# Design and Application of an Automatic Scoring System for English Composition Based on Artificial Intelligence Technology

Fengqin Zhang
School of Foreign Languages, Zhengzhou Tourism College, Zhengzhou 450000, China

***Abstract*—The automatic grading of English compositions involves utilizing natural language processing, statistics, artificial intelligence (AI), and other techniques to evaluate and score compositions. This approach is objective, fair, and resource-efficient. The current widely used evaluation system for English compositions falls short in off-topic assessment, as subjective factors in manual marking lead to inconsistent scoring standards, which affects objectivity and fairness. Hence, researching and implementing an AI-based automatic scoring system for English compositions holds significant importance. This paper examines various composition evaluation factors, such as vocabulary usage, sentence structure, errors, development, word frequency, and examples. These factors are classified, quantified, and analysed using methods such as standardization, cluster analysis, and TF word frequency. Scores are assigned to each feature factor based on fuzzy clustering analysis and the information entropy principle of rough set theory. The system can flexibly identify composition themes in batches and rapidly score English compositions, offering more objective and impartial quality control. The goal of the proposed system is to address existing issues in teacher corrections and evaluations, as well as low self-efficacy in students' writing learning. The test results demonstrate that the system expands the learning material collections, enhances the identification of weak points, optimizes the marking engine performance with the text matching degree, reduces the marking time, and ensures efficient and high-quality assessments. Overall, this system shows great potential for widespread adoption.**

***Keywords*—English composition; automatic scoring; artificial intelligence; text matching degree; natural language processing***

## I. Introduction

For a long time, college English writing teaching has been a ‘short board’. The traditional teaching mode of college English writing is simple: teachers construct a writing framework-analysis model, students imitate writing, and teachers provide correct-writing comments. The disadvantages of this model are particularly evident in English as a public class. Correcting compositions is energy-consuming; the teachers are powerless in the latter part of writing teaching [1]. As time passes, the writing teaching link becomes ‘top-heavy’, and the writing evaluation class often becomes a general model essay appreciation class. Students have a low sense of self-efficacy in English writing learning and even fear and anxiety, and their writing ability does not improve. These ‘stubborn diseases’ are particularly evident among the non-English majors at the author’s independent university. The students’ English foundation is weak, and they have writing problems, but many of them do not know how to improve. This ‘old and difficult’ problem of college English writing needs to be solved with new ideas. In automatic composition grading, statistics, natural language processing, artificial intelligence and other technologies are used to evaluate and grade compositions. The procedure of automatic composition scoring is to preprocess an English article with initial word segmentation, clauses and part-of-speech tagging, then analyse its morphology, grammar, content richness and other characteristics, and finally score the composition according to appropriate standards [2]. Automatic composition grading brings speed and high efficiency and can greatly reduce teachers’ efforts for essay scoring, which leaves them more time to teach students. Simultaneous targeted training can also enable English learners to improve their English writing ability and, when applied to large-scale tests, can greatly reduce manpower and material resources, improve efficiency, and guarantee the impartiality scores. Technology in the field of education, particularly artificial intelligence, comprehensively and profoundly influences the education concept, teaching model and exam method. Based on speech technology in English listening, oral tests have a wide range of applications. Handwriting recognition and natural language understanding, such as artificial intelligence technology, are also being explored and applied in education examination evaluations. Therefore, it is of practical significance to explore an automatic scoring system that is suitable for nonnative English speakers and has low cost, automation and high accuracy [3-5].

Automatic composition grading involves the use of statistics, natural language processing, artificial intelligence and other technologies to evaluate and grade essays. Automatic composition grading has been widely applied to various examinations; indeed, teaching in the classroom has obtained a certain effect, but in the process of using an automatic scoring system, it has been found that even if the input of an article has nothing to do with the thesis topic, as long as there are not too many grammar and vocabulary errors and high scores can be obtained, researchers can realize the importance of track detection to the thesis [6]. Since the 1960s, with the rapid development of natural language processing technology, automatic composition scoring systems have made great progress and have gradually been introduced for a variety of teaching and exam uses. The PEG scoring system does not use natural language processing technology, there is

no study of the composition or content of the chapter structure, and the theme of the thesis is not considered [7]. The IEA scoring system effectively grades sample essays on a variety of topics, automatically judges the content and quality of essays and provides quick feedback. The E-rater grading system, with its complex feature engineering, can better reflect the quality of composition, so its scoring results are highly consistent with those of manual grading. IntelliMetric uses standardized scoring rules and follows the human brain's judgment of points to extract relevant characteristics of essays and then grades essays based on a constructed model [8]. The BETSY scoring system can first extract the characteristics of the quality of a composition and then, according to these characteristics, the thesis is divided into several different levels of people. After manual annotation data are used as the training sample set, the classification of the training to obtain the composition model needs to follow the same method to test the composition, and the extracted features of a classification model can be assigned to the corresponding collection. The above systems mainly use regression and classification methods to score compositions. In recent years, with the rapid development of deep learning technology, many scholars have attempted to use neural networks to grade English compositions [9, 10]. At present, the composition of the track detection method has a certain effect but also has obvious problems: based on the supervised method, the accuracy is higher, but in daily teaching use, it is flexible in composition. Once the system does not include the new theme in the corpus proposition, track detection accuracy will be discounted, and this method is only applicable to large tests. Based on the unsupervised method, user operation is relatively simple and can also adapt to a variety of different propositions, but the current accuracy is not high and cannot meet the requirements of use [11]. Feature extraction may affect the selection of the training model in the process of constructing an automatic composition scoring model. Therefore, the interaction between these two aspects should be fully considered in the process model construction to select a more appropriate feature extraction method and training model. Most automatic composition grading models can achieve good results, but in supervised learning, especially when the number of candidates is large, much effort is needed for composition annotations, and at the same time, the training process is often dependent on the composition title information and has widespread migration problems, so automatic composition grading also needs to be further examined [12-15].

At present, although the momentum of the development and application of artificial intelligence auxiliary to English writing is good, there are few studies in this field on how to combine students' autonomous learning and improve the enthusiasm of students practising writing outside the classroom for autonomous learning, lifelong learning, and learning views; therefore, it is necessary to conduct further research on the English composition score method [16]. This paper discusses how to apply the machine automatic scoring scheme to composition scoring more effectively to ensure the objectivity, fairness and accuracy of composition scoring in various English exams. This paper examines an automatic scoring system based on artificial intelligence technology English composition design and application with the goal of exploring the teaching effect of improving English writing classes, improving students' self-efficacy and innovation and cultivating their autonomous learning and consciousness of lifelong learning. Ensuring that students' English composition examination papers are high quality is a good objective and worth reviewing, which will provide a rigorous basis for teaching improvement and talent selection.

Despite the evolving landscape of English teaching and grading, there remains a critical gap, that of the effective and efficient application of modern day technologies, such as artificial intelligence, in the field of English composition writing. English composition education, especially among non-English majors at independent universities, has long suffered from the problems of traditional rote learning, lack of individualized attention and ineffective qualitative assessment. Moreover, the emphasis on understanding student self-efficacy and motivation for autonomous learning in the context of AI-led teaching has been underexplored. This paper seeks to elucidate this understudied area, with a specific focus on nonnative English speakers.

The above work mainly discusses the application of an automatic English composition scoring system based on artificial intelligence technology in college English writing teaching. There are some defects in the traditional teaching mode; for example, correcting compositions consumes teachers' energy and leads to an imbalance in the writing teaching process. Therefore, the introduction of an automatic grading system can improve the efficiency and quality of writing teaching. In general, this paper first provides a brief overview of the application of artificial intelligence-based automatic scoring systems in college English writing teaching. At the same time, it also highlights some problems that need further research and improvement and emphasizes the importance of improving teaching effects and cultivating students' self-learning consciousness through this technical means.

## II. Technical Theory of Automatic Composition Grading

Automatic composition scoring mainly involves using natural language processing technology to process composition text content and employing statistical methods for analysis and prediction. Among the many types of questions on an English test, English composition questions can most comprehensively reflect an examinee's comprehensive ability to use English. Through writing, examinees can express and transmit their thoughts and opinions and teachers can examine students' logical thinking and language expression ability.

### *A. An Overview of Related Natural Language Processing Techniques*

Word segmentation and clause segmentation are the basis of natural language processing. English word segmentation is relatively simple, generally using spaces and punctuation for natural separation but also a small number of abbreviations. Clause segmentation refers to an article being divided into a single sentence, usually according to punctuation. The technology of English word segmentation and clause

segmentation is currently quite perfect, and there are many open technologies that are very effective in use. A part of speech is the basic grammatical attribute of a word and is generally called part of speech. Part of speech tagging refers to considering the grammatical category of the vocabulary in a sentence, noting its parts of speech and marking it. The common parts of speech of words are nouns, adjectives, adverbs, verbs and so on. Part-of-speech tagging is one of the basic problems of NLP because data preprocessing of many tasks in NLP requires part-of-speech tagging. Word form reduction is an important part of text preprocessing. In English, it generally refers to the restoration of words in any form to the general tense [17,18]. To put it simply, word restoration removes the affixes of a word and keeps only the main part of the word. Generally, restored words also exist in the dictionary. Form reduction is similar to stem extraction, but it is possible that the word from stem extraction does not appear in the dictionary [19].

Generally, English composition questions have the following characteristics:

*1) Lexical features:* English has strict requirements on the form of words; for example, there are clear usage scenarios for singular and plural nouns.

*2) Phrasal features:* English phrases have many types, such as noun phrases, verb phrases, and prepositional phrases. The collocations of these phrases generally have fixed forms and become modules that constitute sentences. This modular structure ensures the dominant characteristics of English forms.

*3) Syntactic features:* The basic sentence structure is subject and predicate, but English also has complex clauses to express rich content, and the existence form of clauses is flexible.

*4) Structural features:* Cohesion between sentences and paragraphs is the basis of coherence in English composition, and coherence is an important prerequisite for rigorous structure.

The general block diagram of the model is shown in Fig. 1. Given the characteristics of a composition, we should consider the following:

*1) Vocabulary of the composition:* According to the number of novel words in the composition, the number of backbone words is used to evaluate whether students' vocabulary is qualified, and according to the spelling of words, their memory of the vocabulary is evaluated.

*2) Composition syntax:* The use of complex clauses is an important indicator of an examinee's command of English.

*3) Whether the composition structure is rigorous:* The main test for students before and after the description needs to be logical, and the context needs to be appropriate.

*4) Composition content:* This mainly investigates whether the composition content is rich and closely related to the topic.

Therefore, we can find English essay scoring with a strong subjectivity, while a large current test will be equipped with corresponding criteria for English composition, but the scoring criteria from the linguistics angle only provide guidance and are not combined with specific question operability, which is given a set of scoring rules specific to the process of evaluation [20,21]. Careful evaluation and discussion by highly specialized experts is often needed to determine operational scoring criteria.

### B. Evaluation Methodology

The Pearson correlation coefficient is a common linear correlation coefficient that can be used to measure the degree of linear correlation between two variables. For linear variables *M* and *N*, for example, the Pearson correlation coefficient can be used to measure the related degree. For *x*, the value of *x* ranges from -1~1, and the absolute value of *x* tends to be closer to l, showing that the strength of the correlation of *M* and *N* is closer to zero, indicating a weaker correlation. A positive *x* suggests a positive association between *M* and *N*, and a negative x indicates a negative correlation [22]. Pearson's correlation coefficient is equal to the covariance between linear variables *M* and *N* divided by the product of their standard deviations, as shown below:

$$\rho_{xy} = \frac{\operatorname{cov}(M,N)}{\sigma_M \sigma_N} = \frac{E(M-\mu_M)(N-\mu_N)}{\sigma_M \sigma_N} \quad (1)$$

First, calculate the standard deviation and covariance of samples, that is, calculate the Pearson correlation coefficient between samples, denoted as *r*:

$$r = \frac{\sum_{i=1}^{n}(M_i-\bar{M})(N_i-\bar{N})}{\sqrt{\sum_{i=1}^{n}(M_i-\bar{M})^2}\sqrt{\sum_{i=1}^{n}(N_i-\bar{N})^2}} \quad (2)$$

*r* can also be estimated by means of the average standard score of sample points, and the calculation result is equivalent to Eq.(2):

$$r = \frac{1}{n-1}\sum_{i=1}^{n}\left(\frac{M_i-\bar{M}}{\sigma_M}\right)\left(\frac{M_i-\bar{M}}{\sigma_M}\right) \quad (3)$$

where, $\frac{M_i-\bar{M}}{\sigma_M}$, $\bar{M}$ and $\sigma_M$ are the standard score, sample mean, and sample standard deviation for sample $M_i$, respectively.

In the process of evaluating the automatic scoring system of composition, we mainly compare the scoring of the system with the manual scoring. In this model, the following three indicators are mainly referenced: the average error of scoring, the average accuracy of scoring, and the relevance of scoring [23]. The formulas are as follows:

$$\text{AverageScoreError} = \frac{\sum_{i=1}^{N}|x_i-y_i|}{N} \quad (4)$$

$$\text{AverageScoreAccurate} = \left(\sum_{i=1}^{N}\frac{|x_i-y_i|}{y_i}\right)\frac{1}{N} \quad (5)$$

$$\text{ScoreRelevance} = \frac{N\sum_{i=1}^{N}x_i y_i - \sum_{i=1}^{N}x_i\sum_{i=1}^{N}y_i}{\sqrt{N\sum_{i=1}^{N}x_i^2-\left(\sum_{i=1}^{N}x_i\right)^2}\sqrt{N\sum_{i=1}^{N}y_i^2-\left(\sum_{i=1}^{N}y_i\right)^2}} \quad (6)$$

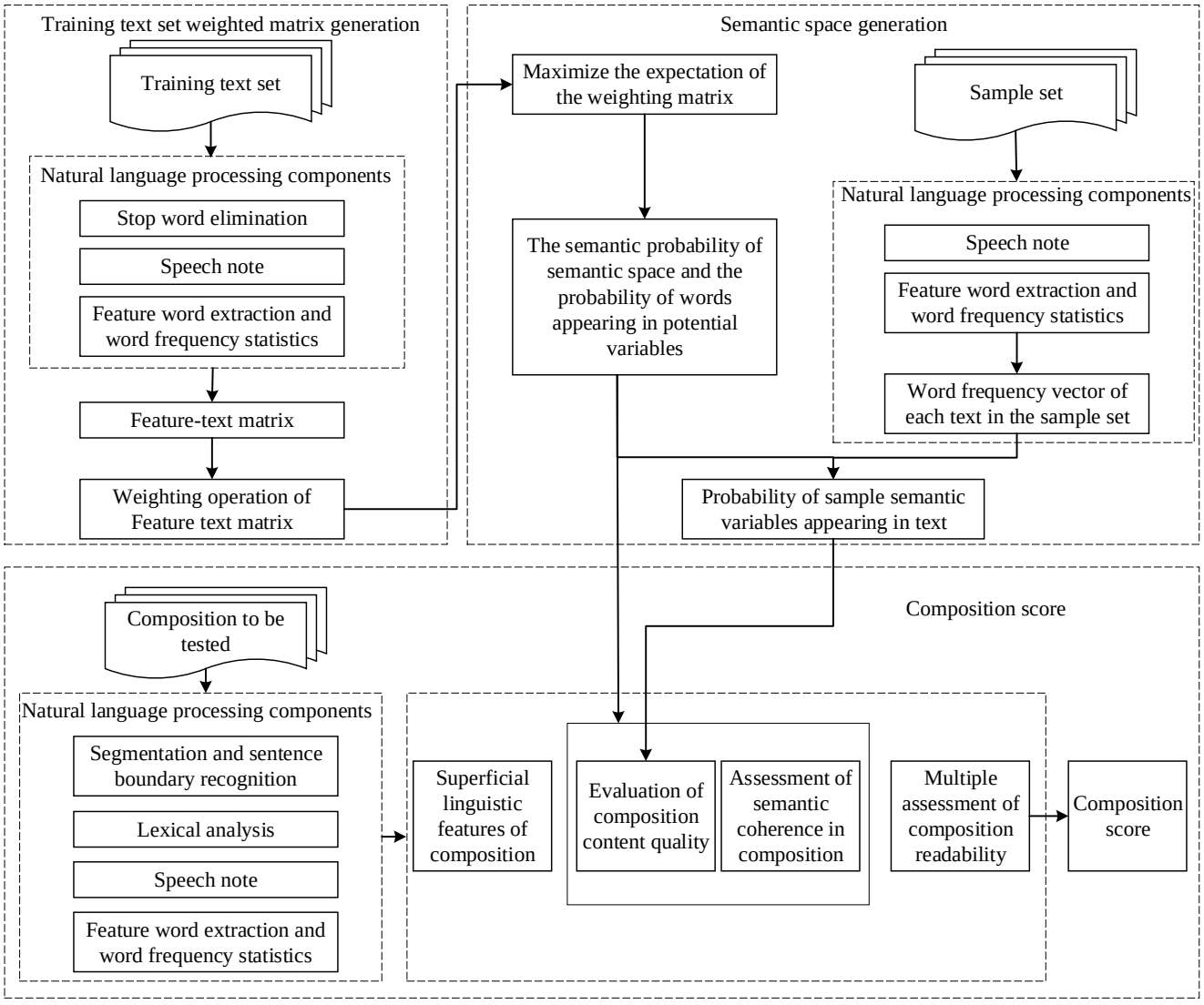


Fig. 1. General block diagram of the model.

### C. Neural Network Applications in Natural Language Processing and Composition Grading

Neural networks have always been a research hotspot in artificial intelligence, cognitive science, nonlinear dynamics and other related fields. Neural networks have been used for academic research in recent years. Over the course of these studies, hundreds of neural network models, including pattern recognition, associative memory, signal processing, control engineering, expert systems, combinatorial optimization, image processing and computer graphics, and many other aspects, have been proposed and have made remarkable progress [24]. At present, deep learning has also been introduced in the natural language field, using the concept of word vectors to enable neural networks to complete the work accomplished in the field of statistics. It can be said that neural networks have begun to show advantages in the field of natural language processing. In the process of automatic composition scoring, the biggest problem that puzzles researchers is how to construct a reasonable model for the characteristics obtained from statistical quantification so that it can evaluate the composition quality well. Because there are many factors that determine the quality of a composition, many different text features can be extracted when the text is processed. These features are complex and changeable, and it is difficult to model them with traditional mathematical theory. Through the understanding of the basic principle of neural networks, combined with the relevant research on automatic scoring technology of composition, the features extracted from composition are processed by artificial neural networks, and the learning, associative memory and distributed parallel information processing functions of neural networks are used to simulate the thinking mode of the human brain [25,26]. Therefore, the constructed neural network model can acquire, learn and reason the expert-rated experience from a large number of expert-rated articles, determine the relationship between the extracted features and the composition score, which is called learning the expert-rated experience, and then grade the text through the learned experience.

After clustering, the word vector has a certain representativeness. Suppose that in the English composition

analysis, after a clustering algorithm has been used to obtain a text clustering, each word in this sentence is in the text clustering, and each different semantic information in the text collection can obtain a word weight value such as frequency and frequency of use [27]. In this paper, by using the automatic grading system of the main statistical, the three characteristics of the term vectors, including word frequency location, size and distribution, the distribution is used to express the sentence, the complexity and diversity of characteristics; for example, in a semantic statement, the author uses more than one word for a description because the authors have a better command of the English language. Based on the above research ideas, this paper extracts and classifies the text features of English compositions.

In the process of grading, teachers can score by item. Usually, there are several indices in grading that are extracted from this model and include the following: score of content quality, semantic coherence, superficial linguistic features of text, and text readability indices. Among them, the semantic score has the largest proportion and indicates whether the central idea of the composition is distinctive and meets the requirements of the topic. If a composition is off topic, then it will have few points. Semantic coherence is also very important; it indicates that the composition of the thought content is consistent, there is a natural smooth transition between statements, text is characterized by shallow linguistics statistics in the text, the word count, the number of sentences, paragraphs and the number of complex words are appropriate, and the main purpose is to evaluate the level of the students to master words, writing skills, etc. The readability index of a text indicates that the text is worth reading. This model uses the features of the existing text to score the text comprehensively.

## III. Design and Implementation of an Automatic Scoring System for English Composition

### A. Requirement Analysis of System

Proposition composition is very common in the daily teaching of a thesis topic. This article designs rating systems, mainly for the proposition composition rate, so the rating system needs to consider not only the composition, such as vocabulary, grammar, and sentences, but also whether the content of the thesis tracks. At the same time, the system's performance and ease of use are also very important [28]. The functional requirements of the system are as follows:

*1)* The system can easily input composition content and can be repeatedly modified and scored.

*2)* The system can score multiple compositions at one time, which is convenient for teachers to score the whole class compositions.

*3)* The interface of the system should be as simple and intuitive as possible so that users can use it quickly and conveniently.

*4)* The system can efficiently identify the topic of the essay and judge whether the content of the essay to be graded fits the topic.

*5)* The system can flexibly configure the topic of the thesis composition, which is convenient for users when scoring different thesis compositions.

The automatic scoring system based on the above requirements will be divided into seven modules: login information maintenance, sample volume, candidate set generation, experts, screening and grading, model training and essay scoring, password changes and user information management, including the composition of model training and essay scoring automatic grading and result output two functions.

In summary, an expert-assisted automatic scoring process of English composition can be obtained. Its main workflow includes test paper cutting, image processing, sample paper candidate set generation, sample paper screening and scoring, scoring model training, automatic scoring of composition and result output. The automatic scoring process of English composition is shown in Fig. 2.

*1) Test paper cutting:* The technical staff adopts cutting software to batch cut all candidates' answer sheet images in accordance with the established cutting scheme, forming a separate English composition answer image test paper library.

*2) Image processing:* The English composition answer image examination paper library is processed and the composition content in the picture is obtained and saved as a text form, which is output into a file according to a specific format.

*3) Sample paper candidate set generation:* The machine adopts certain technical means to traverse the whole composition database and screen out the sample paper candidate set that can fully reflect the whole examination paper library of each grade of examination paper level.

*4) Sample volume screening and scoring:* Experts screen relatively uniform sample volume sets from sample volume candidate sets and score them.

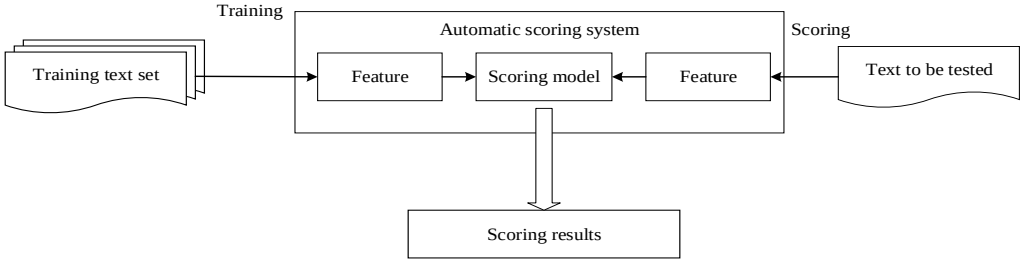


Fig. 2. Automatic scoring process of English compositions.

*5) Training of the scoring model:* The automatic writing scoring model is trained by using the set of selected sample papers and the scores of experts on the set of sample papers as training sets.

*6) Automatic composition scoring:* The automatic composition scoring model obtained by training is applied to paperless marking instead of marking teachers for composition scoring.

*7) Output of scores:* The scores are output.

### B. Preprocessing and Feature Extraction for Automatic Grading of English Composition

Before feature extraction of the document, it is necessary to preprocess the composition to some extent to extract feature items. First, it is necessary to divide the composition into paragraphs, sentences, punctuation and words of the text. The obtained paragraphs, sentences, punctuation marks and words are stored separately for the statistical analysis of lexical features and structural features. Natural language processing (NLP) technology is used to tag the parts of speech in the articles to facilitate the statistical analysis of syntactic features and the detection of word errors and grammatical errors. In this paper, the natural language processing tool developed by Stanford University is used for text segmentation and part-of-speech tagging [29].

Composition feature extraction is the core of automatic grading to solve the English composition problem. Because, unlike humans, a computer cannot understand the connotation of a composition or gift and cannot appreciate and evaluate the merits of a composition, it needs to have quantitative data for calculations. Therefore, in the process of implementing automatic grading composition, we need to extract some quantitative data from the text for the computer to calculate and process these data. These data can reflect the real writing level of students. Through these data, we can establish a certain mathematical model, which can be an effective evaluation of students' writing.

The model mainly covers 64 features of composition, such as vocabulary use, phrase collocation, sentence, coherence, organizational structure and fluency, and then divides these features into three types: morphology, syntax and structure. Training sets are used to train the designed neural network [30]. Then, it evaluates the linguistic quality and organizational structure of the composition from the aspects of morphology, syntax and structure. In addition to the evaluation of the above three aspects, this paper also analyses and evaluates the semantic content of the text. Finally, the characteristics of the above aspects are comprehensively scored.

$p(d_i)$ represents the probability of document $d_i$ appearing in the dataset, and $p(z_k|d_i)$ represents the probability distribution of the topic of document $d_i$; $p(w_j|z_k)$ represents the probability distribution of words in topic $z_k$, each topic follows a polynomial distribution over all terms, and each document follows a polynomial distribution over all topics. Based on the above probability distribution, a document $d_i$ is randomly selected according to the document probability distribution $p(d_i)$ in the dataset, the topic $z_k$ is selected according to the document topic probability distribution $p(z_k|d_i)$, and then the word of the document $d_i$ is selected according to the keyword probability distribution $p(w_j|z_k)$ of the topic $z_k$. The data that we can observe are $(d_i, w_j)$, and $z_k$ represents the implicit variables. The joint distribution of $(d_i, w_j)$ is:

$$p(d_i, w_j) = p(d_i)p(w_j|d_i) \tag{7}$$

$$p(w_j|d_i) = \sum_{k=1}^{K} p(w_j|z_k)p(z_k|d_i) \tag{8}$$

The $p(z_k|d_i)$ and $p(w_j|z_k)$ distributions correspond to two sets of polynomial distributions. To estimate the parameters of these two sets of distributions, the expectation maximization algorithm should be used. The EM algorithm is divided into E steps and M steps, where the E step is used to solve the post probability distribution of the implicit variable $z_k$ when $d_i, w_j$ is known, and the formula is as follows:

$$p(z_k|d_i, w_j) = \frac{p(w_j|z_k)p(z_k|d_i)}{\sum_{k=1}^{K} p(w_j|z_k)p(z_k|d_i)} \tag{9}$$

The left side of the formula represents the probability of the occurrence of the *k*th implied topic under the probability of the occurrence of the *i*th document and the *j*th word. Step M is used to solve the posterior probability distribution $p(w_j|d_i)$ of topic words and topic documents when $p(z_k|d_i, w_j)$ are known. The formula is as follows:

$$p(w_j|z_k) \propto \sum_{i=1}^{N} n(d_i, w_j)p(z_k|d_i, w_j) \tag{10}$$

$$p(z_k|d_i) \propto \sum_{j=1}^{M} n(d_i, w_j)p(z_k|d_i, w_j) \tag{11}$$

It can be found from the above formula that the E and M steps of the expectation maximization algorithm depend on each other, and the three distributions of , , and can be obtained $p(z_k|d_i, w_j)$ $p(w_j|z_k)$ $p(z_k|d_i)$ after a continuous iterative solution. The scoring process of the composition scoring system is shown in Fig. 3.

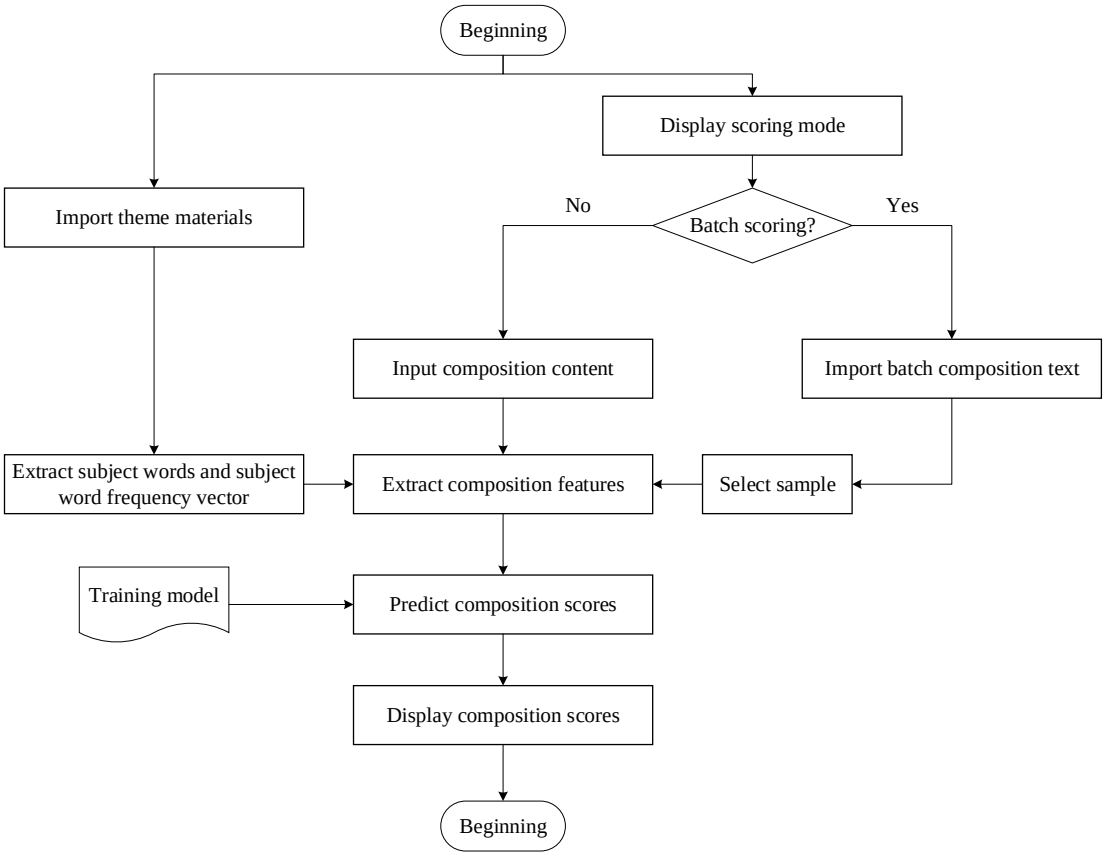


Fig. 3. Scoring process of the composition scoring system.

## IV. System Test and Analysis

### A. Evaluation and Verification of an Automatic Scoring System for Composition Based on Word Vector Clustering

To verify the scoring effect of the automatic scoring system based on word vector clustering, this paper takes the standard data provided in a composition scoring competition as the research object, and the total amount of processing of various samples is shown in Fig. 4. In this verification, all the samples except the two parts of the calibration set and the exception answer were scored by a computer. According to the results, the intelligent score of Chinese composition was 420070, accounting for 99.82% of the total sample size, and the intelligent score of English composition was 418820, accounting for 99.53% of the total sample size. The abnormal samples included high similarity with the Chinese text, high similarity with the current test paper (reading comprehension), and high similarity with each other. There were 235 Chinese compositions, accounting for 0.06% of the total test papers. English composition 1469, accounting for 0.35% of the total examination papers. The subject expert group conducts targeted quality inspection re-evaluation on abnormal samples.

In this paper, 235 Chinese compositions and 1469 English compositions were selected and matched with the standard target text. The comparison of the bit error rate of key word recognition is shown in Fig. 5. The statistical results are as follows: the recognition accuracy of Chinese characters is 97.6%, and the recognition accuracy of English words is 97.3%. This high-precision transliteration recognition has three important factors: first, examinees' attention to the composition of the college entrance examination ensures the standardization of writing. Second, the Chinese composition area is designed in square paper format, and the English composition area is designed in a line-by-line underline format to ensure the writing position of characters. Third, there are advanced recognition algorithms. These three factors can ensure the accurate recognition of all the scoring samples, and the overall transliteration recognition rate should be kept at approximately 97%, which can meet the actual requirements of marking papers.

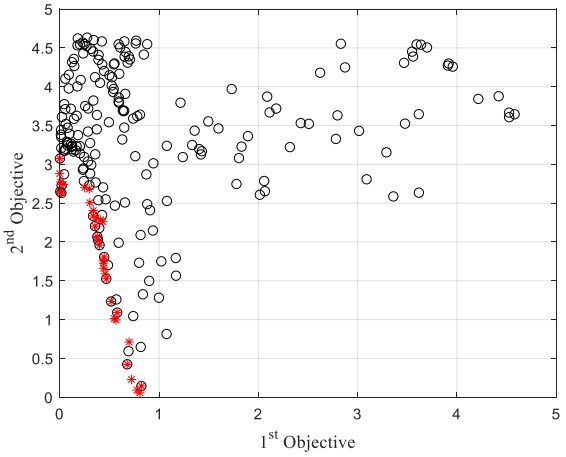


Fig. 4. Total amount of processing of various samples.

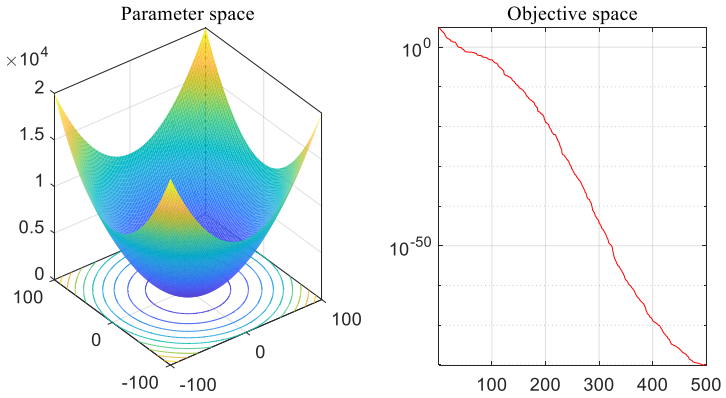


Fig. 5. Comparison of the bit error rate of key word recognition.

### B. *Evaluation and Comparison of Automatic Scoring Algorithms for Composition*

The comparison of the training time of different algorithms is shown in Fig. 6. The average score of the machine score and the average score of the manual score were less than 1 point, and the standard deviation was basically the same. The correlation between the machine score and report score is 0.95, and the consistency rate is 95.24%, which is very close to the correlation and consistency rate of Manual 1 and Manual 2 and is at a high level, which proves that the overall effect of the intelligent score is good. At the same time, it also shows from another angle that the intelligent marking system has a very high learning ability for the calibration set and basically reaches the level of mastering the marking standards of the teachers. It can be seen that most students do not resist the use of correcting network writing, and the correcting network has a good impression. The correcting network writing system makes sense and can stimulate students' desire for writing and continuous improvement. This model provides statistics on the content quality, coherence, readability and basic information of the composition and gives students detailed feedback so that students can better understand their own composition.

As seen from the prediction results of automatic scoring based on the random forest model in the figure, in the scoring results of the composition subset based on the random forest algorithm, the quadratic weighted K value is generally above 0.78, the highest value is 0.905, and the average value is 0.862. The lowest value of weighted K obtained by the international scoring algorithm is 0.654, the highest value is 0.755, and the average value is 0.792. In terms of the prediction results, the calculation method in this paper is obviously better than the existing prediction model, 10%~18% higher than the general algorithm, and can basically achieve the matching effect with the artificial score. Further analysis of the composition sample structure shows that the random forest algorithm based on the bagging method can effectively avoid overfitting error in the case of an insufficient sample size after obtaining accurate clustering vector features, thus reducing the variance value. When the number of samples is less than 1400, the quadratic weighted K value of the conventional model prediction algorithm decreases obviously and is basically lower than 0.7. The average score deviation rate of each formula combination is shown in Fig. 7.

This model not only gives the composition of the overall score as the composition of the machine but also, in this paper, extracts feature feedback to assist students in better understanding their writing level. This process uses internet technology to achieve a short score. With this composition, students can receive reliable information in a timely and effective manner, which is also one of the advantages of intelligent reading systems. As shown in Fig. 8, the score of the essay given by the machine is 8.41; while the score given by the teacher manually is 9.0, indicating that the consistency between the essay and the manual score is very high. The text coherence is 0.25, 0.77, 0.62 and 1.0, indicating that the text coherence is very good. The readability of the text is 11.89, the comprehensibility is 193.98, and the writing level is 10.82. In the linguistic features of the text, the total number of words is 204, word density is the ratio of the number of different words to the total number of words, word density is 0.66, complex word digit number is 36, sentence number is 17, paragraph number is 3, and average sentence length is 12. All these features can effectively give students an intuitive understanding of their own compositions. Therefore, the students know more about their own composition to improve the efficiency of their learning. The impact of different training sets on model performance is shown in Fig. 8.

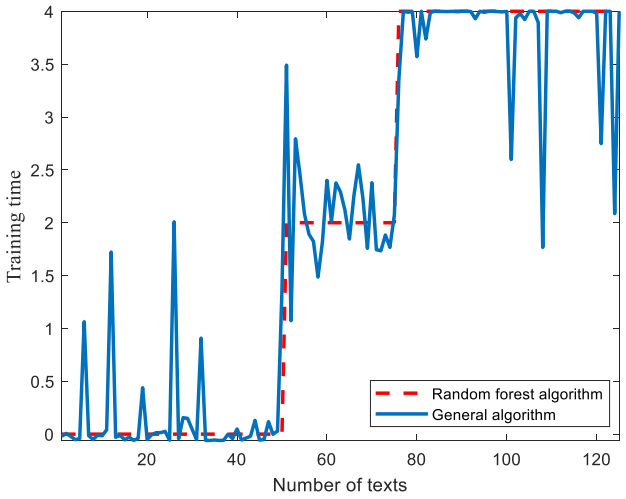


Fig. 6. Comparison of training time of different algorithms.

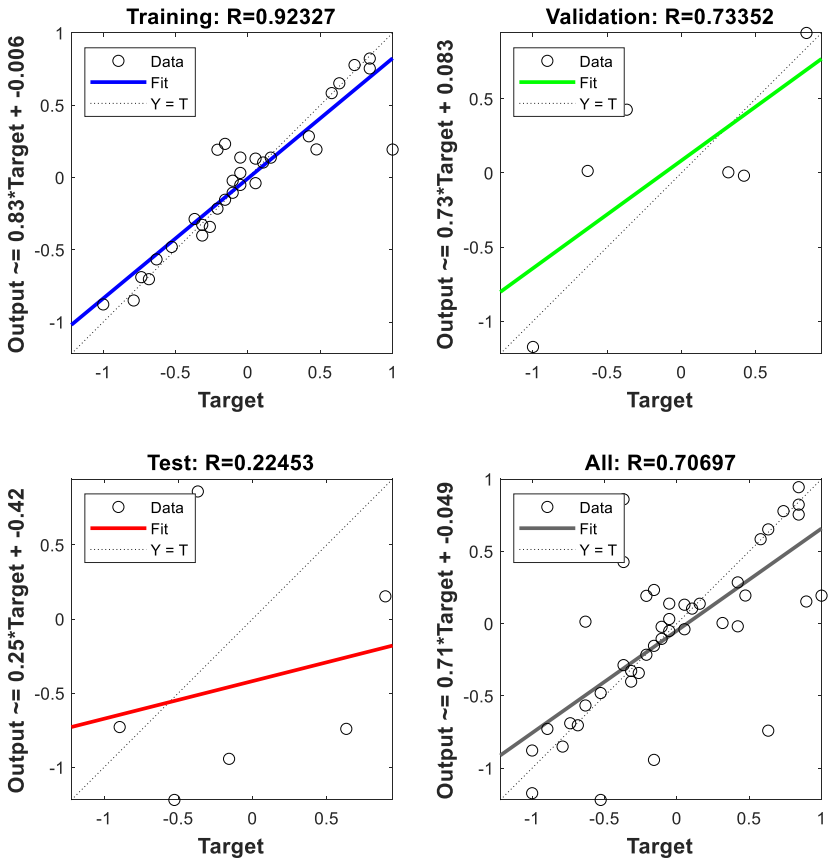


Fig. 7. Average score deviation rate of each formula combination.

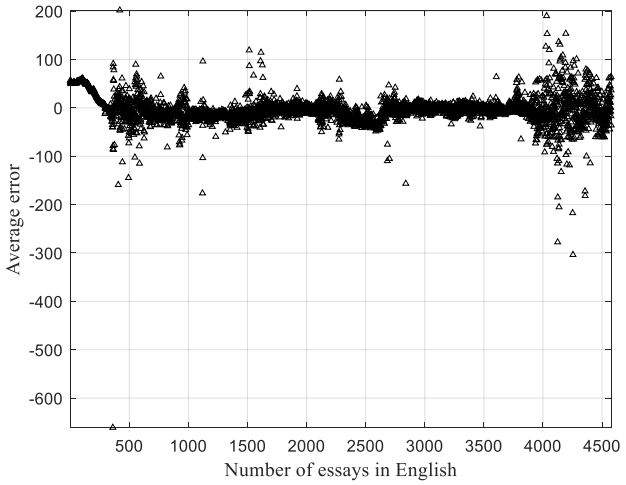


Fig. 8. Impact of different training sets on model performance.

## V. Analysis

An AI-based English essay grading system utilizes artificial intelligence technology to automatically evaluate and score essays written in the English language. This advanced system has gained popularity due to its potential to efficiently assess a large number of essays with consistent accuracy, saving time for educators and providing immediate feedback to students. The advantages of an AI-based English essay grading system lie in its consistency, time efficiency, scalability, immediate feedback, objective evaluation, enhanced learning experience, consistent feedback, and analytics capabilities. While it has limitations in contextual understanding and ethical considerations, integrating this technology with human expertise can lead to a balanced approach that leverages the strengths of both artificial intelligence and human judgment to improve the overall quality of writing assessment and instruction.

## VI. Conclusion

With the further development of global economic integration and internet technology, English has become the mainstream language for people worldwide to communicate. Therefore, an increasing number of governments are paying attention to the learning of English. Early essay scoring was performed manually, that is, when some teachers read the composition of the paper-based test answers directly to score them, the results are affected by the fact that some teachers' subjective factors are quite large, which is almost impossible to achieve in the large exam score standard consistency requirements, and given how much is required by the evaluation objective, impartiality is more difficult to achieve. Over time, the method of writing assessment has reached the stage of combining intelligent correction with teacher correction. To date, the introduction of artificial intelligence into college English writing teaching can effectively assist teachers in completing composition correction and data analysis more efficiently to a certain extent, encourage students to evaluate, learn from each other more conveniently, and cultivate students' awareness of independent English

writing and lifelong learning. Therefore, this paper examines how to apply an automatic scoring scheme to composition scoring more effectively and provides some reference for research on the automatic scoring direction of English composition under the background of artificial intelligence. The main research contents and conclusions are summarized as follows:

*1)* Natural language processing technology is used to analyse the text features of a composition, including part of speech tagging, syntactic analysis and article organization structure analysis.

*2)* Through the correlation analysis of the statistical features, the features related to scoring were extracted, and the relevant features were classified according to the different angles of the features reflecting the quality of the composition.

*3)* Semantic analysis technology was combined to analyse the semantic content of the composition and obtain the content quality characteristics of the composition.

*4)* The neural network is successfully applied in the automatic composition scoring process, and the mapping relationship between text features and composition scoring is obtained.

*5)* A comprehensive score for the composition was given based on its morphology, syntax, organizational structure and content quality.

## VII. Future Work

This paper can not only directly evaluate the language quality and organizational structure of a composition but also evaluate the content quality of the composition and conduct statistical analysis of the lexical use, syntactic use, semantic coherence and readability of the composition. Students and teachers can quickly obtain the quantitative indicators of a composition, providing a powerful reference for teachers and students and not only greatly reducing teachers' workload but also increasing the number of English learners' writing. Due to the large number of English learners in China and the shortage of teacher resources, the English composition scoring system has great application value. Although this model has certain functions, it still has certain deficiencies, which are embodied in the following aspects:

*1)* In this article, the model used in the training process of neural networks included a Chinese English learner corpus, and 1000 compositions were chosen. Among them, 720 were used as the training set, the training set for neural network training was somewhat smaller, and the composition of the training set was not rich. If the training set were larger, the training content would be richer and the performance of the neural network and the accuracy of scoring would be improved.

*2)* In the semantic analysis module, this model uses more than 20 high-scoring compositions with different topics as sample articles. If more sample compositions with different topics were added, the accuracy of semantic scoring would be greatly improved, and the comprehensive scoring accuracy of compositions would also be greatly improved.

## References

[1] Zhou J, Liao H, Liu LX. A new basic thoracoscopic surgical skill training and assessment system using automatic scoring techniques. Surgical Endoscopy and other Interventional Techniques, 2021, 36(5): 3076-3086.

[2] Varban OA, Thumma JR, Dimick JB. Evaluating the effect of surgical skill on outcomes for laparoscopic sleeve gastrectomy a video-based study. Annals of Surgery, 2021, 273(4): 766-771.

[3] Kim JE, Park K, Jung SY. Automatic scoring system for short descriptive answer written in Korean using lexico-semantic pattern. Soft Computing, 2018, 22(13): 4241-4249.

[4] Choi WY. An application of an AI chatbot automatic pronunciation scoring system to elementary school students. English Language Assessment, 2021, 16 (2): 167-185.

[5] Kim DS. Automatic scoring system for picture-based English caption writing test adopting deep learning based word-embedding. The Linguistic Association of Korea Journal, 2021, 29(2): 1-20.

[6] Yuan Z. Interactive intelligent teaching and automatic composition scoring system based on linear regression machine learning algorithm. Journal of Intelligent & Fuzzy Systems, 2021, 40(2): 2069-2081.

[7] Caroline G, Schlemmer P, Rodrigues E. Evaluation of nutritional status and causes of mal-nutrition. Archivos Latinoamericanos De Nutricion, 2019, 69(5): 317-326.

[8] Chen ZG, Wang YJ, Chen J. Value of distal high signal vessel sign of FLAIR sequence in the establishment of secondary collateral circulation after cerebral infarction. Boletin De Malariologia Y Salud Ambiental, 2019, 59(5): 86-91.

[9] Srinivasu PN, Rao TS, Olariu I. A comparative review of optimisation techniques in segmentation of brain MR images. Journal of Intelligent & Fuzzy Systems, 2020, 38(5): 6031-6043.

[10] Hulber T, Kocsis ZS, Pesznyak C. A scanning and image processing system with integrated design for automated micronucleus scoring. International Journal of Radiation Biology, 2020, 96(5): 628-641.

[11] Arora V, Lahiri A, Reetz H, Phonological feature-based speech recognition system for pronunciation training in non-native language learning. Journal of the Acoustical Society of America, 2018, 143(1): 98-108

[12] Fogerty D, Madorskiy R, Dubno JR. Comparing speech recognition for listeners with normal and impaired hearing: simulations for controlling differences in speech levels and spectral shape. Journal of Speech Language and Hearing Research, 2020, 63(12): 4289-4299.

[13] Hong YJ, Hosung N. Evaluating score reliability of automatic English pronunciation assessment system for education. Studies in Foreign Language Education, 2021, 3(1): 91-104.

[14] Bequette M, Cardiel C.L.B., Cohn S. Evaluation capacity building for informal stem education: working for success across the field. New Directions for Evaluation, 2019 32(161): 107-123.

[15] Nikkonen S, Korkalainen H, Toyras J. Automatic respiratory event scoring in obstructive sleep apnea using a long short-term memory neural network. IEEE Journal of Biomedical and Health Informatics, 2021, 25(8): 2917-2927.

[16] Fu SX, Gu HM, Yang B. The affordances of AI-enabled automatic scoring applications on learners' continuous learning intention: An empirical study in China. British Journal of Educational Technology, 2020, 51(5): 1674-1692.

[17] Lee Y. An analysis of the influence of block-type programming language-based artificial intelligence education on the learner's attitude in artificial intelligence. Journal of the Korean Association of Information Education, 2019, 23(2): 189-196.

[18] Wilhelm D, Bouarfa L, Padoy N. Artificial intelligence in visceral medicine. Visceral Medicine, 2020, 36(6): 471-475.

[19] Baek S, Lee HJ, Kim H. Analysis of artificial intelligence's technology innovation and diffusion pattern: focusing on USPTO patent data. The Journal of the Korea Contents Association, 2020, 20 (4): 86-98.

[20] Kim GS. A study on the trust of artificial intelligence. Korean Journal of Local Government & Administration Studies, 2020, 34 (3): 21-41.

[21] Lew M,d Wilbur DC. A novel approach to integrating artificial

intelligence into routine practice. Cancer Cytopathology, 2021, 129(9): 677-678.

[22] Westcott RJ, Tcheng JE. Artificial intelligence and machine learning in cardiology. Jacc- Cardiovascular Interventions, 2019, 12(14): 1312-1314.

[23] Ahmed A, Agarwal S. Teaching an old dog new tricks: three-dimensional visual spatialisation of viscoelastic testing and artificial intelligence. Anaesthesia, 2020, 75(8): 1006-1009.

[24] Jang C, Sung W. A study on policy acceptance intention to use artificial intelligence-based public services: focusing on the influence of individual perception & digital literacy level. Informatization Policy, 2022, 29 (1): 60-83.

[25] Wilhelm D, Padoy N. Artificial Intelligence in Medicine: Passing Hype or the Holy Grail of Solutions? Visceral Medicine, 2020, 36 (6): 425-427.

[26] Kim SA. Research trends in elementary and secondary school artificial intelligence education using topic modeling and problems in technology education. The Korean Journal of Technology Education, 2021, 21(1): 106-124.

[27] Bhattacharya S, Pradhan KB, Singh A. Artificial intelligence enabled healthcare: A hype, hope or harm. Journal of Family Medicine and Primary Care, 2019, 8(11): 3461-3464.

[28] Teramoto A. Application of artificial intelligence in radiology. Gan to Kagaku Ryoho. Cancer & Chemotherapy, 2019, 46(3): 418-422.

[29] Lu HM, Li YJ, Serikawa S. Brain intelligence: go beyond artificial intelligence. Mobile Networks & Applications, 2018, 23(2): 368-375.

[30] Kim T, Jung HR. A study on the utilization of artificial intelligence technology and technology management in criminal justice procedures. Journal of Business Administration & Law, 2020, 31(1): 581-610