques on Logistic Regression in Data Science. *J. Inf. Syst. Appl. Res.* **2019**, *12*, 37.
- Rajeswari, D.; Thangavel, K. The Performance of Data Normalization Techniques on Heart Disease Datasets. *Int. J. Adv. Res. Eng. Technol.* **2020**, *11*, 2350–2357.
- Shanker, M.; Hu, M.Y.; Hung, M.S. Effect of data standardization on neural network training. *Omega* **1996**, *24*, 385–397. [\[CrossRef\]](#)
- Yao, J.; Han, T. Data-driven lithium-ion batteries capacity estimation based on deep transfer learning using partial segment of charging/discharging data. *Energy* **2023**, *271*, 127033. [\[CrossRef\]](#)

13. Kandanaarachchi, S.; Muñoz, M.A.; Hyndman, R.J.; Smith-Miles, K. On normalization and algorithm selection for unsupervised outlier detection. *Data Min. Knowl. Discov.* **2020**, *34*, 309–354. [[CrossRef](#)]
14. Zhu, Q.; Zhong, Y.; Zhang, L.; Li, D. Adaptive Deep Sparse Semantic Modeling Framework for High Spatial Resolution Image Scene Classification. *IEEE Trans. Geosci. Remote Sens.* **2018**, *56*, 6180–6195. [[CrossRef](#)]
15. Singh, D.; Singh, B. Investigating the impact of data normalization on classification performance. *Appl. Soft Comput.* **2020**, *97*, 105524. [[CrossRef](#)]
16. Dialameh, M.; Jahromi, M.Z. A general feature-weighting function for classification problems. *Expert Syst. Appl.* **2017**, *72*, 177–188. [[CrossRef](#)]
17. Wei, P.; Lu, Z.; Song, J. Variable importance analysis: A comprehensive review. *Reliab. Eng. Syst. Saf.* **2015**, *142*, 399–432. [[CrossRef](#)]
18. Zhang, L.; Jiang, L.; Li, C.; Kong, G. Two feature weighting approaches for naive Bayes text classifiers. *Knowl.-Based Syst.* **2016**, *100*, 137–144. [[CrossRef](#)]
19. Nataliani, Y.; Yang, M.-S. Feature-Weighted Fuzzy K-Modes Clustering. In Proceedings of the 2019 3rd International Conference on Intelligent Systems, Metaheuristics & Swarm Intelligence, Male, Maldives, 23–24 March 2019; pp. 63–68.
20. Malarvizhi, K.; Amshakala, K. Feature Linkage Weight Based Feature Reduction using Fuzzy Clustering Method. *J. Intell. Fuzzy Syst.* **2021**, *40*, 4563–4572. [[CrossRef](#)]
21. Zeng, X.; Martinez, T.R. Feature weighting using neural networks. In Proceedings of the 2004 IEEE International Joint Conference on Neural Networks (IEEE Cat. No. 04CH37541), Budapest, Hungary, 25–29 July 2004; pp. 1327–1330.
22. Dalwinder, S.; Birmohan, S.; Manpreet, K. Simultaneous feature weighting and parameter determination of neural networks using ant lion optimization for the classification of breast cancer. *Biocybern. Biomed. Eng.* **2020**, *40*, 337–351. [[CrossRef](#)]
23. Zhang, Q.; Liu, D.; Fan, Z.; Lee, Y.; Li, Z. Feature and sample weighted support vector machine. In *Knowledge Engineering and Management*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 365–371.
24. Wang, J.; Wu, L.; Kong, J.; Li, Y.; Zhang, B. Maximum weight and minimum redundancy: A novel framework for feature subset selection. *Pattern Recognit.* **2013**, *46*, 1616–1627. [[CrossRef](#)]
25. Wang, Y.; Feng, L. A new hybrid feature selection based on multi-filter weights and multi-feature weights. *Appl. Intell.* **2019**, *49*, 4033–4057. [[CrossRef](#)]
26. Singh, D.; Singh, B. Hybridization of feature selection and feature weighting for high dimensional data. *Appl. Intell.* **2019**, *49*, 1580–1596. [[CrossRef](#)]
27. Othman, Z.; Shan, S.W.; Yusoff, I.; Kee, C.P. Classification techniques for predicting graduate employability. *Int. J. Adv. Sci. Eng. Inf. Technol.* **2018**, *8*, 1712–1720. [[CrossRef](#)]
28. Swesi, I.; Abu Bakar, A. Feature Clustering for PSO-Based Feature Construction on High-Dimensional Data. *J. Inf. Commun. Technol.* **2019**, *18*, 439–472. [[CrossRef](#)]
29. Schober, P.; Boer, C.; Schwarte, L.A. Correlation coefficients: Appropriate use and interpretation. *Anesth. Analg.* **2018**, *126*, 1763–1768. [[CrossRef](#)] [[PubMed](#)]
30. Khamis, H. Measures of association: How to choose? *J. Diagn. Med. Sonogr.* **2008**, *24*, 155–162. [[CrossRef](#)]
31. Ratner, B. The correlation coefficient: Its values range between $+1/-1$, or do they? *J. Target. Meas. Anal. Mark.* **2009**, *17*, 139–142. [[CrossRef](#)]
32. Hall, M.A. *Correlation-Based Feature Selection of Discrete and Numeric Class Machine Learning*; Department of Computer Science, University of Waikato: Hamilton, New Zealand, 2000.
33. Saidi, R.; Bouaguel, W.; Essoussi, N. Hybrid Feature Selection Method Based on the Genetic Algorithm and Pearson Correlation Coefficient. In *Machine Learning Paradigms: Theory and Application*; Hassanién, A.E., Ed.; Springer International Publishing: Cham, Switzerland, 2019; pp. 3–24. [[CrossRef](#)]
34. Hsu, H.-H.; Hsieh, C.-W. Feature Selection via Correlation Coefficient Clustering. *J. Softw.* **2010**, *5*, 1371–1377. [[CrossRef](#)]
35. Rahman, G.; Islam, Z. A decision tree-based missing value imputation technique for data pre-processing. In Proceedings of the Ninth Australasian Data Mining Conference, Ballarat, Australia, 1–2 December 2011; pp. 41–50.
36. Chen, X.; Wei, Z.; Li, Z.; Liang, J.; Cai, Y.; Zhang, B. Ensemble correlation-based low-rank matrix completion with applications to traffic data imputation. *Knowl.-Based Syst.* **2017**, *132*, 249–262. [[CrossRef](#)]
37. Sefidian, A.M.; Daneshpour, N. Estimating missing data using novel correlation maximization based methods. *Appl. Soft Comput.* **2020**, *91*, 106249. [[CrossRef](#)]
38. Mu, Y.; Liu, X.; Wang, L. A Pearson's correlation coefficient based decision tree and its parallel implementation. *Inf. Sci.* **2018**, *435*, 40–58. [[CrossRef](#)]
39. Pan, J.; Zhuang, Y.; Fong, S. The Impact of Data Normalization on Stock Market Prediction: Using SVM and Technical Indicators. In *Soft Computing in Data Science*; Berry, M.W., Mohamed, A.H., Yap, B.W., Eds.; Springer: Singapore, 2016; pp. 72–88.
40. Kumari, B.; Swarnkar, T. Importance of data standardization methods on stock indices prediction accuracy. In *Advanced Computing and Intelligent Engineering*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 309–318.
41. Singh, D.; Singh, B. Effective and efficient classification of gastrointestinal lesions: Combining data preprocessing, feature weighting, and improved ant lion optimization. *J. Ambient Intell. Humaniz. Comput.* **2021**, *12*, 8683–8698. [[CrossRef](#)]
42. Ali, N.A.; Omer, Z.M. Improving accuracy of missing data imputation in data mining. *Kurd. J. Appl. Res.* **2017**, *2*, 66–73. [[CrossRef](#)]

43. Henderi, H.; Wahyuningsih, T.; Rahwanto, E. Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *Int. J. Inform. Inf. Syst.* **2021**, *4*, 13–20. [[CrossRef](#)]
44. Shahriyari, L. Effect of normalization methods on the performance of supervised learning algorithms applied to HTSeq-FPKM-UQ data sets: 7SK RNA expression as a predictor of survival in patients with colon adenocarcinoma. *Brief. Bioinform.* **2017**, *20*, 985–994. [[CrossRef](#)] [[PubMed](#)]
45. Jayalakshmi, T.; Santhakumaran, A. Statistical normalization and back propagation for classification. *Int. J. Comput. Theory Eng.* **2011**, *3*, 1793–8201.
46. Patro, S.; Sahu, K.K. Normalization: A preprocessing stage. *arXiv* **2015**, arXiv:1503.06462. [[CrossRef](#)]
47. Dalatu, P.I.; Midi, H. New Approaches to Normalization Techniques to Enhance K-Means Clustering Algorithm. *Malays. J. Math. Sci.* **2020**, *14*, 41–62.
48. Jin, D.; Yang, M.; Qin, Z.; Peng, J.; Ying, S. A Weighting Method for Feature Dimension by Semisupervised Learning With Entropy. *IEEE Trans. Neural Netw. Learn. Syst.* **2023**, *34*, 1218–1227. [[CrossRef](#)]
49. Polat, K.; Sentürk, U. A novel ML approach to prediction of breast cancer: Combining of mad normalization, KMC based feature weighting and AdaBoostM1 classifier. In Proceedings of the 2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), Ankara, Turkey, 19–21 October 2018; pp. 1–4.
50. Poongodi, K.; Kumar, D. Support vector machine with information gain based classification for credit card fraud detection system. *Int. Arab J. Inf. Technol.* **2021**, *18*, 199–207.
51. Niño-Adan, I.; Landa-Torres, I.; Portillo, E.; Manjarres, D. Analysis and Application of Normalization Methods with Supervised Feature Weighting to Improve K-means Accuracy. In Proceedings of the 14th International Conference on Soft Computing Models in Industrial and Environmental Applications (SOCO 2019), Seville, Spain, 13–15 May 2019; Martínez Álvarez, F., Troncoso Lora, A., Sáez Muñoz, J.A., Quintián, H., Corchado, E., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 14–24.
52. Dialameh, M.; Hamzeh, A. Dynamic feature weighting for multi-label classification problems. *Prog. Artif. Intell.* **2021**, *10*, 283–295. [[CrossRef](#)]
53. Liu, X.; Lai, X.; Zhang, L. A Hierarchical Missing Value Imputation Method by Correlation-Based K-Nearest Neighbors. In *Intelligent Systems and Applications: Proceedings of the 2019 Intelligent Systems Conference (IntelliSys), London, UK, 5–6 September 2019*; Springer: Cham, Switzerland, 2019; pp. 486–496.
54. Kim, S.-i.; Noh, Y.; Kang, Y.-J.; Park, S.; Lee, J.-W.; Chin, S.-W. Hybrid data-scaling method for fault classification of compressors. *Measurement* **2022**, *201*, 111619. [[CrossRef](#)]

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.