applied sciences

Article

Development of an Evaluation Method for Deriving the Water Loss Reduction Factors of Water Distribution Systems: A Case Study in Korean Small and Medium Cities

Young Hwan Choi [1], Taeho Choi [2], Do Guen Yoo [3] and Seungyub Lee [4,*]

1 Department of Civil and Infrastructure Engineering, Gyeongsang National University, Jinju 52725, Republic of Korea
2 K-water Research Institute, K-water, Daejeon 34045, Republic of Korea
3 Department of Civil Engineering, The University of Suwon, Hwaseong-si 18323, Republic of Korea
4 Department of Civil and Environmental Engineering, Hannam University, Daejeon 34430, Republic of Korea
* Correspondence: seungyub.lee@hnu.kr

Abstract: This study introduces a method that can evaluate the efficiency of leakage management practices and devises a formula to set leakage management goals. To develop the evaluation method for deriving leakage reduction factors, real data from small- and medium-sized cities in South Korea were collected. With the data collected, four leakage management factors (or activities) that could improve revenue water ratio or reduce leakage ratio were identified. With the leakage management factors, correlation analysis was carried out to identify the relationship between independent and dependent variables and within independent variables. Once the relationships were identified, standardization of the data using T-score conversion was carried out to scale all data with different units into similar ranges. Finally, the efficiency of leakage management actions was determined by the formulation of leakage using various data analysis approaches using multiple linear regression analysis and deep neural networks. As a result, pipe replacement was determined as an essential activity to decrease the leakage ratio or increase the revenue water ratio. In addition, annual water loss management actions of the small cities were more actively performed. Furthermore, the performance of data analysis using DNN is more appropriate in data classification, considering the characteristics of time series rather than independent data analysis. Through comparison of the above data classification approaches, the increase or decrease in the leakage ratio/revenue water ratio by the water loss management activity of local water distribution systems can be used to construct a more effective model for classification considering both local and temporal characteristics.

Keywords: water distribution system; leakage management; correlation analysis; revenue water ratio; multiple regression analysis; deep neural network

check for updates

Citation: Choi, Y.H.; Choi, T.; Yoo, D.G.; Lee, S. Development of an Evaluation Method for Deriving the Water Loss Reduction Factors of Water Distribution Systems: A Case Study in Korean Small and Medium Cities. *Appl. Sci.* **2022**, *12*, 12530. https://doi.org/10.3390/app122412530

Academic Editor: José Miguel Molina Martínez

Received: 13 November 2022
Accepted: 5 December 2022
Published: 7 December 2022

1. Introduction

The water distribution system (WDS) is one of the major infrastructures and supplies water to customers with sufficient pressure and quality. However, recent climate change has increased the frequency of drought events, and water availability has dropped significantly. Therefore, the need for efficient water supply and distribution using WDSs has increased. In most cases, transmission mains are regularly maintained, but distribution mains and service pipes are relatively difficult to maintain. Due to this issue, the rate of water loss is gradually increasing. In addition, as water is recognized as an economic product, water loss leads to economic loss. Therefore, recently, many studies have investigated the reduction of water loss, water loss management, and the factors influencing water loss.

The South Korea Water utility uses the revenue water ratio as a performance indicator for WDSs. If the target of the revenue water ratio in the planning stage is achieved in terms of operation, it is judged that the target of the project has been achieved. Therefore, the revenue water ratio is used as an indicator to evaluate the efficiency of business performance and operational management.

Since 2004, in small and medium cities in South Korea, 22 cities have begun evaluating the project planning and operation process to solve the aging problem and improve the operational efficiency of WDSs. In addition, in this regard, the revenue water ratio was used to evaluate each project. The revenue water ratio is an indicator used to evaluate the performance of waterworks projects, and because there are no other quantitative indicators other than the revenue water ratio, there is no choice but to use the revenue water ratio as a future project goal.

The revenue water ratio is a factor that evaluates the performance of WDSs, and in order to improve their performance, the revenue water ratio should be improved. The representative approach to improve the revenue water ratio is to manage effective water systems and reduce ineffective ones (i.e., leakage). As mentioned earlier, the revenue water ratio is the ratio of the total water supply to the effective water quantity; to increase the revenue water ratio, there are ways to increase the quantity of effective water systems or decrease the quantity of ineffective ones.

First, one effective water management approach to improve the revenue water ratio is to accurately manage the amount produced by the water treatment plant's so-called supply management. For efficient supply management, it is necessary to determine the exact production water volume through regular inspection and replacement of the flow meter, and there is also a method to efficiently increase the supply amount by improving the supply system.

Second, as an effective water management method, an accurate fee should be charged through an accurate meter reading. This method involves regularly checking and replacing old water meters in order to obtain more accurate meter readings, and to reduce reading errors through accurate meter reading by meter-reading personnel. Such accurate meter reading is a way to reduce the amount of water loss by preventing the illegal use of water.

The third approach to improve the revenue water ratio is to manage ineffective water quantity, that is, to reduce water loss. To reduce water loss, leakage reduction activities such as pipe replacement, meter replacement, repair, leak detection, and pressure reduction valve installation are required. However, because the quantitative efficiency of the improvement of the revenue water ratio for each leakage reduction activity cannot be evaluated, it is insufficient as a basis for calculating the project cost for various waterworks projects. In addition, it is difficult to reflect on reality when establishing a strategy to improve the revenue water ratio for the project or when calculating a leak management goal [1–3].

For these reasons, recently, studies have been conducted related to the leakage reduction framework, evaluation of the water loss reduction approach, and evaluation of the factors influencing water loss. Zyoud et al. [4] showed the application of multi-criteria decision analysis (MCDA) based on participatory interaction approaches to select the most appropriate solutions to reduce and manage water losses in water supply systems and make the best consensus decisions in a complex environment. Ndunguru and Hoko [5] evaluated the non-revenue water situation in Harare water in some areas of Zimbabwe and assessed the contribution of water leakages to the water losses based on the monitoring data and SANFLOW model to determine average real losses. According to these studies, evaluating the effect of water loss reduction and the process of evaluating and determining the amount of leakage in the WDS is performed with four different methods: the top-down approach (water balancing), water and wastewater balancing, component analysis of leakage (the background and bursts estimates method), and the bottom-up approach (the minimum night flow method) [6–8]. Moreover, related to the evaluation of leakage reduction, data envelopment analysis (DEA) was performed to assess the effect of water loss reduction activity (e.g., pipe replacement, meter replacement, water loss repair, leakage detection, etc.) [9].

Although the above studies were conducted based on various approaches to reduce leakages, these methods are not practical approaches for several reasons: (a) lack of consideration for the leakage characteristics of real-world systems, (b) insufficient evaluation of leakage reduction methods, and (c) absence of an accurate and systemized method evaluating economic feasibility.

Therefore, in this study, the revenue water ratio improvement factors (investment and output indexes) were determined using real water leakage reduction data. Data were obtained from waterworks projects performed at small- and medium-sized cities in South Korea from 2004 to 2018. This study used statistical methods (i.e., multiple regression analysis; MRA) that have been widely applied to analyze water performance efficiency, along with artificial intelligence techniques (i.e., deep neural network; DNN). The revenue water ratio improvement factor derived through this study is an empirical methodology for evaluating the efficiency of leakage management and can be used to estimate leakage management goals in the future.

2. Materials and Methodology

In this study, evaluation of the efficiency of leakage management and the leakage management model formulation were derived using data related to revenue water ratio improvements according to the leakage of small- and medium-sized WDSs. The evaluation was performed by MRA analysis and DNN, which followed the procedure shown in Figure 1. There are three steps in developing the approach for the evaluation of the water loss reduction factors, which are (1) data collection, (2) pre-processing of data, and (3) evaluation of water loss management efficiency and determination of the formulation of water loss.

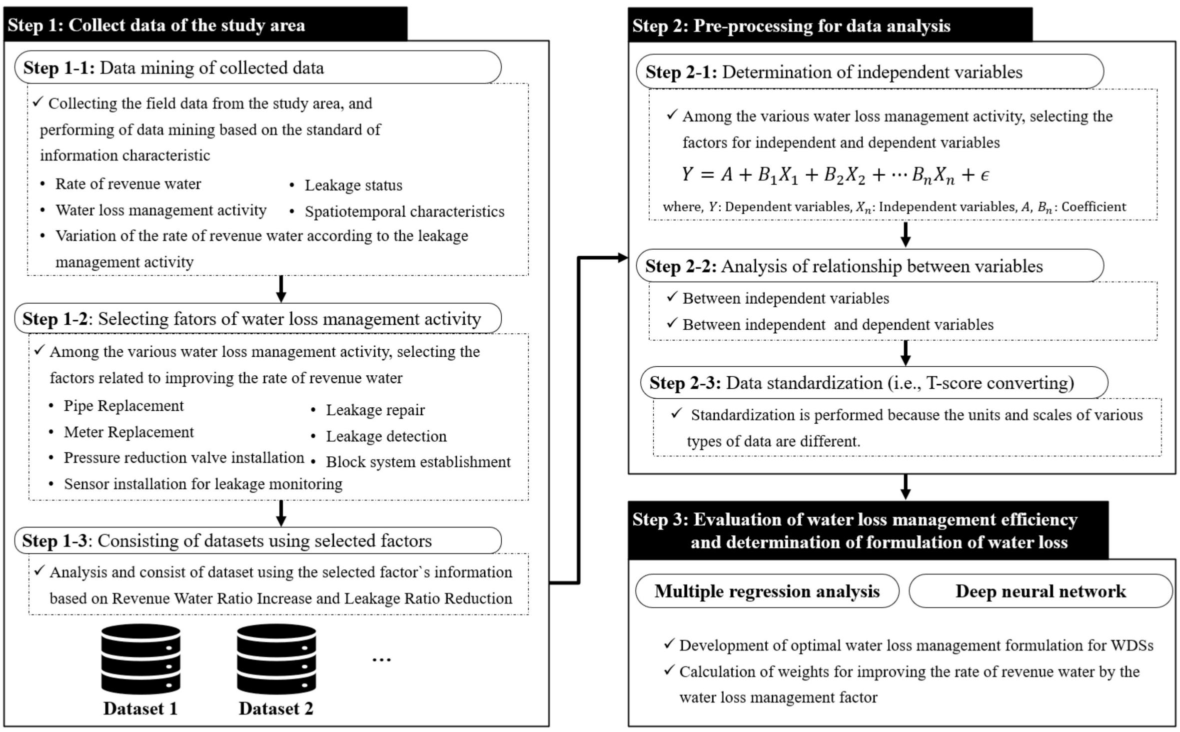

Figure 1. Procedure of this study.

2.1. Collection Data in Study Area

In the first step, field data from the study area were collected, and data mining was performed based on the information characteristic (e.g., revenue water ratio, leakage management activity, variation of the revenue water ratio according to the leakage management activity, leakage status, and the spatiotemporal characteristics of leakage). This study considered the variation of the revenue water ratio according to the leakage management activity as a data-mining standard and then, among the various leakage management activities, the factors that can improve the revenue water ratio were selected.

Leakage can be categorized into burst or background leakage, according to characteristics such as actual quantity of leakage and type of leakage generation. Background leakage is leakage that occurs at the pipe connection or joint, and the amount of leakage is too small to be recognized. Burst leakage is divided into reported and unreported burst leakage. Generally, the reported burst leakage is a case in which leakage can be confirmed by exposure to the surface or by a significant amount greater than normal conditions in terms of supply, whereas unreported burst leakage is leakage that is difficult to detect [10,11].

The International Water Association (IWA) Water Losses Task Force is divided into seven categories, using the concepts of unavoidable annual real loss, current annual real losses, and potentially recoverable real losses, to analyze the change in the amount of water leakage according to the leakage management methods (i.e., pipe replacement, meter replacement, pressure reduction valve installation, sensor installation for leakage monitoring, leakage repair, leakage detection, and block system establishment). Among these, pressure reduction valve installation, block system establishment, and sensor installation for leakage monitoring were excluded in this study because it is difficult to quantify their effects, and these operations are irregular.

Therefore, considering the leakage management factors suggested by the IWA, pipe replacement (PR), meter replacement (MR), leakage repair (LR), and leakage detection (LD) were selected as the water loss management factors to estimate revenue water ratio increase (or reduction of leakage ratio) following those activities. This study applied six raw data elements: four leakage management factors (PR, MR, LR, LD), the revenue water ratio, and the leakage ratio reduction. The variance of the revenue water ratio and the leakage ratio reduction according to the four leakage management factors were then calculated according to those six raw data elements. The data were collected every year for 10 years, comparing the maintenance effect with the leakage management activity. Therefore, based on the characteristics of the applied data, this study performed two types of data mining based on the revenue water ratio or the leakage ratio and time-series characteristics.

2.2. Pre-Processing for Data Analysis

In step 2, the collected data were pre-processed to evaluate the efficiency of leakage management and to determine the water loss formula. First, four leakage management factors (PR, MR, LR, LD) were classified into independent and dependent variables based on the revenue water ratio increase and the leakage ratio reduction. The correlations of variables (i.e., independent and dependent variables) were then evaluated. Finally, the collected data were standardized using standardization scores (e.g., T-score converting) to alleviate the unit and scale differences between factors.

2.2.1. Determination of Independent and Dependent Variables

The first part of step 2 is the categorization of leakage management factors, revenue water ratio, and leakage ratio into independent and dependent variables. The four leakage management factors (PR, MR, LR, LD) determined from step 1 will directly or indirectly affect changes in the revenue water ratio or the leakage ratio. Therefore, the leakage management factors were determined as independent variables, and the revenue water ratio increase and the leakage ratio reduction were determined as dependent variables.

2.2.2. Correlation Analysis between Independent Variables

The correlations were analyzed to prevent multicollinearity between the four influencing factors selected to derive dependent and independent variables. The higher the covariance, the higher the correlation, and when the covariance is completely identical, the correlation becomes 1. The correlation coefficient indicates the degree of such correlation. The Pearson correlation coefficient [12] is used for normally distributed quantitative variables, and Kendall's tau correlation [13] is used to measure the relationship between sequences when they are not normally distributed, or when the order of categories is not specified. The correlation coefficient ranges from −1 (completely negative relationship) to 1 (completely positive relationship); 0 indicates a non-linear relationship, and the correlation between independent variables that does not cause a problem of multicollinearity is conservatively 0.7, as judged below.

2.2.3. Correlation Analysis between Independent and Dependent Variables

In order to minimize the problem of collinearity between independent variables, there should be no correlation between independent variables. However, typically, the correlation between dependent and independent variables should be high. For this reason, it is necessary to select variables that include significant independent variables and remove independent variables without statistical significance through correlation analysis of each dependent–independent variable combination to evaluate the leakage management effectiveness and determination of formulation. Therefore, according to each data classification group, it is excluded from the independent variable, and the target calculation formula is derived.

2.2.4. Data Standardization through T-Score Conversion

Data standardization is a technique for determining the relative position of each variable when it is difficult to evaluate in the same standard because the units and ranges of the acquired data are different. In this study, T-score conversion was applied, taking into account the characteristics of the data. In order to compare the effects on the increase in the flowrate and the decrease in the leakage rate, it was converted into a T-score that was converted into an average of 50 and a standard deviation of 10 and applied. This can be calculated as in Equation (1).

$$Tscore = 10\left(\frac{X-\mu}{\sigma}\right) + 50 \tag{1}$$

where X: variables, μ: average of variable $X_1,\ X_2, \cdots X_N$, σ: standard deviation of variable $X_1,\ X_2, \cdots X_N$.

2.3. Evaluation of Leakage Management Efficiency and Determination of Formulation of Leakage

In this study, the efficiency of leakage management actions was evaluated, and the formulation of leakage was determined using various data analysis approaches, such as MRA analysis and DNN model.

2.3.1. Multiple Regression Analysis

Multiple Regression Analysis is a type of regression analysis that estimates the relationship between variables using a statistical method. In regression analysis, there is an independent variable that is a cause and a dependent variable that is an effect. In this case, multiple regression analysis is a method of performing analysis on a regression model in which there is one dependent variable and two or more independent variables [14].

In this study, the increase in the flow rate and the decrease in the leakage rate according to the data classification criteria were classified according to four input indicators (PR, MR, LR, LD), and the target formulation was derived according to the increase in the flow rate and the decrease in the leakage rate for the derived input indicators.

2.3.2. Deep Neural Network

Another technique applied in this study to evaluate the efficiency of leakage management is DNN [15] (Figure 2). DNN is a technique composed of two or more hidden layers of an Artificial Neural Network (ANN) [16], and DNN have the advantage of learning complex data with fewer neurons compared to ANN. In addition, it is suitable for predicting non-linear data, and the performance can be adjusted by adjusting the number of neurons and the number of hidden layers, which are parameters of the technique. However, DNN has a high possibility of overfitting, and because this problem is directly related to the performance of the model, it is necessary to adjust parameters and select an activation function to prevent overfitting.

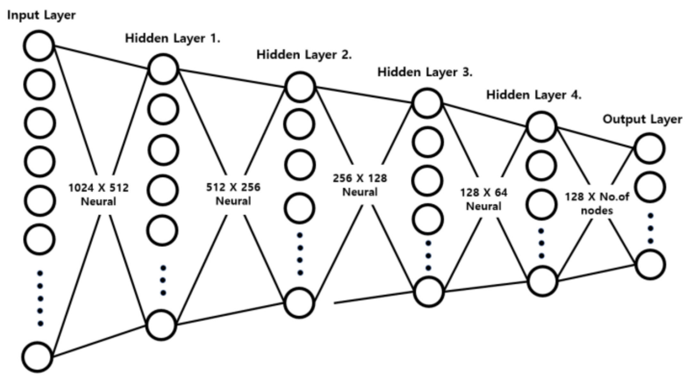

Figure 2. Configuration of Deep Neural Network Model.

In each hidden layer, a function is applied to the value that enters the neuron and is then transmitted. The function used at this time is called the activation function, and the Rectified Linear Unit (Relu) function is applied to each neuron as the activation function. Equation (2) is the expression defining the Relu function.

$$f(x) = \max(0,\ x) \tag{2}$$

Figure 3 is the graph form of the Relu function. If the input value is less than 0, it is output as 0, and if it is greater than 0, the input value is output as it is. The Relu function learns faster than other sigmoid and tanh functions and does not have a gradient loss problem. For the above reasons, the Relu function was applied in each layer when constructing the model.

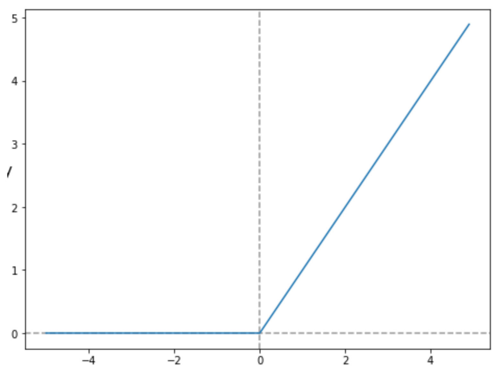

Figure 3. Formulation of Relu function.

When training the model, an error function that calculates the error between the output value and the target value of the model is specified. The Mean Squared Error (MSE) function was used as the error function when constructing the learning model. The MSE function is the mean square error between the predicted value and the actual value. In addition, an optimization function is essential when training a model. The optimization function is a function that finds the parameter that minimizes the value of the loss function. The Adaptive Moment Estimation (Adam) function was applied to the per-rate model. Adam is an algorithm that combines the strengths of $\mathrm{RMS_{Prop}}$ and Momentum, which are used as optimization functions. Like Momentum, it stores the exponential average of the slope, and similarly to $\mathrm{RMS_{Prop}}$, it stores the exponential average of the square of the device. In this study, when constructing a learning model, one input layer, five hidden layers, and one output layer were stacked, and the number of neurons in each hidden layer was composed of 64, 128, 256, and 512.

3. Application and Results

3.1. Study Area and Data

In this study, data on the statistics of waterworks between 2005 and 2018 were used to initially select the study areas. Among the 163 local small and medium cities across South Korea, 22 study areas were selected to ensure there were no missing data or errors in terms of collecting data and that the leakage reduction effect was clear according to the activity of leakage management.

To be defined as small and medium cities, cities had to be of typical type, not too big or too small. In Korea, city size is generally defined by population. The population of local small and medium cities is at least 50,000, and the maximum range of population is 1 million. The selected cities' data were accumulated for up to 14 years. For leakage management factors, PR, LR, and LD were divided by the total pipe length, and MR was used as the value divided by the water supply population in order to offset the difference depending on the size of WDS. In addition, only leakage management factors with either a revenue water ratio less than 80% or a leakage ratio higher than 10% were applied. Examples of data are shown in Table 1.

Table 1. Example of data from the application area.

Study Area	Revenue Water Ratio	Increase Revenue Water Ratio	Leakage Ratio	Decrease Leakage Ratio	Pipe Replacement (PR)	Meter Replacement (MR)	Leakage Repair (LR)	Leakage Detection (LD)
	%	%	%	%	km/km × Million	No./Per. × 1000	No./km × 1000	No./km × 1000
TY 2011	47.1	6.7	40.5	−5.7	28.776	40.958	1.771	1.261
HP 2011	49.3	−0.9	28.5	0.1	0.000	30.599	0.359	0.186
HP 2010	50.2	5.4	28.6	−0.2	19.158	141.390	0.645	0.250
BH 2014	52.2	−15.2	43.1	−15.5	0.229	37.663	0.107	0.000
NS 2005	54.1	−2.6	36.3	0.2	11.594	35.942	2.839	0.516
JD 2014	54.8	9.0	40.7	9.7	22.270	89.247	0.489	0.249
BH 2015	57.6	5.4	31.6	11.5	9.971	114.219	0.455	0.213
SC 2007	58.0	10.9	36.0	−0.4	24.310	24.889	0.920	0.456
GS 2011	58.3	10.6	36.6	3.2	14.601	59.704	0.464	0.630
WD 2015	59.3	8.7	40.7	8.7	7.408	14.482	0.298	0.160
TY 2012	59.6	12.5	35.9	4.6	35.705	27.696	1.934	0.756
HP 2012	59.8	10.5	35.7	−7.2	0.000	30.473	0.346	0.147
CS 2018	59.8	0.1	37.0	3.3	0.000	227.128	0.644	0.319
GR 2008	60.6	9.2	35.0	8.4	8.053	32.365	1.137	0.272
HP 2013	62.5	2.7	28.6	7.1	10.347	1.468	0.457	0.269
JE 2006	62.7	9.7	32.0	8.0	16.094	19.853	0.842	0.072
DY 2009	64.1	2.5	14.1	0.4	13.584	67.877	0.762	0.593
YC 2007	64.2	7.1	27.9	−4.2	40.304	55.074	2.100	0.100
NS 2006	64.5	10.4	29.4	6.9	45.217	10.604	4.322	0.180
BH 2016	64.9	7.3	25.9	5.7	29.730	52.147	0.505	0.207
SC 2008	65.2	7.2	28.9	7.1	15.390	21.729	0.802	0.318
GS 2012	66.2	7.9	29.0	7.6	41.112	36.274	0.659	0.342
GJ 2009	66.4	7.6	27.4	9.4	18.491	5.685	2.846	1.097
JD 2015	66.7	11.9	28.5	12.2	30.748	31.010	0.315	0.293
JE 2007	67.7	5.0	15.1	16.9	32.885	8.071	0.951	0.292
GJ 2010	67.8	1.4	26.9	0.5	19.846	7.281	1.717	0.763
WD 2016	67.9	8.6	27.2	13.5	40.430	30.411	0.336	0.258

Table 1. *Cont.*

Study Area	Revenue Water Ratio	Increase Revenue Water Ratio	Leakage Ratio	Decrease Leakage Ratio	Pipe Replacement (PR)	Meter Replacement (MR)	Leakage Repair (LR)	Leakage Detection (LD)
	%	%	%	%	km/km × Million	No./Per. × 1000	No./km × 1000	No./km × 1000
GeS 2009	69.5	−1.1	25.1	−1.0	32.920	23.754	0.905	0.237
NS 2007	69.8	5.3	27.5	1.9	125.818	10.432	2.604	0.074

3.2. Pre-Processing for Data Analysis

The collected data show the change in the revenue water ratio and the leakage ratio according to the activity of leakage management (PR, LR, LD, MR). To evaluate the leakage reduction factors, these data were pre-processed. Table 2 shows the correlation between independent variables. Because the correlation depends on the data categorization, in this study, to compare the correlation by data classification criteria, revenue water ratio increase has been divided into four sections: (1) less than 60%, (2) 60–70%, (3) 70–80%, (4) higher than 80%, and four sections of leakage ratio: (1) less than 10%, (2) 10–20%, (3) 20–30%, (4) higher than 30%.

Table 2. Results of correlation analysis between independent variables.

Standard of Categorization			PR	MR	LR	LD
All data	Leakage ratio of higher than 20%	PR	1	-	-	-
		MR	−0.103	1	-	-
		LR	0.401	−0.145	1	-
		LD	0.006	−0.044	0.360	1
	Revenue water ratio of less than 70%	PR	1	-	-	-
		MR	−0.268	1	-	-
		LR	0.455	−0.273	1	-
		LD	−0.079	−0.086	0.318	1
The leakage ratio	Less than 10%	PR	1	-	-	-
		MR	0.461	1	-	-
		LR	0.832	0.422	1	-
		LD	0.854	0.457	0.866	1
	10–20%	PR	1	-	-	-
		MR	0.021	1	-	-
		LR	0.416	−0.002	1	-
		LD	0.272	0.064	0.649	1
	20–30%	PR	1	-	-	-
		MR	0.025	1	-	-
		LR	0.420	−0.168	1	-
		LD	−0.118	−0.176	0.243	1
	Higher than 30%	PR	1	-	-	-
		MR	−0.316	1	-	-
		LR	0.497	−0.213	1	-
		LD	0.690	−0.071	0.616	1

Table 2. *Cont.*

	Standard of Categorization		PR	MR	LR	LD
The revenue water ratio	Less than 60%	PR	1	-	-	-
		MR	−0.191	1	-	-
		LR	0.525	−0.195	1	-
		LD	0.710	−0.157	0.664	1
	60–70%	PR	1	-	-	-
		MR	−0.136	1	-	-
		LR	0.386	−0.337	1	-
		LD	−0.373	−0.165	0.151	1
	70–80%	PR	1	-	-	-
		MR	0.026	1	-	-
		LR	0.399	−0.021	1	-
		LD	0.113	0.023	0.431	1
	Higher than 80%	PR	1	-	-	-
		MR	0.106	1	-	-
		LR	0.186	0.003	1	-
		LD	0.240	−0.065	0.671	1

As a result, PR–LR (0.832), PR–LD (0.854), and LD–LR (0.866) of leakage ratio less than 10% showed a high correlation of more than 0.7. Therefore, the leakage ratio of less than 10% data was excluded when generating the prediction equation. In addition, the data with a revenue water ratio of less than 60% showed a high correlation for LD–PR (0.71), but in other data show correlation of less than 0.7.

Moreover, in order to minimize the problem of collinearity between independent variables, there should be no correlation between independent variables. However, the correlation between the dependent and independent variables should be high. Therefore, this study performed an analysis of correlation between the independent and the dependent variable, and Table 3 shows the results.

Table 3 shows the results of analysis of correlation components between dependent–independent variables according to various classification criteria. Only data of the leakage ratio of less than 10% showed high correlation between dependent–independent variables, and low correlation in other cases. In other words, it was confirmed that, even if old pipes and meters were replaced and leakage detection and repair were performed in some areas, the tendency of a large increase in the revenue water ratio or a decrease in the leakage ratio was small. For this reason, in order to derive a leakage management formulation according to the data analysis group, it is necessary to include significant independent variables and remove statistically insignificant independent variables. Therefore, the significance of independent variables was considered when deriving the leakage management formulation according to the data analysis group.

In this study, because the dependent and independent variables used different units, data standardization should be performed for correlation analysis and data analysis. Therefore, in this study, T-score was used, which was converted to an average of 50 and a standard deviation of 10.

Table 3. Results of correlation analysis between independent and dependent variables.

Standard of Categorization		Pipe Replacement (PR)	Meter Replacement (MR)	Leakage Repair (LR)	Leakage Detection (LD)
		km/km × Million	No./Per. × 1000	No./km × 1000	No./km × 1000
All data	Leakage ratio of higher than 20%	0.112	−0.013	0.064	0.059
	Revenue water ratio of less than 70%	0.131	−0.074	−0.026	−0.034
The leakage ratio	Less than 10%	0.776	0.595	0.688	0.695
	10–20%	0.582	−0.060	0.284	0.166
	20–30%	0.097	−0.213	0.094	0.225
	Higher than 30%	0.216	0.199	0.003	−0.119
The revenue water ratio	Less than 60%	0.569	−0.111	0.071	0.374
	60–70%	0.023	0.082	0.136	−0.227
	70–80%	0.463	−0.190	0.287	0.197
	Higher than 80%	0.507	−0.103	0.107	0.248

Table 4 shows the results of converting factors that have a correlation of 0.5 or more to a T-score in the correlation analysis between the dependent and independent variables to obtain the formulation of revenue water ratio of the leakage ratio of less than 10% data. According to the decrease in the leakage ratio, the variable values of PR converted to T-scores were 42.28–91.54, MR, 37.56–86.53; LR, 37.62–91.85; and LD, 40.16–89.58. T-score values of the independent variables were converted to a relatively similar range.

Table 4. Results of converting to T-scores.

Study Area	Leakage Ratio Reduction (%)	Pipe Replacement (PR)	Meter Replacement (MR)	Leakage Repair (LR)	Leakage Detection (LD)	T-Score for PR	T-Score for MR	T-Score for LR	T-Score for LD
NS 2012	0.50	6.45	15.96	0.40	0.12	61.74	51.38	59.54	57.62
GJ 2013	−1.20	0.00	21.01	0.19	0.01	42.28	56.61	46.61	41.32
GJ2017	0.80	0.28	3.63	0.16	0.02	43.14	38.65	44.61	42.20
PJ 2012	0.60	3.20	9.87	0.38	0.12	51.94	45.10	57.83	56.98
DDC 2015	−2.70	2.66	13.01	0.17	0.06	50.32	48.33	45.12	47.41
GJ 2018	0.50	0.33	28.02	0.16	0.04	43.27	63.84	44.93	45.69
YJ 2010	−1.60	1.26	10.88	0.32	0.11	46.07	46.14	54.73	54.95
PJ 2013	0.40	2.45	9.39	0.33	0.04	49.68	44.60	54.85	45.05
YJ 2011	0.20	0.92	15.67	0.28	0.03	45.07	51.09	52.07	44.09
PJ 2014	0.40	2.19	8.99	0.34	0.12	48.90	44.18	55.67	56.45
GJ 2012	1.90	0.28	18.49	0.18	0.02	43.12	54.00	45.65	42.77
PJ 2015	0.40	7.35	10.22	0.42	0.12	64.45	45.45	60.67	57.07
DDC 2017	−0.40	3.36	24.76	0.12	0.05	52.43	60.48	42.43	46.31
YJ 2013	0.00	0.00	15.11	0.17	0.01	42.28	50.51	45.40	40.16
YJ 2012	1.30	0.00	3.26	0.21	0.01	42.28	38.26	47.98	40.18
DDC 2016	1.80	1.15	32.23	0.04	0.02	45.75	68.19	37.62	41.57
YJ 2009	−0.90	1.47	9.03	0.25	0.10	46.72	44.23	49.97	54.14
NJ 2016	2.70	0.18	22.49	0.09	0.09	42.83	58.13	40.73	53.12
PJ 2017	−0.50	4.11	8.80	0.22	0.13	54.69	43.99	48.54	58.75
EC 2008	20.40	16.33	49.98	0.94	0.34	91.54	86.53	91.86	89.58

Table 4. *Cont.*

Study Area	Leakage Ratio Reduction (%)	Pipe Replacement (PR)	Meter Replacement (MR)	Leakage Repair (LR)	Leakage Detection (LD)	T-Score for PR	T-Score for MR	T-Score for LR	T-Score for LD
YJ 2015	−0.30	0.15	6.21	0.14	0.03	42.73	41.32	43.34	43.38
YJ 2014	0.70	0.14	8.25	0.14	0.01	42.71	43.42	43.32	41.22
PJ 2016	1.30	3.87	15.03	0.31	0.15	53.94	50.42	53.65	61.41
DDC 2014	6.00	6.98	2.57	0.17	0.07	63.32	37.56	45.52	49.15
DDC 2018	2.10	2.17	10.66	0.21	0.10	48.83	45.91	47.79	53.94
PJ 2018	1.80	3.23	14.54	0.41	0.11	52.03	49.92	59.78	55.95
YJ 2018	−0.50	1.54	15.93	0.17	0.02	46.92	51.35	45.46	42.69
YJ 2017	−0.20	0.76	11.52	0.11	0.01	44.56	46.80	41.79	40.84
YJ 2016	3.40	1.38	8.42	0.12	0.05	46.46	43.60	42.53	46.02

3.3. Evaluation of Water Loss Management Efficiency and Determination of Formulation of Water Loss

In this study, the increase in the flow rate and the decrease in the leakage rate according to the data classification criteria (higher than 10% of leakage ratio, less than 80% of the revenue water ratio) were analyzed according to four input indices (PR, MR, LR, LM). Through this, a leakage management model formulation was derived according to the increase in the revenue water ratio and the decrease in the leakage ratio for input indicators. In order to derive a reliable leakage management model formulation, MRA and DNN were used, and Table 5 shows the result of deriving the leakage management model formulation according to the data classification criteria.

Table 5. Results of deriving leakage management model formulation.

Standard of Categorization		MRA						DNN
		Leakage Management Model Formulation	Used Variables				R	
			X1	X2	X3	X4		
Leakage ratio	Less than 10% (Group 1)	YG1 = 0.22X1 + 0.12X2 + 0.04X3 − 3.17	O	O	O		0.82	0.78
	10–20% (Group 2)	YG2 = 0.14X1 + 1.06X3 + 0.24	O		O		0.59	0.56
	20–30% (Group 3)	YG3 = 0.03X1 + 7.01X4 + 1.71	O			O	0.32	0.57
	Higher than 30% (Group 4)	YG4 = 0.52X1 + 0.06X2 + 1.47X3 − 3.41	O	O	O		0.58	0.66
Revenue water ratio	Less than 60% (Group 5)	YG5 = 0.42X1 + 3.75X4 + 1.14	O			O	0.64	0.71
	60–70% (Group 6)	YG6 = 0.02X2 + 0.94X3 + 6.99		O	O		0.34	0.45
	70–80% (Group 7)	YG7 = 0.08X1 + 0.85X3 + 4.34X4 + 1.53	O		O	O	0.53	0.55
	Higher than 80% (Group 8)	YG8 = 0.1X1 + 4.35X4 − 0.47	O			O	0.55	0.58

As a result of the leakage management model formulation, the coefficient of determination (R) of the eight classification criteria was distributed from 0.32 to 0.82, and an average value of 0.55 for MRA and 0.61 for DNN. Among the eight groups, the highest coefficient of determination in both analysis through MRA and DNN is Group 1, with a leakage ratio of less than 10%. In the case of Group 1, it depends on PR, MR, LD, etc., to reduce the leakage ratio, and it is confirmed that PR has the greatest influence on reducing the leakage ratio. In addition, if the leakage ratio is less than 10%, the revenue water ratio is high. In addition, it was confirmed that PR among the four leakage reduction activities was essential for reducing the leakage ratio and increasing the revenue water ratio in all classification criteria, except for Group 6, where the revenue water ratio was 60–70%. According to Group 8 in contrast to Group 1, LD was shown to be the most significant factor to increase the revenue water ratio. In the case of Group 8, the revenue water ratio is already high, and because PR, MR, and LR are judged to be in progress, the revenue water ratio is expected to decrease if no leakage reduction activities are carried out. This is because the model formula was derived as a negative number (-0.47), as in Group 1. The average value for the coefficient of determination on each data classification criteria of MRA and DNN were 0.55 and 0.61. In all groups, except for Groups 1 and 2, the DNN model of the coefficient of determination was high. Through the above results, it can be judged that DNN is a suitable model for the formulation of water loss management.

However, the conditions required for performing each model are different, and the results can also be derived differently depending on data processing, such as model parameters and training data mining. As a result of comparison of the models applied in this study, it was found that the application of an artificial intelligence algorithm similar to DNN rather than MRA is more appropriate when there is a large variance, such as water loss management data, among various artificial intelligence algorithms. Therefore, in order to derive better results, the analysis result can be improved through sensitivity analysis of the algorithm parameters and various training data classifications.

The multiple correlation coefficient of MRA analysis is the Pearson correlation coefficient for the relationship between the dependent variable according to the independent variable and the predicted value through the regression model. The average value for the coefficient of determination is 0.55, which is not a high correlation, and it would not be the best data classification for reflecting the characteristics of the data. This result means that it cannot be evaluated that the observed values of the sample are well clustered around the regression line, and that there is a difference between the predicted values and the observed values. Moreover, because the data classification was based on the leakage ratio and the revenue water ratio, it can meet the goal to improve the efficiency of leakage management in the future, but the model formulation predicts the increase in the revenue water ratio and the decrease in the leakage ratio It shows some error with the predicted value.

The prediction values of the model formulation show differences in accuracy depending on the classification of the data, even if the same data are used [17]. In this study, the leakage ratio and the revenue water ratio were classified according to the water loss management activity, which means that each variable is independent of each other in the data classification process. However, because the water loss management activities such as pipe and meter replacement reflect the time-sequential characteristics of the revenue water ratio and the leakage ratio after recovery, it is necessary to consider the spatiotemporal characteristics when classifying the data.

Therefore, in this study, when classifying the data, the increase in the revenue water ratio and the decrease in the leakage ratio affect the water loss management activity of the immediately preceding year. In this study, the time series and regional characteristics were considered, and a formulation was derived. Table 6 shows the data classified based on spatial and temporal characteristics.

As shown in Table 6, there are a total of 22 target areas classified based on spatial and temporal characteristics, and the model formulation was derived for a total of 11 target areas, except for small- and medium-sized cities with less than four datasets. The results are shown in Table 7.

The multiple correlation coefficient of the model formulation considering spatiotemporal characteristics is 0.87 on average (maximum: 0.99, minimum: 0.75), which shows a higher correlation than the data classified based on the revenue water ratio and the leakage ratio in both analysis approaches (i.e., MRA, DNN). This means that the observations of the sample are well clustered around the regression line, and the regression model predicts the dependent variable well. In particular, BH and JE showed a coefficient of determination of almost 1.00. This means that there are no missing values in the data; these two cities are smaller cities relatively, and the annual water loss management actions were more actively performed than in other areas. Accordingly, it is judged that the improvement of the revenue water ratio and the leakage ratio is larger than others. Moreover, in the case of NJ, the gap in the coefficient of determination presented the largest between MRA and DNN. This is because NJ has a distinct time-series characteristic compared to other cities. Therefore, it is thought that in the case of NJ, data analysis considering the characteristics of time series is more appropriate than independent data analysis.

Table 6. Example of data classifying based on spatial and temporal characteristics.

Study Area	Revenue Water Ration	Increase Revenue Water Ration	Leakage Ratio	Decrease Leakage Ratio	Pipe Replacement (PR)	Meter Replacement (MR)	Leakage Repair (LR)	Leakage Detection (LD)
	%	%	%	%	km/km × Million	No./Per. × 1000	No./km × 1000	No./km × 1000
GeJ 2009	66.4	7.6	27.4	9.4	18.49	5.68	2.85	1.10
GeJ 2010	67.8	1.4	26.9	0.5	19.85	7.28	1.72	0.76
GeJ 2011	72.5	4.7	22.1	4.8	4.57	10.23	0.55	0.42
GeJ 2012	75.6	3.1	19.4	2.7	32.32	12.44	0.78	0.51
GeJ 2013	74.8	−0.8	19.2	0.2	20.30	8.91	0.88	0.40
GeJ 2014	80.4	5.6	14.7	4.5	26.43	5.80	0.65	0.33
GeJ 2015	80.5	0.1	14.7	0.0	14.04	9.16	0.68	0.37
GeJ 2016	80.0	−0.5	14.8	−0.1	6.47	11.03	0.54	0.18
GeJ 2017	80.3	0.3	15.1	−0.3	1.43	6.86	0.67	0.17
GeJ 2018	75.9	−4.4	19.5	−4.4	1.73	7.52	0.71	0.21
GR 2008	60.6	9.2	35.0	8.4	8.05	32.37	1.14	0.27
GR 2009	72.1	11.5	12.4	22.6	64.17	26.88	0.83	0.21
GR 2010	72.3	0.2	21.2	−8.8	2.45	15.60	0.60	0.06
GR 2011	76.7	4.4	18.7	2.5	2.62	34.67	0.74	0.15
GR 2012	78.6	1.9	16.6	2.1	3.25	33.51	0.86	0.22
GR 2013	80.0	1.4	15.0	1.6	3.44	10.95	0.80	0.46
GR 2014	80.0	0.0	15.5	−0.5	7.70	14.62	0.68	0.39
GR 2015	80.3	0.3	15.2	0.3	2.34	49.38	0.46	0.28
GR 2016	80.7	0.4	14.8	0.4	2.31	28.57	0.54	0.19
GR 2017	78.2	−2.5	14.9	−0.1	1.05	50.59	0.37	0.14
GR 2018	75.6	−2.6	14.9	0.0	1.48	51.91	0.26	0.13
				…				
GS 2011	58.3	10.6	36.6	3.2	14.60	59.70	0.46	0.63
GS 2012	66.2	7.9	29.0	7.6	41.11	36.27	0.66	0.34
GS 2013	73.0	6.8	22.5	6.5	21.53	35.11	0.46	0.25
GS 2014	80.0	7.0	15.0	7.5	50.51	11.34	0.33	0.18
GS 2015	78.1	−1.9	17.2	−2.2	14.23	18.69	0.32	0.20
GS 2016	80.5	2.4	14.9	2.3	6.99	22.07	0.28	0.18
GS 2017	80.3	−0.2	15.0	−0.1	0.77	24.29	0.29	0.16
GS 2018	79.3	−1.0	15.9	−0.9	2.88	25.73	0.22	0.12

Through the comparison of the above data classification methods, the increase or decrease in the leakage ratio/revenue water ratio by the water loss management activity of local water distribution systems can construct a more effective model for classification, considering both local and temporal characteristics.

However, this study could not be performed due to the lack of accumulated data, but when the data are classified considering both the current state of the water distribution systems (e.g., the revenue water ratio, the leakage ratio) and spatial and temporal characteristics, it is believed that a more reliable predictive model can be developed.

Table 7. Results of deriving leakage management model formulation considering spatial and temporal characteristics.

Standard of Categorization	MRA		DNN
	Leakage Management Model Formulation	**R**	
GeJ	$Y_{GeJ} = -0.34X_2 - 4.59X_3 + 19.28X_4 + 1.68$	0.75	0.76
GR	$Y_{GR} = 0.35X_1 + 0.24X_2 + 11.05X_3 + 14.07X_4 - 18.62$	0.98	0.82
GS	$Y_{GS} = 0.35X_1 + 0.30X_2 - 2.22X_3 - 20.35X_4 - 6.37$	0.91	0.92
NJ	$Y_{NJ} = 0.12X_1 - 0.11X_2 - 7.57X_3 + 16.26X_4 + 1.53$	0.77	0.84
DDC	$Y_{DDC} = -0.06X_1 - 0.10X_2 - 72.42X_3 - 13.54X_4 + 5.50$	0.85	0.86
BH	$Y_{BH} = -0.14X_1 + 0.03X_2 - 10.64X_3 + 13.99X_4 - 15.41$	0.99	0.99
SC	$Y_{SC} = -0.40X_1 - 0.18X_2 + 30.92X_3 - 21.09X_4 - 2.51$	0.76	0.85
SS	$Y_{SS} = 0.22X_1 + 0.03X_2 - 3.59X_3 + 26.57X_4 - 1.13$	0.9	0.86
JE	$Y_{JE} = -0.19X_1 + 0.01X_2 + 22.64X_3 + 14.56X_4 - 6.19$	0.99	0.99
TY	$Y_{TY} = 0.22X_1 - 0.38X_2 + 1.47X_3 - 1.90X_4 + 5.06$	0.82	0.87
HP	$Y_{HP} = 0.50X_1 - 0.09X_2 + 1.67X_3 + 35.71X_4 - 6.39$	0.87	0.91

4. Summary and Conclusions

This study developed an evaluation method for deriving the leakage reduction factors of WDSs in the case of Korean small and medium cities. To develop the evaluation method for deriving leakage reduction factors, water loss management efficiency evaluation and formulation of water loss were derived based on improving the revenue water ratio data according to the water loss activity of WDSs in local small and medium cities. For this goal, this study performed three steps. First, data from the study area were collected, and the significant factors selected based on the information characteristic (e.g., revenue water ratio, water loss management activity, variation of the revenue water ratio according to the leakage management activity, leakage status, spatiotemporal characteristics). Secondly, the collected data underwent pre-processing for the selected factors. In the data pre-processing, independent and dependent variables were determined, and correlation analyses were performed between these independent and dependent variables. Through T-score conversion, the data were then standardized to be matched up with the units of various factors. Finally, the efficiency of leakage management actions was determined by the formulation of leakage using various data analysis approaches using MRA analysis and DNN. To apply the proposed approach, the revenue water ratio and the leakage ratio variation data according to the leakage management activity (PR, LR, LD, MR) for 22 local small and medium cities in South Korea were used.

Among the four leakage reduction activities, PR was highlighted as an essential activity for decreasing the leakage ratio and increasing the revenue water ratio in most of the data classifications, excluding Group 6. In addition, in relatively well-managed WDSs (i.e., the leakage ratio is less than 10% and the revenue water ratio is higher than 80%), the LD was presented as an effective activity to improve WDS network conditions. However, the first analysis did not consider the temporal correlation of each water loss management activity; pipe and meter replacement reflected the time-sequential characteristics of the revenue water ratio and the leakage ratio after recovery. It is necessary to consider the spatiotemporal characteristics when classifying the data. Therefore, in the second analysis, the data were classified by considering the characteristics of time and location, and the effect of water loss management activity was analyzed.

According to the second analysis results, the annual water loss management action of small cities (e.g., BH and JE) was more actively performed. It is thought that this is because a city with a small water supply population is more likely to be more efficient than other regions in distributing support as it is more likely to improve leakage according to the water loss management activity. Lastly, regarding the various analyses in this study, in most of the classification groups, depending on the revenue water ration and the leakage ration, the DNN model of the coefficient of determination was higher than MRA. Moreover, in the second analysis, the performance of data analysis using DNN is more appropriate in the data classification considering the characteristics of time series compared with independent data analysis. Therefore, it can be judged that DNN is a suitable model for the formulation of water loss management. Through the comparison of the above data classification approaches, the increase or decrease in the leakage ratio/revenue water ratio by the water loss management activity of local WDSs can be used to construct a more effective model for classification, considering both local and temporal characteristics.

The results of this study are expected to be highly useful when analyzing leakage management efficiency and determining the leakage management goals of WDSs with similar size and characteristics in the future. In particular, it can be used as a basis for determining project goals and budgets at the planning stage, and it can also be used for prioritization and evaluation of each leakage reduction activity to achieve project goals. In addition, it will be possible to derive a water leakage reduction plan with optimal efficiency by evaluating the current pipeline conditions according to the leakage ratio/revenue water ratio, selecting the pipeline maintenance target, and reducing the water leakage through pressure control. Finally, it will be possible to prioritize pipeline maintenance in order to increase leakage reduction efficiency by calculating the amount of leakage reduction for each pipeline.

Author Contributions: Conceptualization, Y.H.C. and T.C.; data curation, T.C.; methodology, D.G.Y. and Y.H.C.; supervision, S.L.; writing—original draft, Y.H.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) (NRF-2021R1G1A1003295).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are not publicly available.

Acknowledgments: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government.

Conflicts of Interest: There is no conflict of interest, and the source of funding, including the grant number for this paper, have been declared.

References

1. Criminisi, A.; Fontanazza, C.M.; Freni, G.; Loggia, G.L. Evaluation of the apparent losses caused by water meter under-registration in intermittent water supply. *Water Sci. Technol.* **2009**, *60*, 2373–2382. [CrossRef] [PubMed]
2. De Marchis, M.; Fontanazza, C.M.; Freni, G.; La Loggia, G.; Napoli, E.; Notaro, V. A model of the filling process of an intermittent distribution network. *Urban Water J.* **2010**, *7*, 321–333. [CrossRef]
3. Wang, R.; Wang, Z.; Wang, X.; Yang, H.; Sun, J. Pipe burst risk state assessment and classification based on water hammer analysis for water supply networks. *J. Water Resour. Plan. Manag.* **2014**, *140*, 04014005. [CrossRef]
4. Zyoud, S.H.; Shaheen, H.; Samhan, S.; Rabi, A.; Al-Wadi, F.; Fuchs-Hanusch, D. Utilizing analytic hierarchy process (AHP) for decision making in water loss management of intermittent water supply systems. *J. Water Sanit. Hyg. Dev.* **2016**, *6*, 534–546. [CrossRef]
5. Ndunguru, M.G.; Hoko, Z. Assessment of water loss in Harare, Zimbabwe. *J. Water Sanit. Hyg. Dev.* **2016**, *6*, 519–533. [CrossRef]
6. Kayaalp, F.; Zengin, A.; Kara, R.; Zavrak, S. Leakage detection and localization on water transportation pipelines: A multi-label classification approach. *Neural Comput. Appl.* **2017**, *28*, 2905–2914. [CrossRef]
7. Taha, A.W.; Sharma, S.; Lupoja, R.; Fadhl, A.N.; Haidera, M.; Kennedy, M. Assessment of water losses in distribution networks: Methods, applications, uncertainties, and implications in intermittent supply. *Resour. Conserv. Recycl.* **2020**, *152*, 104515.
8. Negharchi, S.M.; Shafaghat, R. Leakage estimation in water networks based on the BABE and MNF analyses: A case study in Gavankola village, Iran. *Water Supply* **2020**, *20*, 2296–2310. [CrossRef]
9. Choi, T.; Kang, K.; Koo, J. Efficiency evaluation of leakage management using data envelopment analysis. *J.-Am. Water Work. Assoc.* **2015**, *107*, E1–E11. [CrossRef]

10. Lambert, A. Accounting for losses: The bursts and background concept. *Water Environ. J.* **1994**, *8*, 205–214. [CrossRef]
11. Boztaş, F.; Özdemir, Ö.; Durmuşçelebi, F.M.; Firat, M.A.H.M.U.T. Analyzing the effect of the unreported leakages in service connections of water distribution networks on non-revenue water. *Int. J. Environ. Sci. Technol.* **2019**, *16*, 4393–4406. [CrossRef]
12. Benesty, J.; Chen, J.; Huang, Y.; Cohen, I. Pearson correlation coefficient. In *Noise Reduction in Speech Processing*; Springer: Berlin/Heidelberg, Germany, 2019; pp. 1–4.
13. Abdi, H. The Kendall rank correlation coefficient. *Encycl. Meas. Stat.* **2007**, *2*, 508–510.
14. Mark, J.; Goldberg, M.A. Multiple regression analysis and mass assessment: A review of the issues. *Apprais. J.* **1988**, *56*, 89–109.
15. Krizhevsky, I.A.; Sutskever, G.E. Hinton. ImageNet classification with deep convolutional neural networks. In Proceedings of the Advances in Neural Information Processing Systems 25 (NIPS'2012), Lake Tahoe, NV, USA, 3–6 December 2012.
16. Wang, S.C. Artificial neural network. In *Interdisciplinary Computing in Java Programming*; Springer: Boston, MA, USA, 2003; pp. 81–100.
17. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [CrossRef]