

Article

Deep Ensemble Learning-Based Sensor for Flotation Froth Image Recognition

Xiaojun Zhou * and Yiping He

School of Automation, Central South University, Changsha 410083, China; yiping.he@csu.edu.cn

* Correspondence: michael.x.zhou@csu.edu.cn

Abstract: Froth flotation is a widespread and important method for mineral separation, significantly influencing the purity and quality of extracted minerals. Traditionally, workers need to control chemical dosages by observing the visual characteristics of flotation froth, but this requires considerable experience and operational skills. This paper designs a deep ensemble learning-based sensor for flotation froth image recognition to monitor actual flotation froth working conditions, so as to assist operators in facilitating chemical dosage adjustments and achieve the industrial goals of promoting concentrate grade and mineral recovery. In our approach, training and validation data on flotation froth images are partitioned in K-fold cross validation, and deep neural network (DNN) based learners are generated through pre-trained DNN models in image-enhanced training data, in order to improve their generalization and robustness. Then, a membership function utilizing the performance information of the DNN-based learners during the validation is proposed to improve the recognition accuracy of the DNN-based learners. Subsequently, a technique for order preference by similarity to an ideal solution (TOPSIS) based on the F1 score is proposed to select the most probable working condition of flotation froth images through a decision matrix composed of the DNN-based learners' predictions via a membership function, which is adopted to optimize the combination process of deep ensemble learning. The effectiveness and superiority of the designed sensor are verified in a real industrial gold–antimony froth flotation application.

Keywords: flotation froth; deep ensemble learning; image recognition; membership function; TOPSIS



Citation: Zhou, X.; He, Y. Deep Ensemble Learning-Based Sensor for Flotation Froth Image Recognition. *Sensors* **2024**, *24*, 5048. <https://doi.org/10.3390/s24155048>

Academic Editor: Kang Ryoung Park

Received: 30 June 2024

Revised: 29 July 2024

Accepted: 2 August 2024

Published: 4 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Froth flotation is one of the most widespread and significant methods for mineral separation, applicable to nearly all types of ores. In the traditional flotation unit, skilled operators adjust chemical dosages by examining visual characteristics on the surface of the flotation froth to achieve more precise separation of minerals from impurities [1]. This not only improves the recovery of minerals, but also avoids wastage of chemicals. Therefore, the recognition of production conditions in the flotation process is of great significance for improving concentrate grade and mineral recovery. However, in actual production, relying on the naked eye to observe the froth condition on the surface of the flotation tank is costly and laborious, and manual errors can easily occur due to workers' slightest carelessness, which can lead to frequent fluctuations in flotation generation indexes, serious loss of raw materials, high consumption of chemicals, and low resource recovery [2]. Thus, researchers began to develop machine vision-based froth flotation working condition recognition systems to automatically monitor the working condition of the flotation process [3], where the working condition is reflected by flotation froth in the flotation cell.

Machine vision is used to automatically extract visual features from flotation froth images and classify them accordingly. Since the surface state of the froth is closely related to the flotation performance index, researchers have developed some statistical distribution feature representations of flotation froth images, such as color features [4], texture features [5], bubble size [6], and a machine learning sensor can be used to classify the

working condition based on the statistical distribution features of the flotation froth images. However, the traditional method is often targeted and not robust enough [7]. Deep learning is the latest progress in machine learning, which can mine deeper feature information from flotation froth images. So deep neural network (DNN) sensors are increasingly applied to the problem of flotation froth working condition recognition [8]. Incorporating state-of-the-art deep learning methods into flotation froth image recognition applications can provide performance improvement. Liu et al. [9] used AlexNet and GoogLeNet in flotation froth image recognition, and they found that the deep learning models AlexNet and GoogLeNet had higher accuracy than the local binary patterns under the same number of synthetic and real images. Zarie et al. [10] employed a convolutional neural network (CNN) sensor to classify flotation froth images captured at different air flow rates, and found that the classification accuracy of the CNN sensor was higher than that of a traditional ANN sensor, which shows the great potential of DNN in the flotation froth image classification task. With the development of transformer in machine vision [11], the ViT model has also been employed in the flotation froth image recognition [12]. It has been shown that froth image features extracted from the customized ViT have good discrimination of froth image classes.

Although deep learning sensors already have good performance in flotation froth image recognition, deeper DNN models are often considered when exploring better predictive performance. They also increase training time and the risk of overfitting of DNNs, which is not expected. At this point, deep ensemble learning is a pretty good solution that decreases the performance requirements for DNN models. In [13], a deep ensemble learning sensor is employed in sintering state recognition, and has superior recognition accuracy compared to DNNs such as VGG16, Inception-v3, ResNet50, and ResNeXt50. Zhou et al. [14] proposed an ensemble learning method based on group decision making. They categorized the working condition into eight different classes on the gold–antimony froth flotation problem, and combined the predictions of the trained LeNet, AlexNet, VGGNet, and GoogLeNet through a group decision-making method.

In this paper, a deep ensemble learning-based sensor is designed for froth flotation condition recognition, which is crucial to guide workers in the control of mineral separation dosages. The practicability and effectiveness of our method is verified on a gold–antimony flotation froth image recognition application. Specifically, pre-trained models of ViT, Swin Transformer, and EfficientNet are used to generate heterogeneous base learners in partitioned and image-enhanced training sets from collected flotation froth images. Then, DNN-based learners predict the flotation froth working condition of the images and their output goes through a membership function according to their performance on the validation set. This improves the recognition accuracy of DNN-based learners. Finally, these members construct the decision matrix and select the most probable working condition by the technique for order preference by similarity to an ideal solution (TOPSIS) based on the F1 score, which is regarded as our combination method in deep ensemble learning. Compared to other methods, our approach has the highest recognition accuracy. Therefore, we build such a deep ensemble learning-based sensor in the approach, so as to achieve industrial purpose of improving concentrate grade and mineral recovery in flotation froth image recognition.

The main contributions of this paper are summarized as follows.

- (1) Heterogeneous base learners are generated with several deep learning algorithms on different training sets from images of flotation froth, which can have better diversity and accuracy.
- (2) Taking advantage of the DNN-based learners' performance differences on the validation set, a membership function of the DNN-based learners' prediction is proposed to form a decision-making problem. It also enhances the recognition accuracy of the base learners.
- (3) The TOPSIS method based on F1 score is put forward to select the most probable flotation froth working condition, which further improves recognition accuracy.

The rest of the study is organized as follows. In Section 2, preliminaries about deep ensemble learning are introduced. In Section 3, the proposed method is described in detail. In Section 4, experiments are conducted and the results are discussed. Finally, the conclusions are summarized in Section 5.

2. Preliminaries

2.1. Deep Neural Network

Recently, deep neural networks (DNNs) have become an important research field in artificial intelligence, and show excellent performance in many tasks [15]. As a branch of deep learning, CNNs play an important role in improving prediction performance in image classification, and significantly promote the development of DNNs. To deploy CNNs in practical applications more easily, Tan [16] proposed EfficientNet in 2019 and explored model scaling methods for convolutional neural networks (ConvNets) to achieve better performance with the lightweight network, which reduced the deployment requirements for applications. To make DNN models understand the overall structure and content of the image [17], researchers started to apply transformers to computer vision [18–20], since transformers can capture long-distance dependencies between pixels in the images. ViT [21] was one of the first successful studies to apply transformers to image classification. It divides and encodes an image into a set of fixed-size blocks, and then, handles these blocks through a transformer model, so as to complete image classification tasks. To achieve local sensing and deal with image information at different scales more effectively, Liu et al. [22] proposed Swin Transformer, that applied a hierarchical structure and designed a windowed self-attention mechanism on top of ViT. Finally, it achieved a good trade-off between accuracy and recognition speed. Since then, visual transformer has become an important branch of deep neural networks in image classification as well.

2.2. Deep Ensemble Learning

Deep ensemble learning refers to a methodology that combines multiple DNN-based learners to build a more powerful single model than its constituents, and improves the generalization ability of models [23]. The main process in deep ensemble learning is to train a few DNN-based learners, and then, fuse their outputs in a combination method.

Suppose we let $\{I_i\}_{i=1,\dots,m}$ denote a set of labels, K is the number of base learners, and m is the number of classes. For an observation sample \mathbf{x} , let $P_k(y_i|\mathbf{x})$ be the probability that the k -th classifier estimates for label y_m . Then, the predictions of the base learners can be summarized as three output forms as follows [24]. In this paper, we adopt the fuzzy labels predicted by DNN-based learners for the combination process of deep ensemble learning.

- Crisp label: This represents a type predicted by the base classifier with $P_k(y_i|\mathbf{x}) \in \{0, 1\}$ and $\sum_{k=1}^K P_k(y_i|\mathbf{x}) = 1$;
- Fuzzy label: This is a probabilistic explanation of the type of output of the base classifier with $\sum_{k=1}^K P_k(y_i|\mathbf{x}) = 1$;
- Possibilistic label: This relaxes the restriction of the sum of probabilities in the fuzzy label, and just needs to be greater than 0, namely, $\sum_{k=1}^K P_k(y_i|\mathbf{x}) > 0$.

2.2.1. Generation of DNN-Based Learners

Generating a base learner is the first step in ensemble learning. Generally speaking, the better the accuracy and diversity of the base learners are, the more obviously the ensemble improves the prediction performance. Deep neural networks usually have excellent accuracy in image classification, so they can be used as base learners to effectively guarantee the initial performance of ensemble learning. As for the diversity requirements of base learners, scholars have carried out a lot of studies on the generation of base learners.

In terms of the generation of base learners, it can be categorized as either serial (boosting [25]) or parallel (bagging [26]). The base learners in the two approaches are homogeneous since a single machine learning algorithm is adopted to generate the base learners. However, the models are still structured similarly, which is not conducive to further improvement of base learner diversity. It is the same for the generation of DNN-based learners [27]. Therefore, an approach to improving the diversity of base learners further is training DNN-based learners with different structures, namely, heterogeneous base learners, where the DNN-based learner can be the classical CNN or the latest ViT based on transformer structures.

Dividing the dataset into training and testing sets is a common aspect of deep learning on datasets, where the training set is mainly provided to train the model to obtain a model with better predictive ability, while the testing set is used to evaluate the generalization performance of the model. The studies by Li et al. [28], Zheng et al. [29], and Thakkar et al. [30] adopted additional bootstrapping to partition of the dataset. This obtains different training sets to train the base learner and further expand the differences between the generated base learners. Then, the dataset on which the base learner is not chosen in the training set is used as the validation set to evaluate the weights of the base learner. Akyol adopted K-fold cross validation on an EEG dataset, splitting it into certain proportions for the training and validation sets, so that the meta-classifier could be trained on the validation set [31]. Zhou et al. [14] split a fixed proportion of one part of the dataset for training, and the other part as a validation set in order to compute the combination weights of the trained base learner. Therefore, the partition method plays an important role in the generation of base learners, and it also guides the combination process of the models.

2.2.2. Combination Methods of Deep Ensemble Learning

After generating the DNN-based learners, their outputs are combined to obtain the final hypothesis in deep ensemble learning. Many combination methods for deep ensemble learning have been researched and developed in recent years, and a complete theoretical system is beginning to be formed [23,32]. The combination method affects the performance of deep ensemble learning to a certain extent, and is mainly summarized in three forms, as follows, in this paper.

- (1) Methods based on meta-classifiers: To obtain the ensemble prediction results, a meta-learner can fuse base learners' predictions and provide final results. This approach performs as a two-stage procedure in the image classification task, and is shown in Figure 1. Firstly, base learners are trained to predict target class labels. Secondly, the predictions of base learners in the validation set are taken as input of a meta-classifier, and then, the meta-classifier is trained with the input and true labels, so that it can predict the final class label accordingly. Consequently, when the deep ensemble model tests new samples, the meta-learner is able to combine the predictions of the previous DNN-based learners and obtain the final class label. In addition, the meta-learner can theoretically be any machine learning model. In this paper, the Gaussian naïve Bayes classifier is used as the meta-learner to facilitate the comparison of the prediction performance in the flotation froth image recognition application [27].
- (2) Methods based on voting: Voting among DNN-based learners themselves is a common combination method, shown in Figure 2. We categorize combination methods based on voting into simple majority voting and weighted majority voting according to whether additional base learner information is used or not. Simple majority voting combines base learners by calculating the number of votes of the base learners of each class label and selecting the one with the most votes as the final class label. Due to its robustness [33], many studies on ensemble learning still use it as the combination method. Simple majority voting can be regarded as using the same voting weight, but weighted majority voting assigns different weights to the base learners, which reflect the reliability of the base learners' predictions based on their performance during training or validation. Therefore, when there is performance

difference between the base learners, weighted majority voting performs better. In this paper, the weights for the base learners are based on the evaluation measures, such as accuracy and precision. In addition, the voting methods are again differentiated based on the soft and hard labels predicted by the base learners, namely, fuzzy labels and crisp labels, which are defined in Section 2.2. They can both obtain final ensemble class labels through voting rules, and simple majority voting uses hard labels, while weighted majority voting can apply any of them.

- (3) Methods based on aggregation rules: In this case, the process is similar to simple majority voting because there is no external information or base learner information during combination. However, instead of using the crisp label output of the base learner in simple majority voting, the fuzzy label output of the base learner is aggregated according to a rule in this approach, which is as shown in Figure 3. In this paper, we use the average, minimum, and medium aggregation methods as comparison methods, and verify the effectiveness of our proposed method in flotation froth image recognition experiments.

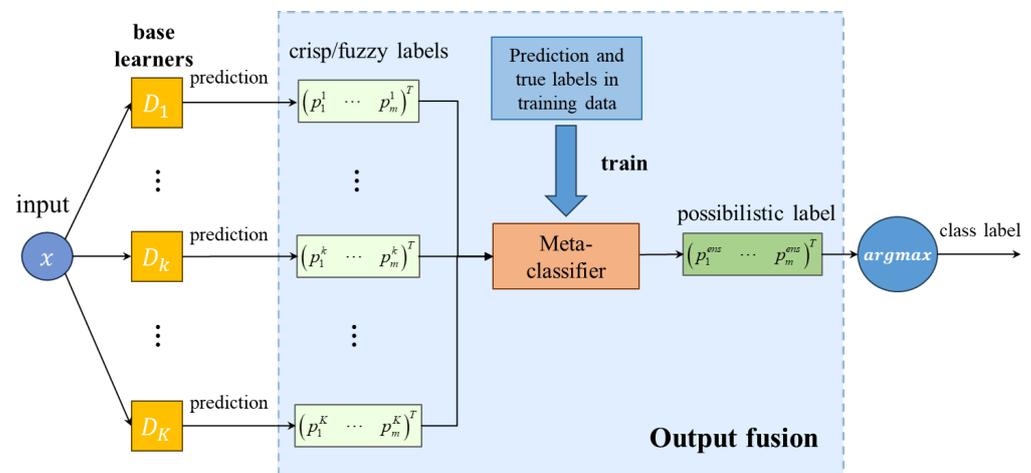


Figure 1. Combination methods based on meta-classifiers.

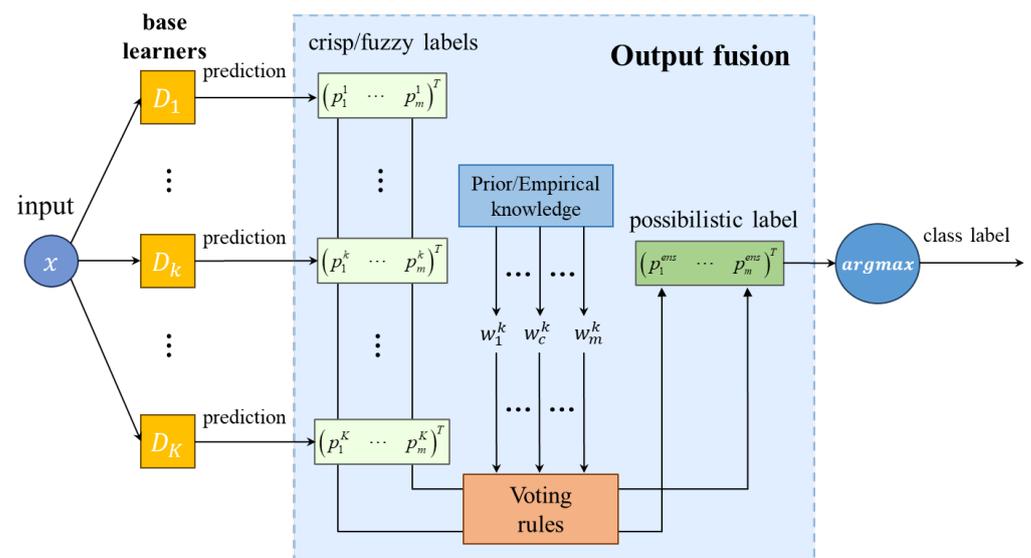


Figure 2. Combination methods based on voting.

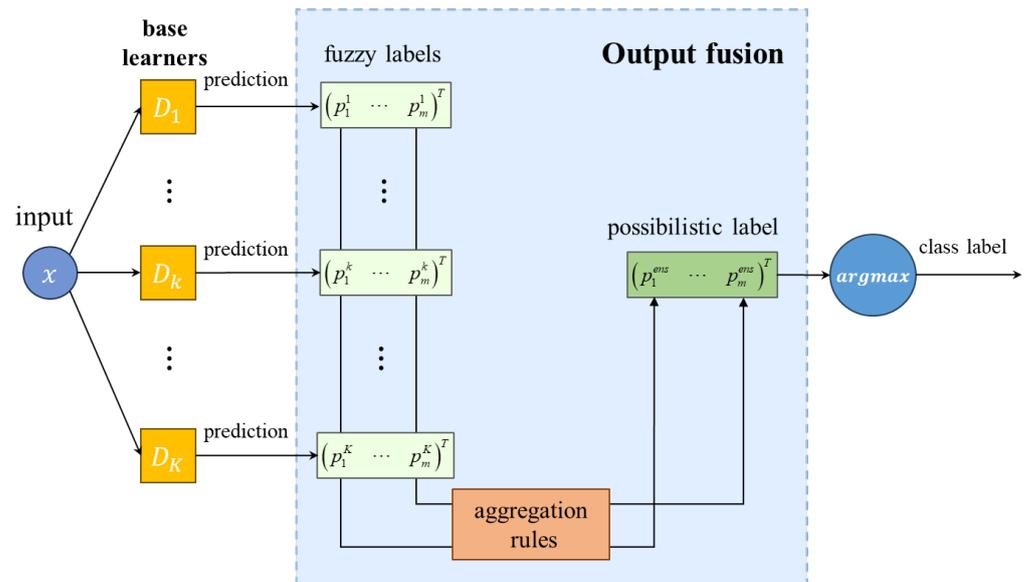


Figure 3. Combination methods based on aggregation rules.

3. Methodology

3.1. Proposed Deep Ensemble Learning-Based Sensor

The flotation froth working condition is one of the important criteria affecting the mineral flotation process, so we recognize the working condition through the captured flotation froth images. A deep ensemble learning-based sensor is used to predict the flotation froth working condition type and guide operators to adjust the chemical dosage in order to improve the separated mineral grade. The scheme of our deep ensemble learning-based sensor is shown in Figure 4.

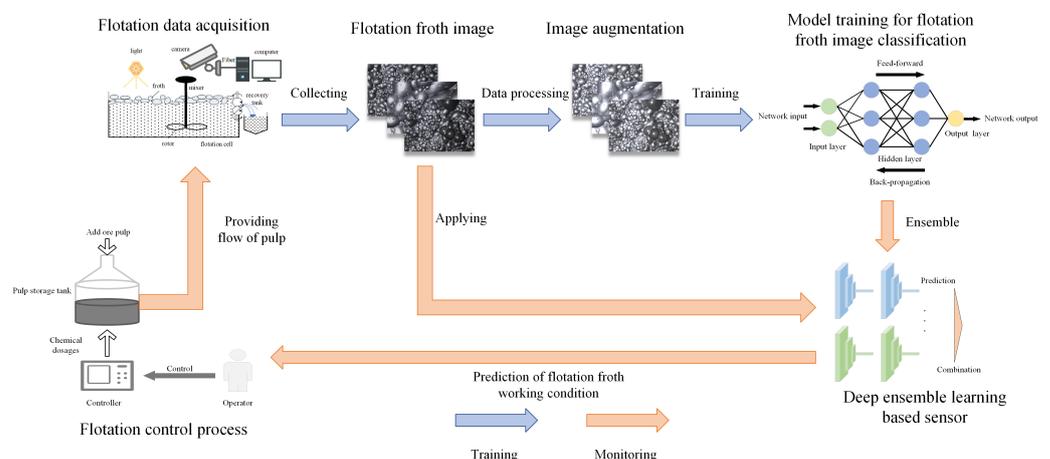


Figure 4. Scheme of flotation froth image recognition in our deep ensemble learning-based sensor.

In the flotation froth image recognition task, we ensemble multiple heterogeneous deep neural networks trained on different training sets, and improve their prediction by a membership function and TOPSIS weighted on the F1 score. This deep ensemble learning approach is proposed to improve the generalization ability in the flotation froth application, and its main steps are as follows.

1. Categorize the flotation froth working condition into m class labels according to actual demand, and use experienced workers to label the flotation froth images.
2. Obtain the training sets and validation sets from the images by the K-fold partitioning, so that the deep neural network can be trained on different training sets via image enhancement.

3. Fine-tune the models with S kinds of pre-trained deep neural networks, then one DNN model with the highest accuracy rate on the validation set is selected to be used to form a total of $K \times S$ DNN-based learners.
4. Obtain fuzzy labels of the DNN-based learners on the flotation froth testing set, and calculate the corrected prediction of the DNN-based learners through a membership function.
5. Construct the decision matrix with corrected prediction and rank the working condition by TOPSIS weighted based on the F1 score. Then, select the most probable flotation froth working condition category.

The flowchart of our deep ensemble learning method for flotation froth image recognition is shown in Figure 5.

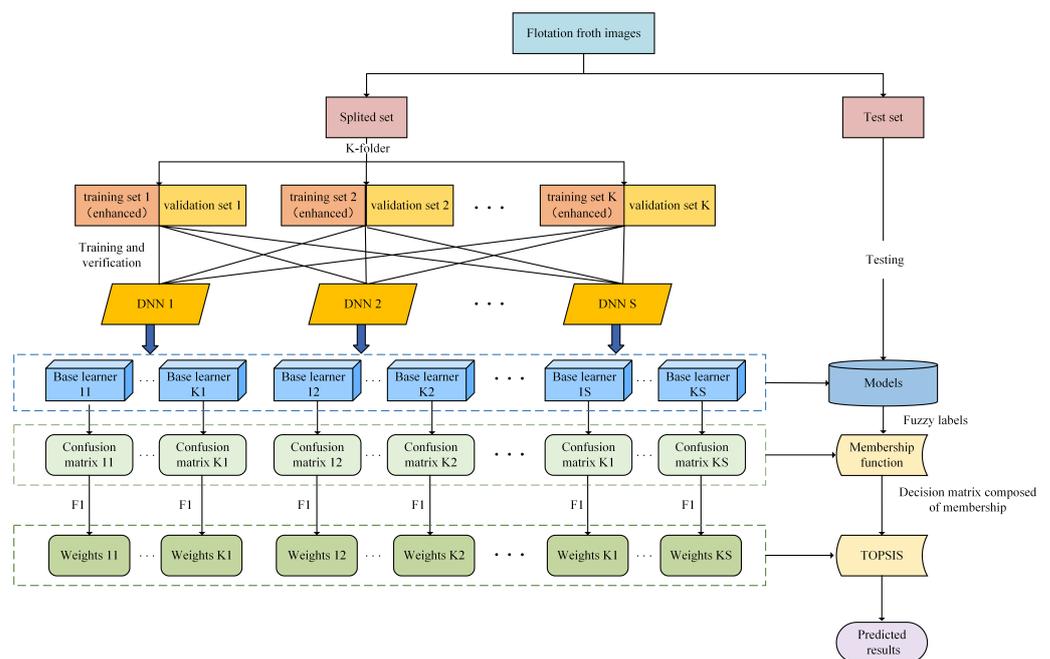


Figure 5. The flowchart of our deep ensemble learning method for flotation froth image recognition.

3.2. Generation of DNN Trained in Partitioned and Enhanced Flotation Froth Images

Training data are one of the most important factors to obtain accurate and diverse DNN-based learners. To ensure that the training data vary as much as possible, we adopt the K -fold cross validation to split the training and validation sets as in stacking, as shown in Figure 6. Specifically, there is a total of K subsets (D_1, \dots, D_K), all of which can be used as validation sets for the DNN-based learners. Meanwhile, the validation set is not only used to select the best DNN model in the iterative process, but also plays an important role in the membership function for the DNN model's prediction correction, because a confusion matrix is obtained in the process of verification. Therefore, the K -fold partition to split flotation froth images is an essential part of our proposed deep ensemble learning method to improve the prediction performance of our deep ensemble learning sensor.

After the k -th subset D_k is selected as the validation set, the subsets, except D_k , are processed for image enhancement (probabilistic flipping, probabilistic clipping, and weak changes in picture attributes). Then, enhanced subsets $D'_i (i \neq k)$ are obtained that constitute a training set together for DNN training. This reduces the sensitivity of the base learner to the flotation froth images, thus improving the robustness of the trained DNN-based learner for the flotation froth application. In addition, the pre-trained model has extracted generic features through a wide variety of images, so using and fine-tuning it on our training data can lead to obtaining a sufficiently satisfactory DNN-based learner in fewer epochs, and the trained DNN models have better generalization ability. Therefore, our trained

DNN models can also perform well when predicting an unknown testing set, even though there are probably some deviations or biases from the training set.

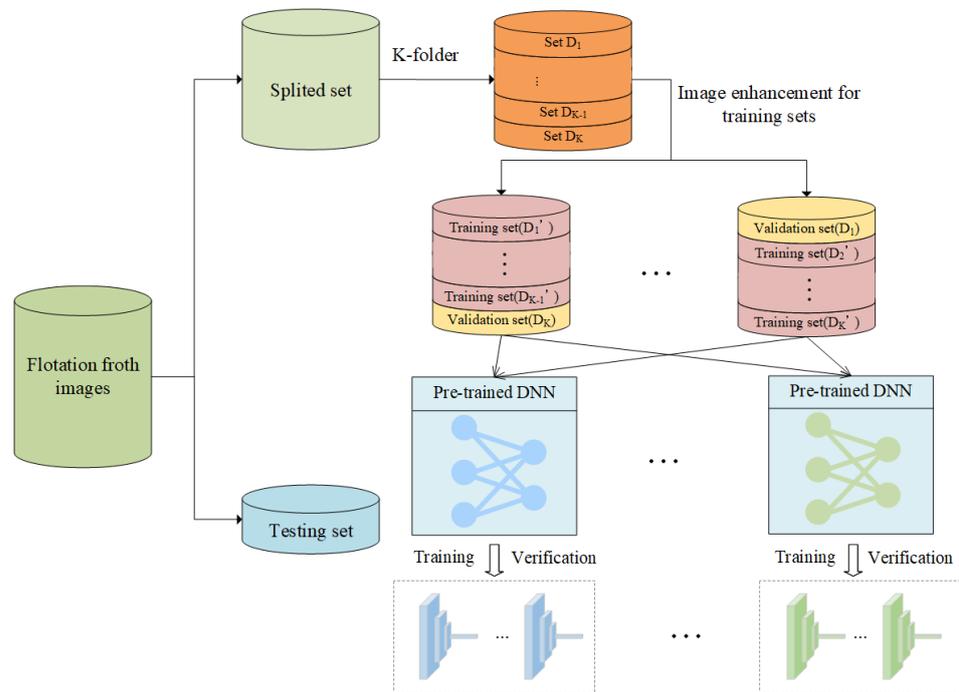


Figure 6. Partitioning of training, validation, and testing sets.

3.3. Combination Method Based on Membership Function and TOPSIS

3.3.1. Design of Membership on Flotation Froth Working Condition

Assuming that l_i is the predicted label and $p_{ks}(l_i|x)$ is the fuzzy label of the test sample x from the DNN-based learner T_{ks} , the confusion matrix CM_{ks} is obtained on the validation set D_k from DNN-based learner T_{ks} .

$$CM_{ks} = \begin{bmatrix} L_{11} & \cdots & L_{1m} \\ \vdots & \ddots & \vdots \\ L_{m1} & \cdots & L_{mm} \end{bmatrix} \quad (1)$$

where m is the number of flotation froth class labels, T_{ks} represents the s -th pre-trained DNN model, and the DNN is validated on the k -th subset as the validation set.

Given that L_{ji} is the number of flotation froth images in the validation set when the true label is l_j and simultaneously the predicted label is l_i , then its prior frequency is calculated as follows:

$$p_{ks}(\hat{l}_j|l_i) = L_{ji} / \sum_{k=1}^m L_{ki} \quad (2)$$

where $p_{ks}(l_i|x)$ is the fuzzy labels predicted by the DNN-based learner T_{ks} .

If there is a sufficiently large number of samples in the validation set, we can regard the frequency as a probability, and then, use a membership function on fuzzy labels of this DNN-based learner in Equation (3).

$$\begin{aligned} p_{ks}(\hat{l}_j|x) &= f_{membership}(p_{ks}(l_1|x), p_{ks}(l_2|x) \cdots p_{ks}(l_n|x)) \\ &= \sum_{i=1}^n p_{ks}(\hat{l}_j|l_i) p_{ks}(l_i|x) \end{aligned} \quad (3)$$

The computation of the membership function on fuzzy labels in this paper draws on the advantage of the weight in the weighted voting method and becoming more flexible. It

not only utilizes all the performance information of the DNN-based learners' confusion matrix, but also considers the original fuzzy label potential links between different categories, in order to represent a better prediction of every DNN-based learner. Finally, the fuzzy labels, via the membership function, construct a decision matrix in Equation (4), which can be solved by means of decision theory.

$$X_{m \times B} = \begin{bmatrix} x_{11} & \cdots & x_{1B} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mB} \end{bmatrix} = \begin{bmatrix} p_1(\hat{l}_1|\mathbf{x}) & \cdots & p_B(\hat{l}_1|\mathbf{x}) \\ \vdots & \ddots & \vdots \\ p_1(\hat{l}_m|\mathbf{x}) & \cdots & p_B(\hat{l}_m|\mathbf{x}) \end{bmatrix} \quad (4)$$

where B is the number of DNN-based learners and $B = K \times S$.

3.3.2. Decision-Making Method for Flotation Froth Working Condition

TOPSIS is a decision-making method commonly used in the decision-making field, and its core idea is the distance between each solution and the positive and negative ideal solutions [34]. We sort the working condition by calculating the proximity of each category to the positive and negative ideal solution according to the membership decision matrix. The closest one has the largest evaluation value and is determined as the final predicted working condition.

The steps of our proposed ranking method based on TOPSIS weighted by the F1 score of the DNN-based learners are as follows.

1. Fuzzy labels, via the membership function, can constitute the decision matrix $X_{m \times t}$, which is shown in Equation (4).
2. Compute the positive ideal solution C_i^+ and the negative ideal solution C_i^- :

$$C_i^+ = \max_{j=1, \dots, KS} (x_{ij}) \quad (5)$$

$$C_i^- = \min_{j=1, \dots, KS} (x_{ij}) \quad (6)$$

3. Calculate of the distance from the working condition l_i to the positive and negative ideal solutions, taking the F1 score for each category of the DNN-based learner as weights.

$$s_i^+ = \sqrt{\sum_{j=1}^{KS} F_{ij} (c_{ij} - c_i^+)^2} \quad (7)$$

$$s_i^- = \sqrt{\sum_{j=1}^{KS} F_{ij} (c_{ij} - c_i^-)^2} \quad (8)$$

where F_{ij} represents the F1 score of DNN-based learner j in the flotation froth working condition l_i .

4. Calculate the value of the ranking indicator f_i for each working condition l_i .

$$f_i = \frac{s_i^-}{s_i^- + s_i^+} \quad (9)$$

5. Let $i_0 = \arg \max_i f_i$, and choose l_{i_0} as the final working condition label.

4. Experimental Results

4.1. Data Description

Gold–antimony froth flotation is an important process that utilizes the principle of bubble buoyancy to separate gold–antimony material from ore slurry [35]. In this paper, gold–antimony froth flotation images were collected to validate the effectiveness

of the proposed method. We set up a computer, lamp, camera, and other devices in a gold–antimony flotation industry in Hunan, China, and then, collected flotation froth images from the flotation tanks for experiments. Workers experienced with flotation froth evaluated the working condition according to the characteristics of the images, and categorized them into eight different types, which are shown in Figure 7. These categories reflect the gold–antimony production condition so as to assist in adjusting the chemical dosages, which can improve the recovery of minerals and reduce the cost. Each category is uniformly distributed in order to verify the effectiveness of the method proposed in this paper.

In addition, each captured flotation froth image is $800 \text{ px} \times 600 \text{ px}$ and clear enough to present its working condition in Figure 7. There are a total of 2400 flotation froth images and a testing set of 480 images is split off. The rest of the images are used for DNN training and verification.

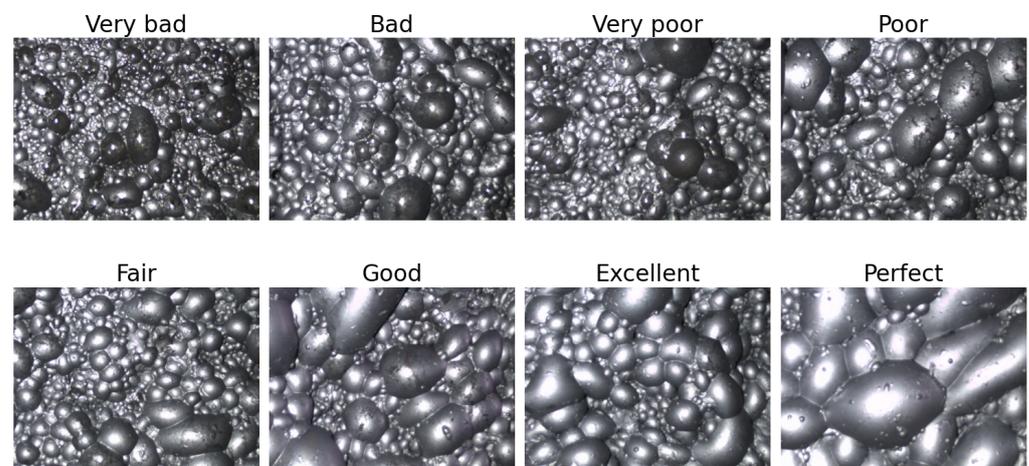


Figure 7. Flotation froth under eight different working conditions.

4.2. Experiment Setting

In our experiment, the pre-trained EfficientNet [16], ViT [21], and Swin Transformer [22] were each used to train a total of 12 heterogeneous DNN-based learners on 4-fold-partitioned flotation froth data, and the trained DNN-based learners were used as our deep ensemble learning sensor. The 4-fold partitioning of the flotation froth data resulted in a quarter of the images being used as the validation set, which could sufficiently represent unknown flotation froth samples and ensure the base learners' generalization ability. In addition, our training set was flipped probabilistically, clipped probabilistically, rotated probabilistically, and jittered in color attributes to enhance the flotation froth images.

The deep neural network settings for training are shown in Table 1. As is shown in Table 1, we attempted to set the same parameters in different DNNs, such as epoch and batch size. Then, we fine-tuned these pre-trained DNNs, and the DNN with the highest accuracy in the validation set in its training epoch process was taken as the DNN-based learner for the deep ensemble learning-based sensor.

Table 1. Experimental settings of DNN.

DNN	EfficientNet [16]	ViT [21]	Swin Transformer [22]
Epoch	50	50	50
Batch size	20	20	20
Optimization function	Adam	Adam	AdamW

Table 1. Cont.

DNN	EfficientNet [16]	ViT [21]	Swin Transformer [22]
Input dimension	224 × 224	224 × 224	224 × 224
Model	efficientnet_v2_s	vit_b_16	swin_tiny_patch4_window7_224
Last layer	Linear(1280,8)	Linear(768,8)	Linear(768,8)

4.3. Comparison with Base Learners

The recognition accuracy of 12 DNN-based learners for the gold–antimony froth flotation application is shown in Figure 8. We can find that it is the same in the validation and testing set of each base learner, which indicates that the data distribution is basically consistent. In addition, the accuracy of the base learners mostly reaches 90% on the validation and testing sets. This illustrates that our K-fold partitioning and image enhancement are useful for training good DNN-based learners.

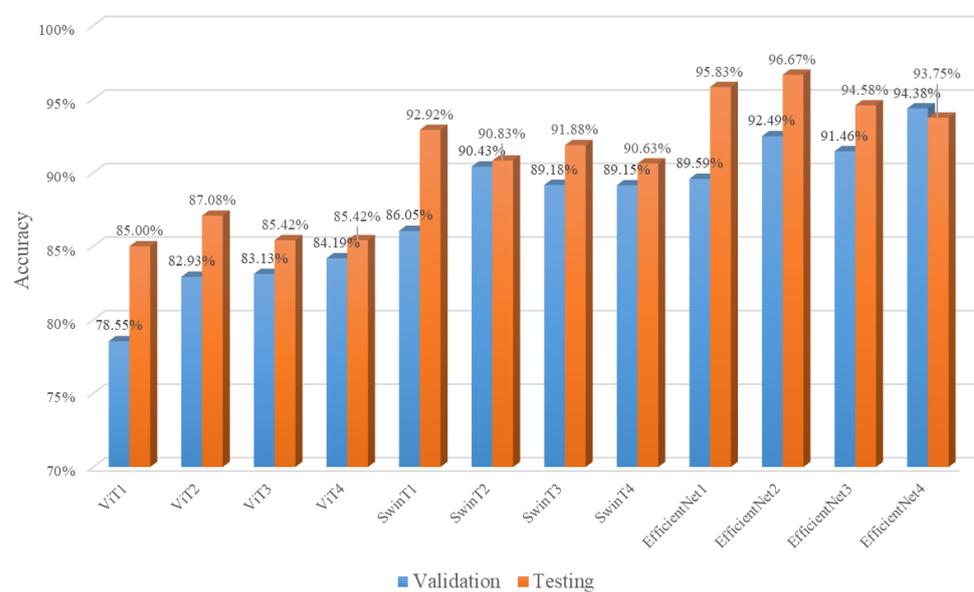


Figure 8. Accuracy in base learners in flotation froth image recognition task.

In Figure 8, the symbols of the DNN-based learner signify which deep learning network was trained under which fold in the flotation froth data. For example, ViT1/SwinT1/EfficientNet1 indicates that ViT/Swin Transformer/EfficientNet is evaluated in first fold of the validation set; after, it is trained on the rest of the subsets. Other DNN-based learners trained on other subsets are also represented in the same way. The best base learner performance on the validation set is EfficientNet3, with the accuracy of 94.38%. However, if the base learner with the highest accuracy on the validation set is selected to be used in the testing set, it has a lower recognition accuracy of 93.75% compared to the other DNN-based learners, one of which achieves 96.67% accuracy in the testing set. This shows that if one of these base learners is selected for the application of gold–antimony froth flotation condition recognition, it is difficult to pick out a DNN with sufficiently good generalization ability according to its performance on the validation set, because an overfitted DNN-based learner is easily chosen, which decreases the recognition accuracy on the testing set.

Therefore, we combine these DNN-based learners using deep ensemble learning, and then, apply the ensemble model to the testing set of gold–antimony froth flotation image recognition. In our method, the membership of the flotation froth working condition constructs the decision matrix and TOPSIS weighted by the F1 score is applied to combine the base learners; its confusion matrix is shown in Figure 9. According to Figure 9, there

are only a total of eight misclassified flotation froth images, and we can calculate that the accuracy of our proposed deep ensemble learning approach reaches 98.33%, which is better than any of the base learners. This demonstrates that our deep ensemble learning-based sensor improves the generalization ability of the model and verifies the effectiveness of our method.

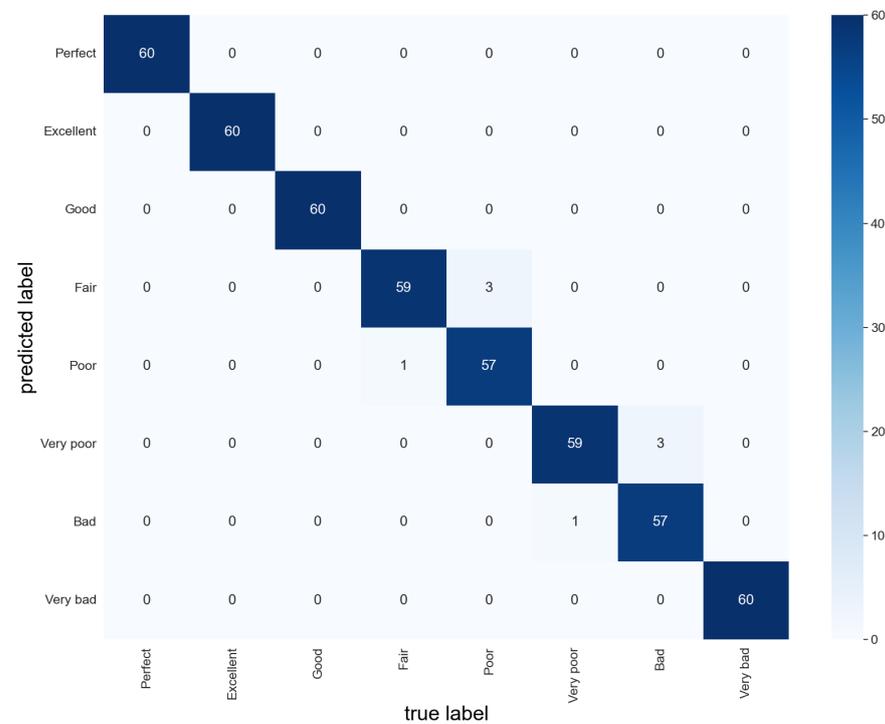


Figure 9. Confusion matrix of our method.

4.4. Comparison with Classic Ensemble Learning Methods

We compare the proposed method with classic ensemble learning methods such as Adaboost [36], gradient boosting [37], bagging [26], XGBoost [38], and random forest [39]. The inputs in the classical ensemble learning method are 38 statistical features extracted from the froth flotation images referring to [40], including the mean of gray, bubble size distribution, length–width ratio, and so on. Finally, the experimental results are shown in Figure 10.

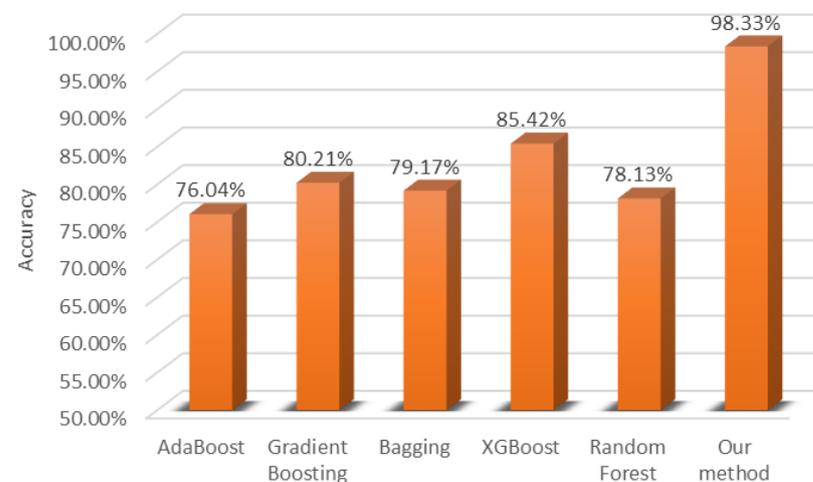


Figure 10. Accuracy of classical ensemble learning.

According to Figure 10, the recognition accuracy of the classical ensemble learning method on the gold–antimony froth flotation testing set is not more than 90%, which is even lower than that of the DNN-based learner shown in Figure 8, indicating that the DNN is more suitable for the recognition task. Classic ensemble learning does not perform well because the extracted statistical features may lose some important information about the flotation froth image or mix with a lot of redundant information, which results in worse generalization ability of the trained ensemble model. This is also an important reason why most studies prefer deep learning methods in image classification.

4.5. Comparison with Other Combination Methods

In order to further verify the superiority of our proposed method in flotation froth image recognition applications, we compare our approach with other combination methods in deep ensemble learning. Specifically, the minimum (MIN), product (PRO), and average (AVR) aggregation rules [41], the soft voting method (SV) [42] and hard voting method (HV) [42], weighted hard voting based on F_{in} (FHV) [14], weighted voting using global accuracy (WVGA) [27], weighted voting using class precision (WVCP) [27], Bayesian meta-classifier using predictions (PBAY) [27], and multilayer perceptron meta-classifier using predictions (PMLP) [43] are taken as combination methods on the testing set and the experimental results are shown in Figure 11. As shown in Figure 11, all of these combination methods clearly improve the flotation froth image recognition accuracy on the testing set compared to the DNN-based learner in Figure 8, which demonstrates the better generalization performance of deep ensemble learning. In addition, the combination method based on the aggregation rules essentially outperforms the voting method, implying that the aggregation rules work well in this application. The stacking method using a Bayesian classifier and MLP as the meta-learner has poor performance in this application, because the meta-learner, with a complicated structure, produces an overfitting phenomenon due to the quite good diversity and accuracy of the DNN-based learners. Therefore, the relatively simpler combination method has better performance. Last, but not least, our approach has the highest recognition accuracy, which verifies the superiority of our combination method in the application.

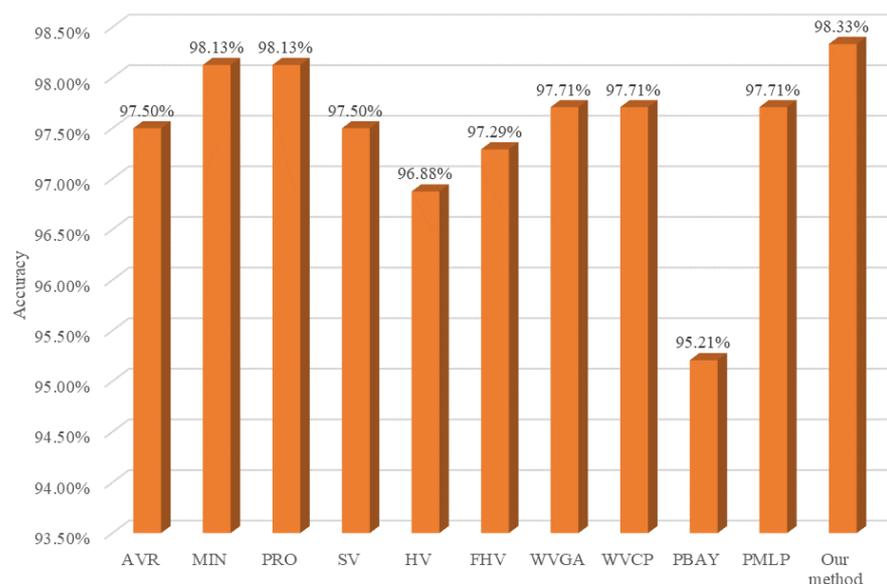


Figure 11. Accuracy of different combination methods on testing data.

To further illustrate the improvement in our combination method, we display the accuracy of the DNN-based learners on the testing set when using the membership function or not in Figure 12. According to Figure 12, most of the 12 DNN-based learners via the

membership function have higher accuracy on the testing set, verifying that the membership function improves the prediction performance of the base learners.

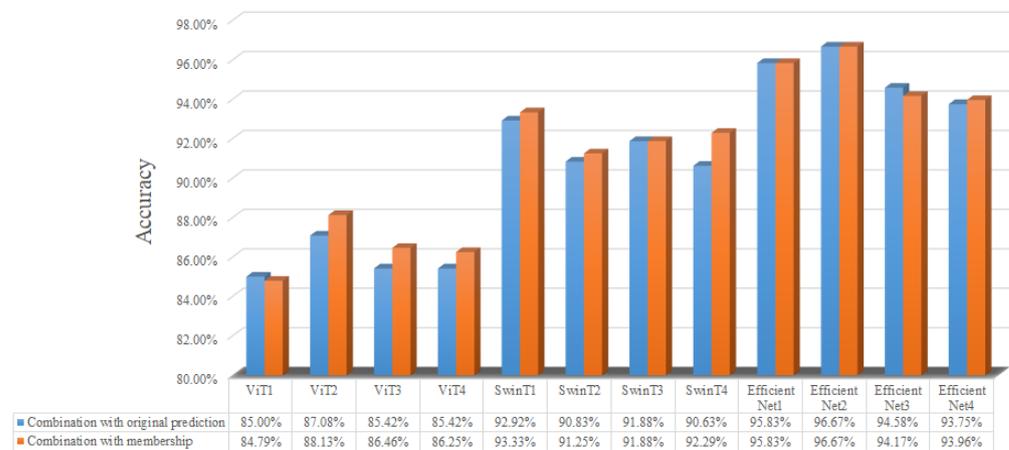


Figure 12. Accuracy of base learners using membership function or not.

In addition, we analyzed a sample of gold–antimony flotation froth images in the testing set for observation. It is hard to be categorized into “good”, but its real working condition type is “good”, because the DNN-based learners prefer categorizing it as “fair”. The fuzzy labels of the DNN-based learners are shown in Table 2, where the bold number is the largest prediction. After the DNN-based learners’ predictions are put through Equation (3), their predictions are closer to the real working condition, according to Table 3. It can be found that most of the base learners do not regard it to be “good”, and thus, cannot correctly predict its working condition even in the soft voting combination method. Afterwards, the DNN-based learners’ predictions are put through Equation (3), and their predictions are closer to the real working condition. Then, the TOPSIS method, weighted by the F1 score, finally makes the evaluation value of the right type, “good”, larger than that of the wrong type, “fair”, and thus, the right flotation froth working condition can be determined successfully.

Finally, we carried out certain image attribute changes and rotations randomly on the flotation froth testing data, which simulate the image interference in the real flotation froth process. The results are shown in Figure 13. It can be found that the accuracy of our proposed method does not decrease obviously and reaches 96.88%, while the accuracy of the other combination methods is not better than it, which verifies the robustness of our method.

Table 2. Original fuzzy labels of base learners.

Type	ViT1	ViT2	ViT3	ViT4	SwinT1	SwinT2	SwinT3	SwinT4	EfficientNet1	EfficientNet2	EfficientNet3	EfficientNet4	AVR
Very bad	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
Bad	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Very poor	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Poor	0.011	0.149	0.038	0.075	0.000	0.000	0.001	0.047	0.001	0.006	0.005	0.000	0.028
Fair	0.125	0.821	0.146	0.666	0.983	0.001	0.659	0.941	0.034	0.933	0.075	0.006	0.449
Good	0.832	0.012	0.714	0.044	0.015	0.997	0.009	0.011	0.770	0.058	0.006	0.994	0.372
Excellent	0.027	0.017	0.089	0.210	0.002	0.000	0.332	0.001	0.194	0.002	0.912	0.000	0.149
Perfect	0.003	0.001	0.012	0.004	0.000	0.001	0.000	0.000	0.000	0.001	0.000	0.000	0.002

Table 3. Fuzzy labels of base learners via membership function.

Type	ViT1	ViT2	ViT3	ViT4	SwinT1	SwinT2	SwinT3	SwinT4	EfficientNet1	EfficientNet2	EfficientNet3	EfficientNet4	Our Method
Very bad	0.001	0.000	0.001	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
Bad	0.000	0.014	0.010	0.017	0.000	0.000	0.000	0.007	0.000	0.000	0.000	0.000	0.000
Very poor	0.001	0.000	0.000	0.000	0.011	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Poor	0.042	0.191	0.044	0.251	0.249	0.016	0.088	0.053	0.005	0.090	0.013	0.017	0.027

Table 3. Cont.

Type	ViT1	ViT2	ViT3	ViT4	SwinT1	SwinT2	SwinT3	SwinT4	EfficientNet1	EfficientNet2	EfficientNet3	EfficientNet4	Our Method
Fair	0.081	0.669	0.122	0.441	0.656	0.050	0.572	0.929	0.031	0.792	0.082	0.006	0.447
Good	0.756	0.052	0.628	0.089	0.082	0.884	0.008	0.011	0.733	0.052	0.006	0.977	0.491
Excellent	0.026	0.069	0.097	0.188	0.002	0.016	0.332	0.001	0.188	0.059	0.897	0.000	0.166
Perfect	0.092	0.004	0.098	0.013	0.000	0.033	0.000	0.000	0.044	0.006	0.000	0.000	0.003

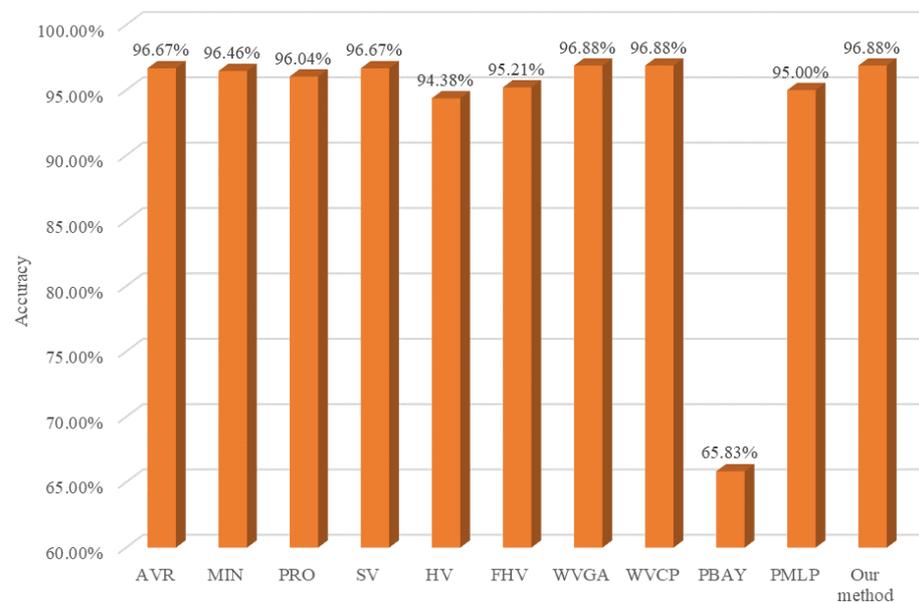


Figure 13. Accuracy of different combination methods on testing data under image interference.

5. Conclusions

In this study, a deep ensemble learning-based sensor is designed to automatically monitor the flotation working condition on the basis of the captured flotation froth images, so that the chemical dosages can be adjusted accordingly. The K-fold partition is used in partitioning training and validation data and image enhancement is performed on the training data to generate DNN-based learners with diversity and accuracy in different deep learning algorithms. With the predictions obtained from these DNN-based learners, a membership function utilizing the performance information of the DNN-based learners during the validation is proposed to improve the recognition accuracy of the DNN-based learners. Then, membership of the flotation froth working condition constructs the decision matrix and TOPSIS weighted by the F1 metric is proposed to combine the base learners. Finally, the most probable working condition is selected to assist operators in adjusting the chemical dosages.

The experimental results show that the sensor based on deep ensemble learning method can effectively recognize the flotation froth working condition and achieves the accuracy of 98.33% on the testing data of the gold–antimony flotation froth application, which demonstrates the superiority of the proposed approach compared to other common methods.

In future work, we will explore different DNN models and more performance metrics of base learners, as well as investigate more deep ensemble learning based on other decision-making methods to improve the applicability and interpretability of our approaches. Moreover, different flotation froth applications under diverse industrial conditions will be explored to enhance the generalizability of our approach, especially in variable flotation froth working conditions. In addition, we also believe that this work has potential to be applied to other recognition tasks like regression and object detection.

Author Contributions: Conceptualization, X.Z. and Y.H.; methodology, X.Z.; software, Y.H.; validation, Y.H.; formal analysis, Y.H.; investigation, X.Z.; resources, X.Z.; data curation, X.Z.; writing—original draft preparation, X.Z. and Y.H.; writing—review and editing, X.Z. and Y.H.; visualization, Y.H.; supervision, X.Z.; project administration, X.Z.; funding acquisition, X.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 62273357, and the Hunan Provincial Natural Science Foundation of China, grant number 2021JJ20082.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Due to privacy or ethical restrictions data are unavailable.

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

References

1. Xu, P.; Tian, L.; Liu, J.; Luo, D.; Jahanshahi, H. MsFFTsGP: Multi-source features-fused two-stage grade prediction of zinc tailings in lead-zinc flotation process via multi-stream 3D convolution with attention mechanism. *Eng. Appl. Artif. Intell.* **2024**, *129*, 107647. [[CrossRef](#)]
2. Jagaba, A.H.; Birniwa, A.H.; Usman, A.K.; Mu'azu, N.D.; Yaro, N.S.A.; Soja, U.B.; Abioye, K.J.; Almabashi, N.M.Y.; Al-dhawi, B.N.S.; Noor, A.; et al. Trend and current practices of coagulation-based hybrid systems for pulp and paper mill effluent treatment: Mechanisms, optimization techniques and performance evaluation. *J. Clean. Prod.* **2023**, *429*, 139543. [[CrossRef](#)]
3. Zhang, J.; Tang, Z.; Xie, Y.; Ai, M.; Gui, W. Generative adversarial network-based image-level optimal setpoint calculation for flotation reagents control. *Expert Syst. Appl.* **2022**, *197*, 116790. [[CrossRef](#)]
4. Uusitalo, S.; Soudunsaari, T.; Sumen, J.; Haavisto, O.; Kaartinen, J.; Huuskonen, J.; Tuikka, A.; Rahkamaa-Tolonen, K.; Paaso, J. Online analysis of minerals from sulfide ore using near-infrared Raman spectroscopy. *J. Raman Spectrosc.* **2020**, *51*, 978–988. [[CrossRef](#)]
5. Bendaouia, A.; Abdelwahed, E.H.; Qassimi, S.; Boussetta, A.; Benzakour, I.; Benhayoun, A.; Amar, O.; Bourzeix, F.; Baina, K.; Cherkaoui, M.; et al. Hybrid features extraction for the online mineral grades determination in the flotation froth using Deep Learning. *Eng. Appl. Artif. Intell.* **2024**, *129*, 107680. [[CrossRef](#)]
6. Jahedsaravani, A.; Massinaei, M.; Zarie, M. Measurement of bubble size and froth velocity using convolutional neural networks. *Miner. Eng.* **2023**, *204*, 108400. [[CrossRef](#)]
7. Yang, Y.; Cao, S.; Wang, X.; Yang, C. Froth velocity extraction method from froth image for flotation with large throughput based on composite deep learning network. *Measurement* **2023**, *223*, 113672. [[CrossRef](#)]
8. Zhang, H.; Tang, Z.; Xie, Y.; Yin, Z.; Gui, W. ES-net: An integration model based on encoder-decoder and siamese time series difference network for grade monitoring of zinc tailings and concentrate. *IEEE Trans. Ind. Electron.* **2023**, *70*, 11819–11830. [[CrossRef](#)]
9. Liu, X.; Aldrich, C. Recognition of flotation froth conditions with k-shot learning and convolutional neural networks. *J. Process. Control.* **2023**, *128*, 103004. [[CrossRef](#)]
10. Zarie, M.; Jahedsaravani, A.; Massinaei, M. Flotation froth image classification using convolutional neural networks. *Miner. Eng.* **2020**, *155*, 106443. [[CrossRef](#)]
11. Han, K.; Wang, Y.; Chen, H.; Chen, X.; Guo, J.; Liu, Z.; Tang, Y.; Xiao, A.; Xu, C.; Xu, Y.; et al. A Survey on Vision Transformer. *IEEE Trans. Pattern Anal. Mach. Intell.* **2023**, *45*, 87–110. [[CrossRef](#)]
12. Liu, X.; Aldrich, C. Flotation Froth Image Recognition Using Vision Transformers. *IFAC-Pap.* **2023**, *56*, 2329–2334. [[CrossRef](#)]
13. Xu, X.; Zhou, X. Deep Learning Based Feature Selection and Ensemble Learning for Sintering State Recognition. *Sensors* **2023**, *23*, 9217. [[CrossRef](#)]
14. Zhou, X.; He, J.; Yang, C. An ensemble learning method based on deep neural network and group decision making. *Knowl.-Based Syst.* **2022**, *239*, 107801. [[CrossRef](#)]
15. Man, T.; Osipov, V.Y.; Zhukova, N.; Subbotin, A.; Ignatov, D.I. Neural networks for intelligent multilevel control of artificial and natural objects based on data fusion: A survey. *Inf. Fusion* **2024**, *110*, 102427. [[CrossRef](#)]
16. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, PMLR, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
17. Touvron, H.; Cord, M.; Douze, M.; Massa, F.; Sablayrolles, A.; Jégou, H. Training data-efficient image transformers & distillation through attention. In Proceedings of the International Conference on Machine Learning, PMLR, Virtual, 18–24 July 2021; pp. 10347–10357.
18. Chen, J.; Zhang, Y.; Pan, Y.; Xu, P.; Guan, C. A transformer-based deep neural network model for SSVEP classification. *Neural Netw.* **2023**, *164*, 521–534. [[CrossRef](#)] [[PubMed](#)]

19. Chopin, B.; Tang, H.; Otberdout, N.; Daoudi, M.; Sebe, N. Interaction transformer for human reaction generation. *IEEE Trans. Multimed.* **2023**, *25*, 8842–8854. [[CrossRef](#)]
20. Rendón-Segador, F.J.; Álvarez García, J.A.; Salazar-González, J.L.; Tommasi, T. CrimeNet: Neural Structured Learning using Vision Transformer for violence detection. *Neural Netw.* **2023**, *161*, 318–329. [[CrossRef](#)] [[PubMed](#)]
21. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv* **2020**, arXiv:2010.11929.
22. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10012–10022.
23. Ganaie, M.A.; Hu, M.; Malik, A.K.; Tanveer, M.; Suganthan, P.N. Ensemble deep learning: A review. *Eng. Appl. Artif. Intell.* **2022**, *115*, 105151. [[CrossRef](#)]
24. Nguyen, T.T.; Nguyen, T.T.T.; Pham, X.C.; Liew, A.W.C. A novel combining classifier method based on variational inference. *Pattern Recognit.* **2016**, *49*, 198–212. [[CrossRef](#)]
25. Freund, Y.; Schapire, R.E. Experiments with a new boosting algorithm. In Proceedings of the Machine Learning: Proceedings of the Thirteenth International Conference, San Francisco, CA, USA, 3–6 July 1996; Volume 96, pp. 148–156.
26. Breiman, L. Bagging predictors. *Mach. Learn.* **1996**, *24*, 123–140. [[CrossRef](#)]
27. Sesmero, M.P.; Iglesias, J.A.; Magán, E.; Ledezma, A.; Sanchis, A. Impact of the learners diversity and combination method on the generation of heterogeneous classifier ensembles. *Appl. Soft Comput.* **2021**, *111*, 107689. [[CrossRef](#)]
28. Li, X.; Jiang, H.; Niu, M.; Wang, R. An enhanced selective ensemble deep learning method for rolling bearing fault diagnosis with beetle antennae search algorithm. *Mech. Syst. Signal Process.* **2020**, *142*, 106752. [[CrossRef](#)]
29. Zheng, X.; Chen, W.; You, Y.; Jiang, Y.; Li, M.; Zhang, T. Ensemble deep learning for automated visual classification using EEG signals. *Pattern Recognit.* **2020**, *102*, 107147. [[CrossRef](#)]
30. Thakkar, A.; Lohiya, R. Attack classification of imbalanced intrusion data for IoT network using ensemble learning-based deep neural network. *IEEE Internet Things J.* **2023**, *10*, 11888–11895. [[CrossRef](#)]
31. Akyol, K. Stacking ensemble based deep neural networks modeling for effective epileptic seizure detection. *Expert Syst. Appl.* **2020**, *148*, 113239. [[CrossRef](#)]
32. Mohammed, A.; Kora, R. A comprehensive review on ensemble deep learning: Opportunities and challenges. *J. King Saud Univ. - Comput. Inf. Sci.* **2023**, *35*, 757–774. [[CrossRef](#)]
33. Clemen, R.T. Combining forecasts: A review and annotated bibliography. *Int. J. Forecast.* **1989**, *5*, 559–583. [[CrossRef](#)]
34. Susmaga, R.; Szczech, I.; Brzezinski, D. Towards explainable TOPSIS: Visual insights into the effects of weights and aggregations on rankings. *Appl. Soft Comput.* **2024**, *153*, 111279. [[CrossRef](#)]
35. Wang, J.; Wang, L.; Cheng, H.; Runge, K. A comprehensive review on aeration methods used in flotation machines: Classification, mechanisms and technical perspectives. *J. Clean. Prod.* **2024**, *435*, 140335. [[CrossRef](#)]
36. Hastie, T.; Rosset, S.; Zhu, J.; Zou, H. Multi-class adaboost. *Stat. Its Interface* **2009**, *2*, 349–360. [[CrossRef](#)]
37. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [[CrossRef](#)]
38. 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.
39. Belgiu, M.; Drăguț, L. Random forest in remote sensing: A review of applications and future directions. *ISPRS J. Photogramm. Remote. Sens.* **2016**, *114*, 24–31. [[CrossRef](#)]
40. Wang, Q.; Huang, M.; Zhou, X. Feature selection in froth flotation for production condition recognition. *IFAC-Pap.* **2018**, *51*, 123–128. [[CrossRef](#)]
41. Kuncheva, L.I.; Bezdek, J.C.; Duin, R.P. Decision templates for multiple classifier fusion: An experimental comparison. *Pattern Recognit.* **2001**, *34*, 299–314. [[CrossRef](#)]
42. Rao, S.; Verma, A.K.; Bhatia, T. Hybrid ensemble framework with self-attention mechanism for social spam detection on imbalanced data. *Expert Syst. Appl.* **2023**, *217*, 119594. [[CrossRef](#)]
43. Di Nunno, F.; Giudicianni, C.; Creaco, E.; Granata, F. Multi-step ahead groundwater level forecasting in Grand Est, France: Comparison between stacked machine learning model and radial basis function neural network. *Groundw. Sustain. Dev.* **2023**, *23*, 101042. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.