Feature-weighted FKNN regression using fuzzy mutual information and Łukasiewicz similarity

***Abstract*—This paper proposes a feature-weighted Minkowski distance-based fuzzy k-nearest neighbor regression method, called FWMD-FKNNreg. The key novelty of FWMD-FKNNreg is its integration of feature weighting directly into the Minkowski distance computation in FKNN regression, enabling adaptive control of each feature's contribution. To achieve this, we introduce four new feature weighting schemes based on relevance, redundancy, and dependency, that are developed using fuzzy mutual information and Łukasiewicz similarity measure. An empirical evaluation identifies the most effective weighting strategy for FKNN regression, with relevance-based weighting outperforming others. Based on this finding, relevance-weighted FWMD-FKNNreg is evaluated on eight real-world datasets and compared with six competitive regression methods, including KNNreg, FKNNreg, Md-FKNNreg, support vector regression (SVR), LASSO, and multiple linear regression (MLR). Experimental results show that FWMD-FKNNreg achieves higher R^2 and lower RMSE values in most cases, indicating that the proposed method offers a robust, simple, and effective alternative for regression problems.**

***Index Terms*—regression, fuzzy k-nearest neighbor, feature weighting, similarity measure, fuzzy entropy**

I. Introduction

In k-nearest neighbor regression (KNNreg) [1], which is based on the principles of KNN [2] classifier, the output for a query instance is determined by averaging the response values of its k nearest neighbors. This method assigns equal weight to all neighbors, which increases its sensitivity to noise. To overcome this limitation, fuzzy k-nearest neighbor regression (FKNNreg), also based on the FKNN [3] classifier, was formally defined in [4] and has since gained interest for its robustness and simplicity. It is noteworthy that although the FKNN method [3], which extends the KNN algorithm [2], has been extensively studied and applied in classification tasks, its use in regression problems has received comparatively less attention [4]. Basically, the FKNNreg computes predictions as weighted averages of neighboring outputs, with weights determined by a fuzzy membership function that quantifies the similarity between the query instance and its neighbors. A significant advantage of nearest neighbor-based regression methods is their non-parametric nature, as they do not require explicit assumptions about the underlying data distribution and remain straightforward to implement and interpret [4]. Consequently, FKNNreg has demonstrated competitiveness with well-known regression techniques, particularly for nonlinear problems, when compared with methods such as support vector regression (SVR) [5] and least absolute shrinkage and selection operator (LASSO) [6].

The Minkowski distance–based fuzzy k-nearest neighbor regression (Md-FKNNreg) [4] extends FKNNreg by using the Minkowski distance instead of the Euclidean distance. This change addresses the limitations of the Euclidean distance and has demonstrated improved predictive performance. In the present study, we further enhance the Md-FKNNreg method by incorporating feature weights into the distance calculation to improve prediction accuracy. We call this approach feature-weighted Minkowski distance-based fuzzy k-nearest neighbor regression (FWMD-FKNNreg).

In feature selection, the concepts of relevance, redundancy, and dependency are often used to identify informative features [7]. Relevance measures the relationship between a feature and the target. Redundancy measures how much features share information. Highly redundant features rarely add value and may cause model's poor performance [8]. Dependency captures how combined features affect the target [7]. In the proposed FWMD-FKNNreg approach, feature weights are determined using these concepts, allowing more informative and less redundant features to influence distance computation. We define relevance-, redundancy-, and dependency-based feature weights using fuzzy mutual information (FMI) [11] and Łukasiewicz similarity [13] measures, which have not previously been applied to this purpose. Feature weighting has been studied in the FKNN classification approach (see e.g., [20]), but not explicitly with FKNNreg. Using feature weights is especially useful for uncertain or noisy data, as it prioritizes informative features and improves distance accuracy when identifying nearest neighbors. Overall, the main contributions of this paper can be summarized as follows:

- We propose four new feature weighting schemes that assess feature relevance, redundancy, and dependency using FMI and Łukasiewicz fuzzy similarity. Each scheme quantifies feature importance for regression. After empirical analysis, we select the best feature weighting scheme for the FWMD-FKNNreg algorithm.
- Based on the derived feature weights, a novel FWMD-FKNNreg algorithm is introduced in which neighbor contributions are weighted according to feature importance.

II. Preliminaries

A. Łukasiewicz similarity and fuzzy entropy

A *fuzzy binary relation* R indicates how strongly two variables are related. When R is a *fuzzy equivalence relation*, it satisfies three key properties: it is reflexive, meaning each element is fully related to itself, $R(x, x) = 1$ for all $x \in X$; it

is symmetric, so $R(x,y) = R(y,x)$ for all $x, y \in X$; and it is transitive, meaning that if x is related to y and y to z, then x is at least as related to z, i.e., $R(x,z) \geq T_L\{R(x,y), R(y,z)\}$ for all $x, y, z \in X$, where T_L denoted Łukasiewics $t-$norm, $T_L(x,y) = \max(x+y-1, 0)$.

Given a finite set $U = \{x_1, x_2, \ldots, x_n\}$, a fuzzy equivalence relation can be represented as an $n \times n$ matrix:

$$R_f = \begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{pmatrix}, \tag{1}$$

where $r_{ij} = R_f(x_i, x_j) \in [0,1]$ represents the degree of similarity between x_i and x_j.

1) Łukasiewicz similarity: One way to define a fuzzy equivalence (or similarity) relation is using the Łukasiewicz algebra [13], [14]. For elements $x_i, x_j \in [0,1]$, the similarity can be defined as

$$R_f(x_i, x_j) = \sqrt[p]{1 - |x_i^p - x_j^p|}, \tag{2}$$

where $p \geq 1$ is a parameter that controls the scaling of the differences. This measure satisfies all three axioms: reflexivity, symmetry, and transitivity.

2) Fuzzy entropy: The fuzzy information entropy of a fuzzy equivalence relation R_f over n elements is defined as [11]:

$$FH(R_f) = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{|R_{f,i}|}{n}, \tag{3}$$

where $|R_{f,i}| = \sum_{j=1}^{n} r_{ij}$ denotes the sum of similarities for the i-th element, and the logarithm base can be specified (commonly natural log or base 2).

For two subsets $f_1, f_2 \subset F$, the *fuzzy joint information entropy* can be defined as

$$FH(f_1, f_2) = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{|R_{f_1,i} \cap R_{f_2,i}|}{n}, \tag{4}$$

where $R_{f_1,i}$ and $R_{f_2,i}$ are the similarity rows corresponding to element i in subsets f_1 and f_2, respectively.

Similarly, given a subset $f_1 \subset F$, the *fuzzy conditional information entropy* of another subset $f_2 \subset F$ can be defined [12] as

$$FH(f_2 \mid f_1) = FH(R_{f_2} \mid R_{f_1}) = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{|R_{f_1,i} \cap R_{f_2,i}|}{|R_{f_1,i}|}, \tag{5}$$

where $|R_{f_1,i}| = \sum_{j=1}^{n} r_{ij}^{(f_1)}$ and $|R_{f_1,i} \cap R_{f_2,i}| = \sum_{j=1}^{n} \min\{r_{ij}^{(f_1)}, r_{ij}^{(f_2)}\}$ denote the fuzzy cardinalities of the i-th row for f_1 and the intersection of f_1 and f_2, respectively. From this, the *fuzzy mutual information* between f_1 and f_2 can be expressed in several equivalent ways:

$$FMI(f_1; f_2) = FH(f_1) - FH(f_1 \mid f_2), \tag{6}$$

$$FMI(f_2; f_1) = FH(f_2) - FH(f_2 \mid f_1), \tag{7}$$

$$FMI(f_1; f_2) = FH(f_1) + FH(f_2) - FH(f_1, f_2), \tag{8}$$

where $FH(f_1, f_2)$ is the *joint fuzzy entropy* of f_1 and f_2. Using fuzzy cardinalities, (3) can be rewritten as

$$FMI(f_1; f_2) = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{|R_{f_1,i}|}{n} - \frac{1}{n} \sum_{i=1}^{n} \log \frac{|R_{f_2,i}|}{n} + \frac{1}{n} \sum_{i=1}^{n} \log \frac{|R_{f_1,i} \cap R_{f_2,i}|}{n} \tag{9}$$

This can be simplified into a compact form:

$$FMI(f_1; f_2) = -\frac{1}{n} \sum_{i=1}^{n} \log \frac{|R_{f_1,i}| \cdot |R_{f_2,i}|}{|R_{f_1,i} \cap R_{f_2,i}|\, n}. \tag{10}$$

B. Feature relevancy, redundancy and dependency

Feature relevance is a key concept in feature selection and weighting [10], [20]. However, some features, while relevant, may not provide additional information for class discrimination. Such features, when combined with other relevant features, can degrade classification performance and are referred to as *redundant features*. Redundancy measures the shared information between features [11]. Formally, for two fuzzy relations R_{f_1} and R_{f_2} of features f_1 and f_2 over $\mathbb{U}$, redundancy can be defined as [11]:

$$R_{\text{red}}(f_1, f_2) = \frac{1}{n} \sum_{x \in X} \log \frac{n\,|R_{f_1} \cap R_{f_2}|}{|R_{f_1}| \cdot |R_{f_2}|}, \tag{11}$$

where R_f denotes the similarity relation. Here, the Łukasiewicz similarity is used. Considering class information together with a feature, the relevance of feature f_1 with respect to class c can be computed as:

$$R_{\text{rel}}(f_1, c) = \frac{1}{n} \sum_{x \in X} \log \frac{n\,|R_{f_1} \cap R_c|}{|R_{f_1}| \cdot |R_c|}. \tag{12}$$

Dependency of f_1 and f_2 with respect to the class c can be computed as

$$R_{dep}(f_1, f_2, c) = \frac{1}{|f_1|} \sum_{x_i \in f_1} \Big(\inf(R_{f_1} \cap R_{f_2} \to R_c) \cup \inf(R_{f_1} \cap R_{f_2} \to R_{\bar{c}}) \Big), \tag{13}$$

where $\bar{c}$ denotes the standard complement, $\inf()$ is the minimum membership value over all pairs, $\cup$ the standard union, and the Łukasiewicz implication is defined as $a \to b = \min\{1, 1-a+b\}$.

C. Feature weighting with relevance, redundancy, and dependency

Feature weighting can be derived from measures of *relevance*, *redundancy*, and *dependency* among features using (11)-(13).

For feature relevance, the weighting is straightforward. The normalized relevance weight for feature f_j is defined as

$$w_j^{rel} = \frac{R_{rel}(f_j, c)}{\sum_{k=1}^{m} R_{rel}(f_k, c)}, \tag{14}$$

where m is the total number of features and c denotes the target variable.

To account for feature redundancy, we evaluate the pairwise redundancy between features. The normalized redundancy weight is defined as

$$w_j^{red} = \frac{1/\sum_{r=1,r\neq j}^{m} R_{red}(f_j, f_r)}{\sum_{k=1}^{m}(1/\sum_{r=1,r\neq k}^{m} R_{red}(f_k, f_r))} \tag{15}$$

This formulation ensures that features with lower redundancy with others receive higher weights.

Feature dependency weights are derived by considering the dependency of feature f_j on other features with respect to the target c:

$$w_j^{dep} = \frac{\sum_{r=1,r\neq j}^{m} R_{dep}(f_j, f_r, c)}{\sum_{k=1}^{m}\sum_{r=1,r\neq k}^{m} R_{dep}(f_k, f_r, c)}. \tag{16}$$

Once the three individual weights are computed, the question arises: how to combine them meaningfully? One simple approach is to sum all three normalized weights:

$$w_j^{rel,red,dep} = \frac{w_j^{rel} + w_j^{red} + w_j^{dep}}{\sum_{k=1}^{m}(w_k^{rel} + w_k^{red} + w_k^{dep})}. \tag{17}$$

The idea here is that combining them will give better results. There are other, more flexible ways to combine strategies, but in this study, we focus on using individual weights and their sum as our weighting methods.

D. Fuzzy k-nearest neighbor regression (FKNNreg) variants

A formal definition of the FKNNreg method [4] is given as follows.

Let $x \in \mathbb{R}^m$ denote a new input vector and $\{(x_i, y_i)\}_{i=1}^n$ be a set of training samples, where $x_i \in \mathbb{R}^m$ and $y_i \in \mathbb{R}$. The Euclidean distance between x and each training point x_i is computed as

$$d_i^{\text{Euc}}(x, x_i) = \sqrt{\sum_{j=1}^{m}(x^j - x_i^j)^2}, \quad i = 1, 2, \ldots, n. \tag{18}$$

Based on these distances, the k nearest neighbors of x are identified, where $k \in \mathbb{N}$ is a predefined parameter. The predicted response value $\hat{y}$ is then obtained as a weighted average of the corresponding output values:

$$\hat{y} = \frac{\sum_{i=1}^{k} w_i^{NN} y_i}{\sum_{i=1}^{k} w_i^{NN}}. \tag{19}$$

The fuzzy weight w_i^{NN} associated with the i-th nearest neighbor is defined by

$$w_i^{NN} = \left(d_i^{\text{Euc}(x,x_i)}\right)^{-\frac{2}{\gamma-1}}, \qquad i = 1, 2, \ldots, k, \tag{20}$$

where $\gamma > 1$ is the fuzzy strength parameter. The exponent $-1/(\gamma - 1)$ controls the degree of fuzziness in the weighting scheme. As $\gamma \to 1$, the weights become increasingly sensitive to distance differences.

The Md-FKNNreg algorithm [4] follows the same procedure as FKNNreg but replaces the Euclidean distance with the Minkowski distance, allowing the model to better match the data's structure and thereby improve predictive accuracy.

III. Proposed FWMD-FKNNreg

This section introduces a novel FWMD-FKNNreg approach for regression problems. Basically, it further improves the Md-FKNNreg approach, allocating feature weighting, which is defined using the relevance, redundancy, and dependence concepts using FMI and Łukasiewicz similarity measure. The goal of the proposed approach is to make more accurate and reliable predictions for new data points by allowing feature importance and the Minkowski distance to play a crucial role in selecting an optimal set of nearest neighbors. The steps of the FWMD-FKNNreg method are presented below:

STEP 1: Determine the weight w_j^F for each feature j using one of the weighting schemes (relevance, redundancy, dependency, or their combination, $w_j^F \in \{w_j^{rel}, w_j^{red}, w_j^{dep}, w_j^{rel,red,dep}\}$) as defined in (14)–(17). The effect of each weighting criterion will be experimentally evaluated, and the best scheme will be selected.

STEP 2: Compute the weighted Minkowski distance between x and x_i:

$$d_i^{\text{Mink}}(x, x_i) = \left(\sum_{j=1}^{m} w_j^F \, |x^j - x_i^j|^p\right)^{1/p} \tag{21}$$

STEP 3: Sort the distances calculated in the previous step and identify the set of k nearest neighbors of x.

STEP 4: Compute the fuzzy weight for each nearest neighbor i using the weighted Minkowski distances:

$$w_i^{NN} = \left(d_i^{\text{Mink}}(x, x_i)\right)^{-\frac{2}{\gamma-1}}, \qquad i = 1, 2, \ldots, k, \tag{22}$$

STEP 5: Estimate the output value of x by the weighted average of the outputs y_i of its k nearest neighbors:

$$\hat{y} = \frac{\sum_{i=1}^{k} w_i^{NN} \, y_i}{\sum_{i=1}^{k} w_i^{NN}}. \tag{23}$$

The above steps are summarized in the pseudocode[1] provided in Algorithm 1.

[1]The MATLAB code of the FWMD-FKNNreg algorithm can be found from: https://github.com/MahindaMK/Feature-weighted-FKNN-regression-using-fuzzy-mutual-information-and-ukasiewicz-similarity.git

Algorithm 1 FWMD-FKNNreg

Require: training data $\{(x_i, y_i)\}_{i=1}^{N}$, test instance x, number of neighbors k, Minkowski parameter p

Ensure: predicted output $\hat{y}$ for x

1: **Step 1:** Compute feature weights w_j^F (using relevance, redundancy, dependency, or their combination using (14)–(17) schemes and select the best)

2: **Step 2:** For each training sample x_i, compute the weighted Minkowski distance:

$$d_i^{\text{Mink}}(x, x_i) = \Big(\sum_{j=1}^{m} w_j^F \, |x^j - x_i^j|^p \Big)^{1/p}$$

3: **Step 3:** Sort the distances $\{d_i^{\text{Mink}}\}$ and select the indices of the k nearest neighbors

4: **Step 4:** Compute fuzzy weights for each neighbor i:

$$w_i^{NN} = \big(d_i^{\text{Mink}}(x, x_i)\big)^{-\frac{2}{\gamma - 1}}, \quad i = 1, \ldots, k$$

5: **Step 5:** Predict output using the weighted average:

$$\hat{y} = \frac{\sum_{i=1}^{k} w_i^{NN} \, y_i}{\sum_{i=1}^{k} w_i^{NN}}$$

6: **return** $\hat{y}$

IV. Empirical setup

A. Datasets

We assessed the effectiveness of the developed weighting schemes and the proposed regression method using eight public multi-dimensional datasets from the UCI and KEEL repositories. Table I summarizes these datasets, including the number of features and instances for each.

TABLE I: Data information.

Dataset	Number of features	Number of samples
AutoMPG	6	392
Anacalt	7	4052
Baseball	16	337
Dee	6	365
Laser	4	993
Stock	9	950
Treasury	15	1049
Yacht	7	308

B. Testing methodology

The empirical analysis consists of two phases: (I) evaluating the performance of the developed feature weighting schemes with FKNNreg, and (II) assessing the proposed FWMD-FKNNreg approach. Both analyses use 50/50 holdout cross-validation, repeated 20 times following [20]. Before applying cross-validation, each dataset was normalized, and the average performance across all runs was reported. The evaluation metrics are the goodness-of-fit value (R^2) and root mean squared error (RMSE), which are standard in regression analysis.

C. Baseline models and parameter settings

We compared the performance of the proposed FWMD-FKNNreg approach with several competitive and widely used regression methods, including support vector regression (SVR) [5], least absolute shrinkage and selection operator (LASSO) [6], and multiple linear regression (MLR) [21]. The key hyperparameters of all models were optimized using a grid search strategy within cross-validation to ensure a fair and unbiased comparison, as in [4].

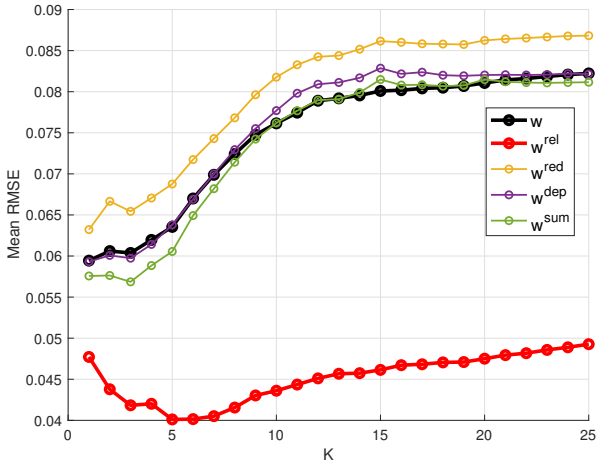
For the KNN and FKNN regression methods, the number of neighbors (k) was tuned within the range of 1 to 25. The Minkowski distance parameter (p) was chosen from the set $\{1, 1.5, \ldots, 5\}$, and the fuzzy strength parameter was fixed at 1.5 for both the FWMD-FKNNreg and MD-FKNNreg approaches, following [22]. For SVR, the kernel function was selected from linear, radial basis function (RBF), and polynomial kernels, as recommended in [23]. The regularization parameter (λ) for the LASSO model was optimized over the set $\{0.001, 0.01, 0.1, 1, 10, 100\}$. For the MLR model, four model types were evaluated: linear, interaction, pure quadratic, and quadratic, according to [4]. All other parameters for SVR, LASSO, and MLR were set to their default values as implemented in the MATLAB *Statistics and Machine Learning* toolbox.

V. Results

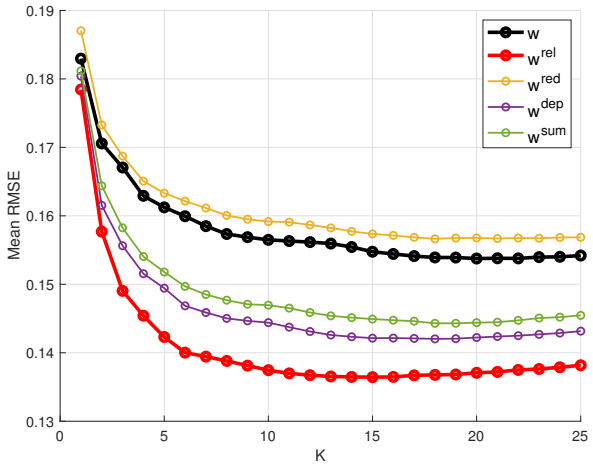
This section first presents the empirical results from the analysis of weighting schemes, followed by a comparison of the proposed FWMD-FKNNreg approach with baseline models.

A. Evaluation of the feature weighting schemes

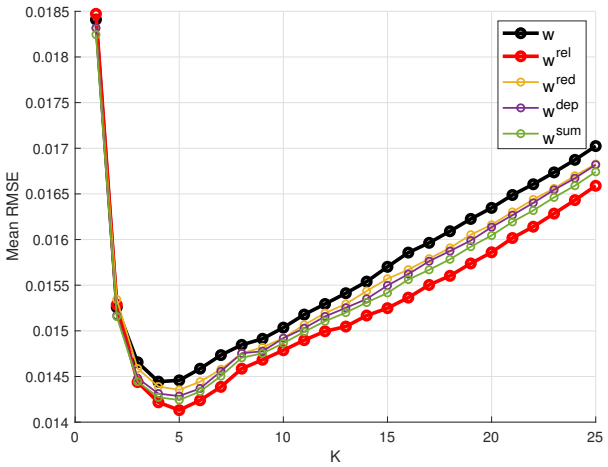
In this experiment, we evaluated the proposed feature weighting schemes within the FKNNreg framework, which served as the baseline. The weighting schemes tested include relevance-based (w^{rel}), redundancy-based (w^{red}), dependency-based (w^{dep}), and their combined sum ($w^{\text{rel+red+dep}}$). The performance was assessed using the average RMSE as a function of the number of nearest neighbors (k) across multiple datasets. Our results indicate that the w^{rel} improved the prediction accuracy of FKNNreg in five out of eight datasets (AutoMPG, Anacalt, Stock, Treasury, and Yacht). The improvements were particularly significant for the Anacalt, Yacht, and Treasury datasets, as illustrated in Fig. 1. For other weighting schemes, FKNNreg with the w^{dep} showed slight improvement on the Baseball dataset, whereas w^{red} performed better on the Dee dataset. In the case of the Laser dataset, no weighting provided the best results. Note that in Fig. 1, w represents the baseline with no specific feature weighting, meaning all features are given equal weight. Overall, the experimental results suggest that w^{rel} is the most effective weighting scheme for FKNNreg. Consequently, we adopted w^{rel} for feature weighting in the proposed FWMD-FKNNreg method.



(a) Anacalt



(b) Yacht



(c) Treasury

Fig. 1: Performance comparison of FKNNreg across different feature weighting methods over selected datasets.

B. Evaluation of the FWMD-FKNNreg performance

In this phase, we evaluated the performance of the proposed FWMD-FKNNreg approach using relevance-based feature weighting.

Table II presents the average R^2 values for each regression method across the tested datasets. FWMD-FKNNreg delivers the best results on five datasets: AutoMPG (87.62%), Anacalt (97.26%), Baseball (64.54%), Stock (98.75%), and Yacht (96.45%). It also performs competitively on the Dee, Laser, and Treasury datasets. Notably, FWMD-FKNNreg increases R^2 by more than 2 pp on Anacalt and 4.6 pp on Yacht compared to FKNNreg. Overall, it achieves the highest average R^2 of 92.33%, especially outperforming MD-FKNNreg (91.52%), FKNNreg (90.91%), and KNNreg (90.39%) methods. The RMSE results in Table III confirm these trends.

Table IV summarizes the optimal parameter settings for each regression method across the datasets. For FWMD-FKNNreg and MD-FKNNreg, small values of k and low-to-moderate Minkowski order p are always chosen. This suggests that local neighborhood structures and robust distance metrics are effective. Parameter choices vary by dataset, reflecting data-dependent model behavior. Selecting p near 1 or 2 highlights the importance of feature-weighted distance modeling.

VI. Conclusions

In this paper, first we developed four feature weighting schemes using relevance, redundancy, and dependency measures based on fuzzy mutual information and Łukasiewicz similarity, and empirically evaluated their effectiveness in FKNNreg. The objective was to determine the most effective feature weighting scheme for improving regression performance and, based on these findings, to propose a new FKNNreg method that incorporates feature weights and Minkowski distance.

Experiments on eight real-world datasets show that relevance-based feature weighting improves FKNNreg performance compared to other schemes. We therefore use relevance-based weights in the proposed FWMD-FKNNreg and compare it with six competitive regression methods, including KNNreg, FKNNreg, Md-FKNNreg, LASSO, SVR, and MLR. FWMD-FKNNreg outperforms baseline approaches in most cases, achieving higher R^2 and lower RMSE values. These results indicate that relevancy-based FWMD-FKNNreg offers a robust, simple, and effective approach for a wide range of regression problems. Despite that, One limitation is that FWMD-FKNNreg depends on grid-search tuning of the Minkowski parameter p. This approach can be suboptimal or require a lot of computing power. Future research could focus on adaptive or locally varying ways to select p.

References

[1] J. K. Benedetti, "On the nonparametric estimation of regression functions," *Journal of the Royal Statistical Society Series B*, vol. 39, pp. 248–253, 1977.

[2] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, pp. 21–27, 1967.

[3] J. M. Keller, M. R. Gray, and J. A. Givens, "A fuzzy k-nearest neighbor algorithm," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 15, pp. 580–585, 1985.

TABLE II: Average R^2 values (%) of seven regression methods over the datasets used. The highest values are shown in bold.

Dataset	FWMD-FKNNreg	MD-FKNNreg	FKNNreg	KNNreg	SVR	LASSO	MLR
AutoMPG	**87.6155**	87.5493	86.8538	86.5672	85.7055	80.3323	86.4333
Anacalt	**97.2563**	95.0460	94.9665	95.3303	89.1145	43.2339	77.3013
Baseball	**64.5420**	63.6132	61.2772	58.8400	59.5156	63.8668	63.1576
Dee	99.4517	99.4436	99.4377	99.3644	99.4419	99.3674	**99.5719**
Laser	95.0678	**95.5007**	94.7010	94.0414	94.0290	75.0196	93.8490
Stock	**98.7471**	98.7451	98.6809	98.4843	97.7994	87.2244	96.8823
Treasury	99.5290	99.5178	99.5178	99.3899	99.4477	99.3733	**99.5803**
Yacht	**96.4462**	92.7293	91.8163	91.1179	92.3428	63.7526	91.3344
Average	**92.3320**	91.5181	90.9064	90.3919	89.6746	76.5213	88.5138

TABLE III: Average RMSE values of seven regression methods over the datasets used.

Dataset	FWMD-FKNNreg	MD-FKNNreg	FKNNreg	KNNreg	SVR	LASSO	MLR
AutoMPG	0.0728	0.0730	0.0750	0.0758	0.0782	0.0918	0.0762
Anacalt	0.0392	0.0527	0.0532	0.0513	0.0784	0.1792	0.1133
Baseball	0.0228	0.0231	0.0238	0.0245	0.0243	0.0230	0.0232
Dee	0.0140	0.0141	0.0141	0.0151	0.0141	0.0151	0.0124
Laser	0.0400	0.0383	0.0415	0.0439	0.0437	0.0912	0.0452
Stock	0.0262	0.0263	0.0269	0.0289	0.0347	0.0839	0.0414
Treasury	0.0129	0.0130	0.0130	0.0147	0.0140	0.0149	0.0122
Yacht	0.0085	0.0124	0.0132	0.0137	0.0125	0.0280	0.0136
Average	0.0296	0.0316	0.0326	0.0335	0.0375	0.0659	0.0422

TABLE IV: Optimal parameter settings for each regression approach.

Dataset	FWMD-FKNNreg	MD-FKNNreg	FKNNreg	KNNreg	SVR	LASSO	MLR
AutoMPG	$(k,p)=(18,1)$	$(k,p)=(12,1)$	$k=9$	$k=6$	Gaussian	0.0010	Pure quadratic
Anacalt	$(k,p)=(4,1)$	$(k,p)=(1,1.5)$	$k=1$	$k=2$	Gaussian	0.0010	Pure quadratic
Baseball	$(k,p)=(8,1)$	$(k,p)=(7,1)$	$k=8$	$k=5$	Linear	0.0010	Linear
Dee	$(k,p)=(5,2)$	$(k,p)=(3,4.5)$	$k=3$	$k=2$	Linear	0.0010	Pure quadratic
Laser	$(k,p)=(3,1)$	$(k,p)=(4,1)$	$k=4$	$k=3$	Polynomial	0.0010	Quadratic
Stock	$(k,p)=(3,1)$	$(k,p)=(3,1)$	$k=3$	$k=2$	Gaussian	0.0010	Quadratic
Treasury	$(k,p)=(5,2)$	$(k,p)=(4,2)$	$k=4$	$k=2$	Linear	0.0010	Pure quadratic
Yacht	$(k,p)=(3,4)$	$(k,p)=(4,4.5)$	$k=4$	$k=3$	Polynomial	0.0010	Pure quadratic

[4] M. M. Kumbure and P. Luukka, "A generalized fuzzy k-nearest neighbor regression model based on Minkowski distance," *Granular Computing*, vol. 7, pp. 657–671, 2022.

[5] H. Drucker and C.J.C. Burges and L. Kaufman and A. Smola and V. Vapnik, "Support vector regression machines," *Neural Information Processing Systems*, vol. 9, pp. 155–161, 1997

[6] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B (Methodological)*, vol. 58, pp. 267–288, 1996

[7] O. A. M. Salem and F. Liu and Y. P. Chen and X. Chen, "Ensemble Fuzzy Feature Selection Based on Relevancy, Redundancy, and Dependency Criteria," *Entropy (Basel)*, vol. 22, pp. 757, 2020

[8] L. Yu and H. Liu, "Efficient feature selection via analysis of relevance and redundancy," *Journal of Machine Learning Research*, vol. 5, pp. 207–228, 2004.

[9] M.-S. Yang and K. P. Sinaga, "Collaborative feature-weighted multi-view fuzzy c-means clustering," *Pattern Recognition*, vol. 119, p. 108064, 2021.

[10] R. Battiti, "Using Mutual Information for Selecting Features in Supervised Neural Net Learning," *IEEE Transactions on Neural Networks*, vol. 5, pp. 537–550, 1994

[11] D. Yu and S. An and Q. Hu, "Fuzzy Mutual Information Based min-Redundancy and Max-Relevance Heterogeneous Feature Selection," *International Journal of Computational Intelligence Systems*, vol. 4, pp. 619–633, 2011

[12] Q. H. Hu and D. R. Yu and Z. X. Xie, "Information preserving hybrid data reduction based on fuzzy rough techniques," *Pattern Recognition Letters*, vol. 27, pp. 414–424, 2006

[13] J. Łukasiewics, "Selected work," *Cambridge University press*, 1970

[14] P. Luukka and K. Saastamoinen and V. Könönen, "A classifier based on the maximal fuzzy similarity in the generalized Łukasiewicz-structure," *In proceedings of* 10^{th} *IEEE International Conference on Fuzzy Systems*, 2001

[15] M. M. Kumbure, P. Luukka, and M. Collan, "A new fuzzy k-nearest neighbor classifier based on the Bonferroni mean," *Pattern Recognition Letters*, vol. 140, pp. 172–178, 2020.

[16] A. De Luca and S. Termini, "A definition of non-probabilistic entropy in the setting of fuzzy set theory," *Information and Control*, vol. 20, pp. 301–312, 1971.

[17] P. Luukka, "Feature selection using fuzzy entropy measures with similarity classifier," *Expert Systems with Applications*, vol. 38, pp. 4600–4607, 2011.

[18] D. Dheeru and E. K. Taniskidou, "UCI machine learning repository," 2017. [Online]. Available: http://archive.ics.uci.edu/ml

[19] L. A. Zadeh, "Fuzzy sets," *Information and Control*, vol. 8, pp. 338–353, 1965.

[20] M. M. Kumbure and P. Luukka, "Local means-based fuzzy k-nearest neighbor classifier with minkowski distance and relevance-complementarity feature weighting," *Granular Computing*, vol. 9, 73, 2024.

[21] D. C. Montgomery and E. A. Peck and G. G. Vining, *Introduction to Linear Regression Analysis*. Hoboken, NJ, USA: John Wiley & Sons, 2012.

[22] M. Arif and M. U. Akram and F. A. Minhas, "Pruned fuzzy k-nearest neighbor classifier for beat classification," *Journal of Biomedical Science Engineering*, vol. 3, pp. 380–389, 2010

[23] S. Ali and K. A. Smith-Miles, "A meta-learning approach to automatic kernel selection for support vector machines," *Neurocomputing*, vol. 70, pp. 173–186, 2006