Prediction of Aquatic Ecosystem Health Indices through Machine Learning Models Using the WGAN-Based Data Augmentation Method

Seoro Lee 1, Jonggun Kim 1, Gwanjae Lee 1, Jiyeong Hong 2, Joo Hyun Bae 3 and Kyoung Jae Lim 1,*

1 Department of Regional Infrastructure Engineering, Kangwon National University, Chuncheon-si 24341, Korea; seorolee91@gmail.com (S.L.); kimjg23@gmail.com (J.K.); Gwanjae2@gmail.com (G.L.)
2 Department of Earth and Environment, Boston University, Boston, MA 02215, USA; jyhong@bu.edu
3 Korea Water Environment Research Institute, Chuncheon-si 24408, Korea; baegop@pusan.ac.kr
* Correspondence: Kjlim@kangwon.ac.kr

Abstract: Changes in hydrological characteristics and increases in various pollutant loadings due to rapid climate change and urbanization have a significant impact on the deterioration of aquatic ecosystem health (AEH). Therefore, it is important to effectively evaluate the AEH in advance and establish appropriate strategic plans. Recently, machine learning (ML) models have been widely used to solve hydrological and environmental problems in various fields. However, in general, collecting sufficient data for ML training is time-consuming and labor-intensive. Especially in classification problems, data imbalance can lead to erroneous prediction results of ML models. In this study, we proposed a method to solve the data imbalance problem through data augmentation based on Wasserstein Generative Adversarial Network (WGAN) and to efficiently predict the grades (from A to E grades) of AEH indices (i.e., Benthic Macroinvertebrate Index (BMI), Trophic Diatom Index (TDI), Fish Assessment Index (FAI)) through the ML models. Raw datasets for the AEH indices composed of various physicochemical factors (i.e., WT, DO, BOD\textsubscript{5}, SS, TN, TP, and Flow) and AEH grades were built and augmented through the WGAN. The performance of each ML model was evaluated through a 10-fold cross-validation (CV), and the performances of the ML models trained on the raw and WGAN-based training sets were compared and analyzed through AEH grade prediction on the test sets. The results showed that the ML models trained on the WGAN-based training set had an average F1-score for grades of each AEH index of 0.9 or greater for the test set, which was superior to the models trained on the raw training set (fewer data compared to other datasets) only. Through the above results, it was confirmed that by using the dataset augmented through WGAN, the ML model can yield better AEH grade predictive performance compared to the model trained on limited datasets; this approach reduces the effort needed for actual data collection from rivers which requires enormous time and cost. In the future, the results of this study can be used as basic data to construct big data of aquatic ecosystems, needed to efficiently evaluate and predict AEH in rivers based on the ML models.

Keywords: aquatic ecosystem health; machine learning models; WGAN; data augmentation

1. Introduction

The recent rapid climate change and the increase in human activity due to urbanization and industrialization have caused sudden variability of river water quantity and quality [1–3]. In particular, the decrease in water quantity and deterioration of water quality in rivers have caused a significant impact on the reduction in biodiversity and degradation of aquatic ecosystem health (AEH) [4,5]. Maintaining and restoring the AEH has a positive effect on ecosystem services provided to humans and creation of a sustainable habitat environment for aquatic organisms [6,7]. Therefore, it is important to establish a systematic
aquatic ecosystem management plan through the efficient evaluation of the ecological health status of river ecosystems in advance.

The aquatic ecosystem health (AEH) of rivers can be evaluated through various biological indices developed to evaluate current characteristics of aquatic organisms, such as benthic macroinvertebrates, diatoms, and fish [8–10]. Since 2008, the Ministry of Environment of Korea has periodically conducted the aquatic ecosystem monitoring survey for most rivers nationwide. In addition, the AEH is divided into five grades (A to E) and evaluated through biological indices such as Benthic Macroinvertebrate Index (BMI), Trophic Diatom Index (TDI), and Fish Assessment Index (FAI) [11]. The results of the AEH evaluation based on these indices can be efficiently used as basic data for selecting impaired rivers that require aquatic ecosystem management and establishing plans for river ecosystem restoration. The AEH is continuously affected by the complex interaction of natural and anthropogenic environmental factors such as flow, water quality, climate, topography, and land development. In particular, physicochemical factors such as the flow and water quality of rivers and water temperature have much influence on the AEH [12]. However, there are limitations to interpret accurately the cause of AEH deterioration and the effect on AEH change using individual factors only. Therefore, there is a need for a method that can efficiently predict and evaluate the AEH by considering the complex interactive impacts of factors such as flow, water quality, and water temperature, etc. The data-driven Machine Learning (ML) models make effective predictions by mining the relevant information between input and output variables inherent when using a larger dataset without the physical process required by conventional numerical models [13]. In many studies, various ML models have been successfully used for the estimation of runoff [13], rainfall erosivity (R-factor) [14], dam discharge [15], sediment [16], water quality [17], and AEH [18,19], etc. However, the predictive performance of an ML model depends greatly on the quantity and quality of datasets, so it is necessary to collect sufficient amounts of data samples to build robust ML models. In S. Korea, AEH monitoring sites and time frames are irregular due to climatic characteristics, with large seasonal rainfall fluctuations [18]. Thus, continuous monitoring for BMI, TDI, and FAI could be time-consuming and labor-intensive tasks. Additionally, most of the AEH grade data samples collected are unbalanced because there are relatively more streams in a good state than streams in a poor state. Due to data imbalance, the ML model could be trained biasedly towards majority classes, which can result in incorrect AEH predictions. This would be a frequent problem using the data collected from various sources in data classification areas. To deal with this problem, the generative adversarial network (GAN) [20] algorithms that generate images and voices that are similar to actual data are being widely used in various research fields as a data augmentation method [21,22]. Despite its efficient applicability to a variety of problems, original GAN is difficult to learn while maintaining a balance between the discriminator and generator. Additionally, mode collapse and gradient loss problems can occur when using the GAN for data augmentation [23]. The recently proposed Wasserstein GAN (WGAN) [24] shows better performance than the original GAN, with progress towards stable training [25]. The applicability of the WGAN for obtaining image data samples and its effectiveness has been demonstrated in several studies. Jiang et al. [26] used WGAN to obtain high-quality CT dental images that are important for oral disease and treatment. Xia and Liu [27] developed and evaluated a deblurring model based on WGAN to augment the remote sensing image dataset. However, very few studies have been conducted to evaluate the applicability of the WGAN for augmenting standardized data samples in hydrology and aquatic ecology fields and its effect on the performance of ML model training and test.

Therefore, the main objectives of this study are (i) to evaluate the applicability of WGAN to solve the imbalanced distribution of data samples required for predicting the grades of the AEH indices in classification field; (ii) to propose a method for building the ML model to effectively evaluate the grades of each AEH index by considering multiple physicochemical factors such as flow, water quality, and water temperature.
In the future, the AEH grade results predicted from the ML model reflecting various physicochemical factors can be utilized in decision making for the selection of impaired rivers and the establishment of restoration strategies. To our knowledge, this is the first study to predict the grades of each AEH index through ML models and WGAN-based data augmentation, which is the main novelty of this study. The overall workflow of this study is shown in Figure 1.

**Figure 1.** Workflow scheme of this study.

### 2. Materials and Methods

#### 2.1. Description of the Study Area

The Han River basin (35,770 km²) is located in the center of the Korean peninsula within the latitudes of 126°35′ to 128°46′ and the longitudes of 36°30′ to 38°56′ (Figure 2). The Han River has a main reach length of about 481.8 km from the source of the river to the estuary, consisting of about 920 rivers flowing into the main stream. In the Han River basin, there are 232 water quality monitoring stations collecting flow and water quality data and 746 biomonitoring stations collecting aquatic ecosystem data. The Han River consists largely of the Bukhan River and the Namhan River, and these two main rivers flow into Paldang-dam, which supplies drinking and living water for the residents of the metropolitan Seoul area. In particular, numerous tributaries flowing into these two main rivers are important in terms of river water quality and aquatic ecosystem management as habitats for various species. Recently, in the Han River, due to urbanization and industrialization, water pollution problems have been occurring frequently due to the influence of non-point sources as well as point source discharge facilities such as water waste treatment plants (WWTPs) [28]. In addition, the recent decrease in river flow due to frequent drought in spring is causing deterioration of river water quality [29]. These natural and anthropogenic influences can have a significant impact on aquatic ecosystems. Therefore, it is necessary to establish appropriate sustainable management plans for the protection of aquatic ecosystems through the efficient prediction of the AEH in the Han River in advance.
2.2. Data Collection

In Korea, AEH monitoring has been performed in most rivers nationwide. The grades of the AEH indices (BMI, TDI, and FAI) are calculated according to the health score evaluation criteria as shown in Table 1. In this study, raw datasets for AEH indices, required for training the ML models, were prepared using observed data collected from 76, 73, and 67 water quality monitoring and biomonitoring stations from 2008 to 2018, respectively. Each raw dataset was composed of physicochemical factors (i.e., flow, dissolved oxygen (DO), biochemical oxygen demand in five days (BOD\textsubscript{5}), total nitrogen (TN), total phosphorus (TP), suspended solids (SS), and water temperature (WT)) and AEH grade data (A to E), which are target data in ML training process. The physicochemical factors were selected as factors that can have a relatively significant effect on the health of aquatic ecosystems by referring to previous research results [30,31]. However, in Korea, the AEH monitoring data, collected twice a year during Spring and Autumn, are insufficient compared with water quality data in terms of the number of datasets and timing of data collection. Thus, the raw datasets were built by classifying physicochemical factors’ data from the same monitoring date based on the AEH monitoring date. The general statistical data such as mean, minimum,
and maximum values for each input datum of raw datasets are presented in Table 2. Table 3 shows the number of data by grade in the raw datasets for each AEH index. In particular, as shown in Figure 3, the number of samples of grades D or E is relatively small compared to the number of samples of grades A, B, and C of each AEH index. In such cases, the ML models tend to be overwhelmed by large classes and ignore small classes [32]. Therefore, to improve performance of the ML models for grade prediction of each AEH index, it is necessary to alleviate the imbalance of the dataset by augmenting the input data for each grade.

Table 1. Evaluation criteria for grades of each AEH index [11].

| Indices | A (Very Good) | B (Good) | C (Fair) | D (Poor) | E (Very Poor) |
|---------|---------------|----------|----------|----------|--------------|
| BMI     | ≥80           | ≥65      | ≥50      | ≥35      | <35          |
| TDI     | ≥90           | ≥70      | ≥50      | ≥30      | <30          |
| FAI     | ≥80           | ≥60      | ≥40      | ≥20      | <20          |

Table 2. The observed input and target data for ML models.

| Target | Description | Flow (m<sup>3</sup>/sec) | BOD<sub>5</sub> (mg/L) | DO (mg/L) | SS (mg/L) | TN (mg/L) | TP (mg/L) | WT (°C) |
|--------|-------------|--------------------------|------------------------|-----------|-----------|-----------|-----------|---------|
| BMI    | Mean        | 41.22                    | 1.92                   | 10.10     | 6.76      | 3.088     | 0.066     | 20.33   |
|        | Min, Max    | 0.006, 544.4             | 0.2, 25.8              | 5.1, 17.74| 0.2, 67.0 | 0.534, 16.0| 0.003, 1.0| 5.2, 30.2|
| TDI    | Mean        | 37.12                    | 1.88                   | 9.87      | 7.00      | 3.297     | 0.079     | 20.17   |
|        | Min, Max    | 0.024, 424.8             | 0.3, 25.8              | 6.34, 14.6| 0.1, 140.5| 0.467, 18.5| 0.004, 2.2| 6.6, 28.7|
| FAI    | Mean        | 39.10                    | 2.18                   | 10.08     | 8.18      | 3.369     | 0.078     | 20.17   |
|        | Min, Max    | 0.012, 342.7             | 0.1, 23.8              | 5.1, 18.7 | 0.1, 126.3| 0.467, 12.9| 0.004, 1.0| 6.8, 29.0|

Table 3. Number of data by grade of raw datasets for each AEH index.

| Indices | A | B | C | D | E | Total |
|---------|---|---|---|---|---|-------|
| BMI     | 171 | 66 | 45 | 35 | 22 | 339   |
| TDI     | 65 | 93 | 68 | 43 | 50 | 319   |
| FAI     | 70 | 63 | 82 | 49 | 18 | 282   |

Figure 3. Distribution of data by grade in the raw dataset for each AEH index.
2.3. Wasserstein Generative Adversarial Network (WGAN)

The Generative Adversarial Network (GAN) is an algorithm that creates data distribution or variance itself through competitive learning of two neural networks, the generator G and the discriminator D [20]. The G is trained to generate data that are similar to the real data, whereas the D is trained to lower the probability of erroneously discriminating the data generated by G. However, the original GAN is sometimes difficult to train due to mode collapse and gradient loss problems [23]. Therefore, we used WGAN as data augmentation method to increase the number of training samples. The WGAN uses the Wasserstein Distance (WD) in Equation (1) as the loss function instead of the Jensen–Shannon Divergence loss function used in the original GAN [24].

\[
W(\mathbb{P}_\gamma, \mathbb{P}_g) = \inf_{\gamma \in \Pi(\mathbb{P}_\gamma, \mathbb{P}_g)} \mathbb{E}_{(x, y) \sim \gamma}[\|x - y\|] 
\] (1)

The WD is the smallest estimate of the expected value of the x and y distances among all the joint probability distributions. This loss function has the advantage in that the slope does not disappear near the optimum point of the parameter and the learning can proceed stably [33]. In WGAN, a newly defined critic is used instead of the discriminator, and when the G is trained once, the critic trains n critic times. The critic updates parameter w based on the RMSProp optimizer and weight clipping for parameter w is performed to satisfy the k-Lipschitz continuous function condition. The TensorFlow (version: 2.3.0) and Keras (version: 2.4.3) libraries were used as a deep learning framework for WGAN. As a parameter of WGAN, the learning rate (\(\alpha\)) based on the RMSProp optimizer was 0.00005, the weight clipping parameter (c) was 0.01, and the \(n_{\text{critic}}\) was 5. The raw datasets of the AEH indices were used as the input data of WGAN, and the epoch of WGAN was set to 15,000 each, and training was performed. Figure 4 shows the model network for G and D (a critic in the paper) constructed for WGAN training. The model network for G was composed of dense layers and activation function LeakyReLU layers, and the model network for D was constructed by adding dropout layers in the middle to prevent overfitting [34].

2.4. Building and Evaluation of ML Models

2.4.1. ML Models Building and Evaluation Process

In this study, we built six ML models (Support Vector Machine [35], Decision Tree [36], K-Nearest Neighbors [37], Random Forest [38], Gradient Boosting [39], and eXtreme Gradient Boost [40]) to predict the grades of each AEH index. Scikit-learn module (version: 0.24.1) and XGboost library (Version: 1.4.2) in the python 3.7 environments were used to build each ML model. A detailed description of ML models can be found in the work of Bae et al. [41]. The function information of the ML models used in the study is presented in Table 4. First, to evaluate the AEH grade prediction performance of the ML model and the applicability of the WGAN, two datasets were constructed. One of the two datasets is a raw dataset with an unbalanced distribution, and the other is a WGAN-based dataset in which a synthetic dataset is augmented through WGAN and raw datasets combined. Before the model training, each dataset was split into 80% for training set and 20% for test set to evaluate the performance of the models. After that, preprocessing was performed on the training set. In general, appropriate preprocessing such as data normalization, abnormal data processing, and data format conversion is required to improve the performance of the ML model [42]. The StandardScaler function was used for a preprocessing step that makes the distribution of the entire input data average 0 and variance 1. Additionally, the dummy function of the Pandas module (version: 1.1.4) was performed to encode the grades of each AEH index as an integer. Second, to evaluate the reliability and stability of the AEH grade prediction of six ML models, the ML models were trained and validated using the raw training set with the K-fold cross-validation (CV) technique. This technique has the advantage of being able to classify training and validation sets while maintaining the proportions of classes, each with a different proportion, and can prevent over-fitting and under-fitting.
on a specific dataset and improve the prediction accuracy of the model [43,44]. For the single parameter, K, representing the number of groups into which a given dataset will be split, we used a value of 10, which is usually recommended to check the generalizability of the model [45]. The training performance of the ML models was validated based on the F1-score performance metric. Additionally, based on the results, three ML models were selected with the best AEH grade prediction performance. Finally, three selected ML models for applicability evaluation of WGAN were trained and validated on the raw and WGAN-based training sets and evaluated on each test set.

![Model networks for WGAN training: (a) generator network; (b) discriminator network.](image)

**Figure 4.** Model networks for WGAN training: (a) generator network; (b) discriminator network.

**Table 4.** The description of ML models used in this study.

| ML Models               | Module            | Function                  | Notation |
|-------------------------|-------------------|---------------------------|----------|
| Support Vector Machine  | Sklearn.svm       | SVC                       | SVM      |
| Decision Tree           | Sklearn.tree      | DecisionTreeClassifier    | DT       |
| K-Nearest Neighbors     | Sklearn.neighbors | KNeighborsClassifier      | KNN      |
| Random Forest           | Sklearn.ensemble  | RandomForestClassifier    | RF       |
| Gradient Boosting       | Sklearn.ensemble  | GradientBoostingClassifier| GB       |
| eXtreme Gradient Boost  | xgboost.xgb       | XGBClassifier             | XGB      |
2.4.2. The Performance Evaluation Metrics of ML Models

Confusion matrix [46], which contains information about the actual and predicted classes obtained by a classification model, is useful for visually understanding the performance of the model. The basic form of the confusion matrix is shown in Figure 5. The confusion matrix is useful for evaluating the performance of ML models to solve various classification problems [47,48]. In general, four evaluation metrics (Accuracy, Precision, Recall, and F1-score) are widely used to evaluate model classification performance based on the confusion matrix. These metrics can be obtained by Equations (2)–(5). Accuracy is a measure of the overall predictive performance of a model as the ratio of the number of correctly classified data to the total number of data. Precision is the ratio of correctly classified data out of the total data predicted by the classification model as positive. Recall is a measure of the completeness of a model and is the ratio of correctly classified data to the total number of positive actual data. F1-score is the harmonic mean which considers both the precision and recall [49]. Precision and Recall have an inherent trade-off relationship in that when one metric increases, the other decreases. Therefore, the F1-score is a special measure that can evaluate the performance of a model through the trade-off between precision and recall. Additionally, the F1-score gives a better view of ML model performance, especially for datasets with an imbalanced class distribution, because F1-score is not biased towards majority classes [50]. This study evaluated the classification performance of ML models using the F1-score based on the confusion matrix.

\[
\text{Accuracy} = \frac{A + D}{A + B + C + D} \quad (2)
\]
\[
\text{Precision} = \frac{A}{A + C} \quad (3)
\]
\[
\text{Recall} = \frac{A}{A + B} \quad (4)
\]
\[
\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)
\]

Figure 5. Diagram of the confusion matrix.

3. Results and Discussion

3.1. Correlation Analysis Results

It is an important process to analyze and understand each input datum before training the ML model [41]. Correlation analysis was carried out on input data (i.e., WT, DO, BOD\(_5\), SS, TN, TP, and Flow) of the raw dataset for AEH indices, which were analyzed using a heat map, as shown in Figure 6. The results showed that BOD\(_5\), TN, TP, and SS concentrations had a strong negative correlation with scores for each AEH index in common compared to other factors. In the results of Woo et al. [51], water quality concentration and scores of AEH indices showed a similar negative correlation. Kim et al. [52] also showed that TDI had a strong negative correlation with BOD\(_5\), TN, and TP concentrations. This indicates that water quality concentration can be used as an important indicator to identify changes in AEH. Additionally, SS concentration had the highest correlation with BMI, followed
by the TDI and FAI score. According to Griffiths and Walton [53], upper tolerance levels for SS concentration are between 80 and 100 mg/L for fish and as low as 10–15 mg/L for bottom invertebrates. SS concentration is a factor that interferes with the feeding or spawning of most benthic organisms and has a greater effect on benthic organisms than any other organisms in the river [54]. On the other hand, the correlation between FAI score and SS concentration was found to be rather small at negative 0.22, which is thought to be because the relatively low-concentration data compared to the high-concentration data were reflected in the correlation analysis results. This indicates that ML models need to be trained on sufficient high-concentration SS data to accurately capture the impact of SS on fish population changes and habitat environment. In addition, since fish and other organisms respond greatly to the duration exposure of sediment as well as the concentration of sediment [55], correlation analysis including variables such as exposure period to SS concentration will be required for accurate sensitivity analysis of physicochemical factors according to AEH changes.

Figure 6. Heat map analysis results for (a) BMI, (b) TDI, (c) FAI.
3.2. Correlation Analysis and WGAN-Based Data Augmentation Results

In this study, the datasets for ML model training were augmented using the raw datasets for the AEH indices through WGAN. In the WGAN training process, we found that the loss function of the discriminator showed large fluctuations in the initial stage (Figure 7). This implies that the generator and discriminator at the initial stage do not recognize anything about basic characteristics of the data, so there is a limit of the discriminator to sufficiently train relevant information related to the real data [56]. However, as training progresses, the loss function stably converges to an optimal value close to zero. In addition, as the loss function was optimized, the synthetic data generated by the generator showed a result similar to the distribution of the real data (Figure 8). This indicates that the generator generates better quality synthetic data because the discriminator’s ability to capture the details of the real data improved as the training iteration. After training, 25,000 synthetic data for the AEH indices were randomly generated through WGAN and raw datasets were combined to construct WGAN-based datasets. Table 5 shows the amount of data by grade in the synthetic dataset and the WGAN-based dataset for the AEH indices. As can be seen in Table 5, the amount of data by grade for each AEH index augmented through WGAN is not the same, but it can be seen that the overall data balance by grade is improved compared to the raw dataset. When the input data of the ML model for solving the classification problem is unbalanced by class, the model may be overwhelmed by the large classes and ignore the small classes [57]. Therefore, the method of increasing the amount of data by grade for the AEH indices through WGAN is a meaningful attempt to improve the classification ability of the ML model. WGAN has the advantage of being able to generate various new data similar to real data through stable learning compared to the original GAN, but it is difficult to always accurately match the distribution of synthetic data generated under the influence of weighted clipping with the real distribution of data [24,58]. Thus, to improve the convergence speed and accuracy of WGAN in the future, it will be necessary to study the WGAN algorithm improvement and parameter calibration.

3.3. Comparison of Validation Results of ML Models

Figure 9 and Table 6 show 10-fold CV results of six ML models using the raw training set in terms of F1-score. As shown in Figure 9, the results showed no significant difference in F1-score for BMI and FAI for each ML model, but among the six models, XGB was the highest with an F1-score of 0.475 for BMI, and F1-score of 0.310 for FAI using the XGB. On the other hand, in the case of TDI, RF and SVM had a significantly higher predictive performance with F1-scores of 0.388 and 0.370, respectively, than other models. The average F1-score for the AEH indices of the six models ranged from 0.349 to 0.376 (Table 6). Especially, the RF had the highest F1-score of 0.376, followed by XGB 0.367 and SVM 0.359. The RF and XGB are decision tree-based ensemble models and have excellent classification capacity of imbalanced datasets [59,60]. SVM is a powerful state-of-the-art algorithm with a strong theoretical foundation, which aims to find the decision boundary or hyperplane, and is widely used to solve high-dimensional nonlinear classification problems using various kernel functions [61]. Additionally, to analyze the effect of improving the grade prediction performance on the AEH indices according to the use of the WGAN-based training set, three ML models (RF, XGB, and SVM) were trained and validated using the WGAN-based training set. Overall, better results were obtained from ML models trained on WGAN-based training set while the worst results were obtained from ML models trained on the raw training set (Figure 10). We also found that the ML model trained on the WGAN-based training set outperformed the model trained on the raw training set with an F1-score of 0.953 for RF, 0.959 for XGB, and 0.953 for SVM (Table 7). This indicates that data augmentation through WGAN alleviated the imbalanced distribution of the raw training set and successfully improved the grade prediction performance of each AEH index of the ML models. In particular, through the above results, the ML model can confirm the possibility of predicting a grade that lacks data relatively better than the ML model trained using limited datasets. Other studies in various fields have also demonstrated that data
augmentation through WGAN improves the performance of training classifiers \[62,63\]. This implies that the data augmentation method through WGAN can be reasonably utilized to enlarge datasets in the fields of hydrology and aquatic ecosystems.

Figure 7. The loss functions for (a) BMI, (b) TDI, (c) FAI according to WGAN training.
Figure 8. Cont.
other studies in various fields have also demonstrated that data augmentation through WGAN can significantly improve the grade prediction performance of each AEH index compared to the ML models trained on the raw training set.

The RF and XGB models were trained as decision tree-based ensemble models, with XGB having the highest F1-score of 0.959, followed by RF and SVM, which had an F1-score of 0.953.

Table 5 shows the number of WGAN-based datasets by grade for each AEH index. The total number of WGAN-based datasets is 25,000.

Table 6 shows the comparison of F1-scores for AEH indices through 10-fold CV by ML model using raw training set. The F1-scores ranged from 0.279 to 0.475.

Figure 8. The synthetic data distribution change for (a) BMI, (b) TDI, and (c) FAI according to loss function optimization.

Figure 9. Comparison of F1-scores for AEH indices through 10-fold CV by ML model using raw training set.
Table 6. Average F1-scores for AEH indices through 10-fold CV using raw training set.

| Metric | Grade | SVM | RF | DT | KNN | GB | XGB |
|--------|-------|-----|----|----|-----|----|-----|
| F1-score | BMI   | 0.412 | 0.430 | 0.448 | 0.449 | 0.448 | 0.475 |
|         | TDI   | 0.370 | 0.388 | 0.313 | 0.279 | 0.335 | 0.308 |
|         | FAI   | 0.296 | 0.310 | 0.306 | 0.318 | 0.290 | 0.317 |
|         | Avg   | 0.359 | 0.376 | 0.356 | 0.349 | 0.358 | 0.367 |

Figure 10. F1-score comparison of (a) RF, (b) XGB, (c) SVM trained on two training sets.

Table 7. Average F1-score by AEH indices of ML models trained using raw and WGAN-based training sets.

| Metric | Grade | RF       | XGB       | SVM       |
|--------|-------|----------|-----------|-----------|
|        | Raw   | WGAN-Based | Raw   | WGAN-Based | Raw   | WGAN-Based |
| F1-score | BMI   | 0.430 | 0.973 | 0.475 | 0.976 | 0.412 | 0.964 |
|         | TDI   | 0.388 | 0.943 | 0.308 | 0.946 | 0.370 | 0.913 |
|         | FAI   | 0.310 | 0.944 | 0.317 | 0.955 | 0.296 | 0.982 |
|         | Avg   | 0.376 | 0.953 | 0.367 | 0.959 | 0.359 | 0.953 |
3.4. Grade Prediction of Each AEH Index for Test Set Using the ML Models

The confusion matrices for the grades of each AEH index for test sets predicted by three ML models trained on the raw and WGAN-based training sets are shown in Figures 11 and 12, respectively, and the performance results through the three evaluation metrics (Precision, Recall, and F1-score) are summarized in Tables 8 and 9. The average F1-scores of the three models trained using the raw training set for the BMI, TDI, and FAI grades were 0.53, 0.35, 0.56 for the RF, 0.44, 0.32, 0.58 for the XGB, and 0.49, 0.24, and 0.61 for the SVM (Table 8). The average F1-scores for the FAI grade predicted by each ML model was greater than those compared with the results of the other indices. This indicates that in the process of training and validating each ML model, the characteristics of the data by grade of the raw training set for FAI are well-reflect. On the other hand, the F1-score of BMI grade A showed a high performance, with 0.80 for the RF, 0.73 for the XGB, and 0.79 for the SVM but the prediction performance for the other grades (B to E) was mostly lower than 0.29. The raw training sets for BMI have an unbalanced distribution characteristic with a proportion of grade A of at least 50%. Therefore, it appears that the ML models trained using the raw training set for BMI overfit the grade A and do not correctly classify all grades in the test set. In the study of Woo et al. [18], the results also showed that the predictive performance of the RF for a small number of grades (B to E) was relatively lower than that of majority grade A due to the effect of data imbalance. Additionally, although the samples by grade of the raw training set for TDI were relatively balanced, each ML model did not correctly classify the grade of TDI as a whole (Figure 11). This implies that each ML model does not capture the characteristics of the samples in the test set well, and it is necessary to train and validate the model with sufficient data for each grade. As can be seen in Table 9, the predictive performance of the ML models trained on the WGAN-based training set was superior compared to the ML model trained on the raw training set. The average F1-scores of the ML models trained on the WGAN-based training set for the BMI, TDI, and FAI grades were 0.92, 0.77, 0.93 for the RF, 0.92, 0.75, 0.93 for the XGB, and 0.93, 0.84, and 0.94 for the SVM. In particular, the results showed that the ML models trained on the WGAN-based training set were not biased toward majority grades and predicted all grades well overall (Figure 12). This is because the synthetic data by grade of the AEH indices augmented through WGAN alleviated the unbalanced distribution characteristics of the raw training set [64]. Through the above results, it was confirmed that the data augmentation method through WGAN can overcome the limitations of limited datasets and improve the overall model performance for grade prediction of each AEH index. Therefore, securing data for each AEH grade through the WGAN-based data augmentation is an important process for building a robust ML model that can efficiently predict all AEH grades.

### Table 8. The classification report of test results using the ML models trained on the raw training set.

| Model | Grade | BMI | TDI | FAI |
|-------|-------|-----|-----|-----|
|       | P     | R   | F1  | S   | P     | R   | F1  | S   | P     | R   | F1  | S   |
|       | 0.75  | 0.86 | 0.80 | 35  | 0.31  | 0.31 | 0.31 | 13  | 0.62  | 1   | 0.77 | 15 |
|       | 0.14  | 0.15 | 0.15 | 13  | 0.31  | 0.36 | 0.36 | 18  | 0.50  | 0.33 | 0.40 | 12 |
|       | 0.40  | 0.22 | 0.29 | 9   | 0.42  | 0.38 | 0.38 | 14  | 0.67  | 0.38 | 0.48 | 16 |
|       | 1     | 0.14 | 0.25 | 7   | 0.25  | 0.11 | 0.15 | 9   | 0.44  | 0.44 | 0.44 | 9  |
|       | 0.38  | 0.75 | 0.50 | 4   | 0.56  | 0.50 | 0.53 | 10  | 0.67  | 1   | 0.80 | 4  |
| Avg   | 0.59  | 0.56 | 0.53 | 68  | 0.36  | 0.36 | 0.35 | 64  | 0.58  | 0.59 | 0.56 | 56 |
|       | 0.77  | 0.69 | 0.73 | 35  | 0.32  | 0.33 | 0.33 | 13  | 0.68  | 0.87 | 0.76 | 15 |
|       | 0.10  | 0.08 | 0.09 | 9   | 0.30  | 0.33 | 0.32 | 18  | 0.60  | 0.25 | 0.35 | 12 |
|       | 0.12  | 0.22 | 0.16 | 9   | 0.36  | 0.29 | 0.32 | 14  | 0.58  | 0.69 | 0.63 | 16 |
|       | 0.20  | 0.14 | 0.17 | 7   | 0.25  | 0.11 | 0.15 | 9   | 0.50  | 0.33 | 0.40 | 9  |
|       | 0.17  | 0.25 | 0.20 | 4   | 0.42  | 0.50 | 0.45 | 10  | 0.57  | 1   | 0.73 | 4  |
| Avg   | 0.46  | 0.43 | 0.44 | 68  | 0.32  | 0.33 | 0.32 | 64  | 0.60  | 0.61 | 0.58 | 56 |
|       | 0.65  | 0.15 | 0.20 | 13  | 0.26  | 0.26 | 0.27 | 13  | 0.68  | 1   | 0.81 | 15 |
|       | 0.29  | 0.11 | 0.18 | 9   | 0.09  | 0.07 | 0.08 | 14  | 0.69  | 0.69 | 0.69 | 16 |
|       | 0.50  | 0.14 | 0.18 | 7   | 0.11  | 0.20 | 0.20 | 9   | 0.50  | 0.56 | 0.53 | 9  |
|       | 0.25  | 0.14 | 0.18 | 7   | 0.11  | 0.20 | 0.20 | 9   | 0.50  | 0.56 | 0.53 | 9  |
| Avg   | 0.48  | 0.57 | 0.49 | 68  | 0.34  | 0.25 | 0.24 | 64  | 0.62  | 0.64 | 0.61 | 56 |

**P**: Precision, **R**: Recall, **F1**: F1-score, and **S**: Support (the number of data in each grade class).
Figure 11. Confusion matrices for AEH indices predicted by (a) RF, (b) XGB, and (c) SVM models trained on the raw training set.
The classification report of test results using the ML models trained on the WGAN-based training set.

4. Conclusions

This study evaluated the applicability of WGAN to augment ML training datasets in the fields of hydrology and aquatic ecosystems and proposed a method to predict the grades of each AEH index using ML models. The main results of this study are as follows. First, among various physicochemical factors, water quality factors such as BOD, TN, TP, and SS concentrations were found to have a relatively significant effect on the AEH indices compared to other factors in common. This implies that water quality factors can be important indicators in predicting AEH changes in rivers through ML models. Second, as the training of the WGAN progressed stably, the synthetic data for the AEH indices similar to the distribution of the real data were generated. Additionally, the 10-fold CV performance of the ML models (RF, XGB, and SVM) reflecting these synthetic data showed an improved result with an average F1-score of 0.9 or more for each AEH index. Finally, as a result of predicting the grade of the test set by AEH indices through the above three ML models, the models trained using the raw training set did not properly classify a small number of grades due to data imbalance, whereas the models trained using the WGAN-based training set classified all AEH indices well overall without being biased towards a large number of grades. Through the results of this study, it was confirmed that the AEH grade classification prediction performance of the ML model can vary greatly depending on the data distribution for each grade; additionally, the synthetic datasets augmented through WGAN can contribute to improving the model performance and reducing the effort needed for real data collection. However, the ML models built in
this study did not reflect the data characteristics of periods such as flood and drought; it is considered that there is a limit to accurately predict changes in AEH due to extreme events. In future research, it will be necessary to continuously train and validate the ML model by acquiring observed data from various rainfall events and supplementing the synthetic datasets of WGAN to accurately predict changes in AEH through the ML model. The ML approach that is proposed in this study could contribute to not only evaluating the achievement of the AEH improvement goal through the introduction of structural and non-structural watershed management plans, but also selecting the impaired rivers that need water quality and aquatic ecology management. In addition, the data augmentation method using the WGAN can be used for building the large amount of data necessary for an integrated water environment management system that organically links the quantity-water quality-aquatic ecosystem.

Author Contributions: Conceptualization, S.L.; formal analysis, G.L. and J.H.B.; data curation, J.H.; writing—original draft preparation, S.L.; writing—review and editing, K.J.L. and J.K.; visualization, S.L. and J.H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by 2018 Research Grant (PoINT) from Kangwon National University and Korea Environment Industry & Technology Institute (KEITI) through Aquatic Ecosystem Conservation Research Program, funded by Korea Ministry of Environment (MOE), grant number 2020003030004.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References
1. Peters, N.E.; Meybeck, M.; Chapman, D.V. Effects of Human Activities on Water Quality. *Encycl. Hydrol. Sci.* 2005. [CrossRef]
2. Delpla, I.; Jung, A.V.; Baures, E.; Clement, M.; Thomas, O. Impacts of climate change on surface water quality in relation to drinking water production. *Environ. Int.* 2009, 35, 1225–1233. [CrossRef] [PubMed]
3. Qiu, J.; Shen, Z.; Leng, G.; Xie, H.; Hou, X.; Wei, G. Impacts of climate change on watershed systems and potential adaptation through BMPs in a drinking water source area. *J. Hydrol.* 2019, 573, 123–135. [CrossRef]
4. Liao, H.; Sarver, E.; Krometis, L.A.H. Interactive effects of water quality, physical habitat, and watershed anthropogenic activities on stream ecosystem health. *Water Res.* 2018, 130, 69–78. [CrossRef] [PubMed]
5. Reid, A.J.; Carlson, A.K.; Creed, I.F.; Eliason, E.J.; Gell, P.A.; Johnson, P.T.J.; Kidd, K.A.; MacCormack, T.J.; Olden, J.D.; Ormerod, S.J.; et al. Emerging threats and persistent conservation challenges for freshwater biodiversity. *Biol. Rev.* 2019, 94, 849–873. [CrossRef] [PubMed]
6. Baron, J.S.; Poff, N.L.; Angermeier, P.L.; Dahm, C.N.; Gleick, P.H.; Hairston, N.G.; Jackson, R.B.; Johnston, C.A.; Richter, B.D.; Steinman, A.D. Meeting Ecological and Societal Needs for Freshwater. *Ecol. Appl.* 2002, 12, 1247. [CrossRef]
7. Zhao, C.; Shao, N.; Yang, S.; Ren, H.; Ge, Y.; Zhang, Z.; Zhao, Y.; Yin, X. Integrated assessment of ecosystem health using multiple indicator species. *Ecol. Eng.* 2019, 130, 157–168. [CrossRef]
8. Karr, J.R. Assessment of Biotic Integrity Using Fish Communities. *Fishes 1981*, 6, 21–27. [CrossRef]
9. Ohio EPA. *Biological Criteria for the Protection of Aquatic Life: Standardized Biological Field Sampling and Laboratory Methods for Assessing Fish and Macroinvertebrate Communities*; Tech. Rept. EAS/2015-06-01; revised 26 June 2015; Ohio Environmental Protection Agency, Division of Water Quality Monitoring and Assessment: Columbus, OH, USA, 1987; Volume III, p. 120.
10. U.S. EPA. *Biological Assessments and Criteria: Crucial Components of Water Quality Programs*; EPA 822-F-02-006; U.S. Environmental Protection Agency Office of Water: Washington, DC, USA, 2002.
11. National Institute of Environmental Research. *Biomonitoring Survey and Assessment Manual*; National Institute of Environmental Research: Incheon, Korea, 2016; p. 372.
12. Chen, H.; Ma, L.; Guo, W.; Yang, Y.; Guo, T.; Feng, C. Linking Water Quality and Quantity in Environmental Flow Assessment in Deteriorated Ecosystems: A Food Web View. *PLOS ONE* 2013, 8, e70537. [CrossRef]
13. Liu, Y.; Zhang, T.; Kang, A.; Li, J.; Lei, X. Research on Runoff Simulations Using Deep-Learning Methods. *Sustainability* 2021, 13, 1336. [CrossRef]
14. Lee, J.; Lee, S.; Hong, J.; Lee, D.; Bae, J.H.; Yang, J.E.; Kim, J.; Lim, K.J. Evaluation of Rainfall Erosivity Factor Estimation Using Machine and Deep Learning Models. *Water* 2021, 13, 382. [CrossRef]
15. Hong, J.; Lee, S.; Bae, J.H.; Lee, J.; Park, W.J.; Lee, D.; Kim, J.; Lim, K.J. Development and evaluation of the combined machine learning models for the prediction of dam inflow. *Water* 2020, 12, 2927. [CrossRef]

16. Nourani, V.; Gokcekus, H.; Gelete, G. Estimation of Suspended Sediment Load Using Artificial Intelligence-Based Ensemble Model. *Complexity* 2021, 2021, 6635760. [CrossRef]

17. Al-adaileh, M.H. Modelling and Prediction of Water Quality by Using Artificial Intelligence. *Sustainability* 2021, 13, 4259. [CrossRef]

18. Woo, S.Y.; Jung, C.G.; Lee, J.W.; Kim, S.J. Evaluation of watershed scale aquatic ecosystem health by SWAT modeling and random forest technique. *Sustainability* 2019, 11, 3397. [CrossRef]

19. Xue, H.; Zheng, B.; Meng, F.; Wang, Y.; Zhang, L. Assessment of Aquatic Ecosystem Health of the Wutong River Based on Benthic Diatoms. *Water* 2019, 11, 727. [CrossRef]

20. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. In Proceedings of the 27th International conference on Neural Information Processing Systems; MIT Press: Cambridge, MA, USA, 2014; Volume 2, pp. 2672–2680.

21. Frid-Adar, M.; Diamant, I.; Klang, E.; Amitai, M.; Goldberger, J.; Greenspan, H. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing* 2018, 321, 321–331. [CrossRef]

22. Lu, C.; Jeric, D.; Rustia, A.; Lin, T.; Rustia, A.; Lin, T.; Lu, C.; Jeric, D.; Rustia, J.A.; Lin, T.; et al. Generative Adversarial Networks (GAN); Image augmentation; Integrated pest management. *IFAC Pap.* 2019, 52, 1–5. [CrossRef]

23. Goodfellow, I. NIPS 2016 Tutorial: Generative Adversarial Networks. *arXiv* 2016, arXiv:1701.00160.

24. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; Volume 70, pp. 214–223.

25. Wei, X.; Gong, B.; Liu, Z.; Lu, W.; Wang, L. Improving the improved training of wasserstein gans: A consistency term and its dual effect. *arXiv* 2018, arXiv:1803.01541.

26. Jiang, C.; Zhang, Q.; Ge, Y.; Liang, D.; Yang, Y.; Liu, X.; Zheng, H.; Hu, Z. Wasserstein generative adversarial networks for motion artifact removal in dental CT imaging. In Proceedings of the Medical Imaging 2019: Physics of Medical Imaging. International Society for Optics and Photonics, San Diego, CA, USA, 17–20 February 2019; Volume 10948. [CrossRef]

27. Xia, H.; Liu, C. Remote Sensing Image Deblurring Algorithm Based on WGAN. In Proceedings of the International Conference on Service-Oriented Computing, Hangzhou, China, 12–15 November 2018; Springer: Cham, Switzerland, 2018; pp. 113–125.

28. Cho, Y.; Park, M.; Shin, K.; Choi, H.; Kim, S.; Yu, S. A Study on Grade Classification for Improvement of Water Quality and Water Quality Characteristics in the Han River Watershed Tributaries. *J. Environ. Impact Assess.* 2018, 28, 215–230.

29. Lee, S.; Shin, J.Y.; Lee, G.; Sung, Y.; Kim, K.; Lim, K.J.; Kim, J. Analysis of water pollutant load characteristics and its contributions during dry season: Focusing on major streams inflow into South-Han river of Chungju-dam downstream. *J. Korean Soc. Environ. Eng.* 2018, 40, 247–257. [CrossRef]

30. Fan, J.; Li, M.; Guo, F.; Yan, Z.; Zheng, X.; Zhang, Y.; Xu, Z.; Wu, F. Priorization of river restoration by coupling soil and water assessment tool (SWAT) and support vector machine (SVM) models in the Taizi river basin, northern China. *Int. J. Environ. Res. Public Health* 2018, 15, 2090. [CrossRef]

31. Kemp, P.; Sear, D.; Collins, A.; Naden, P.; Jones, I. The impacts of fine sediment on riverine fish. *Hydrol. Process.* 2011, 25, 1800–1821. [CrossRef]

32. Chawla, N.V.; Japkowicz, N.; Kotcz, A. Editorial: Special Issue on Learning from Imbalanced Data Sets. *ACM SIGKDD Explor. Newsl.* 2004, 6, 1–6. [CrossRef]

33. Arjovsky, M.; Bottou, L. Towards principled methods for training generative adversarial networks. *arXiv* 2017, arXiv:1701.04862.

34. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *J. Mach. Learn. Res.* 2014, 15, 1929–1958.

35. Vapnik, V.N. An overview of statistical learning theory. *IEEE Trans. Neural Netw.* 1999, 10, 988–999. [CrossRef]

36. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* 1986, 1, 81–106. [CrossRef]

37. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* 1992, 46, 175–185. [CrossRef]

38. LEO Breiman Random forests. *Random For.* 2001, 1–122. [CrossRef]

39. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* 1999, 29, 1189–1232. [CrossRef]

40. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd Acm Sigkdd International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]

41. Bae, J.H.; Han, J.; Lee, D.; Yang, J.E.; Kim, J.; Lim, K.J.; Ne, J.C.; Jang, W.S. Evaluation of Sediment Trapping Efficiency of Vegetative Filter Strips Using Machine Learning Models. *Sustainability* 2019, 11, 7212. [CrossRef]

42. Choi, J. A study on the standardization strategy for building data set for machine learning applications. *J. Digit. Converg.* 2018, 16, 205–212. [CrossRef]

43. Fushiki, T. Estimation of prediction error by using K-fold cross-validation. *Stat. Comput.* 2011, 21, 137–146. [CrossRef]

44. Molinaro, A.M.; Simon, R.; Pfeiffer, R.M. Prediction error estimation: A comparison of resampling methods. *Bioinformatics* 2005, 21, 3301–3307. [CrossRef]

45. Singh, G.; Panda, R.K. Daily sediment yield modeling with artificial neural network using 10-fold cross validation method: A small agricultural watershed, Kapgari, India. *Int. J. Earth Sci. Eng.* 2011, 4, 443–450.
46. Sammut, C.; Webb, G.I. Encyclopedia of Machine Learning; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2011; ISBN 0387307680.
47. Musumba, M.; Fatema, N.; Kibriya, S. Prevention Is Better Than Cure: Machine Learning Approach to Conflict Prediction in Sub-Saharan Africa. *Sustainability* 2021, 13, 7366. [CrossRef]
48. Taner, A.; Öztekin, Y.B.; Duran, H. Performance Analysis of Deep Learning CNN Models for Variety Classification in Hazelnut. *Sustainability* 2021, 13, 6527. [CrossRef]
49. Zheng, A. *Evaluating Machine Learning Algorithms*; O’Reilly, Media Inc.: Sebastopol, CA, USA, 2015; ISBN 9781491932469.
50. Ibrahim, M.; Torki, M.; El-Makky, N. Imbalanced Toxic Comments Classification Using Data Augmentation and Deep Learning. In Proceedings of the 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 17–20 December 2018; pp. 875–878.
51. Woo, S.Y.; Jung, C.G.; Kim, J.U.; Kim, S.J. Assessment of climate change impact on aquatic ecology health indices in Han river basin using SWAT and random forest. *J. Korea Water Resour. Assoc.* 2018, 51, 863–874. [CrossRef]
52. Kim, M.; Yoon, C.G.; Rhee, H.-P.; Soon-Jin, H.; Lee, S.-W. A Study on Predicting TDI (Trophic Diatom Index) in tributaries of Han river basin using Correlation-based Feature Selection technique and Random Forest algorithm. *J. Korean Soc. Water Environ.* 2018, 34, 183–201. [CrossRef]
53. Griffiths, W.H.; Walton, B.D. The Effects of Sedimentation on the Aquatic Biota. In *Alberta Oil Sands Environmental Research Program*; Report No. 35; Oil Sands Research and Information Network; University of Alberta: Edmonton, AB, Canada, 1978. [CrossRef]
54. Kong, D.; Son, S.; Hwang, S.; Won, D.H.; Kim, M.C.; Park, J.H.; Jeon, T.S.; Lee, J.E.; Kim, J.H.; Kim, J.S.; et al. Development of Benthic Macroinvertebrates Index (BMI) for Biological Assessment on Stream Environment. *J. Korean Soc. Water Environ.* 2018, 34, 183–201. [CrossRef]
55. Newcombe, C.P.; Macdonald, D.D. Effects of Suspended Sediments on Aquatic Ecosystems. *N. Am. J. Fish. Manag.* 1991, 11, 72–82. [CrossRef]
56. Sun, Q.; Wang, W.; Gan, A. A method to accelerate the training of WGAN. In Proceedings of the 2018 5th International Conference on Information Science and Control Engineering (ICISCE), Zhengzhou, China, 20–22 July 2018; pp. 2–5. [CrossRef]
57. Longadge, R.; Dongre, S. Class Imbalance Problem in Data Mining Review. *Int. J. Comput. Sci. Netw.* 2013, 2.
58. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved training of wasserstein GANs. *Adv. Neural Inf. Process. Syst.* 2017, 2017, 5768–5778.
59. More, A.S.; Rana, D.P. Review of random forest classification techniques to resolve data imbalance. In Proceedings of the 2017 1st International Conference on Intelligent Systems and Information Management (ICISIM), Aurangabad, India, 5–6 October 2017; IEEE: Manhattan, NY, USA, 2017; pp. 72–78.
60. Zhang, M.; Shi, W.; Xu, Z. Systematic comparison of five machine-learning models in classification and interpolation of soil particle size fractions using different transformed data. *Hydrol. Earth Syst. Sci.* 2020, 24, 2505–2526. [CrossRef]
61. Patle, A.; Chouhan, D.S. SVM kernel functions for classification. In Proceedings of the 2013 International Conference on Advances in Technology and Engineering (ICATE), Mumbai, India, 23–25 January 2013; pp. 1–9.
62. Bhatia, S.; Dahyot, R. Using WGAN for improving imbalanced classification performance. *CEUR Workshop Proc.* 2019, 2563, 365–375.
63. Han, X.; Zhang, L.; Zhou, K.; Wang, X. Deep learning framework DNN with conditional WGAN for protein solubility prediction. *arXiv 2018*, arXiv:1811.07140.
64. Zhang, L.; Yang, H.; Jiang, Z. Imbalanced biomedical data classification using self-adaptive multilayer ELM combined with dynamic GAN. *Biomed. Eng. Online* 2018, 17, 1–21. [CrossRef]