



## Article

# Enhancing Sensor Data Imputation: OWA-Based Model Aggregation for Missing Values

Muthana Al-Amidie <sup>1,2</sup> , Laith Alzubaidi <sup>3,4,\*</sup> , Muhammad Aminul Islam <sup>5</sup> and Derek T. Anderson <sup>2</sup>

<sup>1</sup> Department of Electrical Engineering, University of Babylon, Babylon, Hilla 51001, Iraq; eng.muthana.khalil@uobabylon.edu.iq

<sup>2</sup> Department of Electrical Engineering & Computer Science, University of Missouri, Columbia, MO 65211, USA; andersondt@missouri.edu

<sup>3</sup> School of Mechanical, Medical, and Process Engineering, Queensland University of Technology, Brisbane, QLD 4000, Australia

<sup>4</sup> Research and Development Department, Akunah Company, Brisbane, QLD 4120, Australia

<sup>5</sup> Department of Electrical and Computer Engineering & Computer Science, University of New Haven, West Haven, CT 06516, USA; mislam@newhaven.edu

\* Correspondence: l.alzubaidi@qut.edu.au

**Abstract:** Due to some limitations in the data collection process caused either by human-related errors or by collection electronics, sensors, and network connectivity-related errors, the important values at some points could be lost. However, a complete dataset is required for the desired performance of the subsequent applications in various fields like engineering, data science, statistics, etc. An efficient data imputation technique is desired to fill in the missing data values to achieve completeness within the dataset. The fuzzy integral is considered one of the most powerful techniques for multi-source information fusion. It has a wide range of applications in many real-world decision-making problems that often require decisions to be made with partially observable/available information. To address this problem, algorithms impute missing data with a representative sample or by predicting the most likely value given the observed data. In this article, we take a completely different approach to the information fusion task in the ordered weighted averaging (OWA) context. In particular, we empirically explore for different distributions how the weights/importance of the missing sources are distributed across the observed inputs/sources. The experimental results on the synthetic and real-world datasets demonstrate the applicability of the proposed methods.

**Keywords:** missing data imputation; quadratic programming; OWA operators; data aggregation; measurement learning



**Citation:** Al-Amidie, M.; Alzubaidi, L.; Islam, M.A.; Anderson, D.T.

Enhancing Sensor Data Imputation: OWA-Based Model Aggregation for Missing Values. *Future Internet* **2024**, *16*, 193. <https://doi.org/10.3390/fi16060193>

Academic Editor: Wei Yu

Received: 7 April 2024

Revised: 19 May 2024

Accepted: 28 May 2024

Published: 31 May 2024



**Copyright:** © 2024 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

Daily, large amounts of data are produced and gathered using specialized electronic devices or sensors. However, several factors cause data inconsistency and incompleteness [1]. These factors may include human error (e.g., data handling errors or a lack of strictly defined protocols for data collection), malfunctioning devices and/or sensors, power system failures, noise, environmental factors (e.g., rain, snow), inaction in scientific experiments, issues with data transfer in digital systems, etc. [2,3]. Whenever there are missing data in a dataset, it can significantly affect the accuracy of data interpretation algorithms. Therefore, it is crucial to estimate the missing data value as precisely as possible to obtain better insights from subsequent data analysis tasks mostly based on machine learning algorithms [4]. In other words, efficient techniques should be developed to analyze the dataset using machine learning algorithms to deal with the missing values [5,6].

Nowadays, promising scenarios for collecting and storing large amounts of data have been developed. With the plentiful volume of data points, it is easy to understand the behavior and/or insights of some underlying phenomena. However, such data should

be complete and consistent so that their relevance is not questionable and can provide important information about related phenomena [7]. Different data mining techniques, such as classification, time series analysis, regression, etc., could produce insight into the data. The performance of such methods cannot be optimized if some critical data points are missing. Thus, maintaining the relevance of the collected data is essential.

As stated in [8], the various scenarios involving missing data can be classified into three distinct forms: missing at random (MAR), missing completely at random (MCAR), and missing not at random (MNAR). Suppose the missing values are assumed according to be of type MAR. In that case, these values can be estimated using another technique (i.e., the approach in which values of the missing sample are ignored). MCAR-type missing data indicate that any missing observation is unrelated to the data values of the other variables or the available data values [9]. When missing data values are considered as MNAR-type, the values related to a variable are missed due to the nature of the variable itself. Practically, MCAR-type missing data occur very rarely. Among these three, MAR-type data is a reliable assumption. In this case, data are considered to be MAR-type as long as the data meet the requirement of not depending on the value of another variable. To achieve an effective imputation of missing values, it is crucial to model the mechanism through which these values are calculated accurately. The assumption of data missing at random (MAR) is generally acknowledged and commonly used in various approaches, as defined in [10].

The primary emphasis of previous investigations was the efficacy and precision of the suggested methodologies in restoring randomly absent values for individual values. The model's design to impute missing data is considered an essential problem of great practical importance in many fields, such as machine learning [11]. Missing data imputation and identification of the active aggregation have recently become a topic of interest in research to increase knowledge aggregation in practical applications [12]. In the end, despite the many research studies conducted to solve this task, particularly the imputation of missing data in practical applications such as collecting information from sensors, it is still challenging due to many limitations [13]. Many methods have been invented for missing data imputation; some of these studies are based on a priori-estimated parameters for the model, such as in [14], and others do not require estimated parameters, such as clustering algorithms and the Choquet integral (ChI). In addition, Yager [15] offered a method for calculating the weights of the ordered weighted averaging (OWA) operator using a set of samples that contain the aggregated data required. The present study aims to investigate the problem of OWA aggregation of a set of variables interconnected and constrained by a series of linear inequalities. In his paper, Yager suggests a linear programming (LP) model to solve such a problem. Meanwhile, another study introduced a linear objective programming strategy to estimate the OWA weights depending on a specified level of "orness", as described in [16].

The following line discusses fuzzy aggregation and fuzzy set theory. A group of operators named the fuzzy induced generalized hybrid averaging (FIGHA) operator, the fuzzy generalized hybrid averaging (FGHA) operator, and the quasi-FHA operator were proposed by Merigó and Casanovas in [17]. These operators have the advantage of being flexible, meaning they can incorporate a wide variety of fuzzy aggregation operators. This makes them suitable for various applications, including situations with decision-making challenges. Ref. [18] introduces a new practical operator based on the concept of least mean square errors. The new OWA operator may obtain the OWA weights utilizing fuzzy inference and weight judgments from decision-makers (DMs).

Moreover, missing-value imputation approaches can also be categorized based on techniques that are used for approximating those values. Consequently, imputation methods can be classified into two types [19]. The first type includes approaches in which missing values can be predicted using statistical or mathematical methods. These approaches are very simple, as the mean value of the features can replace the missing data values. However, using such methods can potentially introduce bias into the estimations [20,21]. Some methods are sophisticated, where more advanced statistical tools are used. The second type includes famous machine learning techniques that have been heavily used in the literature

lately. The methods included in this category involve the creation of a model or a fusion of multiple models that acquire knowledge from a substantial amount of data. A well-trained model can make accurate predictions, which helps in accurately imputing missing values. The methods that come under this category could use neural networks [22,23], support vector machines [24], or nearest-neighbor-based strategies [25]. In addition, the current state of machine learning approaches has limitations or deficiencies in complexity and expense, such as the artificial neural network method.

In this work, we propose a new OWA-based missing data imputation method. It is a simple and effectively trained model that obtains a prediction capacity of absent values using a quadratic optimization technique. This methodology assumes that the given data can be represented as an additive input vector that follows a Gaussian distribution. Specifically, we assume that the mean of this distribution is zero and the covariance matrix is deterministic. Without losing generality, this assumption is typical in data aggregation applications. The covariance of the random variables is assumed to be a deterministic matrix, e.g., an identity matrix. Some values in the data points are randomly chosen as missing values. This approach proposes an OWA operator to impute the missing values in the dataset. In particular, a realization dataset was used to evaluate the proposed method. The rest of the sections of the paper are organized as follows. In Section 2, we clarify the detailed methodology of the proposed approach, which includes the description of the modeling of the measurements, the dataset type, and the learning strategy. In Section 3, we discuss the performance of the proposed approach in various simulation environments. Concluding remarks are provided in Section 4.

## 2. OWA Methodology

In this section, we explain the concept of ordered weighted averaging and the learning algorithm that is the heart of our proposed method for missing value imputation.

### 2.1. Ordered Weighted Averaging

The OWA operators were originally introduced in [26] to create a function to aggregate data associated with multiple criteria. Subsequently, this technique has been used as a popular aggregation operator for many different types of applications [27,28]. The primary objective of implementing these aggregation operators is to establish a suitable methodology for assigning importance weights to data from various sources. In particular, there are different OWA families, such as max, min, and average. Thus, these operators are extremely flexible for practitioners to utilize in various applications. These operators can be defined as follows:

**Definition:** OWA operators are a means to associate a set of  $n$  numbers with a single numerical value or equivalently real-valued functional over a finite set  $S = s_1, \dots, s_n$ . In other words, the OWA method is a mapping function ( $F$ ) from multiple dimensions to one dimension,  $R^n \rightarrow R$ , by using an associated weighting value for each observed value as follows:

$$\mathbf{w} = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} \quad (1)$$

where

$$\sum_{j=1}^n w_j = 1 \quad (2a)$$

$$0 \leq w_j \leq 1, \quad (2b)$$

$$F(a_1, a_2, \dots, a_n) = \sum_{j=1}^n b_j w_j \quad (3)$$

where  $\mathbf{F}$  is a column vector consisting of  $n$  observation elements with associated weighted after sorting the observation data, and  $\mathbf{a} = (a_1, a_2, \dots, a_n)$  represents the unordered arguments' values,  $\mathbf{b} = (b_1, b_2, \dots, b_n)^T$ .  $\mathbf{b}$  can be rewritten in a different way, such as  $\mathbf{b} = (a_{\gamma_1}, a_{\gamma_2}, \dots, a_{\gamma_n})$ , by sorting the arguments with  $b_1 \geq b_2 \geq \dots \geq b_n$ , where  $\gamma$  is considered as a permutation of the arguments on  $(1, 2, \dots, n)$ , and  $\mathbf{w} = (w_1, w_2, \dots, w_n)^T$  is the OWA weighting vector.

Now, it can be seen that the essential key of this operator is ordering the argument values. In addition, this operation is idempotent, commutative, and monotonic. In this work, we explain the methodology for incorporating importance into OWA aggregation applications by utilizing a transformation operator. The implementation of various aggregation operators can be achieved by considering the magnitude of the generality of the weights and determining the appropriate emphasis on the arguments based on the order of the weight positions in  $\mathbf{w}$ . By placing high values with the weights, higher aggregation scores can be generated. On the other hand, lower scores in the aggregation can be achieved by setting lower values of the weights. The challenge associated with this problem lies in the extensive range of operators represented by the OWA operator. Specifically, the various components of the OWA family, such as max, min, and average, employ distinct methods for including priority; for more detail, please see [29].

## 2.2. Learning Algorithm

This section provides a description of the learning method and its data-driven learning. Consider  $\{b_1, \dots, b_n\}$  to be a set of finite elements, i.e., the observation data. In other words,  $\mathbf{b}$  can be sensor measurements, experts, or attributes in decision-making problems. Let the training dataset  $O$  consist of  $M$  pairs of both the labels and the observations dataset such that  $O = \{(b_i, y_i)\}, i = 1, 2, \dots, M$ ,  $b_i$  represents the  $i^{th}$  observation, and  $y_i$  represents the corresponding label. Furthermore,  $o_i(b_k)$  is the observed value for the input of the observation data from the instance  $i^{th}$  with the input of  $k^{th}$ . Considering  $\mathbf{w} = [w(\{b_1\}), w(\{b_2\}), \dots, w(\{b_n\})]^T$  is considered to be the weight value vector; the discrete OWA on  $o_j$  with respect to  $\mathbf{w}$  can be rewritten as follows:

$$C(o_j) = \mathbf{b}_j^T \mathbf{w} \quad (4)$$

where  $j^{th}$  represents the index of the observation vector in the dataset, and  $\mathbf{b}$  represents the input column vector. The sum of squared error (SSE) of the proposed approach can be expressed as

$$\begin{aligned} E_j(O, \mathbf{w}) &= \sum_{j=1}^M (C(o_j) - y_j)^2 = \sum_{j=1}^M (\mathbf{b}_j^T \mathbf{w} - y_j)^2 \\ &= \sum_{j=1}^M (\mathbf{w}^T \mathbf{b}_j \mathbf{b}_j^T \mathbf{w} - 2y_j \mathbf{b}_j^T \mathbf{w} + y_j^2). \end{aligned} \quad (5)$$

where  $E_j(O, \mathbf{w})$  is the error. The least-squares minimization formulation can be expressed as follows:

$$(\text{OP1}) \min_{\mathbf{w}} f_O(\mathbf{w}) = \mathbf{w}^T \mathbf{H} \mathbf{w} + \mathbf{d}^T \mathbf{w}, \quad (6)$$

Subject to

$$\sum_{j=1}^n w_j = 1 \quad (7a)$$

$$0 \leq w_j \leq 1, \quad (7b)$$

where

$$\mathbf{H} = \sum_{j=1}^M \mathbf{b}_j \mathbf{b}_j^T, \text{ and } \mathbf{d} = -2 \sum_{j=1}^M y_j \mathbf{b}_j.$$

$\mathbf{H} \in \mathbb{R}^{n \times n}$ ; standard QP solvers can minimize the problem error. In this case, OP1 is used as a model for training according to Equation (6) to determine OWA's operators when inputs exist.

### 3. The Proposed Formulation

#### 3.1. Modeling Missing Inputs

This work's primary goal is to learn the OWA weighting vector where the missing data samples exist, i.e., the estimated OWA weighting vector can be modeled related to the OWA weighting vector with the existing complete data point. In other words. After generating all data, arrange the data values in descending order to provide a decision maker based on the newly assigned OWA weighting vector. Upon assuming the values of the OWA weighting vector such as the softmax OWA weighting vector, the data point labels are obtained according to Equation (4). Next, assign the percentage of missing data volume. For example, 10% or 20% of the total observation dataset has missing values within each data point. These values can be randomly picked for each data point. Since the data point, including the missing samples and the labels, is available, new weighted vector values are produced using the data point with the missing data samples. This label was determined when the complete dataset was available earlier. The estimated weighting vector  $\mathbf{r}$  is now determined, which has the same dimension as in the incomplete dataset. Since two sets of OWA operators are obtained, one where the complete dataset exists and another with an incomplete dataset, the new expression is obtained between the estimated OWA operator and the original OWA weighting vector. This expression can be solved using the quadratic form with specific constraints, as shown in Equation (5), to determine the coefficient parameters. Each estimated OWA operation has an expression in terms of the original values of the OWA weighting vector, and this expression can be used to predict the missing data sample. The missing values within the data point will be regressed on other variables based on the relationship between the estimated OWA weighting vector and the original operators. The missing values will then be replaced by the predictive values obtained. Similarly, each data point with missing values can replace its missing values with predicted ones.

Figure 1 shows more details on the proposal approach. Three steps can be seen in this figure. The first step shows the learning expression to obtain the missing data. Subsequently, the missing values are derived based on the other variables. Ultimately, the mean squared error (MSE) is obtained to analyze the performance of the proposed approach.

Next, the synthetic datasets are utilized in this experiment. Synthetic datasets are generally used because it is thought to be very difficult to identify real-world applications of many complex scenarios. However, it is much easier to build OWA weighting vectors in synthetic data with varying complexity. The second reason is that synthetic datasets allow us to see how the learning approach would work with noise-free datasets. Moreover, carrying out tests also yields computational complexity. The mean square error performance is also obtained in various settings by applying the learned OWA weighting vector and the expected test labels. Table 1 represents the OWA weighting vector with independent variables.

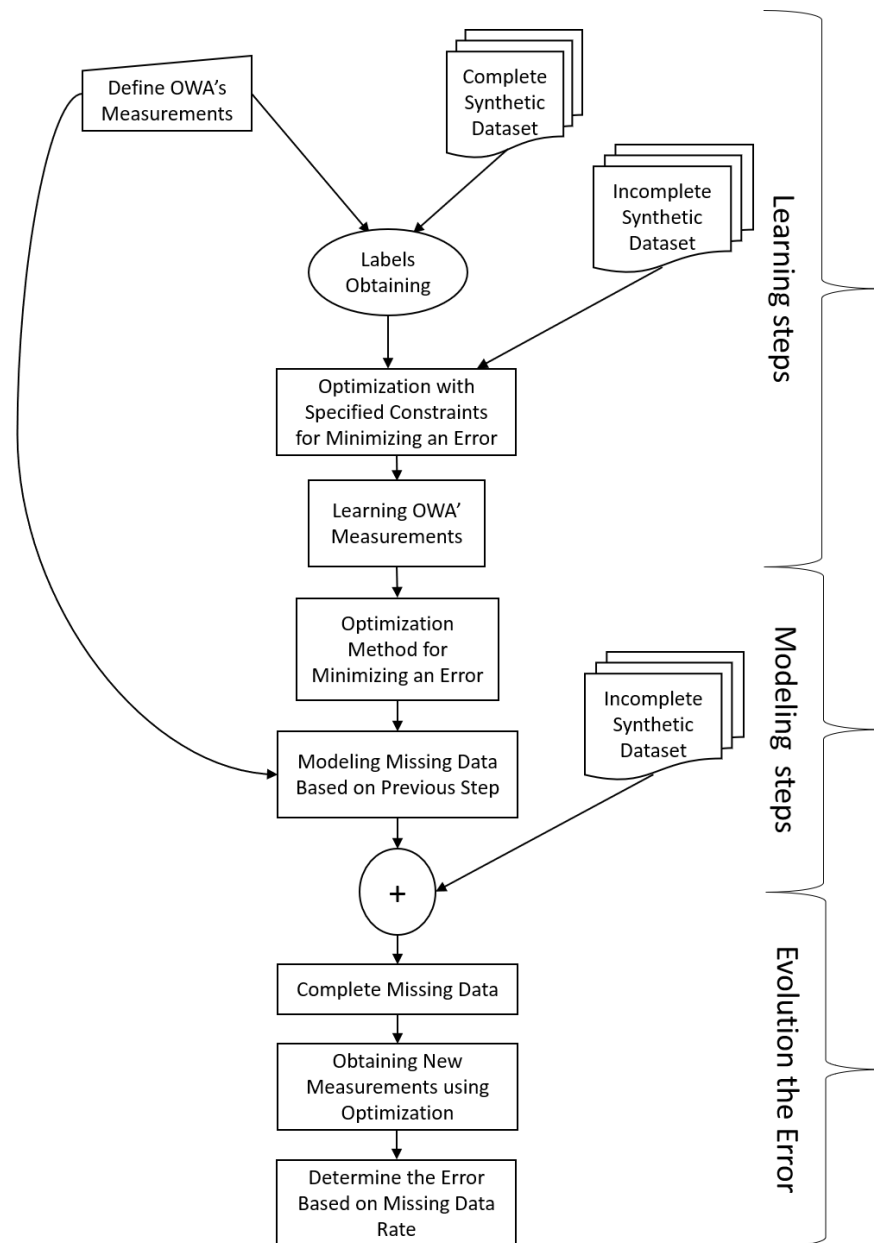


**Figure 1.** OWA approach for incomplete dataset aggregation.

**Table 1.** OWA's value.

| Type of Measurement | $w_1$ | $w_2$ | $w_3$ | $w_4$ |
|---------------------|-------|-------|-------|-------|
| Soft-Min            | 0.1   | 0.2   | 0.3   | 0.4   |
| Soft-Max            | 0.4   | 0.3   | 0.2   | 0.1   |
| Mean                | 0.25  | 0.25  | 0.25  | 0.25  |

### 3.1.1. Synthetic Dataset

In this section, we assume that  $M$  is the observed data and that  $N$  samples are used for each observation. We assume that the observations have an independent and identical distribution value, where each  $x$  is a vector of Gaussian distribution numbers, it can be written as  $x \sim \mathcal{N}(\mu, \sigma^2)$ , where  $\mu$  is the mean of the distribution while  $\sigma^2$  is the variance of the distribution. Since the scaling of the OWA weighting vector has directly affected the aggregation decision, we consider different values of the OWA operators set, such as a soft-min. In the missing data problem, we consider the absence of the input value, denoted

by the “-” symbol. Thus, using Gaussian distribution input, the following OWA operator test of the data aggregation can be expressed as in Equation (4). Moreover,  $\mathbf{x}$  is a synthetic dataset with a real zero-mean random value with a unity covariance matrix. Tractable results are obtained under these typical assumptions; that is, the problem becomes finding a straightforward way to bring robust learning of OWA operators that can be utilized for the real dataset to determine missing values. In general, we specified the different groups of OWA operators, and each group’s sum must be based on the OWA constraints. The main goal of this paper is to regress the missing value using other available datasets. We assume that four datasets are observed and one variable is missing; for this case, the estimated number of OWA values is three-dimensional because one variable is missing. Those new OWA weighting vectors are learned using a quadratic form using the dataset with missing values and the original labels. All the inputs are independent and identically distributed (i.i.d) random variables, following a Gaussian distribution with a mean of zero. According to the modeling formula mentioned in Equation (5), aggregation measurements can be obtained by

$$C(o_j) = \mathbf{b}_j^T \hat{\mathbf{w}} \quad (8)$$

where  $C(o_j)$  is the label corresponding to inputs  $\mathbf{b}_j$ , which is then a column vector, and  $\hat{\mathbf{w}}$  is the estimated OWA weighting vector which is supposed to be calculated in this step.

$$\begin{aligned} E_1(O, \mathbf{u}) &= \sum_{j=1}^M (C(o_j) - y_j)^2 = \sum_{j=1}^M (\mathbf{b}_j^T \hat{\mathbf{w}} - y_j)^2 \\ &= \sum_{j=1}^M (\hat{\mathbf{w}}^T \mathbf{b}_j \mathbf{b}_j^T \hat{\mathbf{w}} - 2y_j \mathbf{b}_j^T \hat{\mathbf{w}} + y_j^2). \end{aligned} \quad (9)$$

Thus, the least-squares minimization solver can be described as follows:

$$(\text{OP1}) \min_{\hat{\mathbf{w}}} f_O(\hat{\mathbf{w}}) = \hat{\mathbf{w}}^T \mathbf{H} \hat{\mathbf{w}} + \mathbf{d}^T \hat{\mathbf{w}},$$

subject to

$$\sum_{j=1}^n \hat{\mathbf{w}}_j = 1 \quad (10a)$$

$$0 \leq \hat{\mathbf{w}}_j \leq 1, \quad (10b)$$

where

$$\mathbf{H} = \sum_{j=1}^M \mathbf{b}_j \mathbf{b}_j^T, \text{ and } \mathbf{d} = -2 \sum_{j=1}^M y_j \mathbf{b}_j.$$

OWA measurement is determined at the location of missing data points. After obtaining a new OWA weighting vector, the model can be determined based on the relationship between the original values of the OWA operators and the new OWA operators.

In this case, we consider four variables with one missing input to model the relationship between the predicted values of the OWA weighting vector and the original OWA weighting vector. The coefficient of this model is estimated using the quadratic optimization technique. The model can be shown as follows:

$$\begin{aligned}
 \hat{w}_1 &= w_1 a_{11} + w_2 a_{12} + w_3 a_{13} + w_4 a_{14} + 0 + 0 + 0 + 0 + \dots \\
 &\quad 0 + 0 + 0 + 0 \\
 \hat{w}_2 &= 0 + 0 + 0 + 0 + w_1 a_{21} + w_2 a_{22} + w_3 a_{23} + w_4 a_{24} + \dots \\
 &\quad 0 + 0 + 0 + 0 \\
 \hat{w}_3 &= 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + 0 + w_1 a_{31} + w_2 a_{32} + \dots \\
 &\quad w_3 a_{33} + w_4 a_{34}
 \end{aligned} \tag{11}$$

It can be seen that the total number of variables depends on the number of samples and the number of variables for each sample. From Equation (11), it can be seen that the number of samples is three due to a missing input, where each sample has four variables. Then, the total number of variables can be obtained by multiplying the number of samples by the number of variables for each sample. For the above example, since the number of samples is three and each sample has four different variables, the total number of variables is twelve. This equation is considered a linear system and can be rewritten as the following equation:

$$\hat{w}_{i,j} = \mathbf{z}_{i,j}^T \mathbf{a}, \tag{12}$$

where  $\mathbf{z}_{1,j} = [w_{1,j} w_{2,j} w_{3,j} w_{4,j} 0 0 0 0 0 0 0]^T$ ,  $\mathbf{z}_{2,j} = [0 0 0 0 w_{1,j} w_{2,j} w_{3,j} w_{4,j} 0 0 0 0]^T$ ,  $\mathbf{z}_{3,j} = [0 0 0 0 0 0 0 w_{1,j} w_{2,j} w_{3,j} w_{4,j}]^T$ , and  $p$  is again considered as the number of different types of measurements and  $\hat{\mathbf{w}}$  is the predicted values of new measurements when missing input exists,  $\mathbf{w}$  is the original values of the OWA weighting vector, and  $\mathbf{a}$  is the column vector coefficient in this linear system that is obtained by using the quadratic programming, as shown in Equation (5). In Equation (13), the final form of optimization can be determined with the corresponding constraints.

$$\begin{aligned}
 E_1(O, \mathbf{a}) &= \sum_{j=1}^P \sum_{i=1}^3 (\mathbf{z}_{i,j}^T \mathbf{a} - \hat{w}_{i,j})^2 \\
 &= \mathbf{a}^T \sum_{j=1}^P \sum_{i=1}^3 (\mathbf{z}_{i,j} \mathbf{z}_{i,j}^T) \mathbf{a} - 2 \sum_{j=1}^P \sum_{i=1}^3 (\hat{w}_{i,j} \mathbf{z}_{i,j}^T) \mathbf{a} + \sum_{j=1}^P \sum_{i=1}^3 \hat{w}_{i,j}^2.
 \end{aligned} \tag{13}$$

where  $w_{1,j}$ ,  $w_{2,j}$ ,  $w_{3,j}$ , and  $w_{4,j}$  are the 'true' weights for the  $j^{th}$  OWA and  $w'_{1,j}$ ,  $w'_{2,j}$ ,  $w'_{3,j}$  are the learned weights for the  $j^{th}$  OWA when one input is missing. The least-squares minimization problem can be obtained by formulating QP as

$$(\text{OP1}) \min_{\mathbf{a}} f_O(\mathbf{w}) = \mathbf{a}^T \mathbf{H} \mathbf{a} + \mathbf{d}^T \mathbf{a}, \tag{14}$$

Subject to:

$$\begin{aligned}
 a_{11} + a_{21} + \dots + a_{p1} &= 1 \\
 a_{12} + a_{22} + \dots + a_{p2} &= 1 \\
 a_{13} + a_{23} + \dots + a_{p3} &= 1 \\
 a_{14} + a_{24} + \dots + a_{p4} &= 1
 \end{aligned} \tag{15}$$

Considering that we have  $N$  variables, the above constraints can be generalized as follows:

$$\begin{array}{ll}
 \text{For first variable} & \sum_{j=1}^P a_{j1} = 1 \\
 \text{For second variable} & \sum_{j=1}^P a_{j2} = 1 \\
 \vdots & \vdots \\
 \text{For } N^{th} \text{ variable} & \sum_{j=1}^P a_{jN} = 1
 \end{array} \quad (16)$$

where

$$\mathbf{H} = \sum_{j=1}^P \sum_{i=1}^2 (\mathbf{z}_{i,j} \mathbf{z}_{i,j}^T), \text{ and } \mathbf{d} = -2 \sum_{j=1}^P \sum_{i=1}^2 (\hat{w}_{i,j} \mathbf{z}_{i,j}).$$

standard QP solvers are capable of minimizing this problem. In this paper, OP1 is used as a training framework, as shown in Equation (14), to estimate the coefficients of the linear system. As a result, the following is how the optimization's solution appears:

$$\begin{aligned}
 \hat{w}_1 &= w_1 + w_2/2 \\
 \hat{w}_2 &= (w_2 + w_3)/2 \\
 \hat{w}_3 &= w_4 + w_3/2
 \end{aligned} \quad (17)$$

We apply the final aggregation formula based on the weighting data vector that shows regular behavior to find the final expression of the estimated OWA vectors. Since the number of data samples is either even or odd, we have studied several experiments in which the sample values are an even number and an odd number to show a more reliable and precise algorithm. For example, in the cases where the number of inputs is even and the missing number is one value, which means the rest of the data sample is five, then the final model of the measurement is obtained as shown in Equation (18).

$$\begin{aligned}
 \hat{w}_{MSD} &= w_{MSD} + w_{MSD+1}/2 \\
 \hat{w}_{Middle} &= (w_{N/2} + w_{N/2+1})/2 \\
 \hat{w}_{LSD} &= w_{LSD} + w_{LSD-1}/2
 \end{aligned} \quad (18)$$

According to Equation (12), the generic form for the measurements when the number of the input samples is an odd number (five digits) can be obtained as shown in Equation (19).

$$\begin{aligned}
 \hat{w}_{MSD} &= w_1 + w_2/2 \\
 \hat{w}_{Middle} &= w_{\lceil N/2 \rceil} \\
 \hat{w}_{LSD} &= w_N + w_{N-1}/2
 \end{aligned} \quad (19)$$

For a more reliable OWA algorithm to predict the missing data, we go further to assume that the observation data sample is four variables with two missing data samples. Again, we assume that the distribution of the set of inputs is normal with existing independent features among the variables. We perform the final aggregation process as shown in Equation (21) concerning the original weighting vector to learn the new OWA vector where there are missing data. In this case, Equation (20) shows the generic expression for the

learned OWA vector in terms of the original OWA values where two missing data are available in four data samples.

$$\begin{aligned}\hat{w}_{MSD} &= w_{MSD} + w_{MSD+1}/4 \\ \hat{w}_{LSD} &= w_{LSD} + w_{LSD-1}/4\end{aligned}\quad (20)$$

Moreover, five data samples are assumed to be an odd number with two missing data samples. For this case, Equation (21) is obtained according to Equation (21) to be the final expression of this case.

$$\begin{aligned}\hat{w}_{MSD} &= w_1 + w_2/4 \\ \hat{w}_{Middle} &= w_{\lceil N/2 \rceil} / 2 \\ \hat{w}_{LSD} &= w_N + w_{N-1}/4\end{aligned}\quad (21)$$

Now, obtaining the new OWA operators in terms of the original OWA values. As shown above, this vector is obtained where the missing samples are available in the data points. The missing data can now be modeled in terms of the other available data samples. In this case, the learned OWA vector can be utilized to model the missing data samples. In case one, we assume that there are four data samples with one missing, and according to Equation (17), the observation dataset becomes

$$\begin{aligned}y_j &= \hat{w}_1 x_1 + \hat{w}_2 x_2 + \hat{w}_3 x_3 \\ &= (w_1 + w_2/2)x_1 + (w_2 + w_3)/2x_2 + (w_4 + w_3/2)x_3\end{aligned}\quad (22)$$

From Equation (22), it can be seen that the new inputs of each observation have missing input. The modeled missing samples can be determined using Equation (23).

$$\begin{aligned}x'_1 &= x_1; \\ x'_2 &= x_1/2 + x_2/2; \\ x'_3 &= x_2/2 + x_3/2; \\ x'_4 &= x_3/2;\end{aligned}\quad (23)$$

Finally, we can obtain each input's relation with other data samples; this relation can be utilized to determine the missing data sample that most probably occurs in data aggregation. Next, we shall show experimental tasks to assess the missing sample in the data points, and then, obtain the evaluation metric to find the error for our proposed method. The sum of squared error (SSE) [30–32] represents the error between the predicted labels after abating the missing data sample and the original labels.

### 3.2. Experimental Results and Discussion

The main purpose of conducting this experiment is to show the effectiveness of the proposed approach when utilizing real data and comparing this approach with other essential methods for treating missing data. In Section 3.1.1, OWA weighting vectors were learned where the missing inputs are available in synthetic datasets. Next, we performed a real experiment using the OWA weighting vectors learned in Section 3.1.1. We assume that real datasets have more than one missing input. We utilized the modeling of the OWA weighting vector based on previous learning to replace the missing inputs in the real datasets. Abalone datasets [33] are used in this experiment. The UCI machine learning repository provides this benchmark. For more information about the real datasets that were used for the experiment, Table 2 is set up to summarize the characteristics of the dataset. It shows that each observation has nine attributes that describe the age of the abalone, categorizing it as an infant, a male, or a female. The age of the abalone is obtained by clipping the shell from the cone and staining it, then counting the number of rings

using a microscope. This process is considered a time-consuming task, especially with missing features.

**Table 2.** Abalone dataset.

| Attribute Information    |                  |                  |                             |
|--------------------------|------------------|------------------|-----------------------------|
| Name                     | Data Type        | Measurement Unit | Description                 |
| Sex Attribute            | Nominal value    | --               | M, F, and I (infant)        |
| Length Attribute         | Continuous value | mm               | Longest shell measurement   |
| Diameter Attribute       | Continuous value | mm               | Perpendicular to length     |
| Height Attribute         | Continuous value | mm               | With meat in shell          |
| Whole Weight Attribute   | Continuous value | grams            | Whole abalone               |
| Shucked Weight Attribute | Continuous value | grams            | Weight of meat              |
| Viscera Weight Attribute | Continuous value | grams            | Gut weight, after bleeding  |
| Rings Attribute          | Integer          | --               | +1.5 gives the age in years |

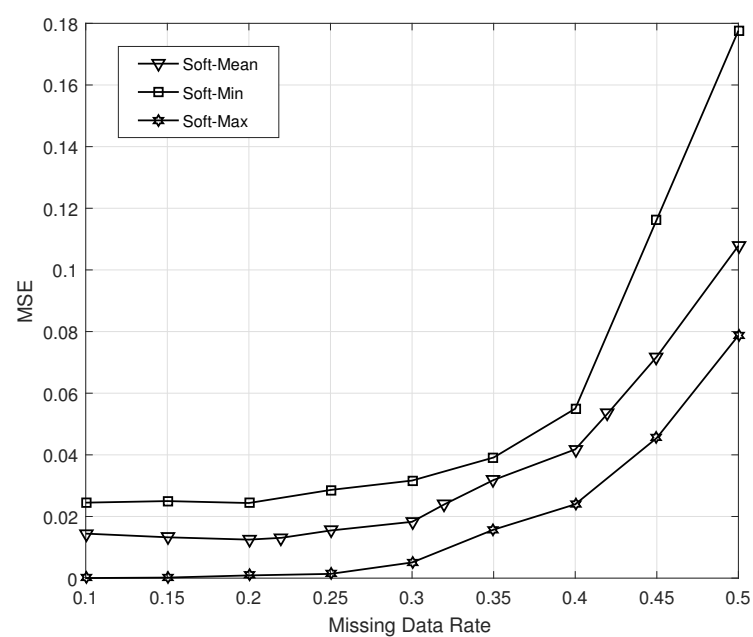
Furthermore, we demonstrate the results with simulated data and illustrate the convergence of our proposed method among different parameter settings. Within the simulation experiment, the constraint of the measurements has been defined as shown in Equation (7a,b). First, we clarify the effect of missing input in data aggregation, where the number of missing data points is in different positions for each observation. As mentioned, the data distribution is assumed to be Gaussian with an i.i.d. correlation feature. For these assumptions, we show the performance of our approach for estimation of the OWA weighting vector from which the model of missing data is obtained. Also, we show its performance when there are a varied number of missing inputs. It can be seen that a Gaussian distribution is utilized to make the approach more precise in data aggregation. As mentioned earlier, this approach considered the number of the observation datasets  $M$  to be 500, the number of simulations is 20, while the number of variables ( $N$ ) is four. At first, a Gaussian distribution is used to model the missing input within the observation dataset. Subsequently, another dataset is generated randomly to produce the missing values within the datasets based on the percentage number of the disappeared inputs, as shown in Table 3. The percentage of disappeared numbers is assumed to be [10%, 20%, ..., 50%] and randomly chosen within the observation. The efficiency of the missing data imputation approach is determined by calculating the mean standard error.

$$\text{MSE} = \frac{\sum_i^L (x_i - \hat{x}_i)^2}{N} \quad (24)$$

where the number of measurement values for OWA is assumed to be  $N$ , and  $i^{th}$  is assumed to be the index of each element in the measurement vector. The following figure illustrates the performance of the proposed approach, showing the error of missed prediction in terms of missing percentages where the number of missing data points is assumed to be one. The result shows that the error increases slightly as the percentage of missing data increases. It also shows that the error will increase as the number of missing inputs increases, meaning that the learned OWA weighting vector will be affected by a large number of missing inputs. In addition, three different real measurements are used in the proposed approach where missing variables are available. Thus, the rate of missing data in the observation also has a significant effect on the error. These results illustrate that the error is directly proportional to the percentage of missing data. This relation can be clearly seen in Figure 2.

**Table 3.** Example of missing inputs within the dataset.

| Length | Diameter | Height | Whole Weight | Rings |
|--------|----------|--------|--------------|-------|
| 0.455  | 0.365    | 0.095  | -            | 15    |
| 0.35   | -        | 0.09   | -            | 7     |
| 0.53   | 0.42     | 0.135  | 0.677        | 9     |
| 0.44   | -        | 0.125  | 0.516        | 10    |
| 0.33   | 0.255    | 0.08   | 0.205        | 7     |
| 0.425  | 0.3      | 0.095  | -            | 8     |
| 0.53   | -        | 0.15   | -            | 20    |
| 0.545  | 0.425    | 0.125  | 0.768        | 16    |
| 0.475  | -        | 0.125  | 0.5095       | 9     |
| 0.55   | 0.44     | 0.15   | 0.8945       | 19    |
| 0.525  | 0.38     | -      | 0.6065       | 14    |
| 0.43   | 0.35     | 0.11   | 0.406        | 10    |
| 0.49   | -        | 0.135  | 0.5415       | 11    |
| -      | 0.405    | -      | 0.6845       | 10    |
| 0.47   | 0.355    | 0.1    | 0.4755       | 10    |
| 0.5    | 0.4      | 0.13   | 0.6645       | 12    |
| 0.355  | 0.28     | 0.085  | 0.2905       | 7     |
| 0.44   | 0.34     | 0.1    | -            | 10    |
| 0.365  | 0.295    | 0.08   | 0.2555       | 7     |
| 0.45   | -        | 0.1    | 0.381        | 9     |
| 0.355  | 0.28     | 0.095  | 0.2455       | 11    |
| 0.38   | 0.275    | 0.1    | 0.2255       | 10    |
| 0.565  | -        | 0.155  | -            | 12    |
| 0.55   | 0.415    | -      | 0.7635       | 9     |
| 0.615  | 0.48     | 0.165  | 1.1615       | 10    |
| 0.56   | 0.44     | -      | 0.9285       | 11    |
| -      | 0.45     | 0.185  | -            | 11    |

**Figure 2.** The approach's performance for synthetic dataset.

In the next scenario, all variables are assumed to be exactly the same as mentioned earlier, except that the number of missing samples is assumed to be one or two. The results can be seen in Figure 3. This clarifies the performance of the proposed method based on error versus the data missing rate; it also shows that the proposed approach outperforms the other methods even in the case of two missing inputs.

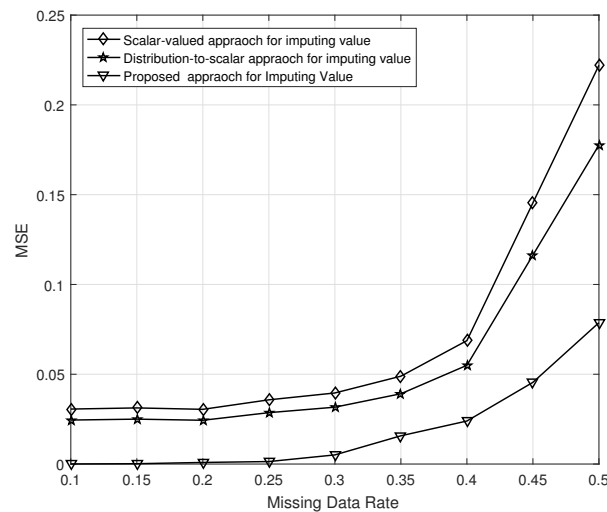


**Figure 3.** OWA approach for incomplete dataset aggregation.

Furthermore, we also show the performance of the proposed approach when the observation is a real-world dataset having nine attributes. In this case, the number of rings represents the label of these observations. Four different features are used to represent the observations in which we assumed that there is missing data. Missing data are randomly chosen where the number of observations is the same as in the first scenario. Different missing numbers are also applied to verify the reliability of our approach. Figure 4 shows the error versus the percentage of missing data. As in the first scenario, a different number of missing inputs is assumed to aggregate the data. From these experiments, it can generally be observed that even when there are significant effects of missing input for most cases, the proposed approach maintains a better performance against missing input in data aggregation, i.e., the proposed method has enough potential as a robust approach to aggregate missing inputs in datasets.

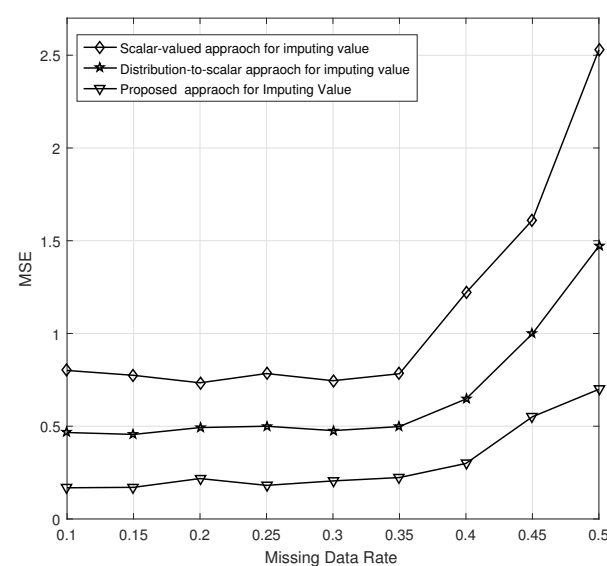


**Figure 4.** Performance analysis for the realization dataset.

#### 4. Conclusions

Determining an aggregation strategy with missing variables is intrinsically difficult since it is subjective to the nature of the data. In this paper, we propose a robust approach based on an OWA algorithm to estimate the measurement when the observations contain missing inputs. Different numbers of missing variables for various types of datasets are investigated and evaluated as an alternative validity metric. As shown in the results, learning with the OWA approach has significantly good performance in data aggregation where the missing data point has a different value in the observation. Therefore, this approach can be utilized as an efficient alternative method for data aggregation based on the learning of OWA operators. In general, we draw two major concluding factors in this paper. Initially, we have demonstrated the procedure for building data models in situations where there are missing inputs in the observed data. It has been argued for certain learning techniques to be applied to large databases because the cost is directly proportional to the number of variables. Then, it can also be seen that the aggregation operators are an efficient method to substitute missing sources in a real environment by learning OWA's operators. For future work, one could investigate optimization methods to identify the most effective parameters for the OWA operator when aggregating missing data. This could include analyzing multiple optimization algorithms, such as particle swarm optimization and the water cycle algorithm, to determine the settings that improve the imputation error.

**Author Contributions:** Conceptualization, D.T.A.; methodology, M.A.I. and M.A.-A.; software, M.A.-A.; validation, M.A.I. and M.A.-A.; formal analysis, M.A.-A.; investigation, M.A.-A.; resources, L.A.; data curation, M.A.I. and M.A.-A.; writing—original draft, M.A.-A. and L.A.; writing—review and editing, L.A.; supervision, D.T.A. and M.A.I. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Data Availability Statement:** The data can be found in [33].

**Conflicts of Interest:** Author Laith Alzubaidi was employed by the company Akunah Company. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

#### Abbreviations

The manuscript employs the following abbreviations, which may be found in following table (Table of Notation):

|                             |   |                                         |
|-----------------------------|---|-----------------------------------------|
| OWA                         | = | Ordered weighted averaging              |
| Min                         | = | Minimum value                           |
| Max                         | = | Maximum value                           |
| FM                          | = | Fuzzy measurements                      |
| QP                          | = | Quadratic programming                   |
| MSE                         | = | Mean square error                       |
| SSE                         | = | Sum square error                        |
| i.i.d.                      | = | Independent and identically distributed |
| Lightface letters           | = | Define a scalar value                   |
| Boldface-lower-case letters | = | Define a vector                         |
| Boldface-upper-case         | = | Define a matrix                         |
| <b>Variables</b>            |   |                                         |
| $\gamma$                    | = | Permutation of the arguments            |
| $O^T$                       | = | Transpose of the vector                 |
| $\hat{O}$                   | = | Estimated or predicated value           |
| $\bar{O}$                   | = | Mean of the variables                   |
| $\lceil O \rceil$           | = | The ceiling of the variable             |
| $O_{MSD}$                   | = | Most significant digit                  |
| $O_{LSD}$                   | = | Least significant digit                 |

## References

1. Ma, L.; Wang, M.; Peng, K. A missing manufacturing process data imputation framework for nonlinear dynamic soft sensor modeling and its application. *Expert Syst. Appl.* **2024**, *237*, 121428. [\[CrossRef\]](#)
2. Zhang, Y.; Thorburn, P.J. Handling missing data in near real-time environmental monitoring: A system and a review of selected methods. *Future Gener. Comput. Syst.* **2022**, *128*, 63–72. [\[CrossRef\]](#)
3. Peng, D.; Zou, M.; Liu, C.; Lu, J. RESI: A region-splitting imputation method for different types of missing data. *Expert Syst. Appl.* **2021**, *168*, 114425. [\[CrossRef\]](#)
4. Sareen, K.; Panigrahi, B.K.; Shikhola, T.; Sharma, R. An imputation and decomposition algorithms based integrated approach with bidirectional LSTM neural network for wind speed prediction. *Energy* **2023**, *278*, 127799. [\[CrossRef\]](#)
5. Li, D.; Zhang, H.; Li, T.; Bouras, A.; Yu, X.; Wang, T. Hybrid missing value imputation algorithms using fuzzy c-means and vaguely quantified rough set. *IEEE Trans. Fuzzy Syst.* **2021**, *30*, 1396–1408. [\[CrossRef\]](#)
6. Huang, Z.; Ng, M.K. A fuzzy k-modes algorithm for clustering categorical data. *IEEE Trans. Fuzzy Syst.* **1999**, *7*, 446–452. [\[CrossRef\]](#)
7. Wang, A.; Ye, Y.; Song, X.; Zhang, S.; James, J. Traffic prediction with missing data: A multi-task learning approach. *IEEE Trans. Intell. Transp. Syst.* **2023**, *24*, 4189–4202. [\[CrossRef\]](#)
8. Hammon, A. Multiple imputation of ordinal missing not at random data. *ASTA Adv. Stat. Anal.* **2023**, *107*, 671–692. [\[CrossRef\]](#)
9. Little, R.J.; Rubin, D.B. *Statistical Analysis with Missing Data*; John Wiley & Sons: Hoboken, NJ, USA, 2019; Volume 793.
10. Islam, M.A.; Anderson, D.T.; Petry, F.; Smith, D.; Elmore, P. The fuzzy integral for missing data. In Proceedings of the 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Naples, Italy, 9–12 July 2017; pp. 1–8. [\[CrossRef\]](#)
11. Zhang, C.; Qin, Y.; Zhu, X.; Zhang, J.; Zhang, S. Clustering-based missing value imputation for data preprocessing. In Proceedings of the 2006 4th IEEE International Conference on Industrial Informatics, Singapore, 16–18 August 2006; pp. 1081–1086.
12. Junninen, H.; Niska, H.; Tuppurainen, K.; Ruuskanen, J.; Kolehmainen, M. Methods for imputation of missing values in air quality data sets. *Atmos. Environ.* **2004**, *38*, 2895–2907. [\[CrossRef\]](#)
13. Tseng, S.M.; Wang, K.H.; Lee, C.I. A pre-processing method to deal with missing values by integrating clustering and regression techniques. *Appl. Artif. Intell.* **2003**, *17*, 535–544. [\[CrossRef\]](#)
14. Schneider, T. Analysis of incomplete climate data: Estimation of mean values and covariance matrices and imputation of missing values. *J. Clim.* **2001**, *14*, 853–871. [\[CrossRef\]](#)
15. Engemann, K.J.; Filev, D.P.; Yager, R.R. Modelling decision making using immediate probabilities. *Int. J. Gen. Syst.* **1996**, *24*, 281–294. [\[CrossRef\]](#)
16. Bai, C.; Zhang, R.; Song, C.; Wu, Y. A new ordered weighted averaging operator to obtain the associated weights based on the principle of least mean square errors. *Int. J. Intell. Syst.* **2017**, *32*, 213–226. [\[CrossRef\]](#)
17. Merigo, J.M.; Casanovas, M. Fuzzy Generalized Hybrid Aggregation Operators and its Application in Fuzzy Decision Making. *Int. J. Fuzzy Syst.* **2010**, *12*, 15–23.
18. Liu, D.; Li, T.; Liang, D. An integrated approach towards modeling ranked weights. *Comput. Ind. Eng.* **2020**, *147*, 106629. [\[CrossRef\]](#)
19. Lee, K.J.; Carlin, J.B.; Simpson, J.A.; Moreno-Betancur, M. Assumptions and analysis planning in studies with missing data in multiple variables: Moving beyond the MCAR/MAR/MNAR classification. *Int. J. Epidemiol.* **2023**, *52*, 1268–1275. [\[CrossRef\]](#) [\[PubMed\]](#)
20. Woods, A.D.; Gerasimova, D.; Van Dusen, B.; Nissen, J.; Bainter, S.; Uzdavines, A.; Davis-Kean, P.E.; Halvorson, M.; King, K.M.; Logan, J.A.; et al. Best practices for addressing missing data through multiple imputation. *Infant Child Dev.* **2024**, *33*, e2407. [\[CrossRef\]](#)
21. Al-Amidie, M.; Al-Asadi, A.; Micheas, A.C.; Islam, N.E. Spectrum sensing based on Bayesian generalised likelihood ratio for cognitive radio systems with multiple antennas. *IET Commun.* **2019**, *13*, 305–311. [\[CrossRef\]](#)
22. Park, J.; Müller, J.; Arora, B.; Faybishenko, B.; Pastorello, G.; Varadharajan, C.; Sahu, R.; Agarwal, D. Long-term missing value imputation for time series data using deep neural networks. *Neural Comput. Appl.* **2023**, *35*, 9071–9091. [\[CrossRef\]](#)
23. Sun, Y.; Li, J.; Xu, Y.; Zhang, T.; Wang, X. Deep learning versus conventional methods for missing data imputation: A review and comparative study. *Expert Syst. Appl.* **2023**, *227*, 120201. [\[CrossRef\]](#)
24. Honghai, F.; Guoshun, C.; Cheng, Y.; Bingru, Y.; Yumei, C. A SVM regression based approach to filling in missing values. In Proceedings of the International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Melbourne, Australia, 14–16 September 2005; Springer: Berlin/Heidelberg, Germany, 2005; pp. 581–587.
25. Zhang, Y.; Wei, X.; Li, C. Manifold clustering optimized by adaptive aggregation strategy. *Knowl. Inf. Syst.* **2023**, *65*, 379–408. [\[CrossRef\]](#)
26. Yager, R.R. On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Trans. Syst. Man Cybern.* **1988**, *18*, 183–190. [\[CrossRef\]](#)
27. Dujmovic, J.J. Continuous preference logic for system evaluation. *IEEE Trans. Fuzzy Syst.* **2007**, *15*, 1082–1099. [\[CrossRef\]](#)
28. Calvo, T.; Mesiar, R.; Yager, R.R. Quantitative weights and aggregation. *IEEE Trans. Fuzzy Syst.* **2004**, *12*, 62–69. [\[CrossRef\]](#)
29. Merigó, J.M.; Gil-Lafuente, A.M. The induced generalized OWA operator. *Inf. Sci.* **2009**, *179*, 729–741. [\[CrossRef\]](#)
30. Grabisch, M.; Sugeno, M. Multi-attribute classification using fuzzy integral. In Proceedings of the [1992 Proceedings] IEEE International Conference on Fuzzy Systems, San Diego, CA, USA, 8–12 March 1992; pp. 47–54.

31. Anderson, D.T.; Price, S.R.; Havens, T.C. Regularization-based learning of the choquet integral. In Proceedings of the 2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Beijing, China, 6–11 July 2014; pp. 2519–2526.
32. Cho, S.B.; Kim, J.H. Combining multiple neural networks by fuzzy integral for robust classification. *IEEE Trans. Syst. Man Cybern.* **1995**, *25*, 380–384.
33. Waugh, S.; Adams, A. Pruning within cascade-correlation. In Proceedings of the ICNN'95-International Conference on Neural Networks, Perth, WA, Australia, 27 November–1 December 1995; Volume 3, pp. 1206–1210.

**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.