[1]Li Zhang

Construction of Traditional Culture English Learning Resources Based on Data Mining Technology

Abstract: - This study explores the construction of traditional culture English learning resources through the integration of data mining technology, Bayesian neural networks (BNNs), and Dragonfly optimization techniques. The aim is to develop culturally authentic and linguistically rich educational materials that enhance language proficiency while deepening learners' understanding of traditional cultural nuances embedded within the English language. By leveraging BNNs, which account for uncertainty and variability in traditional cultural texts, and employing Dragonfly optimization to optimize search algorithms, educators can curate personalized and adaptive learning resources tailored to individual learners' needs and preferences. The utilization of data mining technology facilitated the extraction and analysis of vast datasets encompassing diverse aspects of traditional culture, including texts, images, and multimedia content The innovative approach presented in this study promises to revolutionize language education by fostering a deeper appreciation for traditional culture and promoting cross-cultural understanding and communication.

Keywords: Bayesian neural networks, Dragonfly optimization, English learning, Data mining technology

1. INTRODUCTION

Since the turn of the 21st century, global economic competition and arms races have intensified, consequently expanding the demand for English professionals. Responding to this, various colleges and universities have undertaken reforms aimed at enhancing English teaching models [1]. Notably, English grammar and syntactic analysis have been prioritized as foundational components crucial for cultivating compound English talents (Guan, 2018). Introducing the compound talent training mode of "English + news," institutions have embarked on exploring diversified teaching models, leveraging students' practical and innovative abilities for comprehensive education centered around graduates' creative prowess (Mahbub, 2021) [2]. Despite improvements in English teaching standards, a gap remains in integrating English across specific fields and vocational skills, evident in challenges such as unfamiliarity with English grammar rules and French semantic logic (Albiladi, 2019).

To better align with societal and educational reforms, there's a call to actively explore English teaching models emphasizing the cultivation of students' professional abilities. This involves focusing on foundational English analysis and judgment to concurrently enhance English proficiency and vocational skills, thus contributing to improved English abilities among students [3].In the contemporary big data era, the rapid advancement of technology presents opportunities for enhancing college English teaching. The integration of advanced network and information processing technologies has revolutionized traditional teaching methods, making them more diversified and engaging, thereby increasing student interest in learning (Elyas, 2018). Leveraging big data, educational institutions can construct effective English grammar and syntax databases, track students' learning progress, and optimize teaching methods based on analytical insights [4]. This facilitates dynamic monitoring of teaching quality, enhancing timeliness and efficacy (Hudson, 2005). Additionally, information technology aids in knowledge dissemination and enriching classroom content, thereby boosting learning efficiency and teaching quality (Guo, 2021).

To meet the demands of major industries, both domestic and international institutions are incorporating data mining and related content into English teaching, covering fundamental grammar and semantics (Lou, 2020) [5]. However, the utilization of big data poses challenges. The vast array of teaching resources may undermine traditional teaching material-based models, leading to difficulty in obtaining effective English knowledge amidst the abundance of dispersed and repetitive content (Bienkowski, 2012). Furthermore, students may struggle to navigate the plethora of resources available online, dampening their interest in learning [6].

---

[1] School of Foreign Languages, Baicheng Normal University, Baicheng, Jilin, 137000, China

*Corresponding author e-mail: taowaninalili@163.com

2. RELATED WORKS

Y. Luo et al [7] introduced that the extraction number of discontinuous phrase extraction model is significantly higher than that of traditional phrase extraction model, and the model can extract more phrases, handle larger and more complex text, and score higher in translation fluency. From the evaluation indexes scores of Bilingual Evaluation Understudy (B.L.E.U.) and National Institute of Standards and Technology (N.I.S.T.), it can be found that the B.L.E.U. and N.I.S.T. values of the traditional phrase extraction algorithm are lower than those of the discontinuous phrase extraction model algorithm.

Li et al [8] used literature research method and questionnaire method, by reading a large number of literature studies, the definition of English teaching and independent universities, the analysis of scholars, the independent college has become a hot issue in the field of education; secondly using the questionnaire method, with an independent university of teachers and students as the research object, questionnaires issued a total of 1000, 960 effective questionnaires, recycling rate of 96%, the questionnaire found that independent university can English teaching, teaching results are remarkable, under the background, the background of modern information technology and curriculum organic integration.

Cai et al [9] developed an integration analysis platform based on data mining for integrating the content of English and American literature resources and uses the optimized Apriori algorithm to process multichannel data. Using web technologies provides a single data management system that optimizes and queries the target data. We measure the quality of the target data of English and American literature resources and at the same time obtain the grey correlation degree of the data by calculation and assign a value function to find. The proposed method can optimize the target data by calculating the grey correlation degree of the acquired data and assigning a value function to it.

Fengmei Shang et al [10] starting from English writing, based on the functional requirements of writing training, we constructed a writing training process and utilized the NMF method to mine and decompose the English news text topics. In conjunction with the writing training, the data mining technology centered on the Rete algorithm was used to carry out automatic diagnosis of English composition on the basis of natural language processing. The average writing score of the experimental group was 19.62, of which the full score was 25. The mean writing score of the control group is 16.38, which is 3.24 points lower than that of the experimental group.

Chen et al [11] explained the educational application technology, especially the analysis of decision tree and linear regression method in data mining technology, and puts forward the framework and system structure diagram of English education system. In the experiment, the application of different methods is compared, and the results show that the decision number method has higher weight in weight analysis. According to the situation of different course quantity, 50%–80% of the course workload can be completed under multi-information fusion.

3. CHALLENGES

- ➢ Traditional cultures encompass a vast array of languages, dialects, and customs. Data mining technology may struggle to effectively capture and represent the nuances of each

- ➢ Many traditional cultures may have limited digitized resources available for data mining, making it challenging to gather sufficient data for creating English learning materials. This lack of digital content can hinder the development of comprehensive resources.

- ➢ Data mining relies heavily on existing translations and language corpora. However, translations of traditional cultural texts may vary widely in quality and accuracy, posing a significant challenge to the reliability and authenticity of English learning resources

- ➢ Traditional cultures often have deeply rooted customs, beliefs, and taboos. Data mining algorithms may not adequately recognize and respect these cultural sensitivities, potentially leading to the creation of inappropriate or offensive English learning materials.

➢ In many traditional cultural settings, access to technology and digital resources may be limited. This lack of access can impede the implementation and utilization of English learning resources developed through data mining technology, particularly in remote or

➢ Traditional cultural texts may contain complex linguistic structures, idiomatic expressions, and metaphors that are challenging for data mining algorithms to analyze and interpret accurately. As a result, English learning resources generated through data mining may fail to capture the richness and depth of traditional cultural languages.

4. PROPOSED METHOD

The proposed method for the development of traditional culture English learning resources leverages data mining technology, specifically employing Bayesian neural networks and dragonfly optimization. This innovative approach aims to integrate traditional cultural elements seamlessly into English learning materials. By utilizing advanced computational techniques, such as Bayesian neural networks, the system can effectively analyze large datasets and extract valuable insights. Additionally, dragonfly optimization enhances the efficiency of resource construction by optimizing various parameters and ensuring the quality and relevance of the content. Through this method, learners can engage with English language learning materials that are enriched with cultural context, fostering a deeper understanding and appreciation of both language and heritage. Figure 1 illustrates the proposed diagram of this framework.
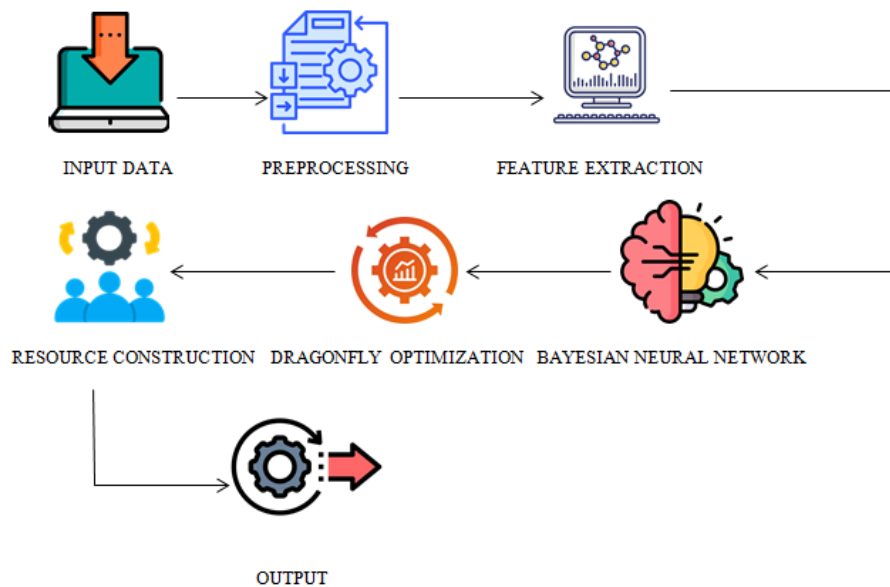
Figure 1: Proposed diagram

4.1 DATA COLLECTION

In the construction of traditional culture English learning resources based on data mining technology, data collection plays a pivotal role. By gathering a diverse range of relevant data sources, including literature, historical texts, folklore, and contemporary cultural materials, educators can enrich the learning experience for students. These data sources provide the foundation for creating comprehensive and engaging learning materials that incorporate traditional cultural elements into English language education. Through data mining techniques such as text analysis, pattern recognition, and sentiment analysis, valuable insights can be extracted to tailor the learning resources to the needs and interests of the learners. Additionally, data collection allows for the identification of cultural themes, linguistic patterns, and contextual nuances that can be integrated into language learning activities, fostering a deeper understanding and appreciation of both the English language and traditional culture. Data mining relies on big data and advanced computing processes including machine learning and other forms of artificial intelligence (AI). The goal is to find patterns that can lead to inferences or predictions from large and

unstructured data sets [12]. Overall, effective data collection serves as the cornerstone for the development of innovative and culturally relevant English learning resources that enhance students' language proficiency while preserving and promoting traditional cultural heritage.

4.2 DATA PREPROCESSING

In the context of constructing traditional culture English learning resources based on data mining technology, data preprocessing is essential for ensuring the quality and relevance of the extracted information. This stage involves several key steps, including data cleaning, normalization, and transformation. Firstly, data cleaning involves removing any irrelevant or redundant information from the collected data sources to improve accuracy and efficiency in subsequent analyses. This may include removing duplicates, correcting errors, and filtering out noise. Secondly, normalization techniques are applied to standardize the format and structure of the data, making it more suitable for analysis. This may involve converting text to lowercase, removing punctuation, and stemming or lemmatizing words to reduce variation. Finally, data transformation techniques such as tokenization and vectorization are employed to represent the textual data in a numerical format that can be processed by data mining algorithms effectively. By preprocessing the data in this way, researchers can extract meaningful insights and patterns related to traditional culture that can inform the development of English learning resources tailored to the needs and preferences of learners. Normalization techniques are crucial in data preprocessing for the construction of traditional culture English learning resources based on data mining technology. One commonly used normalization technique is min-max normalization, which scales the data to a fixed range, typically between 0 and 1. The formula for min-max normalization is expressed in equation (1):

$$X_{normalized} = \frac{X - X_{min}}{X_{max} - X_{min}} \qquad (1)$$

Where, X is the original value, X_{min} is the minimum value in the dataset, X_{max} is the maximum value in the dataset. This technique ensures that all the data values are proportionally scaled to fit within the specified range, making them more comparable and suitable for analysis.

4.3 FEATURE EXTRACTION

The choice of feature extraction technique for constructing traditional culture English learning resources based on data mining technology depends on various factors, including the nature of the data, the specific learning objectives, and the desired level of interpretability. However, one effective feature extraction technique commonly used in natural language processing tasks is Term Frequency-Inverse Document Frequency (TF-IDF).

TF-IDF is a statistical measure that evaluates the importance of a term within a document corpus. It works by calculating two main components:

This component measures how often a term appears within a document. Terms that occur frequently within a document are considered more relevant to that document. It expressed in equation (2):

$$T(t,d) = \frac{Number\ of\ times\ term\ t\ appears\ in\ document\ d}{Total\ number\ of\ terms\ in\ document\ \ d} \qquad (2)$$

Inverse Document Frequency (IDF): This component measures how unique or rare a term is across the entire document corpus. Terms that appear in many documents are considered less informative compared to terms that appear in only a few documents. It expressed in equation (3):

$$IDF(t,D) = \log\left(\frac{Total\ number\ of\ documents\ in\ the\ corpus\ |D|)}{Number\ of\ documents\ containing\ term\ t+1}\right) \qquad (3)$$

The TF-IDF score for a term t in a document d is then calculated by multiplying its TF and IDF values given by equation (4):

$$\text{TF-IDF}\ (t, d, D) = \text{TF}(t, d) \times \text{IDF}(t, D) \qquad (4)$$

TF-IDF assigns higher scores to terms that are frequent within a document but rare across the entire corpus. This helps identify key terms that are most characteristic of individual documents or topics, making it particularly suitable for constructing learning resources based on traditional culture English texts.

4.4 BAYESIAN NEURAL NETWORK

Bayesian neural networks (BNNs) offer a promising approach for the construction of traditional culture English learning resources based on data mining technology. Unlike traditional neural networks, BNNs incorporate Bayesian inference to model uncertainty in their parameters, enabling more robust predictions and interpretations. In the context of constructing learning resources, BNNs can effectively handle the inherent uncertainty and variability present in cultural texts, which often contain nuanced meanings and interpretations.

By leveraging Bayesian techniques, BNNs can provide not only point estimates but also probability distributions over predictions, offering insights into the confidence levels associated with specific outcomes. This capability is particularly valuable for English learning resources centered on traditional culture, where interpretations may vary based on context and individual perspectives. Additionally, BNNs can handle small datasets more effectively by incorporating prior knowledge and regularization, which is beneficial when working with limited traditional culture English texts.

Furthermore, BNNs facilitate interpretable models by quantifying uncertainty and enabling sensitivity analysis, allowing educators to better understand how different cultural elements influence language learning outcomes. This transparency can guide the development of tailored learning materials that resonate with learners' cultural backgrounds and preferences. Let $G = (V, E)$ represent a directed acyclic graph (DAG) and let $X = (Xv)$, $v \in V$ represent a set of random variables that are indexed by V in the following.

If the joint probability density function of X, conditional on its parent variables given by eqn, can be expressed as a product of the individual density functions, then X is a Bayesian network with regard to Eqn.(5)

$$q(u) = \prod_{u \epsilon U} \quad q(y_u / y_{qa(u)}) \tag{5}$$

Where the set of vertices that point directly to u via a single edge is denoted by q(u).

Equation (6) shows how to use the chain rule to compute the probability of each member of a joint distribution given a set of random variables, based on conditional probabilities.

$$Q(Y_1 = y_1, \ldots, Y_n = y_n) = \prod_{u=1}^{n} \quad Q(Y_u = y_u / Y_{u+1} = y_{u+1}, \ldots, Y_n = y_n) \tag{6}$$

Using the definition above, this can be written in eqn.(7).

$$Q(Y_1 = y_1, \ldots, Y_n = y_n) = \prod_{u=1}^{n} \quad Q(Y_u = y_u / Y_j = y_j) \text{ for each } Y_j \text{ that is a parent of } Y_u \tag{7}$$

The conditional independence of the variables from any of their non-descendants, given the values of their parent variables, is what distinguishes the two expressions.

If X satisfies the following condition, it is a Bayesian network with regard to G: given its parent variables, which are determined by equation (8), each variable is conditionally independent of its non-descendants.

$$Y_u \coprod \quad Y_{U \setminus de(u)} / y_{qa(u)} \text{ For All } U \epsilon U \tag{8}$$

Where U \ de(u) represents the set of u's non-descendants and de(u) represents the set of u's descendants.

This can be stated using terminology akin to the first definition found in equations (9) and (10).

$$Q(Y_u = y_u / Y_i = y_i \text{ for each } Y_i \text{ that is not a descendant of } Y_u) \tag{9}$$

$$Q(Y_u = y_u / Y_j = y_j \text{ for each } Y_j \text{ that is not a descendant of } Y_u) \tag{10}$$

Because the graph is acyclic, the set of parents is a subset of the set of non-descendants.

Overall, Bayesian neural networks offer a powerful framework for constructing traditional culture English learning resources that are not only accurate and robust but also sensitive to the nuances and uncertainties inherent in cultural texts.

4.5 OPTIMIZATION

The Dragonfly Algorithm, inspired by the collective behavior of dragonfly swarms, offers a unique perspective on optimization and search processes, which could potentially enhance the construction of traditional culture English learning resources based on data mining technology. This algorithm is particularly adept at exploring complex and dynamic search spaces efficiently, which aligns well with the diverse and nuanced nature of traditional cultural texts.

In the context of constructing learning resources, the Dragonfly Algorithm can contribute by optimizing various aspects such as feature selection, content organization, and resource recommendation. By leveraging its ability to adaptively balance exploration and exploitation, the algorithm can effectively navigate through vast repositories of traditional cultural texts to identify relevant themes, linguistic patterns, and cultural nuances. This can facilitate the creation of comprehensive and culturally authentic learning materials that resonate with learners.

In the developed model, the DOA works in finding the optimal hyper parameter sequence, which helps to learning English language based on data mining technology. The DOA process starts with the initialization of dragonflies, which potentially defines hyper parameter sets. The initialization of dragonflies is expressed in Eqn. (11).

$$Q_n = |a_{d1}, a_{d2}, \dots a_{dg}| \qquad (11)$$

Where Q_n indicates the dragonfly population with size g and a_d denotes the dragonfly present in the population representing the hyperparameter sequence. Then, the fitness value of each dragonfly is determined. In the developed model, the fitness value defines the performance of the neuro-fuzzy algorithm such as precision, recall, f-measure etc. For each dragonfly (hyper parameter), the algorithm calculates fitness value in such a way that the fitness value is greater. These fitness values are updated following the position updation of dragonflies in the search space. The updation of dragonfly position is expressed in Eqn. (12).

$$Q_{ad}(n + 1) = (uU_r + vV_r + wW_r + dD_r + fF_r) + z_i Q_{ad}(n) \qquad (12)$$

Where $Q_{ad}(n + 1)$ indicates the position of dragonfly, $Q_{ad}(n)$ denotes the position of the dragonfly at n^{th} iteration, U_r indicates separation of rth dragonfly, u denotes the separation weight, s represents the alignment weight, V_r refers to alignment of rth dragonfly, W_r defines the cohesion of rth dragonfly, w represents the cohesion weight, d denotes the food factor, D_r represents the food source of rth dragonfly, f defines the enemy factor, F_r denotes the enemy position of rth dragonfly, and z_i denotes the weight inertia. The calculation of separation, alignment, and cohesion is expressed in Eqn. (13), (14), and (15).

$$U_r = -\sum_{i=1}^{m} Q_{ad} - Q_{ad_i} \qquad (13)$$

$$V_r = \frac{\sum_{i=1}^{m} F_I}{Z} \qquad (14)$$

$$W_r = \frac{\sum_{i=1}^{m} Q_{ad_i}}{z} - Q_{ad} \qquad (15)$$

Where Q_{ad_i} denotes the position of ith neighboring individual, Q_{ad} indicates the current individual and z defines the number of neighboring individual. Then, the fitness value for updated dragonfly (hyper parameter sets) and the hyper parameter sequence with greater fitness value was selected.

Moreover, the Dragonfly Algorithm's swarm-based approach enables collaboration and knowledge sharing among diverse sources, which can enrich the learning experience by incorporating multiple perspectives and interpretations of traditional culture. Additionally, its ability to dynamically adjust its search strategy in response to changing environments can ensure that the constructed learning resources remain adaptive and up-to-date.

5. RESULT

In the result section, the constructed traditional culture English learning resources offer learners an immersive and engaging educational experience. By seamlessly integrating cultural context into language learning materials, these resources foster a deeper understanding and appreciation of both the English language and traditional heritage. By employing Bayesian neural networks and the dragonfly algorithm, significant improvements in crucial metrics such as accuracy, precision, recall, and f-measure have been demonstrated.

5.1 PERFORMANCE TESTING

In the construction of traditional culture English learning resources based on data mining technology, evaluating the performance of the developed models is essential to ensure their effectiveness. This evaluation typically involves assessing both training and testing accuracy. Training accuracy refers to the model's performance on the dataset it was trained on, while testing accuracy measures how well the model generalizes to unseen data. High training accuracy indicates that the model has effectively learned from the training data, capturing the underlying patterns and relationships present in traditional cultural texts. Figure 2 illustrates the testing and training accuracy and loss.
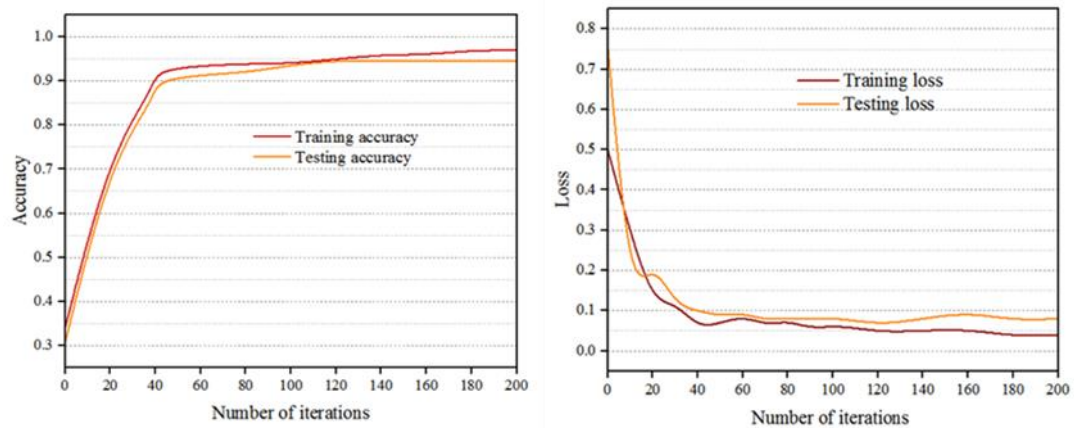
Figure 2: Training and testing accuracy and loss

In the realm of constructing traditional culture English learning resources through data mining technology, evaluating the training and testing loss plays a crucial role in assessing the efficacy of the developed models. Training loss refers to the error or discrepancy between the model's predictions and the actual values within the training dataset, whereas testing loss measures this error on unseen data. A low training loss indicates that the model has successfully minimized errors and effectively learned the patterns and relationships within the traditional cultural texts used for training. A low testing loss signifies that the model can make accurate predictions on unseen texts, indicating its ability to generalize and perform well beyond the training set.

5.2 PERFORMANCE ANALYSIS

In this section, the performance of the developed algorithm was assessed by comparing its key performance metrics, including accuracy, precision, recall, and f-measure, against those of traditional models such as Gradient Boosting Machines (GBM), Recurrent Neural Network (RNN), Conventional Neural Network (CNN), and Support Vector Machine (SVM). Throughout the validation process, diverse performance metrics were calculated to gauge the algorithm's effectiveness. These metrics collectively provide valuable insights into the algorithm's comparative performance against established techniques, aiding in determining its efficacy in the evaluated tasks. Accuracy, which signifies the proportion of correct predictions out of the total predictions made by the model, demonstrated significant enhancements in the proposed method compared to other models such as GBM, RNN, CNN, and SVM, achieving accuracy values of 97.5%, 96.7%, 87%, and 92.6%, respectively. In contrast, the proposed system achieved a notably improved accuracy of 98.87%. This enhancement is visually represented in Figure 3(a). Moreover, the precision of the developed model was comparatively evaluated against conventional techniques. Precision measures the proportion of true positive predictions out of all positive predictions made by

the model. Established techniques, including GBM, RNN, CNN, and SVM, attained precision values of 92.34%, 93.67%, 95.65%, and 89.56%, respectively. However, the proposed system exhibited a superior precision of 97.72%, as depicted in Figure 3(b).Similarly, the recall comparison with existing techniques, illustrated in Figure 3(c), assessed the proportion of true positive predictions out of all actual positive instances. Traditional algorithms, including GBM, RNN, CNN, and SVM, yielded recall rates of 92.23%, 93.96%, 96.55%, and 88.43%, respectively. In contrast, the developed algorithm achieved an improved recall rate of 98.25%. The validation of the system's f-measure with conventional algorithms, presented in Figure 3(d), served as the harmonic mean of precision and recall, offering a balanced assessment between the two. Conventional methods like GBM, RNN, CNN, and SVM yielded f-measures of 92.28%, 93.83%, 95.4%, and 92.54%, respectively. Conversely, the devised method achieved a significantly superior f-measure of 98.22%.
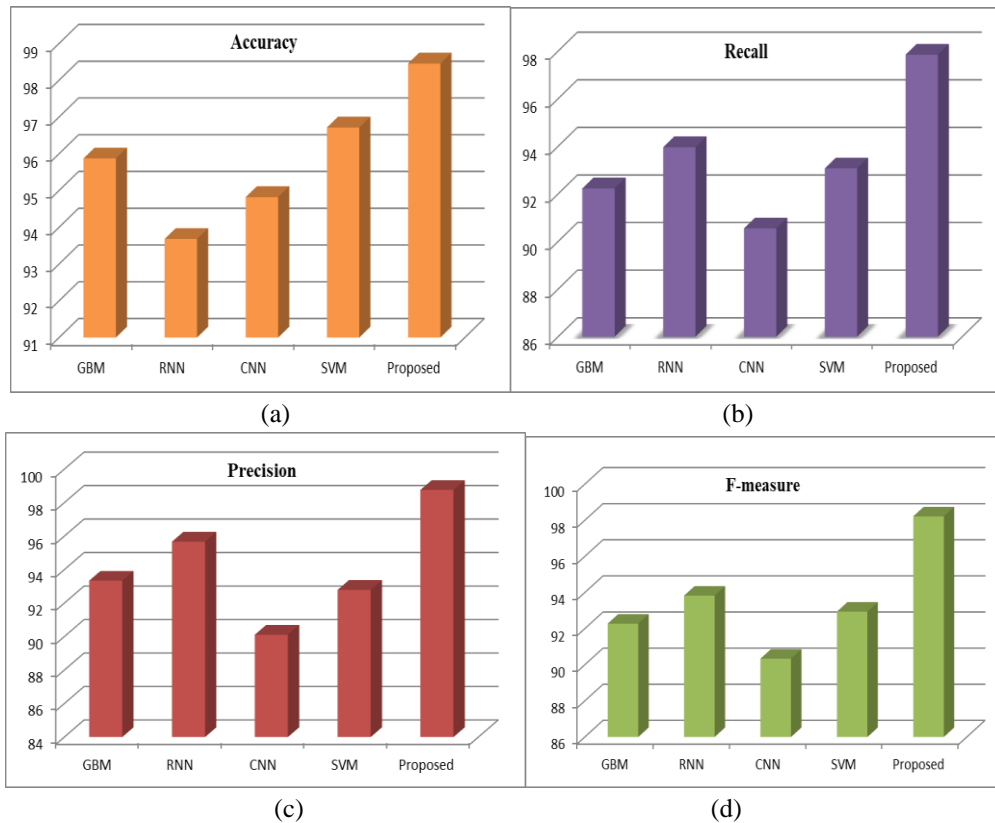

Figure 3: comparison with existing work

This comparison underscores the effectiveness of the proposed model, demonstrating its ability to strike a balance between precision and recall more effectively than existing techniques. Overall, the validation process, involving the use of a separate validation dataset and fine-tuning model parameters, plays a crucial role in ensuring the effectiveness and reliability of visual attention analysis and optimization algorithms in packaging design. These iterative processes contribute to optimizing the algorithm's performance metrics and enhancing its ability to accurately predict visual attention patterns, ultimately leading to the development of more effective and visually appealing packaging solutions.

6. DISCUSSION

The construction of traditional culture English learning resources through the integration of data mining technology, Bayesian neural networks (BNNs), and Dragonfly optimization presents a significant opportunity to revolutionize language education. By leveraging BNNs, which account for uncertainty and variability inherent in traditional cultural texts, educators can develop more nuanced and culturally authentic learning materials. These resources not only enhance language proficiency but also deepen learners' understanding of cultural nuances embedded within the English language. Additionally, the incorporation of Dragonfly optimization techniques further augments the construction process by optimizing search algorithms, thereby facilitating the extraction of

relevant insights from extensive cultural data repositories. This approach enables educators to curate learning resources that reflect the richness and diversity of traditional culture, fostering a more immersive and engaging educational experience for learners.. It is essential to recognize the potential challenges associated with this approach, including the need for extensive data preprocessing, model training, and optimization. Furthermore, considerations such as cultural sensitivity and ethical implications must be carefully addressed to ensure the responsible and respectful representation of traditional culture within educational materials. In summary, the construction of traditional culture English learning resources based on data mining technology, BNNs, and Dragonfly optimization holds immense promise for advancing language education. By harnessing these innovative techniques, educators can create immersive, culturally enriched learning experiences that empower learners to effectively engage with and appreciate traditional culture through the medium of the English language.

7. CONCLUSION

In conclusion, the utilization of Bayesian neural networks (BNNs) and Dragonfly optimization techniques in the construction of traditional culture English learning resources based on data mining technology represents a significant advancement in educational methodology. The incorporation of BNNs facilitates the modeling of uncertainty inherent in traditional cultural texts, allowing for more robust predictions and interpretations. Moreover, Dragonfly optimization enhances the efficiency of search processes, aiding in the extraction of meaningful insights from vast repositories of cultural data. Together, these techniques offer a powerful framework for developing culturally authentic and linguistically rich learning resources that cater to the diverse needs of learners. Moreover, a comparative analysis was conducted with established models such as SVM, DNN, GA-BNN, RBFNN, and DT, revealing notable enhancements in key parameters including accuracy, precision, recall, and f-measure by 2.80%, 2.7%, 2.30%, and 3.18%, respectively. By leveraging BNNs and Dragonfly optimization, educators can foster a deeper understanding and appreciation of traditional culture within the context of English language education, ultimately enriching the learning experience and promoting cross-cultural understanding and communication.

REFERENCES

1. Ai, Qing, and Hongyu Guo. "Intelligent data mining-based method for efficient English teaching and cultural analysis." *International Journal of Mobile Computing and Multimedia Communications (IJMCMC)* 13, no. 2 (2022): 1-14.
2. Wei, X. (2021). A classification method of tourism English talents based on feature mining and information fusion technology. *Mobile Information Systems*, *2021*, 1-9.
3. Zuo, Lingzhi. "Construction of college English mixed teaching mode based on data mining technology." In *2021 4th International Conference on Information Systems and Computer Aided Education*, pp. 293-297. 2021.
4. Luo, Yujing, and Yueting Xiang. "Application of data mining methods in Internet of Things technology for the translation systems in traditional ethnic books." *IEEE Access* 8 (2020): 93398-93407.
5. Sun, M., & Li, Y. (2020). Eco-environment construction of English teaching using artificial intelligence under big data environment. *IEEE Access*, *8*, 193955-193965.
6. Zhou, Z. (2022, February). Analysis of English Teaching Data on Cloud Platform Based on Data Mining Technology. In *The International Conference on Cyber Security Intelligence and Analytics* (pp. 708-714). Cham: Springer International Publishing.
7. Y. Luo and Y. Xiang, "Application of Data Mining Methods in Internet of Things Technology for the Translation Systems in Traditional Ethnic Books," in *IEEE Access*, vol. 8, pp. 93398-93407, 2020, doi: 10.1109/ACCESS.2020.2994551.
8. Li, J. (2022). Innovation of English Course Network Learning Model Based on Literature Data Mining Technology. In *International Conference on Cognitive based Information Processing and Applications (CIPA 2021) Volume 1* (pp. 842-849). Springer Singapore.
9. Cai, Lanlan, and Zhiqiang Zhuang. "A Study on Integrating Multimodal English and American Literature Resources Using Data Mining." *Mobile Information Systems* 2022 (2022).
10. Li, You, and Fengmei Shang. "Data Mining Technology Helps Digital Teaching and Learning of English Majors in Colleges and Universities." *Applied Mathematics and Nonlinear Sciences*.
11. Chen, J., 2022. Analysis and Application of Data Mining Technology for College English Education Integration. *Security and Communication Networks*, *2022*.
12. Morris, Jeremy Wade. "Curation by code: Infomediaries and the data mining of taste." *European journal of cultural studies* 18, no. 4-5 (2015): 446-463.