



Article Deep Learning Based Feature Selection and Ensemble Learning for Sintering State Recognition

Xinran Xu¹ and Xiaojun Zhou^{2,*}

- School of Computer Science and Engineering, Central South University, Changsha 410083, China; xinranxu19@gmail.com
- ² School of Automation, Central South University, Changsha 410083, China
- Correspondence: tiezhongyu2010@gmail.com

Abstract: Sintering is a commonly used agglomeration process to prepare iron ore fines for blast furnace. The quality of sinter significantly impacts the blast furnace ironmaking process. In the vast majority of sintering plants, the judgment of sintering quality still relies on the intuitive observation of the cross section at sintering machine tail by operators, which is susceptible to the external environment and the experience of operators. In this paper, we propose a new sintering state recognition method using deep learning based feature selection and ensemble learning. First, features from the infrared thermal images of sinter cross section at the tail of the sinterer are extracted based on ResNeXt. Then, to eliminate the irrelevant, redundant and noisy features, an efficient feature selection method based on binary state transition algorithm (BSTA) is proposed to find the truly useful features. Subsequently, an ensemble learning (EL) method based on group decision making (GDM) is proposed to recognize the sintering states. Novel combination strategies considering the varying performance of the base learners are designed to further improve recognition accuracy. Industrial experiments conducted at a steel plant verify the effectiveness and superiority of the proposed method.

Keywords: sintering state; deep learning; feature selection; binary state transition algorithm; ensemble learning; group decision making

check for **updates**

Citation: Xu, X.; Zhou, X. Deep Learning Based Feature Selection and Ensemble Learning for Sintering State Recognition. *Sensors* 2023, 23, 9217. https://doi.org/10.3390/ s23229217

Academic Editor: Marcin Woźniak

Received: 18 September 2023 Revised: 8 November 2023 Accepted: 9 November 2023 Published: 16 November 2023



Copyright: © 2023 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

The blast furnace ironmaking process is currently the dominant process worldwide for providing raw materials for steelmaking. It is the main method of modern ironmaking, contributing to more than 95% of the world's total iron production. In the blast furnace ironmaking industry, iron ore plays an extremely important role as the primary feed material. As high-grade natural iron ore reserves decrease, artificial iron ore is becoming more and more critical [1]. Sintering is the primary way to produce high-quality artificial iron ore from low-grade iron ore, which has been widely used around the world. The production of high-quality sinter is crucial for assuring consistent, stable furnace productivity with a low consumption of reductants.

Estimation of the sintering state plays a vital role in improving the quality of sintering. The state of sintering depends on an important parameter known as burn-through point (BTP). It is a position located on the sinter strand, where the mixed materials thoroughly burn for the first time [2]. By judging the BTP, the sintering states can be recognized. The accurate recognition of sintering state is the guarantee of high-quality and high-yield sinter production, which is of great significance in improving sintering productivity and preventing equipment damage. Nowadays, many sintering plants still rely on experienced operators to observe the sintering machine tail to recognize sintering states with their knowledge and experience. Obviously, this method has many drawbacks. Manual judgment is often affected by many factors such as working experience, working environment, physical stamina, and emotions, resulting in potential errors.

Different from the traditional methods relying on operator judgments, methods based on computer vision depend on the sintering machine tail sectional images to recognize sintering states. Over the years, there has been a remarkable and rapid evolution in visionbased object recognition and is applied in many fields. Significant contributions have been made in various studies such as reconstructing and recognizing human motion with 3D motion analysis [3], supporting disabled individuals through computer-based gesture recognition [4], and exploring novel applications like merging anime with face recognition technology [5]. Similarly, computer vision-based recognition techniques are widely used in the iron ore sintering industry. A series of intelligent modeling methods based on computer vision are established for sintering state recognition. Sun T Q et al. proposed a online sintering quality prediction method based on machine vision and artificial neural network (ANN) [6]. Liu et al. proposed a BTP prediction system based on gradient boosting decision tree (GBDT) algorithm and decision rules [7]. A fuzzy neural network prediction model is presented by Wang et al. [8], in which the strand velocity determines the final model. Li Jiangyun et al. used generative adversarial network (GAN) to expand the sample of sintering data set, and proposed a classification model combining attention mechanism and ResNet [9].

However, many of the proposed methods for sintering state recognition still rely on conventional machine learning methods to extract features from images. In view of the digital image processing technology, a series of features that can comprehensively reflect the sintering state are extracted manually from the images of sinter cross section at the machine tail, such as the average brightness of the red fire layer. This method requires a considerable amount of engineering skills, domain expertise and effort. Deep learning can solve the drawbacks of manual feature extraction by automatically discovering the representations needed for recognition [10–12]. However, deep neural networks usually yield a large number of extracted features, including features that are irrelevant, redundant or even noisy for sintering state estimation [13]. These features can bring about high computational complexity and poor learning performance. In addition, it is difficult to determine the best model for recognition tasks without sufficient information, so the recognition accuracy can not be effectively guaranteed with a single model.

In order to deal with the above issues, a novel sintering state recognition method using deep learning based feature selection and ensemble learning is proposed in this paper. A deep neural network ResNeXt pre-trained on ImageNet is used to extract features from infrared thermal images of sinter cross section at the tail of the sintering machine. For the features extracted by deep learning, a feature selection method based on the binary state transition algorithm (BSTA) with a novel training strategy based on feature decomposition is proposed to select the optimal subset of features, which can greatly increase computational efficiency and build better generalization models. Finally, a new ensemble learning (EL) method is proposed to recognize sintering states based on the obtained optimal subset. Considering the limitations of an individual learner and shortcomings of the current EL schemes such as majority voting, the framework of group decision making (GDM) is introduced, which aims to find an optimal alternative considering various suggestions of decision-makers [14]. The main contributions of the proposed method are as follows:

- A feature extractor based upon ImageNet-pretrained ResNeXt50 is used to automatically extract fixed features from images of sinter cross section at sintering machine tail.
- An efficient feature selection method using the binary state transition algorithm (BSTA) and feature decomposition strategy is proposed to eliminate features that are irrelevant, redundant or even noisy, which not only reduces the difficulty of training but also enhances recognition accuracy.
- An ensemble learning framework based on group decision making (GDM) is put forward to further improve recognition accuracy, where new combination strategies are introduced to efficiently fuse base learners.

The remainder of the paper is organized as follows. The industrial background is introduced in Section 2. Section 3 describes the proposed method in detail. In Section 4,

experiments are conducted to verify the effectiveness of the proposed method. Section 5 concludes this paper.

2. Background

2.1. Sintering Process and Problem Description

Iron ore sintering is a crucial step in the ironmaking process [15]. Its primary goal is to produce a material with specific thermal, mechanical, physical, and chemical properties suitable for feeding into the blast furnace. The sintering process, depicted in Figure 1, involves the application of heat to fine iron ore particles, transforming them into coarser grains [16]. Firstly, various raw materials, including iron ore, coke, fluxes, and recycled sintered ore, are blended in a mixer to create a mixture, which is then transported to the mixture bunker. Subsequently, this mixture is evenly spread on a moving trolley to form a sintering bed [1]. The ignition hood is responsible for igniting the surface of the bed and preserving heat, while a blower located beneath the trolley creates a negative pressure zone below the bed through exhaust. As the trolley slowly moves towards the tail of the sintering machine, combustion progresses downward. Gradually, the raw ore powder transforms into sintered ore with specific particle sizes, becoming a valuable feedstock for subsequent stages of blast furnace ironmaking production.



Figure 1. The flowchart of sintering process and feedback control system.

The recognition of sintering state is a pivotal part in assessing sintering quality. Excessive sintering temperatures and prolonged sintering durations can lead to a decline in the ultimate product performance, a phenomenon known as oversintering. Conversely, when sintering temperatures are too low and sintering time is too short, the product fails to attain the necessary performance criteria, a condition referred to as undersintering. Obviously, different sintering states will result in products of varying quality, and only the normal sintering state is desired. Hence, it is necessary to develop an efficient method for accurately assessing the sintering states to ensure the production of high-quality sintered ore.

2.2. Image Acquisition

The tail of the sintering machine marks the conclusion of the sintering process, and it is an important location where a plethora of sintering parameters can be captured from cross-sectional images of the sinter. We deployed the Fluke TiX1000 infrared thermal imager positioned at the viewing port at the end of the sintering machine. This imager boasts a temperature measurement range spanning from -40 °C to 2000 °C and a high-resolution capability of 1024×768 pixels. This device was connected to a computer through a fiber optic network for seamless real-time transmission of the collected infrared thermal images.

The obtained images of sinter cross section at the tail of the sintering machine are shown in Figure 2.



Figure 2. Infrared thermal images of sinter cross section at the tail of the sintering machine.

3. Proposed Method

3.1. Overview of the Proposed Method

This section illustrates the proposed sintering state recognition method in details. The overall framework of the proposed method is shown on Figure 3. This framework includes three modules: feature extraction, feature selection, and sintering state recognition based on ensemble learning.



Figure 3. Framework of the proposed method.

Traditional feature extraction methods based on computer vision suffer from drawbacks of high cost and low efficiency. In this paper, the features from the images of sinter cross section at the sintering machine tail are extracted by a feature extractor built on a ResNeXt50 architecture that has been pretrained on the ImageNet dataset using transfer learning. Currently, ImageNet transfer is the most effective and economical choice. Representations from deep networks are generic and can support transfer learning across domains [17,18].

The number of features obtained is often very large after deep learning-based image feature extraction. Such high-dimensional data is likely to contain redundant, irrelevant, or even noisy features, which significantly increase the training difficulty of the recognition model and lead to reduced recognition accuracy. Feature selection can effectively tackle this problem by identifying the optimal features from the original feature set. This paper proposes a new feature selection method using binary state transition algorithm. A novel training strategy based on feature decomposition is introduced to reduce the computation time of BSTA and the number of selected features.

Following the data preprocessing steps of feature extraction and feature selection, the optimal features that encapsulate the key information from the sinter cross section images are obtained. The sintering state is recognized utilizing this feature set. Considering the limitations of a single individual model, this paper proposes a novel ensemble learning method based on group decision making to judge sintering states. Novel base learner combination strategies that integrate multiple evaluation metrics are put forward to address the shortcomings existing in previous ensemble learning methods.

3.2. Feature Extraction Based on Deep Learning

Conventional machine learning methods have limitations when it comes to processing raw natural data. In the case of sintering image recognition, traditional techniques of extracting features are able to acquire shallow features including the red fire layer area, the average brightness of the red fire layer, etc. But this method requires careful engineering and considerable domain expertise to manually design a feature extractor that converts the raw data into a suitable representation [10]. In fact, this can be circumvented by deep learning which automatically learns useful features through a generic learning procedure. The ResNet-based architecture is a widely adopted deep neural network in image recognition [19]. By introducing the residual connections to the CNN [20], ResNet is capable of solving the degradation problem in very deep networks and improving model performance. ResNeXt [21] was proposed as a variant of ResNet with the redesigned building blocks. It develops an aggregated transformation strategy combining the block stack strategy of ResNet with group convolution techniques of the Inception architecture. A ResNeXt block follows the split-transform-merge strategy. Instead of performing a single transformation as in the ResNet block, it performs a series of transformations. Highly competitive recognition performance is achieved in this way without increasing the complexity of the model.

In this work, the features from the images of sinter cross section at the sintering machine tail are extracted by a feature extractor built upon a ResNeXt50(32x4d) architecture that has been pretrained on the ImageNet dataset. In experiments conducted on Imagenet datasets, ResNeXt achieved high accuracy with a relatively low number of flops compared to existing models. Using an pretrained network as a fixed feature extractor is a form of transfer learning [22], which leverages the model's prior learning of general visual features in a large dataset. Such a property can be very important in fields such as iron ore sintering, where the availability of labels involves extended effort and cost for acquisition. By adopting this strategy, the time-intensive training process is circumvented, saving significant time and computational resources while enhancing performance with a strong capability to capture meaningful features. In each layer of the ResNeXt, there is a new representation of the input image by progressively extracting meaningful information. In this work, the final linear layer has been removed and the features from the last average pooling layer are selected. The architecture of the ResNeXt-based feature extractor is shown in Figure 4. The features extracted by the neural network can be called deep features which contain much more information than the shallow features.

3.3. Feature Selection Based on BSTA

Feature selection is an important step in selecting the most relevant features from the original set, which helps to enhance a model's focus on key information and reduce computational costs. For classification tasks, feature section can greatly improve classification accuracy. In this section, a novel feature selection method based on binary state transition algorithm with feature decomposition strategy is proposed.



Figure 4. The architecture of the proposed feature extractor based on ResNeXt50.

3.3.1. Overview of Feature Selection

Suppose S is a dataset containing *M* samples and *N* features. Feature selection aims to find the best feature subset that contains n ($n \le N$) features that can maximize the classification accuracy while using as few features as possible [23]. In this work, we represent a solution to the feature selection problem using a binary encoding vector denoted as **x**. The encoding vector is described as follows:

$$\mathbf{x} = (x_1, x_2, \dots, x_n), x_i \in \{0, 1\}, i = 1, 2, \dots, n$$
(1)

where $x_i = 1$ means that the *i*th feature is selected, while $x_i = 0$ indicates that the feature is not selected. For instance, Figure 5 illustrates a 9-dimensional vector **x**. The solution $\mathbf{x} = [1, 0, 0, 1, 0, 1, 1, 0, 0]$ corresponds to the selection of 1st, 4th, 6th, and 7th features.



Figure 5. A binary encoding 9-dimensional vector x.

The feature selection problem can be represented as the follows:

$$\max \quad f_1 = Acc(\mathbf{x}) \\ \min \quad f_2 = \|\mathbf{x}\|_0$$
s.t. $x = (x_1, x_2, \dots, x_n), x_i \in \{0, 1\}, i = 1, 2, \dots, n, 1 \leq \|\mathbf{x}\|_0 \leq n$
(2)

where $Acc(\mathbf{x})$ denotes the classification accuracy of the model established based on the corresponding \mathbf{x} . $\|\mathbf{x}\|_0$ represents the number features selected from \mathbf{x} .

3.3.2. Feature Selection Based on BSTA and Feature Decomposition

In this study, a feature selection method using binary state transition algorithm (BSTA) is proposed for higher classification accuracy and shorter training time. The state transition algorithm (STA) was firstly proposed by Zhou (the co-author of this paper) [24] for continuous optimization problems, which has demonstrated outstanding performance in real-world applications. STA was inspired by state space representation from control theory. It is an individual-based optimization algorithm whose main idea is to produce candidate solutions using four intelligent search operators. These candidates are evaluated,

and the algorithm chooses the current best solution, which is extremely efficient in practical applications for the search for the global optimal solution. In STA, a candidate solution is defined as a state, and changes in solutions are described as state transition. To solve integer optimization challenges, the discrete state transition algorithm (DSTA) is proposed as a discrete variant of STA [25]. Binary State Transition Algorithm (BSTA) is a novel intelligent optimization method designed for boolean integer optimization problems, representing the binary adaptation of the DSTA.

Four specialized state transformation operators have been introduced in BSTA to generate candidates for both local and global search.

>

1

(1) Swap transformation:

$$\kappa_{k+1} = A_k^{swap}(m_a)\mathbf{x}_k \tag{3}$$

where, $A_k^{swap} \in \mathbb{R}^{n \times n}$ represents a swap transformation matrix, and m_a is a swap transformation factor which is a constant integer used to control the maximum number of swaps. Swap transformation is illustrated in Figure 6.

(2) Shift transformation:

$$\mathbf{x}_{k+1} = A_k^{shift}(m_b)\mathbf{x}_k \tag{4}$$

where, $A_k^{shift} \in \mathbb{R}^{n \times n}$ is a shift transformation matrix, and m_b is a shift transformation factor which is a constant integer used to control the maximum length of the moved position. Shift transformation is illustrated in Figure 7.

(3) Symmetry transformation:

$$\mathbf{x}_{k+1} = A_k^{sym}(m_c)\mathbf{x}_k \tag{5}$$

where, $A_k^{sym} \in \mathbb{R}^{n \times n}$ is a symmetry transformation matrix, and m_c is a symmetry transformation factor which is a constant integer used to control the maximum length of symmetric sequence. Symmetry transformation is illustrated in Figure 8.

(4) Substitute transformation:

$$\mathbf{x}_{k+1} = A_k^{sub}(m_d)\mathbf{x}_k \tag{6}$$

where, $A_k^{sub} \in \mathbb{R}^{n \times n}$ is a substitute transformation matrix, and m_d is a substitute transformation factor which is a constant integer used to control the maximum number of substitute. Substitute transformation is illustrated in Figure 9.



Figure 6. Illustration of swap transformation.



Figure 7. Illustration of shift transformation.



Figure 8. Illustration of symmetry transformation.



Figure 9. Illustration of substitute transformation.

Algorithm 1 provides the procedure of binary state transition algorithm. swap(.), shift(.), symmetry(.) and substitute(.) are transformation operator functions. In each iteration, these operator functions are performed to generate the candidate solutions. *Best* represents the best candidate solution so far. The algorithm continues until the predefined maximum number of iterations is reached, and the termination condition is satisfied. The current best solution is then returned.

Algorithm 1 Binary State Transition Algorithm.

1:	function BSTA(s)
2:	repeat
3:	$Best \leftarrow swap(Best, *)$
4:	$Best \leftarrow shift(Best, *)$
5:	$Best \leftarrow symmetry(Best, *)$
6:	$Best \leftarrow substitute(Best, *)$
7:	until the termination condition is met
8:	Return Best
9:	end function

A hybrid feature selection method called ReliefF-BSTA was proposed in pursuit of higher classification accuracy and lower computational resources [26]. ReliefF-BSTA is a hybrid feature selection method that combines the strengths of the feature weighting algorithm ReliefF and the intelligent optimization approach BSTA. In this method, ReliefF narrows down the search space and provides valuable insights into features, and BSTA searches for the optimal feature subset on the basis of feature ranking and feature weights. ReliefF is a well-known filter-based feature selection method, which seeks the best feature subset by computing the features' weights [27]. It assigns different weights to features according to the correlations between features and categories. Features with weights greater than a predefined threshold will be selected. In multi-label problems, assuming that the labels of the training dataset are $L = \{l_1, l_2, \dots, l_n\}$, the ReliefF randomly selects a sample R_i from the training dataset. Then it searches for k nearest neighbors (called near Hits) of R_i with the same label, referred to as H_i (j = 1, 2, ..., k). It also searches for k nearest neighbors (called near Misses) of R_i with different label, which is denoted by $M_i(l)(j = 1, 2, ..., k)$. This process is iterated a total of *m* times by ReliefF. The weight assigned to feature *X* is updated as follows:

$$W(X) = W(X) - \sum_{j=1}^{k} \frac{\operatorname{diff}(X, R_i, H_j)}{(m)(k)} + \sum_{\substack{l \notin \operatorname{class}(R)}} \frac{\left[\frac{p(l)}{1 - p(\operatorname{class}(R))} \sum_{j=1}^{k} \operatorname{diff}(X, R_i, M_j(l))\right]}{(m)(k)}$$

$$(7)$$

where *m* is the number of iterations and $diff(X, R_i, R_j)$ denotes the difference of samples R_i and samples R_j on feature *X*, which is defined as:

$$\operatorname{diff}(X, R_i, R_j) = \begin{cases} \frac{|R_i[X] - R_j[X]|}{\max(X) - \min(X)} & X \text{ is continuous} \\ 0 & X \text{ is discrete} \\ & \text{and } R_i[X] = R_j[X] \\ 1 & X \text{ is discrete} \\ & \text{and } R_i[X] \neq R_j[X] \end{cases}$$
(8)

Considering the industrial demand for high efficiency, feature selection, as a crucial step in data preprocessing, should prioritize shorter processing time. Recognizing that it takes a lot of time to traverse the complete dataset with BSTA in each iteration, especially when dealing with large datasets in the feature selection problem, this paper proposes a new training strategy called feature decomposition (FD). The dataset is divided into several parts s_1, s_2, \ldots, s_n based on features, and each part represents a sub-dataset containing a portion of the features in the original training set. Training is then performed separately for each sub-dataset and the corresponding solutions X_1, X_2, \ldots, X_N are obtained. Upon completing the training of all sub-datasets, all solutions are merged and formed into a new dataset termed the final sub-dataset. To derive the ultimate optimal solution, training is conducted on the final sub-dataset. Figure 10 shows the process of the FD strategy.



Figure 10. Feature decomposition when data is decomposed into three sub-datasets.

Our feature selection method comprises a two-stage process. First, the filter method ReliefF is used to narrow the search space and calculate feature weights and feature rankings. Next, the wrapper method BSTA searches for the best feature subset containing the most useful, relevant, and non-redundant features leveraging the important information provided by ReliefF. Feature decomposition strategy is proposed to greatly reduce both the number of selected features and the overall running time while maintaining high classification accuracy. The pseudocode of our proposed feature selection method is illustrated in Algorithm 2.

Algorithm 2	Pseudocode of ReliefF-BS	A with feature	decomposition.
-------------	--------------------------	----------------	----------------

1:	Input: Dataset D
2:	$\{D_1, D_2, \ldots, D_n\} \leftarrow Decompose(D)$
3:	for each $i \in [1, n]$ do
4:	Calculate features ranking and feature weights using ReliefF
5:	Get the initial s_i solution for D_i
6:	$Best_i \leftarrow BSTA(s_i)$
7:	end for
8:	$s \leftarrow Concat(Best_1, \dots, Best_n)$
9:	$Best^* \leftarrow BSTA(s)$
0:	Return Best*

3.4. Proposed Ensemble Learning Method Based on Group Decision Making

This section introduces how to perform recognition tasks based on the feature subset obtained by feature selection. Considering the limitations of a single model and the shortcomings existing in traditional ensemble learning methods, we propose a novel ensemble learning method based on group decision making framework with novel base learner combination strategies.

3.4.1. Group Decision Making

It is common for us to consider multiple suggestions when making a decision. The process of considering and integrating recommendations from peers or experts into the decision-making process is referred to as group decision making (GDM). GDM aims to determine the best alternative among a set of options, taking into account recommendations from multiple advisors [28].

In a GDM problem with a total of *m* alternatives, we denote these alternatives as Q_1, Q_2, \ldots, Q_m , which represent the available options. There are *n* decision criteria, called C_1, C_2, \ldots, C_n [29]. A set of *K* decision makers(DMs) E_1, E_2, \ldots, E_K are assembled, and each decision maker evaluates every alternative independently according to different criteria. The decision making scores generated from DM *k* constitute the decision matrix D_k :

$$\begin{bmatrix} d_{11}^k & \dots & d_{1j}^k & \dots & d_{1m}^k \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{i1}^k & \dots & d_{ij}^k & \dots & d_{im}^k \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{n1}^k & \dots & d_{ni}^k & \dots & d_{nm}^k \end{bmatrix}$$
(9)

where d_i^k (i = 1, 2, ..., m; j = 1, 2, ..., K) is called performance rating, which represents the assessment provided by decision maker E_k for alternative Q_i regarding criterion C_j . The performance rating d_{ij}^k serves as a metric to quantify the extent to which Q_i satisfies C_j from the perspective of E_k .

The important weight of each criterion and each decision maker can be denoted as $w_1, w_2, ..., w_n$ and $W_1, W_2, ..., W_K$ respectively. Performance ratings and important weights range from 0 to 1. Decision matrices $D_1, D_2, ..., D_k$ are aggregated with important weights in the decision-making process. In this study, GDM is utilized to combine the results generated by the base learners in ensemble learning.

3.4.2. Ensemble Learning Based on GDM

Ensemble learning is often taken as the embodiment of crowd intelligence in machine learning and typically exhibits superior generalization capabilities when compared to individual models [30,31]. It refers to the process of generating and combining multiple base learners in order to complete a specific task. In this work, a new ensemble learning

method based on GDM (GDM-EL) with novel combination strategies is proposed. An ensemble learning method can be seen as a decision-making process, where multiple individual learners act as decision makers to collectively determine the final result.

Under the framework of GDM-EL, base learners can be considered as decision makers (DMs), and various categories can be treated as alternatives [32]. For a multi-classification problem with *m* categories and *k* individual learners, a group of *K* decision makers and a set of *m* alternatives are created. Then the performance ratings generated by each of the *K* decision makers are collected and organized into matrices D_k . In this paper, we use the membership degree (MD) as the criterion to reflect the performance of the alternatives [33], which refers to the degree to which an element belongs to a particular set or category in fuzzy logic and fuzzy set theory. The decision matrix D_K produced by decision maker *k* is described as:

$$D_k = [d_1^k, d_2^k, \dots, d_m^k] \qquad (k = 1, 2, \dots, K)$$
(10)

where the performance rating d_i^k (i = 1, 2, ..., m; j = 1, 2, ..., K) is a numerical value between 0 and 1 that indicates the extent to which a sample belongs to Q_i for decision maker E_k .

When *K* base learners are used in GDM-EL and they evaluate each category based on the membership degree, the important weight assigned to each of them can be represented as $W = [W_1, W_2, \dots, W_K]$. The weights of decision makers are important to combine multiple decision matrices into a single decision matrix. Among all the combination strategies in ensemble learning, the majority vote is by far the simplest for implementation, which gives every base learner the same weight and derives single ground truth labels from multiple base learners [34]. This voting method treats each voter as a completely equal individual, ignoring the different performance of each base learner. However, the classification ability of each classifier for each category differs from one another. Hence, it is essential to incorporate the classification performance of each base learner as its weight in the final decision, necessitating a quantitative measure of their classification power. Precision, recall, and accuracy are metrics that effectively evaluate the performance of a classification method. In this study, instead of relying on a single metric, the performance is assessed by considering a combination of precision, recall, and accuracy measures. By utilizing these multiple metrics, a more comprehensive evaluation of the base learners' performance is achieved. The precision, recall and accuracy are presented below.

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$R = \frac{IP}{TP + FN} \tag{12}$$

$$A = \frac{Cr}{Total}$$
(13)

where *P*, *R*, *A* are the precision, recall and accuracy; *Cr* is the number of correct predictions, and *Total* is the total number of predictions. A class could be identified as a positive class and the rest of the classes belong to negative classes. *TP*, *FP*, *FN* can be found in in Table 1.

Table 1. Classification result confusion matrix.

Terrth	Pred	iction
IIuui	Positive Class	Negative Class
Positive Class	TP	FN
Negative Class	FP	TN

Accuracy is the most commonly used index to measure the performance of a classification model on a dataset. The accuracy obtained by classification method k is represented as A_k . Precision and recall are designed for a certain category, which are used

to evaluate the model's performance for that category specifically. For each classification method, if the dataset has a total of *m* categories, we calculate the precision and recall of the classification method *k* for each label, then two vectors can be obtained as $[p_1, p_2, ..., p_m]^T, [r_1, r_2, ..., r_m]^T$. In this study, we compute the average of each vector and denote them as P_k, R_k , calculated as:

$$P_k = \frac{1}{m} \sum_{i=1}^m p_i$$
 (14)

$$R_k = \frac{1}{m} \sum_{i=1}^m r_i \tag{15}$$

where p_i, r_i represents the precision and recall of classification method *k* for category *i*. For a classification problem with *K* base learners, three vectors of length *K* can be obtained as above, respectively described as $P = [P_1, P_2, ..., P_K], R = [R_1, R_2, ..., R_K], A = [A_1, A_2, ..., A_K].$

Evaluating the overall performance of a model requires consideration of all relevant metrics. Specifically, precision, recall and accuracy should be appropriately integrated to collectively assign weights to the base learners and achieve the best possible outcome. In this paper, three base learner combination strategies are proposed, which combine *P*, *R*, *A* in three different ways to generate weights for decision makers. The combination strategies are inspired by arithmetic mean, geometric mean, and harmonic mean respectively. The 3 weights for DM *k* are W_k^1, W_k^2, W_k^3 , defined as follows:

$$W_k^1 = P_k + R_k + A_k \tag{16}$$

$$W_k^2 = P_k R_k A_k \tag{17}$$

$$W_k^3 = \frac{3}{\frac{1}{P_k} + \frac{1}{R_k} + \frac{1}{A_k}}$$
(18)

From Equations (16)–(18), it is obvious that the larger the value of the three indicators, the greater the weight of this base learner. Finally, the score for each option is based on the ratings from all base learners, and these ratings are combined to derive the final score, considering their respective weights. The alternative Q_j with the highest score is the final alternative Q^* , shown as follows:

$$Q^* = Q_{\arg\max\sum_{i=1}^{K} W_k d_j^k}$$
(19)

Our method GDM-EL is also compared with the majority vote EL method in Table 2. It is evident that the majority vote adds the votes of each base learner linearly, while GDM-EL performs a weighted combination of the base learners' results.

Table 2. Comparison of GDM-EL and majority vote.

Method	Calculation Formula
Majority vote	$Q^* = Q_{rg\max_i \sum_{k=1}^{K} rac{1}{K} \cdot d_j^k}$
GDM-EL	$Q^* = Q_{rg\max_j \sum_{k=1}^{K} W_k d_j^k}$

The flowchart for GDM-EL is shown in Figure 11. The dataset is divided into training set, validation set and testing set. Several models are generated on the training set to create base learners. Then the base learners are tested on the validation set to obtain performance indexes based on *P*, *R*, *A*. According to these knowledge, the important weights of the base learners can be acquired. Next, for the testing set, GDM is introduced to fuse outputs from

these base learners on the testing set. Base learners give each alternative evaluating scores known as performance ratings, and the decision matrices are structured according to these performance ratings. Finally, the best alternative is chosen based on the weight of base learners and decision matrices.

Figure 11. The flowchart for ensemble learning method based on group decision making.

4. Experiments and Results

In order to verify the effectiveness of the recognition method proposed above, we conducted industrial experiments on the infrared thermal images of sinter cross section collected at the tail of the sintering machine. The original data came from a sintering plant in Hunan, China. The experimental results demonstrate the effectiveness and feasibility of the proposed method.

The ImageNet-pretrained ResNeXt50 model is used as a fixed feature extractor to extract features from images captured at the sintering machine tail with PCA to further reduce the dimensionality of the data. The architecture of the ResNext-based feature extractor is shown in Figure 4, where fixed feature representations are extracted. Next, a based feature selection method with feature decomposition strategy chooses the best features from those initially extracted by ResNeXt50. Table 3 shows the detailed parameter settings of the feature selection algorithm based on BSTA. *Population* represents the total number of iterations. Since the BSTA belongs to the individual-based algorithm, *SE* represents the number of generated candidate solutions per iteration. (*pct, p,* and *q*) are user-specified parameters to control the generation of the initial solution [26]. The number of sub-datasets during feature decomposition and the total number of rounds for feature selection are also presented. Table 4 shows the specific information of the dataset we obtain through feature extraction and feature selection.

In this work, 5 common classification methods are employed as base learners of the EL method to judge the sintering quality based on the dataset described in Table 4. The base learners are Support Vector Machine(SVM) [35], Adaboost [36], Logistic Regression(LR) [37], k-Nearest Neighbor(KNN) [38], Random Forest(RF) [39]. On the basis of the optimal dataset after feature selection, these classical machine learning methods, serving as base learners for ensemble learning, can efficiently and rapidly yield satisfactory recognition results. In this paper, they are employed to ensure the diversity of base learners. Moreover, seven real-world datasets are used to compare the three base learner combination strategies in Equations (16)–(18). These public datasets are all available at the UCI Machine Learning Repository, namely: Letter Image Recognition Data, Blocks Classification, Dry Beans Dataset, Musk, ISOLET (Isolated Letter Speech Recognition), Pen-Based Recognition of Handwritten Digits, Waveform Database Generator. Details about these datasets can be found in Table 5. Each dataset is split into three subsets: a training set, a validation set, and a testing set. The training set is used to generate base learners. The validation set is applied to generate the priori knowledge of base learners from which the weights of each base learner can be calculated by three combination strategies. Next, for each sample in

the testing set, each decision maker assigns a score for every alternative, and 5 decision matrices can be obtained as Equation (10). At last, GDM is used to fuse the information and determine the final label, and results are calculated as Equation (19).

Table 3. Parameter settings for feature selection based on BSTA.

Population	SE	pct	р	q	Number of Sub-Datasets	Total Rounds
50	20	0.5	0.9	0.1	4	10

Table 4. Description of the sintering dataset obtained by feature extraction and feature selection.

Initial Features Extracted by ResNeXt	Features Selected Using BSTA	Samples	Classes
512	133	1651	3

Table 5. Description of public datasets.

Datasets	Samples	Classes	Attributes
Letter image	20,000	26	16
Blocks Classification	5473	5	10
Dry Bean	13,611	7	17
Musk	6598	2	168
Isolet	7797	26	617
Pen Digits	10,992	10	16
Waveform	5000	3	21

Based on the idea of decision-making, each dataset chooses the best combination strategy according to the classification result on the testing set. Finally, the best combination strategy to generate weights can be selected by voting. The results are listed in Table 6, which presents the classification accuracy (%) for each combination strategy on each dataset. It can be seen from the results that $W^2 = P_k R_k A_k$ achieves the highest accuracy on more datasets than the other two strategies.

Table 6. Classification accuracy(%) for each combination strategy on different datasets.

Datasets	$W_k^1 = P_k + R_k + A_k$	$W_k^2 = P_k R_k A_k$	$W_k^3=rac{3}{rac{1}{P_k}+rac{1}{R_k}+rac{1}{A_k}}$
Letter image	87.863	88.338	87.875
Blocks	94.386	95.753	95.345
Dry Bean	90.046	90.689	90.284
Musk	97.120	96.139	96.970
Isolet	93.716	93.844	92.885
Pen Digits	96.884	97.100	96.884
Waveform	85.750	85.600	86.150

Based on the above results, in our experiment of sintering state recognition, we choose $W^2 = P_k R_k A_k$ as our combination strategy to assign weights to the base learners. The dataset obtained by feature extraction and feature selection from sintering images (described in Table 4) is divided into training set, validation set and testing set. The weight of each decision maker is calculated based on the validation set with combination strategy $W = P_k R_k A_k$. The weights of base learners for ensemble learning are shown in Table 7.

0						
	Adaboost	RF	KNN	LR	SVM	
Non-FS	0.316848	0.23589	0.067454	0.125859	0.253953	-
FS-BSTA	0.25157	0.23587	0.121407	0.173229	0.217915	

Table 7. Weights of base learners.

'Non-FS' shows the weights of base learners without feature selection; 'FS-BSTA' shows the weights with feature selection method based on BSTA.

The final results are calculated as Equation (19). Figure 12 presents the recognition performance of the base learners, the EL method based on majority vote, and our proposed EL method based on GDM (GDM-EL). The effectiveness of the feature selection method based on BSTA is also evaluated in Figure 12. It is evident in Figure 12 that our proposed method demonstrates outstanding performance in sintering state recognition with a recognition accuracy of 97.579%. Specifically, GDM-EL exhibits superior recognition accuracy compared to both the majority vote and individual base learners. Additionally, our feature selection method based on BSTA proves to be effective in enhancing the recognition performance of both base learns, Majority vote, and GDM-EL. The accuracy of each method increases significantly after feature selection. Figure 13a plots the confusion matrix of GDM-EL method without feature selection. It can be found that for each category, the number of correctly recognized samples is increased after BSTA-based feature selection. Figure 13b shows that the proposed method demonstrates excellent recognition performance for each category.

Figure 12. Ablation experiment: Recognition performance of EL methods and base learners. The results above are based on features extracted by ResNeXt50. The recognition accuracy without feature selection and the recognition accuracy with feature selection method based on BSTA are in orange and blue respectively.

To further verify the superiority of the proposed method, we also conducted experiments with popular recognition methods based on deep learning. We examined the recognition performance in two settings: (1) training a VGG16 [40], Inception-v3 [41], ResNet50 and ResNeXt50 model from scratch with randomly initialized weights, (2) finetuning the ImageNet pre-trained ResNeXt50 model. These settings are commonly used in deep learning and transfer learning for training end-to-end models. In the first setting, we examined VGG16, Inception-v3, ResNet50 and ResNeXt50 trained from random initialization using the sinter cross section images and labels for 200 epochs, with a cosine decay learning rate schedule at a batch size of 64. In the second setting, we initialized ResNeXt50 from the ImageNet weights and fine-tuned using a similar training setup. The recognition results are presented in Table 8, where the accuracy and F1-score are reported. As can be

16 of 18

observed, our proposed method achieves the highest accuracy and F1-score. The outcome demonstrates the superior effectiveness of our method compared to popular deep learning models that rely solely on supervised training for sintering state recognition.

Figure 13. Confusion matrices of GDM-EL for sintering state recognition. (**a**) GDM-EL without feature selection; (**b**) GDM-EL with fearure selection.

Table 8. Performance comparison of different methods for sintering state recognition.

	VGG16	Inception-v3	ResNet50	ResNeXt50	Fine-Tuning	Proposed Method
Accuracy(%)	85.714	88.377	87.409	90.073	92.252	97.579
F1-score	0.820	0.857	0.834	0.885	0.911	0.975

Based on the experimental results presented above, it is evident that the proposed method in this paper achieves accurate recognition of the sintering states. Utilizing a feature extractor built upon the ImageNet-pretrained ResNeXt50, relevant features from infrared thermal images are efficiently extracted. Significantly, the BSTA-based feature selection technique and the ensemble learning method based on GDM contribute to a remarkable advancement in enhancing the accuracy of sintering image recognition.

5. Conclusions

In this study, a new sintering state recognition method using deep learning-based feature selection and ensemble learning is proposed to improve the quality of sintered products and reduce energy consumption. A pretrained deep neural network ResNeXt50 is used to extract features from the infrared thermal images of the sinter cross section at sintering machine tail. To handle the high-dimensional features, a feature selection method using binary state transition algorithm and feature decomposition strategy is proposed. With the acquired feature subset, a novel ensemble learning method based on group decision making with novel base learner combination strategies is proposed to estimate sintering states.

The results of industrial experiments showed that the proposed method efficiently recognizes the sintering states with an impressive accuracy of 97.579%. Pretrained ResNeXt is an effective architecture for constructing feature extractors. Feature selection method based on BSTA has proved effective in providing a useful feature subset for higher accuracy and efficiency. The ensemble learning method based on GDM achieves higher accuracy than individual learners and conventional EL methods. Additionally, the proposed method for sintering state recognition demonstrates better performance compared with popular deep learning models. In future work, we will explore more evaluating criteria in the decision-making process and use more advanced machine learning algorithms to enhance the interpretability and applicability of our methods. We believe that the impact of our research extends beyond immediate application in industrial image recognition, with the

potential to have broader applications in other vision-based recognition tasks and diverse areas like object detection and autonomous systems.

Author Contributions: Conceptualization, X.X. and X.Z.; methodology, X.Z.; software, X.X.; validation, X.X. and X.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.Z.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; 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 is unavailable.

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

References

- 1. Du, S.; Wu, M.; Chen, L.; Pedrycz, W. Prediction model of burn-through point with fuzzy time series for iron ore sintering process. *Eng. Appl. Artif. Intell.* **2021**, *102*, 104259. [CrossRef]
- 2. Du, S.; Wu, M.; Chen, L.; Zhou, K.; Hu, J.; Cao, W.; Pedrycz, W. A fuzzy control strategy of burn-through point based on the feature extraction of time-series trend for iron ore sintering process. *IEEE Trans. Ind. Inform.* **2019**, *16*, 2357–2368. [CrossRef]
- 3. Kerdvibulvech, C.; Yamauchi, K. 3D Human Motion Analysis for Reconstruction and Recognition. In *Articulated Motion and Deformable Objects (AMDO)*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2014; Volume 8563.
- Kerdvibulvech, C. A review of computer-based gesture interaction methods for supporting disabled people with special needs. In Proceedings of the ICCHP 2016, Linz, Austria, 13–15 July 2016; pp. 503–506.
- 5. Amasiri, P.; Kerdvibulvech, C. Anime face recognition to create awareness. Int. J. Inf. Technol. 2023, 15, 3507–3512.
- 6. Sun, T.Q.; Yin, Y.X.; Dong, J.; Tu, X. Sintering quality prediction based on machine vision and ANN. *Comput. Eng.* **2008**, *34*, 240–242.
- Liu, S.; Lyu, Q.; Liu, X.; Sun, Y.; Zhang, X. A prediction system of burn through point based on gradient boosting decision tree and decision rules. *ISIJ Int.* 2019, 59, 2156–2164. [CrossRef]
- 8. Wang, J.; Li, X.; Li, Y.; Wang, K. BTP prediction of sintering process by using multiple models. In Proceedings of the 26th Chinese Control and Decision Conference (2014 CCDC), Changsha, China, 31 May–2 June 2014; pp. 4008–4012.
- 9. Li, J.Y.; Yang, Z.F.; Zheng, J.F. Applications of iron and steel industry with deep learning technologies. Iron Steel 2021, 56, 43.
- 10. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef]
- 11. Abdou, M.A. Literature review: Efficient deep neural networks techniques for medical image analysis. *Neural Comput. Appl.* **2022**, *34*, 5791–5812. [CrossRef]
- Samek, W.; Montavon, G.; Lapuschkin, S.; Anders, C.J.; Müller, K.R. Explaining deep neural networks and beyond: A review of methods and applications. *Proc. IEEE* 2021, 109, 247–278. [CrossRef]
- 13. Khaire, U.M.; Dhanalakshmi, R. Stability of feature selection algorithm: A review. J. King Saud Univ.-Comput. Inf. Sci. 2022, 34, 1060–1073. [CrossRef]
- 14. Zhang, Z.; Gao, Y.; Li, Z. Consensus reaching for social network group decision making by considering leadership and bounded confidence. *Knowl.-Based Syst.* 2020, 204, 106240. [CrossRef]
- 15. Du, S.; Wu, M.; Chen, X.; Cao, W. An intelligent control strategy for iron ore sintering ignition process based on the prediction of ignition temperature. *IEEE Trans. Ind. Electron.* **2019**, *67*, 1233–1241. [CrossRef]
- Fernández-González, D.; Ruiz-Bustinza, Í.; Mochón, J.; González-Gasca, C.; Verdeja, L.F. Iron ore sintering: Process. *Miner. Process. Extr. Metall. Rev.* 2017, 38, 215–227. [CrossRef]
- 17. Tammina, S. Transfer learning using vgg-16 with deep convolutional neural network for classifying images. *Int. J. Sci. Res. Publ.* (*IJSRP*) **2019**, *9*, 143–150. [CrossRef]
- 18. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H. A comprehensive survey on transfer learning. *Proc. IEEE* 2020, 109, 43–76. [CrossRef]
- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
- Wen, S.; Chen, J.; Wu, Y.; Yan, Z.; Cao, Y.; Yang, Y.; Huang, T. CKFO: Convolution kernel first operated algorithm with applications in memristor-based convolutional neural network. *IEEE Trans. Comput.-Aided Des. Integr. Circuits Syst.* 2021, 40, 1640–1647. [CrossRef]

- 21. Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; He, K. Aggregated residual transformations for deep neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1492–1500.
- 22. Weiss, K.; Khoshgoftaar, T.M.; Wang, D.D. A survey of transfer learning. J. Big Data 2016, 3, 1–40. [CrossRef]
- Zhang, Y.; Gong, D.; Gao, X.; Tian, T.; Sun, X.Y. Binary differential evolution with self-learning for multi-objective feature selection. *Inf. Sci.* 2020, 507, 67–85. [CrossRef]
- 24. Zhou, X.; Yang, C.; Gui, W. State transition algorithm. J. Ind. Manag. Optim. 2012, 8, 1039–1056. [CrossRef]
- 25. Zhou, X.; Gao, D.Y.; Yang, C.; Gui, W. Discrete state transition algorithm for unconstrained integer optimization problems. *Neurocomputing* **2016**, *173*, 864–874. [CrossRef]
- 26. Huang, Z.; Yang, C.; Zhou, X.; Huang, T. A hybrid feature selection method based on binary state transition algorithm and RELIEF. *IEEE J. Biomed. Health Inform.* **2018**, 23, 1888–1898. [CrossRef]
- Kononenko, I. Estimating attributes: Analysis and extensions of RELIEF. In Proceedings of the European Conference on Machine Learning, Catania, Italy, 6–8 April 1994; Volume 94, pp. 171–182.
- 28. Yu, D.; Chen, C.L.P.; Xu, H. Intelligent decision making and bionic movement control of self-organized swarm. *IEEE Trans. Ind. Electron.* **2020**, *68*, 6369–6378. [CrossRef]
- Zavadskas, E.K.; Mardani, A.; Turskis, Z.; Jusoh, A.; Nor, K.M. Development of TOPSIS method to solve complicated decisionmaking problems—An overview on developments from 2000 to 2015. *Int. J. Inf. Technol. Decis. Mak.* 2016, 15, 645–682. [CrossRef]
- 30. Sagi, O.; Rokach, L. Ensemble learning: A survey. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. 2018, 8, e1249. [CrossRef]
- 31. Dong, X.; Yu, Z.; Cao, W.; Shi, Y.; Ma, Q. A survey on ensemble learning. Front. Comput. Sci. 2020, 14, 241–258. [CrossRef]
- 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]
- Luo, H.; Li, L.; Zhang, Y.; Fang, S.; Chen, X. Link prediction in multiplex networks using a novel multiple-attribute decisionmaking approach. *Knowl.-Based Syst.* 2021, 219, 106904. [CrossRef]
- Davani, A.M.; Díaz, M.; Prabhakaran, V. Dealing with disagreements: Looking beyond the majority vote in subjective annotations. *Trans. Assoc. Comput. Linguist.* 2022, 10, 92–110. [CrossRef]
- Koklu, M.; Unlersen, M.F.; Ozkan, I.A.; Aslan, M.F.; Sabanci, K. A CNN-SVM study based on selected deep features for grapevine leaves classification. *Measurement* 2022, 188, 110425. [CrossRef]
- 36. Schapire, R.E. Explaining adaboost. In *Empirical Inference: Festschrift in Honor of Vladimir N. Vapnik;* Springer: Berlin/Heidelberg, Germany, 2013; pp. 37–52.
- 37. Connelly, L. Logistic regression. Medsurg Nurs. 2020, 29, 353-354.
- 38. Uddin, S.; Haque, I.; Lu, H.; Moni, M.A.; Gide, E. Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. *Sci. Rep.* **2022**, *12*, 6256. [CrossRef] [PubMed]
- Balyan, A.K.; Ahuja, S.; Lilhore, U.K.; Sharma, S.K.; Manoharan, P.; Algarni, A.D. A hybrid intrusion detection model using ega-pso and improved random forest method. *Sensors* 2022, 22, 5986. [CrossRef]
- 40. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
- Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.

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.