A Study on the Teaching Design of a Hybrid Civics Course Based on the Improved Attention Mechanism

Wenwu Miao

Abstract: As an important vehicle for moral education, the moral indicators of civics and political science textbooks are naturally some of the most important criteria for revising textbooks. However, the textbook text dataset has too much textual information, ambiguous features, unbalanced sample distributions, etc. To address these problems, this paper combines a novel data enhancement method to obtain classification results based on word vectors. Additionally, for the problem of unbalanced sample sizes, this paper proposes a network model based on the attention mechanism, which combines the ideas of SMOTE and EDA, and uses a self-built stop word list and synonym word forest to conduct synonym queries, achieve a few categories of oversampling, and randomly disrupt the sentence order and intra-sentence word order to build a balanced dataset. The experimental results also show that the data augmentation method used in this paper’s model can effectively improve the performance of the model, resulting in a higher boost in the F1-measure of the model. The model incorporating the attention mechanism has better model generalization compared to the one without the attention mechanism, as well as a significant advantage compared to the reference model in other settings. The experimental results show that, compared with the original text classifier, the scheme of this paper effectively improves the evaluation effect and the reliability design for teaching a civics course.

Keywords: moral indicators; Chinese text classification; attention mechanism; civics course; teaching design

1. Introduction

China’s Ministry of Education has proposed “eliminating ‘useless courses’ and creating ‘useful courses’” , with “one degree of gender (high order, innovation and challenge)” as a guide to effectively improve teaching quality. Among them, the “innovation” of the Golden Course requires the modernization of the content of civic science courses, with an advanced teaching form and the personalization of the learning results [1]. In a speech at China’s Higher Education Director Conference, the Ministry of Education further proposed to promote a “quality revolution” through a “learning revolution”, and to use “the degree of achievement of course objectives, the degree of adaptation to social needs, the degree of support of technical conditions, the degree of effective quality assurance operation and student learning satisfaction” as the criteria for testing the effectiveness of courses [2]. With the new “Internet + teaching” information, in a student-centered education context, this is necessary in order to overcome the misconception of pure “technology worship” and to ensure students’ roles as active learners and independent constructors through the effective teaching design of civics and political science courses [3]. In recent years, MOOC applications have enabled the sharing of high-quality resources on a large scale, promoting cross-campus and cross-regional online learning, flipped classrooms, online and offline blended learning, and other sharing and application models [4]. New technologies, the Internet, artificial intelligence, and information technology tools are constantly changing and reshaping modern education, fundamentally promoting the development of new forms and modes of teaching civics courses, as well as driving a change in learning styles [5].
Information technology tools are mental models and devices that support, guide, and expand learners’ thinking processes.

The real role of technology is to dilute the learners’ burden of constructing cognition, expand their ability to construct knowledge, and develop higher order thinking skills through the use of technological tools [6]. In order to improve the teaching quality of civics courses, teachers need to use intelligent information technology tools to focus on a new form of teaching wherein teachers and students generate knowledge via interaction, which leads to the creation and transfer of knowledge in the context of an increasingly prominent form of knowledge creation learning [7]. To this end, based on online and offline civics courses, the author conducted an all-round reflection on the course in line with the requirements of the “Golden Course”, seriously exploring how to truly return the course construction to the value of “student-centeredness”. In the context of “Internet + Education”, we have explored how to make full use of the value of information technology tools to empower the connotation of the “Golden Course”, create a new hybrid teaching model for students’ knowledge generation and creation, and move from the “Learning Revolution” to the “Quality Revolution”.

Teaching materials provide the basic guidelines for education and teaching, and the selection of the content of teaching materials plays a key role in the development of students’ values. Meanwhile, there is a structural problem with the simple and repetitive content in the teaching materials of university civics courses [8,9]. Therefore, in order to better evaluate the content of teaching materials, experts of the project team have formulated corresponding moral indicators according to the requirements of today’s socialist core value system in China, including four first-level moral indicators, namely, political identity (A), national consciousness (B), cultural confidence (C), and civic personality (D). In order to evaluate the moral indicators of teaching materials in a more detailed and rigorous way, each primary moral indicator is subdivided into four secondary moral indicators, namely, party leadership (A1), scientific theory (A2), political system (A3), development path (A4), national interest (B1), concept of national conditions (B2), national unity (B3), international perspective (B4), national language (C1), history and culture (C2), revolutionary tradition (C3), spirit of the times (C4), healthy body and mind (D1), law-abiding and equality (D2), honesty and responsibility (D3), and self-improvement and cooperation (D4). This paper investigates the automatic classification of secondary moral indicators.

Guided by activity theory and constructivism theory, this paper—based on the empirical study of teaching practice, teaching data analysis, and questionnaires of civics courses—investigates and summarizes the value-enabling path of information technology tools in the process of creating an online and offline hybrid “Golden class”, and explores the “three-dimensional” approach to deep learning in order to produce a constructive hybrid teaching mode for civics and government.

As an important vehicle for moral education, the moral indicators of civics and political science course materials naturally become some of the most important criteria for revising the materials [10]. However, the textbook dataset has rich textual information, ambiguous feature performance, unbalanced sample distribution, etc. To address these problems, this paper combines a novel data enhancement method to obtain classification results based on word vectors. Additionally, for the problem of unbalanced sample size, this paper proposes a network model based on the attention mechanism [11,12]. The experimental results also show that the data augmentation method used in this paper’s model can effectively improve the performance of the model, resulting in a higher boost in the F1-measure of the model. The model incorporating the attention mechanism has better model generalization compared to the one without the attention mechanism, as well as a significant advantage compared to the reference model in other settings.
2. Related Work

2.1. Deep Learning

In 1976, the authors of [13] published, in which they first proposed the theory of deep learning. This was followed by other scholars’ theoretical development of the concepts of deep learning and shallow learning, which gradually formed an influential learning concept. The authors of [14] argue that deep learning is mainly about the ability of learners to construct resources for learning, come up with learning strategies that are appropriate for them, integrate resources, communicate actively, apply what they have learned in practice, and truly master the essence of learning.

Along with the emergence of MOOCs, SPOCs, and microphoning, blended learning, mobile learning, and ubiquitous learning have gradually developed into new learning modes that are becoming influential [15]. The role of these new information technology tools in promoting deep learning has become a new research proposition of deep learning theory in the 21st century. The authors of [16] give a general definition of deep learning that focuses on critical learning of new knowledge, emphasizing the “integration” and “transfer” of new and old knowledge, promoting the construction of students’ knowledge so that they have the ability to move from learning to problem solving. The authors of [17] argue that for learners, information technology is an intelligent companion and a tool for knowledge construction that helps them to break through temporal and spatial constraints in a truly effective way. The PISA (Program for International Student Assessment) also designed specific requirements for the measurement of students’ core competencies for deep learning [18]. The authors of [19] suggest that attention should be paid to students’ core deep learning competencies, i.e., students’ ability to understand knowledge, apply knowledge to solve problems, media communication, peer collaboration, and self-reflection. The researchers in [20] used the four aspects of “preparation for deep learning”, “integration of old and new knowledge”, “fine processing”, and “assessment” to design a complete blended learning framework based on deep learning theory, and conducted an empirical study pointing out that orientation based on curriculum goals, creating a positive blended learning environment, promoting learners’ self-reflection, and carrying out learner-differentiated instruction are effective ways of promoting deep learning [21].

In summary, in the 21st century, theories of learning have essentially returned to the essence of “student-centeredness”, and scholars have emphasized the deep integration and holistic construction of cognitive, social, and technological aspects of learning from the perspectives of learning goals, motivation, attitudes, and learners’ relationships with knowledge, self, and environment [22,23]. For teachers, they should provide “scaffolding” in the selection of teaching strategies to promote deeper learning, actively create learning tasks in authentic contexts, promote the transfer and externalization of students’ learned knowledge, and pay close attention to learners’ self-reflection and metacognitive development. Deep learning is an innovative way to deliver rich educational content to students, and to apply it to their learning.

2.2. Enabling Exploration of Information Technology Tools

The author of [24] argues that information technology plays the role of an effective tool, an information acquisition tool, a cognitive tool, a contextual tool, a communication tool, and an evaluation tool. These tools can contribute positively to the different dimensions of deep learning through their respective functional characteristics. In a broad sense, along with the development of advanced technologies such as cloud platforms, online learning tools, big data, and various learning apps, IT tools can facilitate students’ deep learning, knowledge construction, knowledge transfer, and problem solving in ways that point to the development of higher order thinking skills, thus helping them to construct lifelong learning concepts through IT, helping students to enrich the core knowledge needed to solve complex problems, and enhancing metacognitive development [25]. The author of [26] argued that various types of information technology tools undergo a process of change in the different integration processes with teaching and learning—including planning.
applications, integrated targeted applications, transformation applications of learning styles and methods, and intelligent applications—and provided an in-depth discussion of the idea of building a smart classroom using information technology. The author designed an information-based teaching model that is assessment-oriented and incorporates the integration of teaching methods and technological tools in the information technology environment, and found that it effectively promotes student engagement and independent learning ability through empirical studies. The author of [27] explored the interaction strategies supported by various information technology tools from the perspective of interaction, and found that the teaching interaction strategies supported by technology can effectively enhance learners’ deep learning ability from an empirical perspective. The authors of [28] explored a new deep-learning-oriented instructional model in a blended learning environment, with a focus on a collaborative knowledge construction model, and found that students’ motivation to learn was enhanced under this model, and that there were significant improvements in the effectiveness of the cognitive, interpersonal, and personal domains of deep learning.

The aforementioned studies explored the blended learning model and its practice under the deep learning orientation from the perspectives of integration of changes in IT tools and teaching, evaluation-oriented deep learning, deep teaching interaction, and collaborative knowledge construction, but most of these studies focused on one unidirectional dimension of the blended learning model, and there have not yet been any empirical studies from a multi-perspective, 3D, or ecological perspective. Therefore, this paper focuses on how to achieve blended learning supported by information technology tools in the new era, with more connotations from a multidimensional perspective under the deep learning orientation, and integrate the changes in teaching methods, promotion of learning evaluation, teacher–student and student–student interaction, and collaborative construction of knowledge, so as to integrate more innovative designs into the construction of blended Golden Courses and improve the quality of course teaching.

3. Model Design

3.1. Data Enhancement

The quality of the dataset in deep learning has a great impact on the performance of the model, and an extremely unbalanced dataset will cause distress to the training of the model and may lead to overfitting. As shown in Figure 1, there is a data imbalance in the text data of college textbooks, and the quantity of each category varies greatly [29]. Therefore, this paper needs to use data enhancement techniques to obtain a more balanced dataset.

![Figure 1. Distribution of each category before data enhancement.](image)

Data augmentation is an effective method to expand the sample size of data, and commonly used methods include oversampling, undersampling, back-translation, and disrupting the order of words within sentences and sentence order. Among these, oversampling is used to balance the dataset by duplication, autonomous sampling, or the synthetic
minority class oversampling technique (SMOTE) [30] on the categories with insufficient data, and SMOTE is also used to create new instances of minority classes by constructing convex combinations of neighboring instances without using duplicate samples and without excessive overfitting. Meanwhile, the authors of [31] proposed a simple and general NLP data enhancement technique called EDA, which consists of four simple operations to perform data enhancement, including Synonym Replacement (SR), Randomly Insert (RI), and Randomly Swap (RS). EDA uses the NLTK for English synonym queries to generate the same number of new texts for instances in the dataset, and achieves outstanding performance when training on smaller datasets, significantly improving the performance of the model.

Therefore, to address the problem of the unbalanced text datasets of college textbooks, this paper combines the ideas of SMOTE and EDA to propose a new data enhancement method, i.e., using self-built deactivated word lists and synonym word forests to conduct synonym queries, achieving oversampling on categories with fewer indicators, and randomly disrupting the order of words within sentences as well as the order of samples. As shown in Figure 2, the dataset after data augmentation essentially meets the demand, effectively preventing the occurrence of overfitting and improving the generalization ability of the model.

3.2. Attention Mechanisms

The attention mechanism was first used by the Google Mind team [14] in conjunction with RNN models for image classification tasks. Later, the authors of [32] used a similar attention mechanism to accomplish the task of machine translation, enabling the model to perform translation and alignment simultaneously, which was the first application of the attention mechanism in NLP. Since then, the attention mechanism has been widely applied to various scenario tasks, and the Google machine translation team [21] has proposed the concept of a transformer whose network structure consists entirely of the attention mechanism, and no longer uses CNNs or RNNs to construct classification models, which performs very well in English–German and English–French translation tasks in the WMT2014 corpus; it also outperforms the traditional CNN and RNN models in training speed. In traditional CNNs for text classification tasks, each channel corresponds to a sentence, and the feature representation of the sentence is obtained and then input to the classifier without considering the connections between the sentence pairs before the feature representation of the text is input to the classifier.

An attention function can be essentially considered as a mapping from a query to a set of key–value pairs, where query, key, value, and output are represented by vectors, as shown in Figure 3. The source can be regarded as the content stored in the memory, the elements of which consist of the address key and the value. We can suppose that there are Lx queries, and that the value of the corresponding value in the memory—i.e., the value
of attention—is finally obtained by $L_x = ∥Source∥$, representing the length of the input text sequence.

![Diagram](image.png)

**Figure 3.** Mapping from a query to a set of key–value pairs.

The functions that are often used to calculate the similarity of query and key are the inner product formula, the cosine similarity formula, and the perceptron network formula, respectively:

\[
\text{Sim}(\text{Query}, \text{Key}_i) = \text{Query} \cdot \text{Key}_i \quad (1)
\]

\[
\text{Sim}(\text{Query}, \text{Key}_i) = \frac{\text{Query} \cdot \text{Key}_i}{∥\text{Source}∥ ∥\text{Key}∥} \quad (2)
\]

\[
\text{Sim}(\text{Query}, \text{Key}_i) = \text{MLP}(\text{Query} \cdot \text{Key}_i) \quad (3)
\]

The similarity calculation result is then normalized using softmax to obtain the probability distribution, which denotes the similarity probability distribution of the $i$-th query, and the formula is as follows:

\[
a_i = \text{Softmax} \left( \frac{e^{\text{Sim}_i}}{\sum_{j=1}^{L_x} e^{\text{Sim}_j}} \right) \quad (4)
\]

Finally, the value is weighted and summed according to the weighting factor, as follows:

\[
\text{Attention( Query, Source) = ∑}_{i=1}^{L_x} a_i \cdot \text{Value}_i \quad (5)
\]

The essence of the attention mechanism is actually an addressing process, which calculates the attention value by computing the attention distribution with the key and attaching it to the value. The attention mechanism also reduces the complexity of the neural network model by focusing the perspective on the information that is more relevant to the current task, and reducing or even ignoring other irrelevant information [25]. In doing so, it can effectively solve the problem of information overload of the neural network model, and at the same time improve the computational efficiency and generalization ability of the model [17,29].

### 3.3. Our Model

As shown in Figure 4, the modeling process of this paper is divided into text preprocessing, data enhancement, feature extraction, and training.

1. **The goal of text preprocessing** is to turn the text into an analyzable and predictable form, and there are a large number of special symbols, tone auxiliaries, and other information irrelevant to classification in the textbook dataset. These noises increase the computational overhead of the model, so they should be removed as much as possible; thus, this paper uses the Jieba [18] word division tool and self-built deactivation word list to remove the noise from the textbook;

2. **A balanced dataset** has a very important impact on the classification performance of the model. In view of the serious imbalance of the Shanghai textbook dataset, this paper combines the ideas of SMOTE and EDA to achieve the oversampling of
a few categories, and to randomly disrupt the order of sentences and words within sentences with a certain probability, so as to obtain a balanced new dataset;

(3) In this paper, we use the negative sampling model Skip-Gram in Word2vec [19,20] to generate word vectors and map the reprocessed words into low-dimensional vectors;

(4) The model structure of this paper is shown in Figure 5. Unlike the normal TextCNN structure, the attention layer is introduced between the input layer and the convolutional layer in order to create a context vector for each word. Then, the context vector and the original pre-training of the word are spliced as the new representation of the word, which is input to the convolutional layer and then passed through the pooling layer and the full concatenation layer to obtain the class of the text.

![Figure 4.](image1.jpg) The modeling flow of this paper.

![Figure 5.](image2.jpg) Model structure of this paper.

The objective function in this paper uses the cross-entropy loss function with a faster convergence rate. The loss function is as follows:

\[
\text{Loss} = -\sum_{i=1}^{n} y_i \log(\hat{y}_i) \tag{6}
\]
where \( y_i \) is the true classification result of the \( i \)-th label and \( \hat{y}_i \) is the prediction result. Then, the loss of each batch is:

\[
\text{Loss} = \sum_{j=1}^{m} \sum_{i=1}^{n} \left( -y_{ji} \log(\hat{y}_{ji}) - (1 - y_{ji}) \log(1 - \hat{y}_{ji}) \right)
\]  

(7)

where \( m \) is the number of samples in the current batch, \( n \) is the number of categories, and \( \hat{y}_{ji} \) is the predicted label of each node.

Denoting the context vector in the graph by \( g_i \), the attention mechanism is additionally an MLP (multilayer perceptron network) that determines which words should receive more attention when predicting sentences, i.e.,:

\[
g_i = \sum_{j \neq i} \alpha_{i,j} \cdot x_j
\]  

(8)

where \( \alpha_{i,j} \) is called the attention weight and requires \( \alpha \geq 0 \) and \( \sum_{j \neq i} \alpha_{i,j} \cdot x_j = 1 \), which can be achieved here by softmax normalization:

\[
\alpha_{i,j} = \frac{\exp(\text{score}(x_i, x_j))}{\sum_{j} \exp(\text{score}(x_i, x_j))}
\]  

(9)

where the score value is calculated by the MLP mentioned above.

4. Experimental Results and Analysis

4.1. Experiment Preparation

The dataset of this paper was derived from the text of college civics textbooks, with a total of 33,360 data, containing 16 categories and 23,083 terms. After data enhancement, the balanced dataset contains 110,665 data, 16 categories, and 28,966 words. A randomly selected 8:2 split from the balanced dataset was used for the training and test sets, respectively. Examples of the sample data are shown in Table 1.

Table 1. Examples of some sample data.

| Category                        | Textbook Text Example                                                                                                                                 |
|---------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| Party leadership (A1)           | It expresses the Chinese people’s incomparable pride and excitement at the birth of new China, and shows the leadership style of the founders of the people’s Republic of China, especially Mao Zedong. |
| National interest (B1)          | Cultivates students’ preliminary speech ability, enhances students’ understanding of modern Chinese humiliation history, cultivates students’ national spirit and patriotic enthusiasm, and enhances students’ sense of historical responsibility. |
| National language (C1)          | Understanding the subtilty of language and characters, developing the rigorous spirit of not relaxing every word, so as to improve the ability of reading and writing, and correctly understand and use the language and characters of the motherland. |
| Physical and mental health (D1) | Understanding Zhuangzi’s “relativism” theory and understanding that in the vast universe, personal cognition and actions are restricted by various subjective and objective conditions. This is very limited, and cannot be complacent or limited by personal insight. |
4.2. Experimental Setup

The experimental environment of this work is based on Python 3.6, and the Keras deep learning framework was used to complete the construction of both this work and the reference model.

4.3. Word Vector Pre-Training

In this work, we used Google’s open-source Skip-Gram with Negative Sampling model to generate pre-trained word vectors. The training window size of the Skip-Gram model was set to 15 (the farthest distance between the current word and the predicted word), the learning rate of the model was set to 0.05, the negative value was set to 5, the min_count was set to 1 (indicating that no words would be discarded), and the number of iterations was set to 10. The parameters of the pre-trained word vector were set as shown in Table 2. The word similarity of the pre-trained word vector is shown in Table 3.

Table 2. Pre-trained word vector parameters.

| Parameter Name                                    | Parameter Value                                      |
|--------------------------------------------------|------------------------------------------------------|
| Model name (model)                               | Skip-Gram                                            |
| Training dimension (100, 150, 200, 250, 300)      |                                                      |
| Window size (window)                             | 10                                                   |
| Minimum word frequency (min count)               | 1                                                    |
| Learning rate (alpha)                            | 0.05                                                 |
| Random downsampling threshold (sample)            | $1 \times 10^{-4}$                                   |
| Iterations (ITER)                                | 10                                                   |

Table 3. Examples of similarity of pre-trained word vectors.

| Environment                          | Similarity | Internet                  | Similarity | Politics                          | Similarity |
|--------------------------------------|------------|---------------------------|------------|-----------------------------------|------------|
| Environmental pollution              | 0.9009     | Network service           | 0.8214     | International relations           | 0.8854     |
| Contaminated                         | 0.8583     | The Beacon Tower          | 0.7460     | Political economy                 | 0.8824     |
| Agricultural and pastoral areas      | 0.8526     | Science and technology    | 0.7449     | Social transformation             | 0.8746     |
| Water pollution                      | 0.8488     | Computer network          | 0.7020     | Politicians                       | 0.8530     |
| Resources                            | 0.8446     | Industrial Revolution    | 0.6724     | Those in power                    | 0.8471     |

4.4. Experimental Design

Experiment 1 verified the effectiveness of the data enhancement method proposed by the model in this paper.

Experiment 2 verified whether the parameter selection in this paper was reasonable.

Experiment 3 verified the effectiveness of the proposed model in this paper, and compared the classification performance with Bi-LSTM [7], Bi-GRU [9], GRNN [21], IoMET, and BERT-Base [22]—provided by Google as reference models. The hypernatremia settings of each model are shown in Tables 4 and 5.

4.5. Experimental Evaluation Criteria

Precision (P) represents the proportion of items found that are relevant. Recall (R) is the proportion of relevant items that are found. F1-measure (F1) is a score that combines P and R, and is defined as the summed average of P and R.
Where TP (true positive) means that the relevant item is correctly identified as relevant, FP (false positive) means that the irrelevant item is incorrectly identified as relevant, and FN (false negative) means that the relevant item is incorrectly identified as irrelevant.

**Table 4. Model hyperparameter settings.**

| Parameter               | Our Method | IoMET | Bi_LSTM | Bi_GRU | GRNN |
|-------------------------|------------|-------|---------|--------|------|
| Convolution kernel size  | (2, 3, 4)  | (2, 3, 4) | -       | -      | -    |
| Regularization          | L2 (0.001) | L2 (0.001) | -       | -      | L2 (0.001) |
| Drop out                | (0.1)      | 0.5   | 0.2     | 0.2    | 0.2  |
| Optimization method     | Adam       | Adam  | SGD     | SGD    | Adam |
| Batch size              | 128        | 128   | 256     | 256    | 128  |
| Epoch                   | 10         | 10    | 10      | 10     | 10   |

**Table 5. BERT model training parameter settings.**

| Parameter Name                          | Parameter Value |
|-----------------------------------------|-----------------|
| Dropout random deactivation rate         | 0.1             |
| Number of model iterations              | 5               |
| Learning rate                           | $5 \times 10^{-5}$ |
| Data volume of each training set         | 24              |

4.6. Analysis of Results

**Experiment 1.** In order to verify the effectiveness of the data enhancement method used in this paper, the textbook text and the original text after data enhancement were input into the TextCNN text classifier and evaluated by the F1-measure; the experimental results are shown in Figure 6.

![Figure 6. Comparison of F1-measure curves for each category.](image)

**Experiment 2.** Because different hyperparameter settings have different effects on the model, the parameters need to be optimized.
The word vector matrices generated by different word vector dimensions are different, and the text feature information expressed is also different. In order to find the optimal word vector dimension under the Shanghai textbook text dataset, multiple groups of control experiments are set, and F1-measure is used as the model performance evaluation index. The performance of the model in different dimensions is shown in Figure 7.

![Figure 7. Effect of word vector dimension on model performance.](image1)

Model overfitting often occurs in the process of model training. Because the loss function of overfitting the model in the test set and the training set is very different, the overfitting model cannot be used in practical applications. Crossover et al. [23] proposed that in the process of training the model some hidden layer nodes can be ignored in each training block in order to reduce the interaction of these hidden layer nodes, so as to reduce the occurrence of overfitting. Therefore, in this paper, multi-group control experiments were carried out in the (0, 1) interval to observe the classification performance of the model under different dropout values. The performance of the model under different dropout values is shown in Figure 8.

![Figure 8. Effect of different dropout values on model performance.](image2)
Experiment 3. In order to verify the validity of this paper on the classification task of the textbook's moral indicators, multiple control experiments were set up, and the results of the experiments are shown in Table 6.

Table 6. Classification results of different classification models.

| Model Name          | p/\text{%} | R/\text{%} | F1-Measure/\text{%} |
|---------------------|------------|------------|---------------------|
| AMMFNet + SE        | 77.66      | 77.36      | 77.51               |
| IoMET               | 61.46      | 61.10      | 61.28               |
| Bi_lstm             | 76.42      | 74.83      | 75.65               |
| Bi_GRU              | 75.03      | 74.61      | 74.82               |
| GRNN                | 74.71      | 75.67      | 75.19               |
| BERT                | 78.38      | 78.32      | 78.35               |

For Q1, it can be seen from Figure 6 that the F1-measure values of each category of the textbook text after data enhancement are higher than those of the original text, verifying the effectiveness of the data enhancement method used in this paper, which can effectively solve the problem of unbalanced text datasets.

For Q2, it can be seen from Figure 7 that the best classification performance of this paper is between 100 and 150 word vectors, which may be due to the decrease in the quality of the pre-trained word vectors from the corpus when the word vector dimension is set too high. From Figure 8, we can see that there is no significant change in the dropout value (0, 0.5), and that the best classification performance is achieved when dropout = 0.3.

For Q3, it can be seen from Table 7 that, in the task of classifying the morphological indicators of college textbooks, the method proposed in this paper achieves the highest F1-measure value of 77.51%, with the word vector dimension of 150 dimensions and dropout = 0.3. The reference models obtain 61.28% (IoMET), 75.62% (Bi_LSTM), 74.82% (Bi_GRU), 75.19% (GRNN), and 78.35% (BERT) of the F1-measure value. The F1-measure values of BERT are less than 1% higher than those of this paper and the convergence speed is much slower than that of this paper because of the larger number of parameters to be trained. The original IoMET model cannot be used as a model, as it does not make good use of the linkage and location information between words, so when the input utterance is too long, some important semantic information may be lost. In this paper, the proposed text classification model is combined with the attention mechanism to generate a context vector for each word. This is then combined with the pre-trained word vector input to the convolutional neural network, which effectively improves the generalization ability of the model and achieves better performance in the classification task of university textbook indicators.

Table 7. Comparison of evaluation parameters of different optimization models on tumor grading.

| Model              | Accuracy | Sensitivity | Specificity | Positive Predictive Rate | Negative Predictive Rate | Average F1 |
|--------------------|----------|-------------|-------------|--------------------------|--------------------------|------------|
| AMMFNet            | 0.913    | 0.928       | 0.88        | 0.946                    | 0.844                    | 0.916      |
| AMMFNet + Res      | 0.894    | 0.893       | 0.897       | 0.951                    | 0.788                    | 0.908      |
| AMMFNet + Res + SE | 0.906    | 0.937       | 0.837       | 0.928                    | 0.855                    | 0.901      |
| AMMFNet + SE       | 0.922    | 0.929       | 0.905       | 0.956                    | 0.850                    | 0.928      |

4.7. Model Performance

In this paper, an adaptive multimodal and multiscale feature fusion classification network is proposed. Under the condition of considering multimodal information and multiscale information at the same time, the accurate prediction of tumor classification can be achieved without sketching the lesion area. As shown in Figure 9, its AUC can reach 0.965, which is higher than the existing mainstream deep learning classification model.
When using the focus area prediction, its AUC is as high as 0.997, which is better than the latest model of tumor grading.

![Figure 9. Comparison of ROC curves of optimizable models.](image)

As shown in Figure 10, the multimodal data can complement one another, and have a richer feature search space than the unimodal data, while the multimodal feature extraction module designed in this paper can effectively explore the semantic features useful for classification. The adaptive multimodal multiscale feature fusion module can increase the adaptability of the fusion rules to a certain extent by using the learnable convolutional approximation fusion rules. However, in practice, there are still some open issues in the experimental process. For example, although our method can classify tumors by the differences in their multimodal features, it is difficult to further improve the classification accuracy due to the inconsistent training speed of multiple modalities during training. In addition, among the four modalities, the tumor area was not effectively identified from the T1 and Flair modal images, and their inclusion in the multimodal model would reduce the prediction accuracy of the model. This suggests that while using the multimodal adaptive fusion strategy, stronger constraints should be added to each modal datum in order to ensure that the training speed of each modal feature extraction module is essentially the same, so as to reduce the influence of the multimodal model on individual modalities. From the results of the model optimization analysis, it is clear that the introduction of different submodules requires consideration of the specific problem to be solved by the network. In addition to the introduction of the SE module, the introduction of a newer image attention mechanism in the learning module of the adaptive fusion rules is also one of the options for future optimization. In addition to the optimization of the fusion module, the depth of the network model designed in this paper is only 11 layers, and the network can be deepened. The deeper the network is, the more semantic it is, and the better it can be expressed. However, as the depth of the network increases, the number of network parameters gradually increases, and it is also worth exploring how to improve the hierarchical performance of the model by reducing the network parameters while deepening the network based on the design ideas of frontier networks.

4.8. Effect of Different Combination Methods of Sample Pairs on Performance

This paper introduces the advantages and disadvantages of two sample pair combination methods—differential image and synthetic image—and combines their characteristics to propose the use of aligned synthetic samples as the sample pair combination method in this paper, i.e., the two samples in a sample pair are firstly aligned to find the common region, and then combined into a synthetic image as the input of the network. In this
section, a comparative experimental analysis of these three combination methods is carried out, and their performance is individually tested on three publicly available datasets for training. The networks are all deep tree networks as proposed in this paper, and the data are divided into training, validation, and test sets using the closed-set test approach; none of them are augmented with online data.

Figure 10. Equivalent error rate curves of test results for different combinations of sample pairs in three publicly available datasets: (a) FV-USM; (b) MMCBNU_6000; (c) SDUMLA-HMT.
Table 8 shows the test results of different combinations of sample pairs on the three public datasets, and Figure 5 shows the corresponding equal error rate curves (the legend “composite” corresponds to “synthetic image”, while “difference” corresponds to “composite image” and “difference” in the table). The legend “composite” corresponds to “composite image”, “difference” corresponds to “difference image”, and “align + difference” corresponds to “align + difference” (align + difference” corresponds to “align + composite” in the table).

| Parameter Name         | Parameter Value          |
|------------------------|--------------------------|
| Operating system       | Windows 10 (64 bit)      |
| Processor              | i5-9300HF (2.40 GHz)     |
| Memory                 | DDR4 (16 GB)             |
| Graphics card          | GTX1660Ti (6 GB)         |
| Parameter name         | Parameter value          |
| Operating system       | Windows 10 (64 bit)      |

As can be seen from Table 9, the aligned synthetic samples proposed in this paper outperform the two sample pair combination methods of differential and synthetic images on all three datasets under the same network structure and datasets, exhibiting a great improvement. When comparing the fourth and fifth rows of the table to the second and fifth rows, respectively, it can be seen that the performance is improved after adding the alignment method to both the original differential and synthetic images, which verifies the effectiveness of the alignment method used in this paper against the vein pattern shift. Meanwhile, the comparison of the specific values in the third and fifth rows of the table shows that the proposed aligned synthetic image has the most obvious effect on SDUMLA-HMT, followed by MMCBNU_6000, and the least improvement on FV-USM. The alignment of the original data greatly improves the vein offset due to finger rotation between the intraclass samples, but has less improvement on the light and dark samples. The improvement is small for the difference between light and dark, due to different illumination, and the most severe grain shift is found in SDUMLA-HMT, but the difference between light and dark between samples is very small, so the improvement is the largest after adding the alignment method. In FV-USM, except for a small number of samples with grain offset, there are large differences between the samples of all fingers, despite it being a closed-set test, because the division method in this paper first randomly forms sample pairs and disrupts them, and then randomly divides them into training, validation, and test sets proportionally, which increases the difficulty of the test to some extent. Therefore, the alignment-only method has less improvement in illumination, and the performance improvement on FV-USM is relatively small.

|                         | FV-USM | MMCBNU_6000 | SDUMLA-HMT |
|-------------------------|--------|-------------|------------|
| Differential image      | 1.745  | 1.296       | 2.096      |
| Composite image         | 0.225  | 0.333       | 0.943      |
| Alignment + differential image | 0.503  | 1.086       | 0.384      |
| Align composite samples | 0.215  | 0.1852      | 0.174      |

5. Conclusions

In this paper, we take the study of moral indices of university teaching materials as the experimental background, and propose a fusion model to achieve a high accuracy of moral index prediction, which can provide an effective reference basis for the study of
moral indices and replace part of the manual work. To address the problem of unbalanced sample size, this paper combines the ideas of SMOTE and EDA, and uses the self-built stop word list and synonym word forest to conduct synonym queries in order to achieve a few categories of oversampling while randomly disrupting the sentence order and intra-sentence word order to build a balanced dataset. The experimental results show that the data augmentation method used in this paper can effectively improve the performance of the model, resulting in a greater improvement of the F1-measure of the model. The model with the attention mechanism has a better model generalization ability than the IoM-ET without the attention mechanism, and also has a significant advantage over the reference model for other settings.

The next research direction is to improve the text pre-processing method and optimize the model structure so that the performance of the model can be further improved.

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

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

**References**

1. Chiarelli, B. The bioecological bases of global bioethics. *Glob. Bioeth.* 2014, 25, 19–26. [CrossRef]
2. Brodsky, J.E.; Brooks, P.J.; Scimeca, D.; Todorova, R.; Galati, P.; Batson, M.; Grosso, R.; Matthews, M.; Miller, V.; Caulfield, M. Improving college students’ fact-checking strategies through lateral reading instruction in a general education civics course. *Cogn. Res. Princ. Implic.* 2021, 6, 23. [CrossRef]
3. Bi, X.-A.; Hu, X.; Xie, Y.; Wu, H. A novel CERNNE approach for predicting Parkinson’s Disease-associated genes and brain regions based on multimodal imaging genetics data. *Med. Image Anal.* 2021, 67, 101830. [CrossRef]
4. Mollahazemi, M.J.; Atyabi, S.A.; Ghaffari, A. Heart beat detection using a multimodal data coupling method. *Physiol. Meas.* 2015, 36, 1729–1742. [CrossRef] [PubMed]
5. Marchman, B.K. Teaching About Homophobia in a High School Civics Course. *Theory Res. Soc. Educ.* 2002, 30, 302–305. [CrossRef]
6. Shapiro, S.; Brown, C. A Look at Civics Education in the United States. *Am. Educ.* 2018, 42, 10.
7. Wang, Y.; Yue, J.; Dong, Y.; Hu, Z. Review on Kernel based Target Tracking for Autonomous Driving. *J. Inf. Process.* 2016, 24, 49–63. [CrossRef]
8. Fatima, I.; Fahim, M.; Lee, Y.-K.; Lee, S. A Unified Framework for Activity Recognition-Based Behavior Analysis and Action Prediction in Smart Homes. *Sensors* 2013, 13, 2682–2699. [CrossRef]
9. Murphy-Chutorian, E.; Trivedi, M.M. Head Pose Estimation in Computer Vision: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 2008, 31, 607–626. [CrossRef]
10. Zhang, C.; Xie, T.; Yang, K.; Ma, H.; Xie, Y.; Xu, Y.; Luo, P. Positioning optimisation based on particle quality prediction in wireless sensor networks. *IET Netw.* 2019, 8, 107–113. [CrossRef]
11. Fernandez, A.; Garcia, S.; Herrera, F.; Chawla, N.V. SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *J. Artif. Intell. Res.* 2018, 61, 863–905. [CrossRef]
12. Langton, K.P.; Jennings, M.K. Political Socialization and the High School Civics Curriculum in the United States. *Am. Political Sci. Rev.* 1968, 62, 852–867. [CrossRef]
13. Letermann, D. Teaching Hybrid Principles Of Finance To Undergraduate Business Students Can It Work? *J. Coll. Teach. Learn. (TLC)* 2008, 5, 43–53. [CrossRef]
14. Rafaillidis, D.; Manolopoulos, S.; Daras, P. A unified framework for multimodal retrieval. *Pattern Recognit.* 2013, 46, 3358–3370. [CrossRef]
15. Chen, H.-S.; Tsai, W.-J. Incorporating frequent pattern analysis into multimodal HMM event classification for baseball videos. *Multimed. Tools Appl.* 2016, 75, 4913–4932. [CrossRef]
16. Jiang, S.; Zheng, Y.; Solomatine, D. Improving AI System Awareness of Geoscience Knowledge: Symbiotic Integration of Physical Approaches and Deep Learning. *Geophys. Res. Lett.* 2020, 47, e2020GL086229. [CrossRef]
17. Zhao, J.; Chen, Y.; Zhuang, X.; Xu, Y. Posterior Probability Based Multi-classifier Fusion in Pedestrian Detection. In *Advances in Intelligent Systems and Computing*; Springer: Cham, Switzerland, 2014; pp. 323–329.
18. Zhang, B.; Velmayil, V.; Sivakumar, V. A deep learning model for innovative evaluation of ideological and political learning. *Prog. Artif. Intell.* 2021, 1–13. [CrossRef]
19. Zhang, Z.; Zhang, C.; Li, M.; Xie, T. Target Positioning Based on Particle Centroid Drift in Large-Scale WSNs. *IEEE Access* 2020, 8, 127709–127719. [CrossRef]

20. Chen, Z.; Huang, W.; Lv, Z. Towards a face recognition method based on uncorrelated discriminant sparse preserving projection. *Multimed. Tools Appl.* 2015, 76, 17669–17683. [CrossRef]

21. Paugam-Moisy, H.; Bohte, S.M. *Computing with Spiking Neuron Networks*; Springer: Berlin/Heidelberg, Germany, 2012; Chapter 10; pp. 335–376.

22. Sowan, A.K.; Jenkins, L.S. Use of the Seven Principles of Effective Teaching to Design and Deliver an Interactive Hybrid Nursing Research Course. *Nurs. Educ. Perspect.* 2013, 34, 315–322. [PubMed]

23. Sagha, H.; Bayati, H.; Millan, J.D.R.; Chavarriaga, R. On-line anomaly detection and resilience in classifier ensembles. *Pattern Recognit. Lett.* 2013, 34, 1916–1927. [CrossRef]

24. Xu, Z. Globalization, culture and ELT materials: A focus on China. *Multiling. Educ.* 2013, 3, 6. [CrossRef]

25. Ito, M.; Soep, E.; Kligler-Vilenchik, N.; Shresthova, S.; Gamber-Thompson, L.; Zimmerman, A. Learning connected civics: Narratives, practices, infrastructures. *Curric. Inq.* 2015, 45, 10–29. [CrossRef]

26. Yuanhong, X.U. Research on the Reform of Humanities English Teaching Based on Ideological and Political Theories teaching in All Courses. *Can. Soc. Sci.* 2021, 17, 52–56.

27. Otsu, K. Civics education in Japan: Values promoted in the school curriculum. *Asia Pac. J. Educ.* 2000, 20, 53–62. [CrossRef]

28. Li, W.; Zhang, Q.; Liu, P.; Liu, W. An Improved Approach for Text Sentiment Classification Based on a Deep Neural Network via a Sentiment Attention Mechanism. *Future Internet* 2019, 11, 96. [CrossRef]

29. Qing, Y.; Liu, W. Hyperspectral Image Classification Based on Multi-Scale Residual Network with Attention Mechanism. *Remote Sens.* 2021, 13, 335. [CrossRef]

30. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* 2002, 16, 321–357. [CrossRef]

31. Kang, T.; Perotte, A.; Tang, Y.; Ta, C.; Weng, C. UMLS-based data augmentation for natural language processing of clinical research literature. *J. Am. Med. Inform. Assoc.* 2021, 28, 812–823. [CrossRef]

32. Zhang, B.; Xiong, D.; Xie, J.; Su, J. Neural Machine Translation With GRU-Gated Attention Model. *IEEE Trans. Neural Netw. Learn. Syst.* 2020, 31, 4688–4698. [CrossRef]