# Constitutive Artificial Neural Network for the Construction of an English Multimodal Corpus

Junhua Li
Teaching Department of Basic Courses, Hebei Vocational University of Industry and Technology
China
junhuaphd@gmail.com

Yuehua Li
Tianjin Foreign Studies University
China

Lihao Han
Tianjin Foreign Studies University
China

**Abstract:** *Multimodal corpus is a novel multimedia teaching tool in social development and educational reform process. It uses a range of multimedia components to build a wide-ranging English corpus and is mostly focused on computer and network technologies. The use of corpus in multimodal English-Chinese instruction is growing. The meaning, usage, and set of English and Chinese multimodality will be better understood with the aid of contemporary information technology, which will also enhance the initiative of autonomous learning. This paper builds a Multimodal English Corpus created on optimized-Constitutive Artificial Neural Network-Honey Badger Algorithm (MEC-CANN-HBA). The input data is collected via the dataset of Gutenberg Literary English Corpus (GLEC). The data are fed to pre-processing to remove the noise and enhance the input data utilizing Multivariate Fast Iterative Filtering (MFIF). The pre-processing output is given to the Feature extraction segment. The three significant features, such as text, audio and video are extracted based on Deep Wavelet Scattering Transform (DWST). After that, the extracted features are given to the multimodal fusion vector. The multimodal feature vectors are employed as the input data for categorization and to obtain the English poetry feature representation that integrates context characteristics. Finally, the features of output are used as the input data of Constitutive Artificial Neural Network (CANN) effectively categorizes as ideographic, phonetic, rhetorical and contextual. Honey Badger Algorithm (HBA) utilized for improving the weight parameter of CANN to check the classification of English poetry is current utterance. The proposed MEC-CANN-HBA approach attains 24.36%, 23.42%, 30%, 10.25% and 16.27% higher accuracy, and 26.61%, 28.50%, 23%, 18.33% and 21.24% greater precision rate, compared with existing methods, like Construction for Multiple Modal Corpus of College Students' Spoken English Using Semantic Concepts (CMC-CSSE-SC), Construction of multimodal poetry translation corpus under AdaBoost method (CMPTC-AdaBoost),Construction with Application of English-Chinese Multimodal Emotional Corpus utilizing Artificial Intelligence (CECMC-AI) respectively*.

**Keywords:** *Constitutive artificial neural network, deep wavelet scattering transform, honey badger algorithm, multivariate fast iterative filtering*.

*Received April 12, 2024; accepted November 14, 2024*
*https://doi.org/10.34028/iajit/22/2/15*

## 1. Introduction

English Poetry has played a significant role in defining world culture throughout China's lengthy history and seen as the core of Chinese culture. Poetry consumes an unfathomable beauty and vitality because of its rhythmic richness, creative and vivid language, author's complete sentiments and thoughts. Modern people are still interested in poetry recitation and admiration [17, 27, 32]. Literary researchers have also been interested in researching literature based on classical poetry [8, 9]. Ancient English poetry differs from contemporary Chinese poetry in terms of phonetic and metrical qualities [2, 12, 28]. Poems frequently possess word constraints include allusions, metaphors, and other literary devices that make them challenging to translate [4, 7, 10, 16].

Individuals are trained to produce translation procedures and criticism from a poetics standpoint for the genre of poetry [15, 28]. New translation ideas or trends take some time to catch on and become widely used in the translation of ancient poetry [3, 25, 29]. The structural list linguistic approach to translation is questioned and deconstructed throughout the entire translation community in the late 1980s, which is when this translation style first gained prominence in poetry translation field [11, 19]. Over ten years, a variety of translation theories establish application points in the translation of poetry, and competing development routes have formed [18]. English poetry has many differences from translations of more generic genres [4]. According to the literature [30], Chinese poetry is differed from English poetry under literary imagery, phonetic rhythm, genre, feeling, form [1, 21, 33]. Literature [6, 13, 31] examined the aspects of translating poetry from Russian to English, discovered the challenges encountered and their solutions, and contrasted the two languages' parallels and contrasts [23, 26].

Custom Scorpus Texts for Analysis and Literature analyzes how culturally bound lexical elements are handled in poetry translations between different languages. Poetry translation is a challenging task much more challenging than the translation of any other genre. Poetry translation necessitates the creation of a second text because of the poetic and unique nature of the language used and translator was accomplished in Chinese literature and translates the flow of unique poetry in other language. Poetry translation cannot be an inflexible or challenging process. It needs familiarity with modern English idioms, the capacity to adapt poetry language to everyday discourse, and the use of effective restitution tactics. There is enough rhyme to carry along the original text's melody [5, 38, 39]. English Poetry is impossible to translate in terms of rhyme and sound since two languages require different phonological systems and rhyme structures. Chinese and English possess dissimilar musical patterns that are impossible to translate in Chinese-English translations because Chinese emphasizes the meaning while English highlights the form. By using addition and subtraction to make up for the grammatical differences between the two languages, the translation is made to follow the rules of the target language.

The two fundamental groups into which present machine translation systems separated are rule created and statistical created machine translation technique. Machine translation systems use rule-based models translate in accordance with the rules that linguists have developed for various languages. The variety and complexity of languages make it difficult for linguists to develop translation rules that achieve high coverage rates, which cause translation systems to fall short of expectations [37, 41]. Setting up a machine translation system take years as well because creating translation rules needs a lot of work. The advancement of machine translation is constrained by these constraints. Because of this, academics have proposed a machine translation method established on statistics. Rather than being created by linguists to optimize translation rules, a statistically produced machine translation technique learns translation rules from a present corpus. The statistical created translation system has outperformed the rule created machine translation technique as translation corpus increased and the technique developed. As a result, it is gradually replacing it as the industry standard for machine translation.

The main contribution of this paper is,

- This study intends to create a multimodal corpus of English texts for researchers and poetry translation fans to easily access.
- This corpus's search reveals with 83.2 and 80.9 mean evaluation scores, five-line stanzas and ancient poems' ideographic translations perform best. With an average rating of 73.2, seven-line stanzas having better phonetic presentation. With the average score of 63.5, five-verse poetry had the greatest rhetoric.
- The CANN technique's multimodal translation corpus is an effective tool for translating poetry, providing substantial data support for studies on English poetry and holding substantial value for the study of Chinese poetry culture.

Remaining manuscript is structured as follows: section 2 provides related work on multimodal English corpus. The multimodal English corpus using machine learning method is proposed in the section 4. Section 5 evaluates experimental results. The final section provides the conclusion.

## 2. Related Work

A number of researches were suggested in the literature on multimodal English corpus categorization using deep learning; certain recent studies are reviewed here,

Liang *et al*. [22] have suggested Deep Learning-base construction of multimodal corpus in mobile edge computing for IoT devices. A deep learning model (English analysis model) was designed and trained on MEC servers using the built-in corpus. The experimental results from IoT-collected data supported the efficacy of the English sentiment analysis algorithm. While the method demonstrated high accuracy, it lacked precision in results.

Tian *et al*. [34] have presented semantic concepts-based construction of multiple modal corpus for college pupils' spoken English. The researchers considered at a number of factors by analyzing the conjunctions distribution in pupils' oral discourse, such as conjunctions types employed, how frequently they occur, and typical mistakes. This work underscores the importance of semantic cohesion in language learning and highlights the potential for multimodal corpora to inform instructional strategies. It has high computational time. It provides high recall.

Li and Shan [21] have developed Construction of a multimodal poetry translation corpus depending upon AdaBoost. Here, the ELM-AdaBoost technique suggested showcasing the effectiveness of integrating weak predictors to enhance performance in poetry translation. This work illustrated the significance of ensemble methods in multimodal corpus construction, demonstrating how different models can work synergistically to produce robust outputs. It has high precision but less f-score.

Zhou *et al*. [40] have presented artificial intelligence dependent construction with application of English-Chinese multimodal emotional corpus. the exploration of semantic representation technology in this study highlighted the challenges in capturing implicit emotions and language traits. The work emphasized the role of representation learning in improving model effectiveness, contributing to the broader discourse on multimodal data analysis and its implications for language processing. It has high f-core but less precision.

Yang *et al*. [36] have presented a Multimodal Aspect-Category Sentiment Analysis dataset (MACSA) along multimodal fine-grained aligned annotations. The MACSA database developed by this study marked a significant advancement in multimodal analysis, offering extensive annotations for text-image pairs. Their Multimodal Graph-Basis Aligned Model (MGAM) shows the effectiveness of fine-grained cross-modal fusion techniques. It has high computational time. It has high specificity.

Li and Okada [20] have presented Interpretable multimodal sentiment analysis using textual modality descriptions with large-scale language modes. This study translated visual elements into text descriptions then used large language models for English predictions. The method provided a novel approach to improving model interpretability, especially in applications involving facial and audio modalities. While the approach offered high accuracy, it lacked precision in classification outcomes.

## 2.1. Critical Analysis

The reviewed literature highlights significant gaps that this research aims to address, particularly the balance between precision and recall in multimodal corpus performance, as many studies achieved high accuracy at the expense of other metrics. Furthermore, issues of high computational time reported in several works limit scalability and practical application, necessitating exploration of optimization techniques for improved efficiency. Additionally, challenges in capturing implicit emotions indicate a need for enhanced representation learning methods to boost accuracy in emotion detection. While efforts to improve model interpretability exist, there remains a demand for frameworks that ensure both interpretability and precision in real-world applications. Lastly, the integration of diverse modalities is often overlooked; this study emphasizes a holistic approach, combining text, audio, and visual elements into a cohesive framework to enrich understanding and enhance the applicability of multimodal data analysis.

# 3. Proposed Methodology

In this section, describe the multi modal English corpus deep learning technique of CANN applied to the English poetry classification experiment. First, it outlines a system for categorizing the significant features using textual, visual and audio, for categorizing the poetry content of the text and images, and a listeners' need for a multimodal corpus fusion technique classifying poetry. Figure 1 shows the block diagram for proposed methodology and Figure 4 shows the stages of data processing.

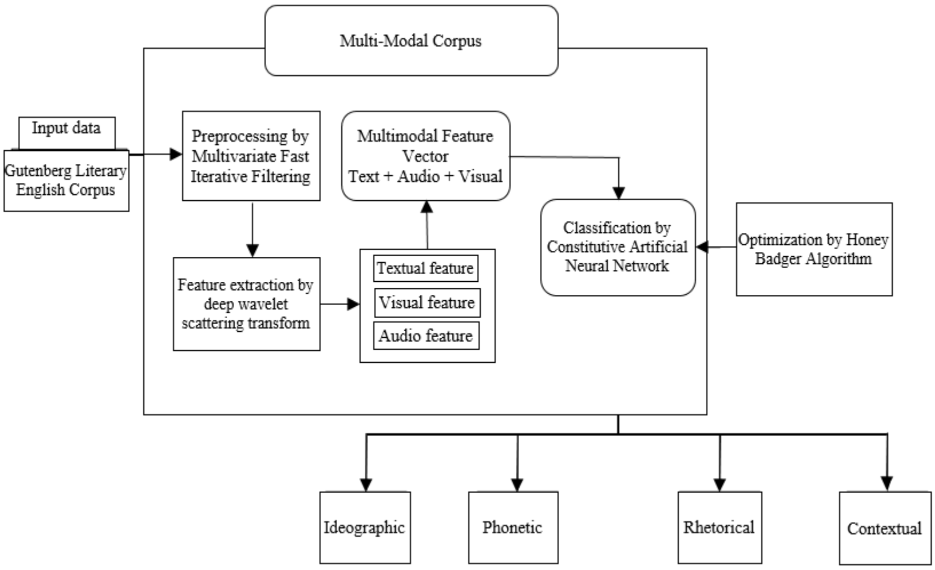


Figure 1. Block diagram of proposed English multimodal English corpus.

## 3.1. Data Collection

The GLEC [24] is a collection of texts from Gutenberg (https://web.eecs.umich.edu/lahiri/gutenberg_database.html), supplemented with texts from Shakespeare Online (http://www.shakespeareonline.com/sonnets/sonnetintroduction.html; see Jacobs *et al*., 2017). GLEC has almost three thousand English texts in the Gutenberg scheme. The fiction, non-fiction and works through authors like Austen, Byron, Bronte, Coleridge, Dickens, Darwin, Eliot, Einstein, Poe, Twain, Woolf Wilde are among the many genres covered by these texts.

## 3.2. Data Preprocessing using Multivariate Fast Iterative Filtering

In this section, MFIF [35] performs the data pre-processing. Multivariate Fast Iterative Filtering (MFIF) offers significant advantages in data preprocessing, particularly for multimodal datasets. It excels in noise removal, enhancing data quality by eliminating irrelevant information and adapting effectively to non-stationary data, which is common in modalities. Consider filter length $M$ that represents a half support length of a filter function of length of the MFIF. Calculate the filter length utilizing $a(g)$, the rotation angle of that vectors with respect to time in Equation (1).

$$a(g) = \arccos\left(\frac{s(g)}{\|s(g)\|} \cdot \frac{s(g-1)}{\|s(g-1)\|}\right) \tag{1}$$

where, $s(g)$ is the column vector, $g$ is the time, $a(g)$ is the angle of rotation. The input data contains label noise and interference. Now assume the sample $p$ is tested over time at $q$ is given in Equation (2),

$$p = \lfloor s_1 s_2 \ldots\ldots s_q \rfloor \tag{2}$$

where, $p$ is the sampled signal, $sq$ is the vector matrix. Preprocessing step is presented for eliminating noise from the data input. Finally, filtering for all fixed $j$=1, 2 …..$m$ it come after the Equation (3) is given by,

$$\begin{gathered} z_j^l - z_j^{l+1} = (Y-B)^l z_j - (Y-B)^{l+1} z_j = \\ Z(Y-X)^l (Y-(Y-X)) Z^T Z_j = ZX(Y-X)^l \tilde{z}_j \to 0 \end{gathered} \tag{3}$$

where, $Z$ signifies unit matrix, $X$ signifies diagonal matrix, $B$ signifies circulate matrix. This method removes the noise from input data. Finally, preprocessed output is set to feature extraction phase.

### 3.3. Feature Extraction using Deep Wavelet Scattering Transform (DWST)

In this step, the significant features extracted with help of Deep Wavelet Scattering Transform (DWST) [18]. DWST offers significant advantages for feature extraction in multimodal data analysis. DWST facilitates multiscale analysis, enabling the extraction of features at different resolutions, and helps reduce dimensionality, speeding up training and mitigating overfitting risks. WST was presented as an iteratively acquired deep representation operator that cascades the convolution and modulus of wavelets non-linearity, which is given in Equation (4),

$$\hat{\chi}\delta_1(\varepsilon) = \hat{\chi}\left(\delta^{-1}\varepsilon\right) \tag{4}$$

where, $\hat{\chi}\,\delta_1(\varepsilon)$ represents Fourier transform mother wavelet, $\hat{\chi}\,(\delta^{-1}\varepsilon)$ is the concentrated wavelet transform. By averaging the wavelet modulus coefficients which are given in Equation (5),

$$R_1 a(l, \log\delta_1) = a_1 * \theta G = |a * \chi\delta_1| * \theta G \tag{5}$$

where, $\theta G$ represents low pass filter coefficient of size $G$, $R_1$ is the single coefficient. The uses of scattering coefficient of order $j$ are given in Equation (6),

$$R_j a(l, \log\delta_1 \ldots\ldots \delta_j) = B_j a(\log\delta_1 \ldots\ldots \delta_j) * \theta(l) \tag{6}$$

where, $R_j$ signifies octave resolution. The final of the highest order $t$ dissipative decomposition the scattering vector sums all the scattering factors to zero j>1. The features extracted from DWST are discussed below,

MFCC indicate the short span power spectrum of audio signal and used for speech and audio analysis at Equation (7),

$$ci = \sum_{n=1}^{nf} dn \cos[i(n-0.5)(\frac{\pi}{nf})] \;, i = 1,2,\ldots l \tag{7}$$

here, $ci$=$c_y(i)$=*ith mfcccoefficient*, $n_f$ implies triangular filters amount at the filter bank, $dn$ implies $n^{th}$ filter coefficient log energy output.

Spectral centroid indicates the “center of mass” of the audio spectrum and used for detecting the brightness of the sound in Equation (8),

$$ai = \sum h = 1wflwxi(h) \sum h = 1wflxi(h) \tag{8}$$

where, $ai$ is the value of spectral centroid.

Zero-crossing rate represents the rate of zero crossings in audio signal and used to detect changes of the signal in Equation (9),

$$\operatorname{sgn}[xi(n)] = 1, xi(n) \geq 0, -1, xi(n) < o \tag{9}$$

here, sgn(.) denotes sign function and $x$ denotes variable.

RMSE represents the entire energy of the audio signal in Equation (10),

$$rms = \sqrt{(3rt/m)} \tag{10}$$

where, $r$ signifies universal gas constant, $t$ signifies gas temperature and $m$ signifies gas molecular mass.

The extracted video features are shown here optical flow is the motion of objects in consecutive frames. It is denoted in Equation (11),

$$d_a u + d_y v + d_s = 0 \tag{11}$$

here, $d$ denotes subscripts, $u$ and $v$ signifies the components of the optical flow vector.

Histogram of Oriented Gradients (HOG) represents the distributions of gradients in an image, which can help detect object shapes and edges.

The Haralick texture features are operations on the normalized gray level co-occurrence matrix, which represents various facets of the ROI’s gray-level distribution in Equation (12)

$$F(L,s) = -\sum_m \sum_{n=h} \left[u(m-h, n-x)\log^2(u(m-l, n-q))\right] \tag{12}$$

Where $m$-$h$ represents pixels spatial relationship and $n$-$q$ represents intensity value, $x$ represents coordinates of points.

The extracted textual features are shown in below. The Equation (13) represents calculating the angular second moment (energy).

$$\sum_i \sum_j \{(f(i,j)\}^2 \tag{13}$$

here, $f(i, j)$ signifies Normalized gray-tone spatial dependency matrix, entry $(i, j)$. The Information Measure of Correlation is calculated by the Equation (14).

$$(1-\exp[-2.0(huv2-huv)]^{\frac{1}{2}} \tag{14}$$

The inverse difference moment is calculated by the Equation (15).

$$\sum_i \sum_j \frac{f(i,j)}{1+(i-j)^2} \tag{15}$$

here, $f(i, j)$ represents Normalized gray-tone spatial dependency matrix, entry $(i, j)$. These extracted features are fed into CANN.

### 3.4. Classification Using Constitutive Artificial Neural Network (CANN)

A deep learning method called CANN [24] can be used to solve categorization and regression issues. The CANN presents several advantages for classification tasks, particularly in complex multimodal datasets. CANN is designed to effectively learn hierarchical depictions of data, allowing it to capture intricate patterns and relation within the input features. Its architecture facilitates better generalization, making it resilient to overfitting while improving classification accuracy. CANN also benefits from adaptive learning mechanisms, enabling it to fine-tune weights dynamically based on the data it processes, which enhances its performance over traditional neural networks.

Here using multimodal English corpus for classifying using the Gutenberg Literary English Corpus dataset categorization is done using CANN. The CANN classifier is used to moderate training skill and time to acquire the constitutive behavior of even complicated nonlinear and anisotropic materials without human supervision. Consider six broadly employed methods and methodically compare their material parameters with network weights. It is given in Equation (16),

$$\phi = \frac{1}{2}\lambda[J_1-3] \tag{16}$$

here, $J_1$ represents first invariant, $\lambda$ represents shear modulus. For completely materials incompressible, $J_3$=1 which solves to the Equation (17),

$$\phi = \frac{1}{2}\lambda[J_2-3] \tag{17}$$

where, $J_2$ signifies second in variant. Mooney Rivlin method is given in Equation (18),

$$\phi = \frac{1}{2}\lambda_1[J_1-3]+\frac{1}{2}\lambda_2[J_2-3] \tag{18}$$

where, $J_3$ is the third invariant, $\lambda_1$, $\lambda_2$ are the shear modulus. Demiray model is given in Equation (19),

$$\phi = \frac{1}{2}\frac{c}{d}[\exp(d[J_1-3])-1] \tag{19}$$

where, $c$ and $d$ are constants. Gent's model uses linear logarithms of the first invariant $J_1$-3 with respect to two parameters $\gamma$ and $\psi$, which is given in Equation (20),

$$\phi = \frac{1}{2}\frac{\gamma}{\psi}\ln(1-(\psi[J_1-3])) \tag{20}$$

where, $\gamma$ and $\psi$ are logarithmic parameters. Holzapfel model is given in Equation (21),

$$\phi = \frac{1}{2}\frac{c}{d}\left[\exp\left(d[J_1-3]^2\right)-1\right] \tag{21}$$

With the help of these straightforward examples, it can show how to recover the common core operations for which the network weights take on perfect physical meaning.

The aim of the CANN is to decrease the loss function $M$ which penalized the error among method as well as data to study the network parameters $\omega$, the network weights of 1st and 2nd layers. To decrease potential over fitting, as well study the effects of L1 as well as L2 regularization is given in Equation (22),

$$M(\omega;E)=\frac{1}{m_{trn}}\sum_{j=1}^{m_{trn}}\left\|Q(E_j)-\hat{Q}_j\right\|^2+\gamma_1\|V\|_1+\frac{1}{2}\gamma_2\|V\|_2^2 \rightarrow \min \tag{22}$$

here, $V$ signifies weighted Euclidian, $\gamma_1$ and $\gamma_2$ signifies the weighting coefficients. The CANN model is built on the premise that there are classifies the English poetry such as ideographic, phonetic, rhetorical and contextual. Figure 2 shows the schematic diagram of CANN.

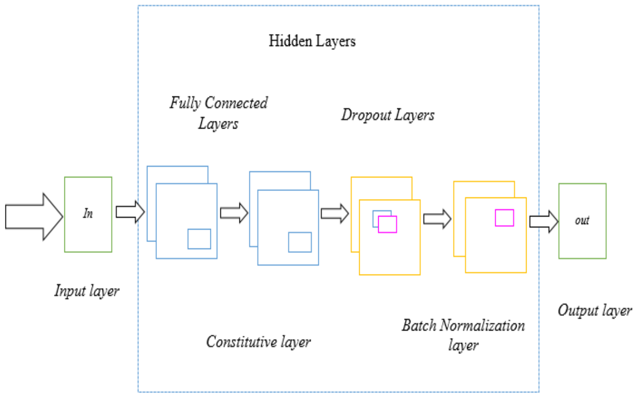


Figure 2. Schematic diagram of CANN.

Using the Honey Badger Algorithm (HBA), CANN's weight and bias parameters are adjusted. Constraint generation typically involves using techniques, such as grid, manual and random explorations. No well-known inquiry based on a trick exists, however these investigations all have an interesting flaw in terms of repetition time. HBA is thus employed to draw attention to this problem.

### 3.5. Honey Badger Algorithm (HBA)

The weight parameters ($\lambda_1$ and $\lambda_2$) of CANN are optimized using HBA [14]. HBA offers several advantages that enhance its effectiveness in optimizing weight and bias parameters in neural networks like CANN. HBA is stimulated by the foraging behavior of honey badgers, which enables it to efficiently explore the solution space and escape local optima through its aggressive exploration strategies. Overall, HBA significantly improves the optimization process in neural network training, leading to more accurate and reliable model performance. The stepwise process of Honey Badger technique is assumed below.

#### 3.5.1. Stepwise Processing of Honey Badger Algorithm (HBA)

HBA is regarded as a global optimization strategy on account of its exploration and exploitation phases. Stepwise process of HBA deliberated further. Figure 3 shows the flow chart of HBA.

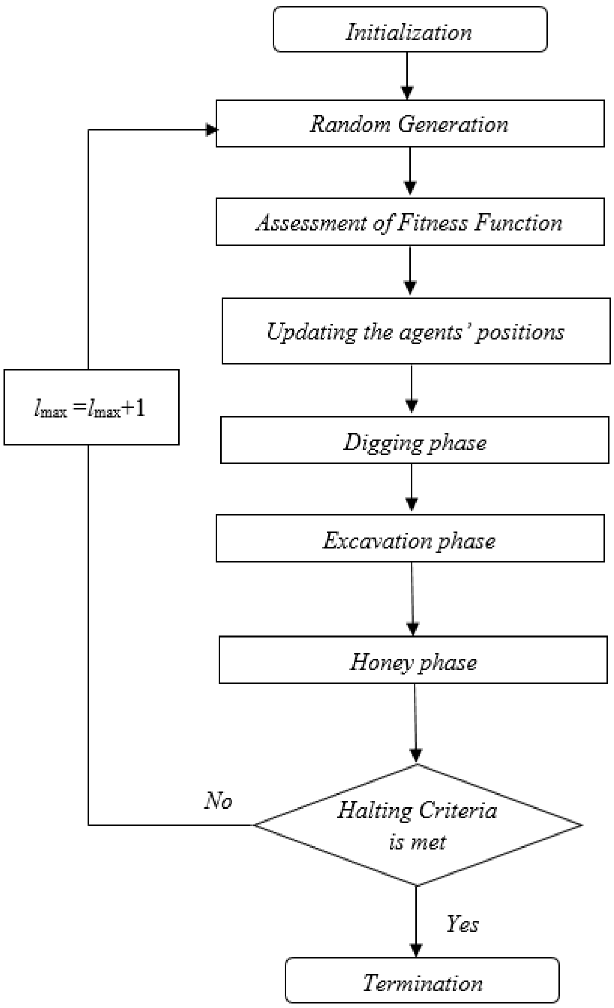


Figure 3. Flow chart of proposed HBA algorithm for optimizing CANN classifier.

- *Step* 1: Initialization

The candidate solutions population in HBA is reflected in Equation (23).

$$\text{Population of candidate solutions} = \begin{bmatrix} y_{11} & y_{12} & y_{13} & \cdots & y_{1D} \\ y_{21} & y_{22} & y_{23} & \cdots & y_{2D} \\ & \ldots\ldots & & & \\ y_{n1} & y_{n2} & y_{n3} & \cdots & y_{nD} \end{bmatrix} \tag{23}$$

From Equation (23), initialization populace of HBA is specified as $y$, and $y_n$. denotes position.

- *Step* 2: Random generation

The input parameters have generated after that initialization. The optimum fitness value is preferred under explicit hyper parameter circumstance.

- *Step* 3: Fitness Function evaluation

The initial calculations generate the random solution. This is accessed with the value of parameter optimization for increasing the CANN weight parameter, and is exhibited in Equation (24),

$$fitnessfunction = optimization(\lambda_1 and\ \lambda_2) \tag{24}$$

where, $\lambda_1$ is for increasing error, $\lambda_2$ is for decreasing error rate.

- *Step* 4: Updating the positioning of agents

The earlier studies update the procedure for the HBA location separated as two phases: "digging phase" and "honey phase."

- *Step* 5: Digging phase

The honey badger performs a process associated with Cardioid shape. It is computed by Equation (25)

$$\begin{aligned} y_{new} &= y_{prey} + F \times \beta \times I \times y_{prey} + \\ & F \times r_3 \times \alpha \times d_i \times \left|\cos(2\Pi r_4) \times \left[1 - \cos(2\Pi e_5)\right]\right| \end{aligned} \tag{25}$$

Here $y_{prey}$ signifies prey location. $\beta \geq 1$ (Default=6) signifies honey badger capacity to acquire food, $d_i$ signifies distance of prey, *ith* honey badger $r_3$, $r_4$, $r_5$, implies random digits 0 to 1.

- *Step* 6: Honey phase for optimizing of CANN

The point aids the bear in deciding on the appropriate course of action for the next position that handles CANN for controlling the virtual network operations. The Probability Odor Components (POC) is calculated in Equation (26),

$$POC = \frac{R_a}{\max(R_a)} \tag{26}$$

here, *POC* denotes Probability Odor Components, $R_a$ denotes threshold values odour dimension as well as sets created on average value of odour's information.

- *Step* 7: Termination Criteria

The weight parameter of the CANN is enhanced with the HBA in this stage, otherwise step 3 is repeated until satisfy the halting condition $y = y + 1$.

As shown in Figure 4.

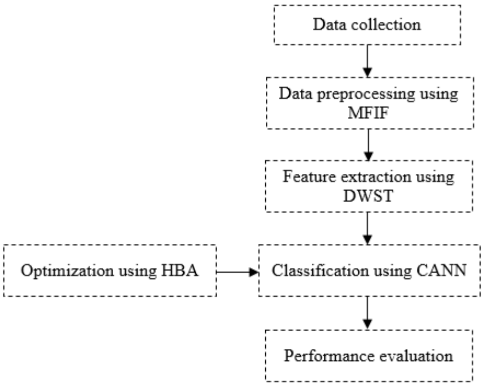


Figure 4. Flow chart for stages of data processing.

# 4. Result and Discussion

The simulation outputs of MEC-CANN-HBA are discussed in this section. The MEC-CANN-HBA technique is compared with existing CMC-CSSE-SC [34], CMPTC-AdaBoost [21], CECMC-AI) [40] respectively.

## 4.1. Experimental Setup

To ensure reproducibility of this experiments, here utilized specific computational resources, including an Intel i7-9700K processor, 32GB RAM, NVIDIA GeForce GTX 1660 Ti GPU, running on either Windows 10 or Ubuntu 20.04. The experiments were conducted using Python 3.8 with key libraries such as TensorFlow (2.6.0), Keras, NumPy (1.19.5), Pandas (1.2.4), SciPy (1.7.0), and Matplotlib (3.4.3) for deep learning, data manipulation, and visualization. The dataset, the Gutenberg Literary English Corpus, is publicly available, and preprocessing scripts are included in the supplementary materials to facilitate easy replication. This comprehensive setup allows other researchers to accurately reproduce the proposed methodology and results in multimodal corpus analysis and neural network optimization.

### 4.1.1. Dataset Description

The GLEC database is utilized, it comprises over 10,000 literary texts from various genres, including poetry, novels, and essays. The corpus provides a rich source of English literature, facilitating a wide-ranging analysis of multimodal features. Prior to the application of the methodology, preprocessing step was performed:

Table 1. Parameter setting.

| Parameter | Setting |
|---|---|
| Dataset | Gutenberg literary English corpus |
| Dataset size | 10,000 literary texts |
| MFIF parameters | Iterations: 100, Threshold: 0.1 |
| DWST parameters | Wavelet: Daubechies (db4), Scattering orders: 2, Scales: 4 |
| HBA parameters | Population size: 30, Max iterations: 200, Exploration rate: 0.1 |
| CANN architecture | Hidden layers: 3, Neurons per layer: 128, Activation function: ReLU |
| CANN training parameters | Learning rate: 0.001, Batch size: 64, Epochs: 100 |

Noise Reduction: Application of Multivariate Fast Iterative Filtering (MFIF) to eliminate any irrelevant data points, enhancing the dataset's overall quality. Table 1 displays the parameter setting.

## 4.2. Performance Metrics

The MEC-CANN-HBA approach is compared with existing techniques like CMC-CSSE-SC, CMPTC-AdaBoost, and CECMC-AI respectively. The mentioned metrics are analysed to examine the robustness of the MEC-CANN-HBA technique.

### 4.2.1. Accuracy

This is the rate of appropriately predicted samples count to the total samples in the dataset. This is calculated by Equation (27),

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{27}$$

Let *TP* represents True Positive, *TN* represents True Negative, *FP* represents False Positive, *FN* represents False Negative.

### 4.2.2. Precision

Precision is a crucial performance metric used to evaluate the effectiveness of classification modes, particularly in the context of the proposed approach involving CANN and HBA. It quantifies the accurateness of the positive predictions made by the method. The precision is calculated by Equation (28)

$$Precision = \frac{TP}{TP + FP} \tag{28}$$

### 4.2.3. Sensitivity

Sensitivity is the rate of true positives to the total actual positives. It is determined through Equation (29),

$$Sensitivity = \frac{TP}{TP + FN} \tag{29}$$

### 4.2.4. Specificity

Specificity is the rate of true negative instances properly recognized by the model out of all true negative samples in the dataset using the Equation (30),

$$Specificity = \frac{TN}{TN+FP} \tag{30}$$

### 4.2.5. F-Score

The single score balances the trade-off among the harmonic mean of precision along recall. The F-score is calculated by Equation (31).

$$F - Score = 2 \times \frac{Precision \times Recall}{Precision \times Recall} \tag{31}$$

## 4.3. Performance Analysis

Figures 6 to 9 and Tables 2 and 3 portrays simulation results of proposed MEC-CANN-HBA technique.

Proposed MEC-CANN-HBA is analyzed with existing CMC-CSSE-SC, CMPTC-AdaBoost, CECMC-AI respectively.

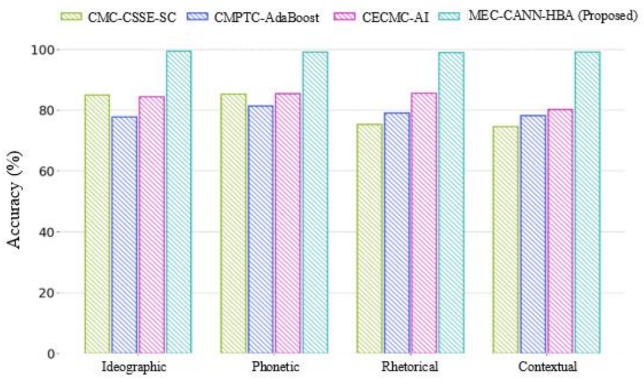


Figure 5. Accuracy analysis.

Figure 5 shows the accuracy analysis. In this research, accuracy is particularly relevant because it helps assess the effectiveness of CANN in categorizing various texts within the corpus after employing data preprocessing techniques, feature extraction and optimization. High accuracy implies that the method can generalize from the training data to unseen data, which is essential for practical applications in poetry translation and multimodal analysis. Here the MEC-CANN-HBA attains 27.86%, 25.85%, and 20.56% greater accuracy for ideographic; 22.85%, 19.07%, and 31.86% greater accuracy for phonetic, 22.85%, 20.07%, and 12.86% greater accuracy for rhetorical, and 22.85%, 10.07%, and 25.86% greater accuracy for contextual compared with existing methods like CMC-CSSE-SC, CMPTC-AdaBoost, CECMC-AI respectively.

Table 2. Performance analysis of precision.

| **Methods** | **Precision (%)** | | | |
|---|---|---|---|---|
| | **Ideographic** | **Phonetic** | **Rhetorical** | **Contextual** |
| **CMC-CSSE-SC** | 79.5 | 73.5 | 74.2 | 78.3 |
| **CMPTC-AdaBoost** | 87.6 | 88.65 | 89.65 | 83.75 |
| **CECMC-AI** | 83.65 | 89.55 | 84.55 | 85.95 |
| **MEC-CANN-HBA (proposed)** | 97.85 | 97.99 | 98.1 | 98.2 |

Table 2 displays the performance of precision. In this research, high precision indicates that when the CANN model predicts a text or data point as belonging to a specific class, it is likely to be correct. This is particularly important in applications like poetry translation, where false positives can lead to misleading interpretations and decrease the overall reliability of the system. Here the proposed MEC-CANN-HBA method attains 29.86%, 25.85%, and 24.56% greater precision for ideographic; 10.85%, 9.07%, and 30% greater precision for phonetic, 6.85%, 22.07%, and 33.76% greater precision for rhetorical, and 17.85%, 10.07%, 25.86% better precision for contextual than the existing CMC-CSSE-SC, CMPTC-AdaBoost, CECMC-AI respectively.

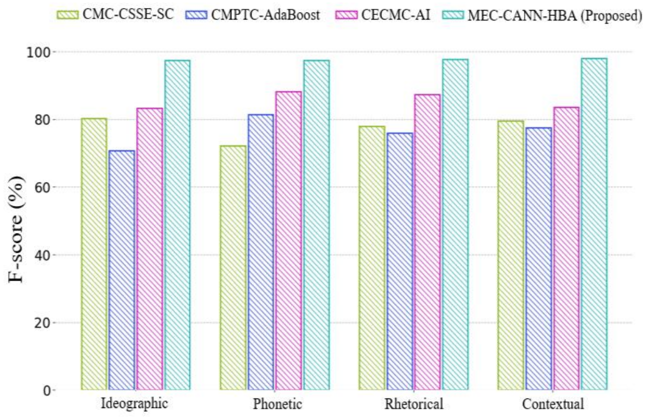


Figure 6. F-score assessment.

Figure 6 exhibits f-score assessment. In the context of MEC-CANN-HBA approach, HBA for optimizing CANN and utilizing deep learning strategies for multimodal corpus analysis, the F-Score serves as a key performance indicator. The proposed MEC-CANN-HBA method attains 28.86%, 23.85%, and 20.56% greater f-score for ideographic; 31.85%, 24.07%, and 23.86% greater f-score for phonetic, 26.85%, 27.07%, and 19.86% greater f-score for rhetorical, 22.85%, 30.07%, 25.86% greater f-score for contextual than the existing CMC-CSSE-SC, CMPTC-AdaBoost, CECMC-AI.

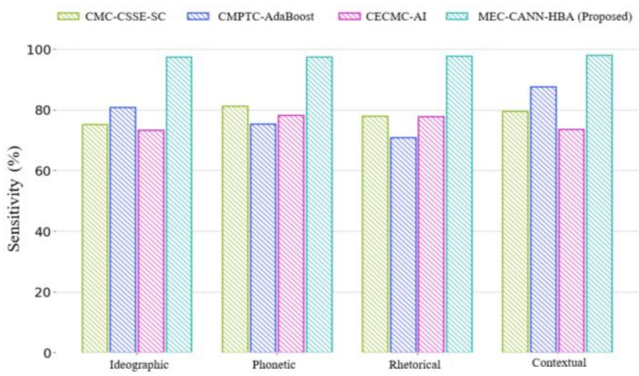


Figure 7. Sensitivity analysis.

Figure 7 displays sensitivity assessment. High sensitivity indicates that the model is effective at capturing the essence of the underlying data, which is particularly important in applications like poetry translation, where missing a positive instance could lead to significant misinterpretations. The proposed MEC-CANN-HBA method attains 26.86%, 29.85%, and 22.56% greater sensitivity for ideographic; 30%, 26.07%, and 30.76% greater precision for phonetic, 27.19%, 24.07%, and 19.76% greater sensitivity for rhetorical, and 28.85%, 22.07%, and 16.86% greater sensitivity for contextual, compared with existing methods like CMC-CSSE-SC, CMPTC-AdaBoost, CECMC-AI respectively.

Table 3 tabulates the performance of Specificity. The proposed approach involves the integration of deep learning techniques for multimodal corpus analysis, specificity acts as a crucial metric to scale the ability to accurately identify non-target classes. The suggested MEC-CANN-HBA method attains 23.65%, 21.98%, 20.76% greater specificity for ideographic; 10.85%, 27.07%, and 32.76% greater specificity for phonetic, 21.85%, 26.07%, and 22.76% greater specificity for rhetorical, and 25.85%, 20.07%, and 11.86% greater specificity for contextual compared with existing methods like CMC-CSSE-SC, CMPTC-AdaBoost, CECMC-AI respectively.

Table 3. Specificity analysis.

| Methods | Specificity (%) | | | |
|---|---|---|---|---|
| | Ideographic | Phonetic | Rhetorical | Contextual |
| **CMC-CSSE-SC** | 73.5 | 74.2 | 78.3 | 72.7 |
| **CMPTC-AdaBoost** | 89.65 | 79.65 | 73.75 | 82.55 |
| **CECMC-AI** | 78.55 | 84.55 | 85.95 | 79.85 |
| **MEC-CANN-HBA (proposed)** | 98.1 | 98.1 | 98.25 | 98.35 |

Figure 8 depicts the computational time analysis. The efficiency of the MEC-CANN-HBA method, which combines data preprocessing, feature extraction, and classification optimized with HBA, was thoroughly assessed in terms of execution time. The proposed MEC-CANN-HBA method attains 26.01%, 30%, 17.44% lesser computational time than the existing CMC-CSSE-SC, CMPTC-AdaBoost, CECMC-AI respectively.

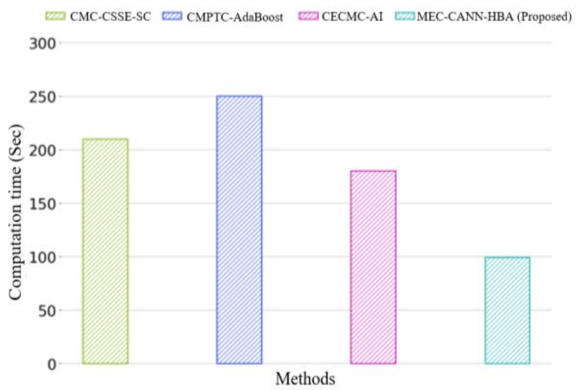


Figure 8. Computational time analysis.

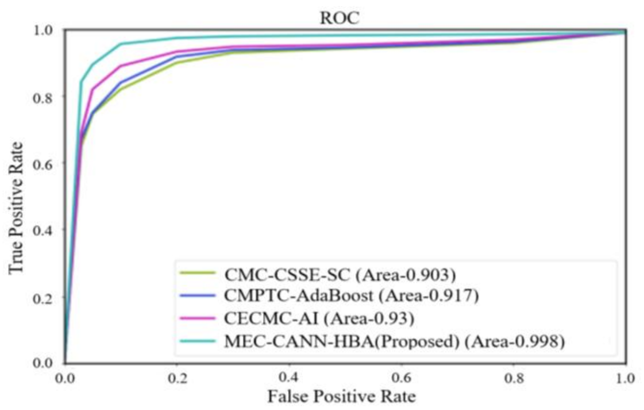


Figure 9. ROC assessment.

Figure 9 shows ROC curve analysis. The Area Under the ROC Curve (AUC) is a key indicator to distinguish between classes. A greater AUC score indicates better classification performance, with values nearer to 1 representing superior model accuracy. Here, the proposed MEC-CANN-HBA technique provides 13.49%, 26.45%, and 28.78% higher ROC than the existing CMC-CSSE-SC, CMPTC-AdaBoost, and CECMC-AI respectively.

## 4.4. Statistical Test

To confirm the effectiveness of the MEC-CANN-HBA approach, conducted both paired t-tests and Wilcoxon signed-rank tests in terms of accuracy, specificity, sensitivity, precision, F1-score, and computational time. The statistical analysis revealed p-values below 0.05 for both tests across all metrics, indicating significant improvements over the baseline models. Specifically, the proposed approach showed notable gains in accuracy and reduced computational time. These findings confirm that the enhancements in performance are statistically significant, strengthening the credibility of the results. Table 4 shows the statistical significance tests on performance metrics for the proposed approach compared to existing methods.

Table 4. Statistical significance tests.

| Metrics | Existing method mean | Proposed method mean | Paired t-test (p-value) | Wilcoxon signed-rank test (p-value) |
|---|---|---|---|---|
| **Accuracy (%)** | 89.4 % | 94.3% | 0.0022 | 0.0023 |
| **Specificity (%)** | 87.1 % | 91% | 0.0033 | 0.0039 |
| **Sensitivity (%)** | 85.9 % | 90.4% | 0.0027 | 0.0034 |
| **Precision (%)** | 86.4 % | 92% | 0.0041 | 0.0055 |
| **F1-Score (%)** | 85.8 % | 93.54% | 0.0019 | 0.0025 |
| **Computational time (sec)** | 125.8 | 94,2 | 0.0048 | 0.0066 |

## 4.5. Discussion

This paper research and construct a multimodal English corpus using the method optimized CANN (MEC-CANN-HBA). English Poetry is a challenging task than the translation of any other genre. Poetry translation necessitates the creation of a second text because of the poetic and unique nature of the language used, and the translator must be highly accomplished in Chinese literature.

The MEC-CANN-HBA method outperforms existing methods due to the effective integration of Multivariate Fast Iterative Filtering for noise reduction, which enhances the quality of input data critical for improving classification accuracy. The Deep Wavelet Scattering Transform captures intricate patterns and hierarchical structures, enabling the model to leverage both local and global features for better class discrimination. Additionally, the Constitutive Artificial Neural Network (CANN) optimized by the Honey Badger Algorithm allows for more efficient adjustment of weight and bias parameters, leading to faster convergence and higher accuracy while avoiding local minima. Observational trends indicate that the proposed method consistently achieves superior performance across various metrics, such as accuracy and precision. This robust integration of advanced preprocessing, feature extraction, and optimization techniques significantly contributes to the method's effectiveness in real-world applications.

## 5. Conclusions

In this paper, using the MEC-CANN-HBA technique to categorize and learn present poetry translation corpus to create multiple modal English translation corpuses. A multimodal poetry corpus created once the model has been utilized to acquire and categorize the prevailing corpus of poem translations. The final step is an analysis and evaluation of the 4 components of poetry translation like ideographic, phonetic, rhetorical contextual. Even though a multimodal English corpus is successfully created as a result of this work, its accuracy and completeness still need to be improved, and more research needs to be done on how to create the ideal system for evaluating poem translations. The proposed MEC-CANN-HBA approach achieves 25.36%, 23.42% and 19.27%, higher f-score and 26.61%, 18.50%, 25% higher precision compared with existing CMC-CSSE-SC, CMPTC-AdaBoost and CECMC-AI respectively.

The limitations of the study primarily relate to the scalability of the proposed approach when applied to larger datasets or different types of multimodal data. As the dataset size increases, computational resource requirements, particularly in terms of memory and processing power, may become a bottleneck. The current approach has been tested on a specific corpus, and its generalizability to other domains with diverse modalities, such as video or complex audiovisual data, remains to be explored. Additionally, the Honey Badger Algorithm (HBA) used for optimization may exhibit slower convergence or performance degradation with significantly larger datasets. Future work could focus on improving the scalability of the model through parallel processing, optimizing neural network architecture, and applying the method to more varied multimodal datasets. Integration of more efficient optimization algorithms or hybrid approaches may also help in enhancing scalability and robustness.

## Acknowledgement

The Research on the Campus Network Security System and Its Implementation in the Background of Big Data Natural Science Key Project of Jining Normal University jsky202209

## References

[1] Baldry A. and Kantz D., *Corpus-Assisted Approaches to Online Multimodal Discourse Analysis of Videos*, *Analysing Multimodality in Specialized Discourse Settings: Innovative Research Methods and Applications*, Springer, 2022. DOI: 10.1057/9781137431738_4

[2] Brunner M. and Diemer S., "Multimodal Meaning Making: The Annotation of Nonverbal Elements in Multimodal Corpus Transcription," *Research in Corpus Linguistics*, vol. 9, no. 1, pp. 63-88, 2021. https://doi.org/10.32714/ricl.09.01.05

[3] Cambria M., "Learning about Schools in the British Isles through a Video Corpus: Reflections on an Online Project for Digital Literacy and Multimodal Corpus Construction," *Journal of Elementary Education*, vol. 16, pp. 117-135, 2023. DOI:10.18690/rei.16.Spec.Iss.2988

[4] Cang H. and Feng D., "Construction of English Corpus Oral Instant Translation Model Based on Internet of Things and Deep Learning of Information Security," *Journal of Computational Methods in Sciences and Engineering*, vol. 24, no. 3, pp. 1507-1522, 2024. https://doi.org/10.3233/JCM-247183

[5] Carrió-Pastor M., "Teaching Multimodal Metadiscourse in Academic English as a Foreign Language," *Porta Linguarum*, vol. 4, pp. 155-172, 2022. https://doi.org/10.4000/lidil.10575

[6] Chakravarthi B., Muralidaran V., Priyadharshini R., and McCrae J., "Corpus Creation for Sentiment Analysis in Code-Mixed Tamil-English Text," *arXiv Preprint*, vol. arXiv:2006.00206, pp. 1-9, 2020. https://doi.org/10.48550/arXiv.2006.00206

[7] Chen J., Liu S., and Shao D., "The Construction and Practice of a Multimodal Corpus for the Ideological and Political Education in Aerospace-themed College English Curriculum," *International Journal of Social Science and Education Research*, vol. 7, no. 5, pp. 258-263, 2024. https://doi.org/10.6918/IJOSSER.202405_7(5).0036

[8] Chen Y., Harrison S., Stevens M., and Zhou Q., "Developing a Multimodal Corpus of L2 Academic English from an English Medium of Instruction University in China," *Corpora*, vol. 19, no. 1, pp.1-15, 2024.

[9] Chen Z., Subtitling as Multimodal Representation: A Corpus-Based Experimental Approach to Text-Image Relations, Ph.D. Thesis, The Hong Kong Polytechnic University, 2024. https://theses.lib.polyu.edu.hk/bitstream/200/12821/3/7272.pdf

[10] Deng L., "Culture's Representation in Van Gulik's Transcreated Novel the Chinese Maze Murders: A Multimodal Corpus Approach," *SAGE Open*, vol. 14, no. 1, pp. 1-12, 2024. https://doi.org/10.1177/21582440231223481

[11] Dong S., "Intelligent English Teaching Prediction System Based on SVM and Heterogeneous Multimodal Target Recognition,"

*Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 6, pp. 7145-7154, 2020. DOI:10.3233/JIFS-179792

[12] Eijk L., Rasenberg M., Arnese F., Blokpoel M., Dingemanse M., Doeller C., Ernestus M., Holler J., Milivojevic B., Özyürek A., and Pou W., "The CABB Dataset: A Multimodal Corpus of Communicative Interactions for Behavioural and Neural Analyses," *NeuroImage*, vol. 264, pp. 119734, 2022. https://doi.org/10.1016/j.neuroimage.2022.119734

[13] Harb H., Multimodal Emotion Recognition Using Temporal Convolutional Networks, Master Thesis, University of Ottawa, 2023. https://ruor.uottawa.ca/items/155adedf-62a6-4ef4-a254-08ed190d1d75

[14] Hashim F., Houssein E., Hussain K., Mabrouk M., and Al-Atabany W., "Honey Badger Algorithm: New Metaheuristic Algorithm for Solving Optimization Problems," *Mathematics and Computers in Simulation*, vol. 192, pp. 84-110, 2022. https://doi.org/10.1016/j.matcom.2021.08.013

[15] Hoemann K. and Leuven K., Modeling the Meaning of Emotion Words Using Multimodal Features of Real-World Contexts, https://iclc16.github.io/abstracts/ICLC16_paper_486.pdf, Last Visited, 2024.

[16] Hu G., "The Model Construction of Human-Computer Interaction Teaching in College English Multimodal Teaching," *Frontiers in Educational Research*, vol. 7, no. 2, pp. 178-188, 2024. https://doi.org/10.25236/FER.2024.070228

[17] Huang L., "Toward Multimodal Corpus Pragmatics: Rationale, Case, And Agenda," *Digital Scholarship in the Humanities*, vol. 36, no. 1, pp. 101-114, 2021. https://doi.org/10.1093/llc/fqz080

[18] Khushaba R. and Hill A., "Radar-Based Materials Classification Using Deep Wavelet Scattering Transform: A Comparison of Centimeter Vs. Millimeter Wave Units," *IEEE Robotics and Automation Letters*, vol. 7, no. 2, pp. 2016-2022, 2022. DOI:10.1109/LRA.2022.3143200

[19] Kumari R., Ashok N., Agrawal P., Ghosal T., and Ekbal A., "Identifying Multimodal Misinformation Leveraging Novelty Detection and Emotion Recognition," *Journal of Intelligent Information Systems*, pp. 1-22, 2023. DOI:10.1007/s10844-023-00789-x

[20] Li S. and Okada S., "Interpretable Multimodal Sentiment Analysis Based on Textual Modality Descriptions by Using Large-Scale Language Models," *arXiv Preprint*, vol. arXiv:2305.06162, pp. 1-10, 2023. https://doi.org/10.48550/arXiv.2305.06162

[21] Li S., "Construction of a Multimodal Poetry Translation Corpus Based on AdaBoost Model," *Applied Mathematics and Nonlinear Sciences*, vol. 9, no. 1, pp. 1-15, 2023. https://doi.org/10.2478/amns.2023.1.00102

[22] Liang C., Xu J., Zhao J., Chen Y., and Huang J., "Deep Learning-Based Construction and Processing of Multimodal Corpus for IoT Devices in Mobile Edge Computing," *Computational Intelligence and Neuroscience*, vol. 2022, no. 241310, pp. 1-10, 2022. https://doi.org/10.1155/2022/2241310

[23] Lin Y., Chi Y., Han H., Han M., and Guo Y., "Multimodal Orthodontic Corpus Construction Based on Semantic Tag Classification Method," *Neural Processing Letters*, vol. 54, no. 4, pp. 2817-2830, 2022. https://doi.org/10.1007/s11063-021-10558-y

[24] Linka K., Hillgärtner M., Abdolazizi K., Aydin R., Itskov M., and Cyron C., "Constitutive Artificial Neural Networks: A Fast and General Approach to Predictive Data-Driven Constitutive Modeling by Deep Learning," *Journal of Computational Physics*, vol. 429, pp. 110010, 2021. https://doi.org/10.1016/j.jcp.2020.110010

[25] Merritt A., Chu C., and Arase Y., "A Corpus for English-Japanese Multimodal Neural Machine Translation with Comparable Sentences," *arXiv Preprint*, vol. arXiv:2010.08725, pp. 1-6, 2020. https://doi.org/10.48550/arXiv.2010.08725

[26] Nan C., "Multidimensional Psychological Model Construction of Public English Teaching Based on Deep Learning from Multimodal Perspective," *Mobile Information Systems*, vol. 2022, no. 1, pp. 1653452, 2022. https://doi.org/10.1155/2022/1653452

[27] Pinto S. and Mubaraki A., "Multimodal Corpus Analysis of Subtitling: The Case of Non-Standard Varieties," *Target*: *International Journal of Translation Studies*, vol. 32, no. 3, pp. 389-419, 2020. https://doi.org/10.1075/target.18085.ram

[28] Reece A., Cooney G., Bull P., Chung C., Dawson B., Fitzpatrick C., Glazer T., Knox D., Liebscher A., and Marin S., "Advancing an Interdisciplinary Science of Conversation: Insights from a Large Multimodal Corpus of Human Speech," *arXiv Preprint*, vol. arXiv:2203.00674, pp. 1-116, 2022. https://doi.org/10.48550/arXiv.2203.00674

[29] Rodríguez-Peñarroja M., "Corpus Pragmatics and Multimodality: Compiling an ad-hoc Multimodal Corpus for EFL Pragmatics Teaching," *International Journal of Instruction*, vol. 14, no. 1, pp. 927-946, 2020. DOI:10.29333/iji.2021.14155a

[30] Siever C., and Siever T., *Shifts Towards Image-Centricity in Contemporary Multimodal*

*Practices*, Routledge, pp. 177-203, 2020.

[31] Singh N. and Kapoor R., "Multi-modal Expression Detection (MED): A Cutting-Edge Review of Current Trends," *Challenges and Solutions Engineering Applications of Artificial Intelligence*, vol. 125, pp. 106661, 2023. https://doi.org/10.1016/j.engappai.2023.106661

[32] Stevens M., Chen Y., and Harrison S., *Variation in Time and Space*, De Gruyter, 2020. https://doi.org/10.1515/9783110604719-015

[33] Tian M., "Construction of Computer English Corpus Assisted by Internet of Things Information Perception and Interaction Technology," *Computational Intelligence and Neuroscience*, vol. 2022, no. 1, pp. 6803802, 2022. https://doi.org/10.1155/2022/6803802

[34] Tian X., "Construction of a Multimodal Corpus of College Students' Spoken English Based on Semantic Concepts," *Mobile Information Systems*, vol. 2022, pp. 1-10, 2022. https://doi.org/10.1155/2022/5270408

[35] Wang Y., Wang Z., Kang X., and Luo Y., "A Novel Interpretable Model Ensemble Multivariate Fast Iterative Filtering and Temporal Fusion Transform for Carbon Price Forecasting," *Energy Science and Engineering*, vol. 11, no. 3, pp. 1148-1179, 2023. https://doi.org/10.1002/ese3.1380

[36] Yang H., Si Z., Zhao Y., Liu J., Wu Y., and Qin B. "MACSA: A Multimodal Aspect-Category Sentiment Analysis Dataset with Multimodal Fine-Grained Aligned Annotations," *arXiv Preprint*, vol. arXiv:2206.13969, pp.1-10, 2022. https://doi.org/10.48550/arXiv.2206.13969

[37] Zadeh A., Zellers R., Pincus E., and Morency L., "MOSI: Multimodal Corpus of Sentiment Intensity and Subjectivity Analysis in Online Opinion Videos," *arXiv Preprint*, vol. arXiv:1606.06259, pp. 1-10, 2016. https://doi.org/10.48550/arXiv.1606.06259

[38] Zeng Z. and Li Y., "Multi-modal Chinese Text Emotion Metaphor Computation Based on Mutual Information and Information Entropy," *ACM Transactions on Asian and Low-Resource Language Information Processing*, vol. 23, no. 8, pp. 1-18, 2023. https://doi.org/10.1145/3605211

[39] Zhang L., "Ideological and Political Empowering English Teaching: Ideological Education Based on Artificial Intelligence in Classroom Emotion Recognition," *International Journal of Computer Applications in Technology*, vol. 71, no. 3, pp. 265-271, 2023. https://doi.org/10.1504/IJCAT.2023.132103

[40] Zhou W. and Gao B., "Construction and Application of English-Chinese Multimodal Emotional Corpus Based on Artificial Intelligence," *International Journal of Human-Computer Interaction*, pp. 1-12. 2023.

[41] Zhu W., Hessel J., Awadalla A., Gadre S., Dodge J., Fang A., Yu Y., Schmidt L., Wang W., and Choi Y., "Multimodal c4: An Open, Billion-Scale Corpus of Images Interleaved with Text," *in Proceedings of the 37th Conference on Neural Information Processing Systems*, New Orleans, pp. 1-17, 2023. https://doi.org/10.48550/arXiv.2304.06939