Received: November 6, 2024. Revised: November 30, 2024.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Hippopotamus Optimization Algorithm with Long Short-Term Memory for Music Genre Classification

G. Indiravathi[1]* **D. Rajesh[1]**

[1]*Department of Computer Science and Engineering,*
Vel Tech Rangarajan Dr.Sagunthala R&D Institute of Science and Technology, Chennai, India
* Corresponding author's Email: vtd1049@veltech.edu.in

Abstract: Music genre is a traditional method of categorizing music, playing a crucial role in music information retrieval. Music genre classification encompasses various types like Disco, Pop, Jazz, and Rock, with some genres sharing similar characteristics, making the model prone to overlap or misclassification. To address this issue, this research introduces the Hippopotamus Optimization Algorithm with Long Short-Term Memory (HOA-LSTM) for genre classification. The HOA enhances convergence and reduces computational time by effectively exploring and exploiting the search space, thereby improving classification accuracy. LSTM, known for capturing temporal dependencies in audio signals, is resilient to vanishing gradient issues, leading to precise classification. Performance metrics such as accuracy, sensitivity, specificity, Positive Predictive Value (PPV), F-measure, error rate, and computation time are used to evaluate HOA-LSTM. The proposed HOA-LSTM achieves an accuracy of 98.47% and an error rate of 1.53% on the GTZAN dataset, outperforming the Bidirectional Long Short-Term Memory-VGG-16 Network (BiLSTM-VGG-16 Net).

Keywords: Hippopotamus optimization algorithm, Long short-term memory, Music genre classification, Positive predictive value, Temporal dependencies.

1. Introduction

Music is one of the significant and integral aspects of people's everyday routines which express social interaction, entertainment, emotions and so on. There are numerous distinct music genres in which each offering unique qualities which makes people have various music preferences [1]. The music genre is a piece of music which reflects its cultural or historical roots and also indicating specific types of instruments used in the piece [2]. The music genre is selected as the focus of this research because it is closely linked to individual personality and musical preferences [3]. It allows user to receive personalized music recommendations based on their listening habits whether occasional, frequent and new [4-6]. Music genres play a crucial role by enhancing and facilitating music search, highlight cultural interactions and significance of cultural features [7]. Within same genre leads various music and started from shared cultural background which utilizing a common music language [8]. However, humans are most crucial role in genre classification through their ears and they accurately identifying genres up to 250ms of audio [9]. This demonstrates that auditory process with brain high-level cognitive mechanisms helps in genre classification through compact explanation of musical exterior rather than by general compact explanation [10]. Music is an integral part of people's daily lives, playing a key role in social interaction, entertainment, and emotional expression. Various music genres exist, each with distinct characteristics, reflecting their cultural or historical roots and often featuring specific instruments [1]. The focus of this research is on music genres because they are closely tied to individual personality and musical preferences [2]. Music genres enable personalized recommendations based on listening habits, whether occasional, frequent, or new [3-5]. Genres also play a significant role in music search, highlighting cultural interactions and the importance of cultural elements [6]. Music within the same genre evolves from a shared cultural foundation, using a common musical language [7]. Humans are key to genre classification, as they can accurately identify genres within just 250 milliseconds of audio [8]. This demonstrates how auditory processing and high-level cognitive mechanisms in the brain assist in genre classification through an understanding of musical characteristics rather than general information [9]. Music genre classification involves two main steps: feature extraction and classification [10].

Feature extraction is crucial, as it identifies and extracts meaningful characteristics from music tracks. The design of classification algorithms to manage these acoustic features is another key component [11]. Traditional machine learning (ML) algorithms, such as Random Forest (RF), Naïve Bayes (NB), and Support Vector Machine (SVM), have been widely used [12]. However, recent research indicates that these algorithms face challenges when dealing with large-scale data, especially with diverse distributions [13]. With the advancement of deep learning (DL) techniques, classification algorithms based on Recurrent Neural Networks (RNN) and Convolutional Neural Networks (CNN) have effectively captured latent data within acoustic features [14, 15]. These methods have significantly improved music genre classification by leveraging hand-crafted features [16, 17]. The contribution of this research is given as follows:

- The Hybrid Optimization Algorithm (HOA) improves convergence and reduces computation time by efficiently exploring and exploiting the search space, thus selecting relevant features and enhancing classification accuracy.
- Long Short-Term Memory (LSTM) networks effectively capture and utilize temporal dependencies in audio signals. LSTM can recall long-term context and identify patterns, resulting in accurate classification. It is also robust against vanishing gradient problems, making it ideal for processing complex and lengthy audio sequences.

This paper is arranged as follows: Section 2 summarizes the relevant literature review, and Section 3 describes the proposed framework in a detailed way. Section 4 gives implemented results for the proposed framework and Section 5 provides a conclusion with future work.

2. Literature review

In this section, the existing methods utilized for music genre classification are investigated and explained with its advantages and limitations.

Sunil Kumar Prabhakar and Seong-Whan Lee [20] presented a BiLSTM cum Attention model with a Graphical Convolution Network (GCN) (BAG) for music genre classification. The BiLSTM handled the sequential data along with the sigmoid function to overcome the gradient vanishing issues Stochastic Gradient Descent (SGD) optimized the features of the data during the training process that enhanced the performance of the model. However, the training of the model was affected due to irrelevant features of the signals that created overfitting issues and decreased the model performance.

Changjiang Xie [21] implemented a res-gated Convolutional Neural Network (CNN) with an attention mechanism for the music genre classification. The 1D res-gated CNN was used to extract the significant details of the spectrogram. The LSTM model was used as an encoder to convert the features into vector form which are passed to decoder. The global-pooling aggregated the vector data which improved the accuracy of music genre classification. Though, the some of the gradients vanished during the training process that created misclassification of data and reduced the classification accuracy.

Mousumi Chaudhury [22] developed Apache spark based multi scalable Machine Learning (ML) model for the analysis and classification of music genre. Naïve Bayes (NB), Decision Tree (DT), Logistic Regression (LR), and Random Forest (RF) are used as a classifier. The audio signal features were selected based on the numtree and maxdepth hyperparameter optimization technique that increased the accuracy of music genre classification. But the model failed to obtain the optimal hyperplane for the classification of music genre due to poor boundary recognition that reduced the performance of classification.

Xin Cai and Hongjuan Zhang [23] presented a spectral and acoustic features-based music genre classification. The Support Vector Machine (SVM), K-Nearest Neighbour (KNN), and Sparse Representation-based classification (SRC) were used for the classification of music genre. The Local Binary Pattern (LBP) was used to extract the important details from the audio signals. The spectral and acoustic features were extracted from the signals that enhanced the classification accuracy of the music genre. Nevertheless, the model failed to handle the dimensionalities of the features due to variable

Received: November 6, 2024. Revised: November 30, 2024.

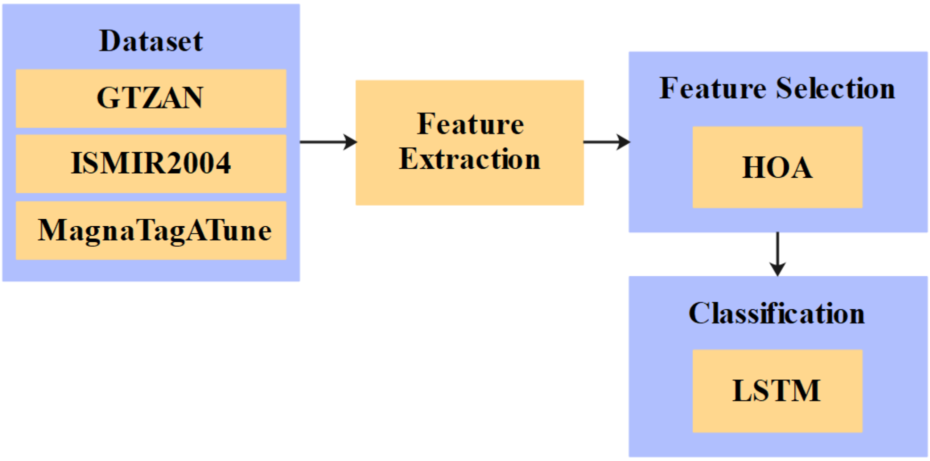


Figure. 1 Framework of proposed methodology

constraints that decreased the classification accuracy of the music genre.

Wang Hongdan [24] implemented a feature extraction-based music genre classification using VGG 16 DL model. The signal data was converted into vector form to increase the classification of music genre. The BiLSTM was used to extract the significant features from the music audio signals based on contextual information that increased the model performance for the music genre classification. However, the accuracy of the classification was affected due to the presence of irrelevant noise signals that minimized the efficiency of the model.

From the investigations, the existing methods have limitations of overfitting issues, vanishing gradients during the training process, misclassification, failure to handle the dimensionalities of the features and the presence of irrelevant noise. By using various feature extraction, the temporal and structural features are extracted from raw signal to distinguish the genre classes. Then, HOA based feature selection is utilized to remove the inappropriate features form feature set and enhancing the model performance. Additionally, LSTM is utilized to classify the genre classes through mitigating issue of vanishing gradient thereby enhancing high classification accuracy.

3. Proposed method

This research introduces an efficient method for music genre classification through a combination of feature selection and classification techniques. Using the GTZAN dataset, which includes 10 genre classes, meaningful features are extracted from audio signals to differentiate between genres. These features are then selected using the HOA algorithm and classified using LSTM. The overall framework of the proposed methodology is shown in Fig. 1.

3.1 Dataset

The GTZAN [25] dataset is widely utilized for music genre classification which contains 1000 audio clips. It contains 10 different music genres such as Rock, Reggae, Pop, Metal, Jazz, Hip-hop, Disco, Country, Classical and Blues. Each genre contains 100 excerpts which last about approximately 30s and stored at sample rate of 22,050Hz. This dataset is split into training of 80% and testing of 20%.

The ISMIR2004 [20] dataset contains 1458 music pieces and 6 various genres such as punk, classicals, electronics, pop, world and blues. The training and test set defined by 50% for training and 50% for testing.

In MagnaTagATune [20] dataset, audio data us in mp3 format which is categorized by 32kbps and 16KHz for sample rate. This dataset is split into training of 70% and testing of 30%. This dataset contains multi-label annotations of mood, instrumentation and genres for 25,877 audio segments.

3.2 Feature extraction

The dataset signals are extracted through using various features extraction techniques: Zero Crossing Ratio (ZCR), Root Mean Square (RMS), Amplitude features, Mean curve length, Statical movement, Autocorrelation, Mean teager energy, Energy feature, Phonation rate, Mean duration passes, Total Duration, Long pass count, Short pass count, Skewness, Kurtosis, Mean cross correlation, Max cross correlation, Min cross correlation, Variance, Linear Predictive Coding (LPC) kurtosis, LPC error, LPC

Received: November 6, 2024. Revised: November 30, 2024.

Skewness, Normalized first difference, Mel-Frequency Cepstral Coefficient (MFCC), Spectral kurtosis, Spectral centroid, Pitch and Linear Spectrum. The detailed explanation is given below:

ZCR: The ZCR is a temporal feature which captures data about time-domain characteristics of audio signal. It efficiently differentiates among genres with distinct percussive and rhythmic elements.

RMS: The RMS extracts average energy or power of signal which presents its intensity and loudness. It helps to differentiate genres according to its energy levels because it captures entire amplitude characteristics.

Amplitude features: It extracts data based on signal intensity which measures volume variations and loudness in signals. It captures difference genre dynamics in which certain genres have different amplitude patterns. It has a capability to capture various genres according to their loudness characteristics thereby contribute to enhance the genre classification.

Mean curve length: It calculates average length of signal curve through time frame which captures signal complexity. It helps to differentiate musical genres by analyzing complex details which makes it effective for identifying unique patterns.

Statical movement: It captures temporal patterns which are specific genre characteristics and its analysis dynamic changes in music signals. It can differentiate among genres according to their unique progressions thereby enhancing classification accuracy.

Autocorrelation: It extracts temporal and periodic structures in audio signals by calculating the similarity of the signal with its delay across various timesteps. This approach is effective for identifying rhythmic patterns which contributes to distinguishing the unique signatures of different music genres.

Mean teager energy: It captures non-linear energy of audio signal through calculating instant energy of waveform. It has capability to highlight rhythmic and dynamic characteristics of genre classification which is helpful for identifying genre-specific patterns.

Energy feature: It captures how soft and loud piece of music that is useful for differentiate among genre with different dynamics. It reflects entire loudness patterns in genre classification based on energy distribution.

Phonation rate: It captures temporal and rhythmic aspect of vocal delivery that is distinctive among genres which assist to identify genre through analyzing patterns in vocal timing and intensity.

Mean duration passes: It is a temporal feature that calculates average time duration among zero crossing in audio signal. This feature captures temporal and rhythmic structure of music which assist to differentiate genres with distinct rhythmic patterns.

Total Duration: By analyzing typical length of zero crossing for every genres which enhances the classification accuracy. Moreover, it captures temporal characteristics which are crucial for recognizing rhythmic features which helps to differentiate genres.

Long pass count: It captures harmonics or constant sounds in audio signals which is helpful for distinguishing genres with lengthy and constant notes. It has ability to capture and highlight temporal features of various genres.

Short pass count: It captures micro-temporal patterns and significant for identifying rhythmic features and elements which helps to distinguish genres according to its rhythmic intensity and complexity.

Skewness: It extracts statistical features and captures imbalance in amplitude distribution that denotes dominance of particular frequencies elements. It assists to differentiate among genres with various characteristics.

Kurtosis: It captures outlier presence and helps to identify distinctive features of various genres in signal intensity distribution. It helps to distinguish genre with various dynamic contrast levels which is significant for accurate genre classification.

Mean cross correlation: It extracts data about similarity among various segments of signals and focuses on its time. It assists in capturing temporal and rhythmic patterns that helps identify genre-specific patterns.

Max cross correlation: It captures similarity within music and helps to identify rhythmic structures and repetitive patterns. It can detect temporal patterns which are significant to differentiate among various music genres with melodic features and characteristic rhythmic.

Min cross correlation: It captures degree of similarity among segments, periodic patterns within music which assist to identify structural similarities and rhythmic patterns thereby enhancing model’s ability.

Variance: It captures variability of audio features over time and reflects complexity of sound which assist to differentiate genres based on variations and improves classification accuracy.

LPC kurtosis: This captures subtle variations in spectral which includes audio signals and assist to identifying unique features of various music genres thereby enhancing classification accuracy.

LPC error: It captures residuals generated by LPC when predicting audio signals which contribute to genre classification through providing spectral and texture data.

LPC Skewness: It measures asymmetry in LPC coefficient distributions which captures spectral features of music in frequency domain. It provides understanding into texture content for enhancing music genre classification which differentiating among genres with different spectral properties.

Normalized First difference: It extracts the rate of change in the audio signal which highlights variations that capture dynamic music features and effectively highlight temporal changes within audio.

MFCC: It captures short-term power spectrum of audio signal through relevant features which makes useful for music genre classification. It provides compressed presentation of audio features which helps to differentiate among various genres.

Spectral Kurtosis: It assists to identify temporal components in audio signals through gaussian distribution in spectral domain. It has ability to distinguish among genres based on its unique spectral features thereby enhancing classification accuracy.

Spectral Centroid: It is used to define sound characteristics which assist to differentiate among music genres based on booth textual and spectral features. By capturing variations in perceived sound brightness which enhances genres differentiation.

Pitch: It analyzes pitch variations to capture musical frequencies and structures that enhance classification performance.

Linear Spectrum: It extracts signals across different frequencies and provides a detailed presentation of frequency content in audio signals. It can differentiate textures by distinguishing among genres with similar rhythmic structures.

Each method captures various features of audio signal for music genre classification and these features are fused by using concatenation which enhances the accuracy and reliability in differentiating music genres. From this technique, 560 features are extracted and given to feature selection stage for selecting relevant features.

3.3 Feature selection

The HOA is a population-based algorithm which is used for selecting features in that search agents are hippopotamuses. The Hippopotamus Optimization Algorithm (HOA) demonstrates better balance among exploration and exploitation to avoid local optima and ensure global convergence than Carpet Weaver Optimization (CWO) [27], Sculptor Optimization Algorithm (SOA) [28], Swarm Bipolar Algorithm (SBA) [29], Migration-Crossover Algorithm (MCA) [30] and Potter Optimization Algorithm (POA) [31]. The HOA has three unique behaviors in three stages such as position updating, defensive behavior against predators and evasion from predators which improves its ability to achieve global convergence. The defensive behavior phase integrates random movements based on Levy distribution that improves the position adjustments for enhanced global search abilities. Additionally, it reduces the computational time by effective convergence and adaptive population to avoid early convergence. The location update of every hippopotamus in search space represents the decision adjustable variables. Hence, every hippopotamus is presented as vector and its population is categorized by matrix. The initialization stage includes randomized initial solution generation [26]. At this time, the decision variable vectors are produced by Eq. (1),

$$X_i : x_{ij} = lb_j + r \cdot \left(ub_j - lb_j\right), i = 1,2,\dots,N; j = 1,2,\dots,m \qquad (1)$$

Where, X_i is a location of ith solution, r is a random number in range of $[0, 1]$, lb_j and ub_j are lower and upper bounds of jth decision variable, N and m are population size and number of decision variables.

3.3.1. Phase 1: Position update (Exploration)

The dominant hippopotamus is defined according to the iteration of objective function value. Generally, hippopotamuses manage to gather near proximity to each other and its leading male hippopotamuses defend territory and herd from probable threats. Moreover, numerous females are located around male hippopotamuses. Once maturity is reached, male is removed from herd through leading male. Then, the removed male individuals are essential to fascinate females or involve in domination challenges by recognized male of herd to create own domination. The position of male hippopotamus members within a herd in a pond or lake is determined by Eq. (2),

$$X_i^{Mhippo} : x_{ij}^{Mhippo} = x_{ij} + y1 \cdot \left(Dhippo - I_1 x_{ij}\right) \qquad (2)$$
$$for\ i = 1,2,\dots,\left[\frac{N}{2}\right] \quad and \quad j = 1,2,\dots,m$$

Where, X_i^{Mhippo} is a location of male hippopotamus, $Dhippo$ is dominant hippopotamus location, $\vec{r}_{1,\dots,4}$ is a random vector among $[0, 1]$, r_5 is a random number among $[0, 1]$. I_1 and I_2 is integer among 1 and 2. The MG_i is a mean score of randomly selected hippopotamus in same probability of involving present evaluated hippopotamus (X_i) and $y1$ is a random number among $[0, 1]$. The formula for herd and territory are given in Eq. (3) and (4), the female or immature hippopotamus new location is given in Eq. (5) and (6),

$$h = \begin{cases} I_1 \times \vec{r}_1 + (\sim \varrho_1) \\ 2 \times \vec{r}_2 - 1 \\ \vec{r}_3 \\ I_2 \times \vec{r}_4 + (\sim \varrho_2) \\ r_5 \end{cases} \tag{3}$$

$$T = \exp\left(-\frac{t}{T}\right) \tag{4}$$

$$X_i^{FBhippo}: x_{ij}^{FBhippo} = \begin{cases} x_{ij} + h_1 \cdot (Dhippo - I_2 MG_i) & T > 0.6 \\ \Xi & \text{else} \end{cases} \tag{5}$$

$$\Xi = \begin{cases} x_{ij} + h_2 \cdot (MG_i - Dhippo) & r_6 > 0.5 \\ lb_j + r_7 \cdot (ub_j - lb_j) & else \end{cases} \tag{6}$$

$$for\ i = 1,2,\dots,\left[\frac{N}{2}\right] \quad and \quad j = 1,2,\dots,m$$

Where, ϱ_1 and ϱ_2 are integer random numbers in 0 or 1. The immature hippopotamus are near to their mothers but sometimes it separated away from their mothers because of curiosity. If T is higher than 0.6 means it has distanced from its mother. The r_6 and r_7 are random number among $[0, 1]$. If r_6 is greater than 0.5 means it is separated from its mother however it is adjacent to herd, otherwise it signifies partition from the herd. The h_1 and h_2 are randomly selected vectors from five different scenarios. The male and female hippopotamus location is updated within the herd by Eq. (7) and (8),

$$X_i = \begin{cases} X_i^{Mhippo} & F_i^{Mhippo} < F_i \\ X_i & else \end{cases} \tag{7}$$

$$X_i = \begin{cases} X_i^{FBhippo} & F_i^{FBhippo} < F_i \\ X_i & else \end{cases} \tag{8}$$

By using h vectors, I_1 and I_2 improves the global search and improves exploration of HOA. The HOA leads optimal convergence and minimized computation time by effectively exploring and exploiting search space. It balances between exploration and exploitation which helps to select relevant features for enhancing classification accuracy.

3.3.2. Phase 2: Defense against predators (Exploration)

The main defensive mechanism of hippopotamuses is to quickly turn and face predators while producing loud vocalizations to prevent predators from imminent closely. At this stage, it exhibits approaching predator behavior to persuade which effectively preventing potential threat. The predator location in search space is given in Eq. (9), the predator distance is given in Eq. (10),

$$Predator: Predator_j = lb_j + \vec{r}_8 \cdot (ub_j - lb_j), \; j = 1,2,\dots,m \tag{9}$$

$$\vec{D} = \left| Predator_j - x_{ij} \right| \tag{10}$$

Where, $\vec{r}_8$ is a random number among $[0, 1]$. At this time, the hippopotamus assumes defense behavior based on $F_{Predator}$ factor to protect it from predators. If $F_{Predator}$ is lesser than F_i which indicates the presence of a predator in close proximity to hippopotamus. In this scenario, the hippopotamus turns quickly to predator and moves to make it retreat. If $F_{Predator}$ is higher than F_i indicates the predator entity is at higher distance from its territory. In this scenario, the hippopotamus turns quickly to predator with less movement range as Eq. (11).

$$X_i^{HippoR}: x_{ij}^{HippoR} = \begin{cases} \overrightarrow{RL} \oplus Predator_j + \left(\frac{f}{(\sigma d \times \cos(2\pi g))}\right) \cdot \left(\frac{1}{\vec{D}}\right) & F_{Predator\,j} < F_i \\ \overrightarrow{RL} \oplus Predator_j + \left(\frac{f}{(\sigma d \times \cos(2\pi g))}\right) \cdot \left(\frac{1}{2 \times \vec{D} + \vec{r}_9}\right) & F_{Predator\,j} \geq F_i \end{cases}$$

$$for\ i = \left[\frac{N}{2}\right] + 1, \left[\frac{N}{2}\right] + 2, \dots, N\ and\ j = 1,2,\dots,m \tag{11}$$

Where, X_i^{HippoR} is a location that faces to predator, f, D, g and σ are uniform random numbers among $[2, 4], [2,3], [-1,1]$ and $[1, 1.5]$. $\vec{r}_9$ is a random vector with $1 \times m$ dimension. The $\overrightarrow{RL}$ is a random vector with Levy distribution which is applied for quick changes in predator location at the hippopotamus attack. The levy movement is estimated by Eq. (12) and (13).

$$Levy\,(\vartheta) = 0.05 \times \frac{w \times \sigma_w}{|v|^{\frac{1}{\vartheta}}} \tag{12}$$

$$\sigma_w = \left[\frac{\Gamma(1+\vartheta)\sin\left(\frac{\pi\vartheta}{2}\right)}{\Gamma\left(\frac{(1+\vartheta)}{2}\right)\vartheta 2^{\frac{(\vartheta-1)}{2}}}\right]^{\frac{1}{\vartheta}} \tag{13}$$

Where, w and v are random numbers in $[0,1]$, ϑ is a constant, Γ is gamma function abbreviation. The updated new position is given in Eq. (14),

$$X_i = \begin{cases} X_i^{HippoR} & F_i^{HippoR} < F_i \\ X_i & F_i^{HippoR} \geq F_i \end{cases} \tag{14}$$

Based on the above Eq. (14), if F_i^{HippoR} is higher than F indicates that the hippopotamus is hunted and others in the herd change their positions accordingly. Otherwise, the hunter escapes and hippopotamus return to the herd. Important improvements are observed in the global search procedure at second phase. The initial and second phases balance each other to prevent from getting trapped in local minima.

3.3.3. Phase 3: Escaping from predators (Exploitation)

In this scenario, the hippopotamus enters its predator phase when it encounters a set of predators with its defensive behavior. This behavior enhances its ability to exploit local search space by finding a safer location near its present location. To stimulate this behavior, random location is generated near the present location of the hippopotamus based on Eq. (15)-(18),

$$lb_j^{local} = \frac{lb_j}{t}, ub_j^{local} = \frac{ub_j}{t}, \quad t = 1,2,\dots,T \tag{15}$$

$$X_i^{Hippo\mathcal{E}}: x_{ij}^{Hippo\mathcal{E}} = x_{ij} + r_{10} \cdot \begin{pmatrix} lb_j^{local} + s1 \cdot \\ \left(ub_j^{local} - lb_j^{local}\right) \end{pmatrix}$$
$$for\ i = 1,2,\dots,N\ and\ j = 1,2,\dots,m \tag{16}$$

$$s = \begin{cases} 2 \times \vec{r}_{11} - 1 \\ r_{12} \\ r_{13} \end{cases} \tag{17}$$

$$X_i = \begin{cases} X_i^{Hippo\mathcal{E}} & F_i^{Hippo\mathcal{E}} < F_i \\ X_i & F_i^{Hippo\mathcal{E}} \geq F_i \end{cases} \tag{18}$$

Where, t and T are present and maximum iteration, $X_i^{Hippo\mathcal{E}}$ is a hippopotamus location which search to locate the nearest safer place, $s1$ is a random vector that selects between three scenarios as Eq. (17). $\vec{r}_{11}$ is a random vector among $[0,1]$, r_{12} and r_{13} are a random number among $[0,1]$. If newly generated location enhances the cost function means the hippopotamus find safe location nearby its present location. After every iteration, all population members are updated based on three phases and continues until the final iteration. During the algorithm execution, the optimal solution is tracked and stored. The parameters for HOA-based feature selection include population size of 30, a maximum of 100 iterations and threshold of 0.50. From this technique, 474 features are selected and given to classification stage for classifying music genres.

3.4 Classification

The LSTM is a type of Recurrent Neural Network (RNN) which records long and short-term data at similar period and resolve gradient disappearance which occurs at the RNN training. The LSTM has three various gate such as forget, input and output gate which are applied to learn long-term data and remove pointless data. The X_t is an input sequence which is a system state to monitor data for music genre classification, h_t is a hidden layer output which is a learning result of every LSTM. The f_t, i_t and O_t is a forgetting, input and output. σ is a sigmoid function, $tanh$ is an activation function, W is a weight matrix, × is point pair product.

3.4.1. Forget gate

The LSTM process time-series data in a particular direction and data contains certain time period. The forget gate function determines which data to retain and which data to discard directly. The forget gate output is given in Eq. (19),

$$f_t = \sigma\left(W_f \cdot [h_{t-1}, x_t] + b_f\right) \tag{19}$$

Where, W_f and b_f are forget gate weight and bias.

3.4.2. Input gate

The data provides input gate after forget gate selection and its function determines which parameter requires to update. The input gate output is given in Eq. (20)-(22),

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{20}$$

$$C_t = \tanh\left(W_C \cdot [h_{t-1}, x_t] + b_C\right) \tag{21}$$

$$C_t = f_t \times C_{t-1} + i_t \times C_t \tag{22}$$

Where, W_i, W_C, b_i and b_C are its respective weights and bias correspondingly, C_t is a present cell state value.

3.4.3. Output gate

The data extends the output gate following the forget and input gates and its function is to determine which data is passed as output. The output gate result is given in Eq. (23) and (24),

$$O_t = \sigma(W_O \cdot [h_{t-1}, x_t] + b_O) \tag{23}$$

$$h_t = O_t \cdot \tanh(C_t) \tag{24}$$

Where, W_O and b_O are output gate weight and bias, h_t is a current unit output value. The LSTM can effectively capture temporal dependencies in audio signal. Moreover, it recognizes and retains long-term patterns, making it suitable for sequential data and enhancing classification performance.

4. Result evaluation

The HOA-LSTM is implemented in MATLAB R2020b with configuration of processor intel i5, operating system windows 10 and 16GB RAM. The performance of HOA-LSTM is estimated using accuracy, sensitivity, specificity, PPV, f-measure, error rate and computation time and its numerical formula is given in Eq. (25)-(29),

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{25}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{26}$$

$$Specificity = \frac{TN}{TN+FP} \tag{27}$$

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

$$F - Measure = 2 \times \frac{PPV \times Sensitivity}{PPV + Sensitivity} \tag{29}$$

Where, TP, TN, FP and FN are True Positive, True Negative, False Positive and False Negative respectively.

Table 1 shows the evaluation of various feature selection techniques to compare performance with proposed HOA for GTZAN dataset. The conventional techniques such as Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Gorilla Optimization Algorithm (GOA) and Pelican Optimization Algorithm (POA) performance are taken and compared with proposed HOA. The HOA achieves 98.47%, 95.99%, 98.76%, 91.86%, 93.95% and 1.53 of accuracy, sensitivity, specificity, PPV, f-measure and error rate respectively.

Table 2 shows the evaluation of various classifiers with actual features to compare performance with the proposed LSTM for GTZAN dataset. The conventional techniques such as autoencoder, RNN, Deep Neural Network (DNN), BiLSTM and Gated Recurrent Unit (GRU) performance are taken and compared with the proposed HOA. The LSTM with actual feature achieves 95.68%, 92.35%, 96.56%, 90.25%, 91.29% and 4.31% of accuracy, sensitivity, specificity, PPV, f-measure and error rate respectively.

Table 1. Evaluation of various feature selection for GTZAN dataset

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	F-measure (%)	Error rate
PSO	89.57	87.35	85.68	83.56	85.42	10.4
ACO	85.60	86.45	83.56	85.65	86.05	14.3
GOA	92.65	90.36	89.54	90.26	90.31	7.35
POA	95.68	92.35	90.23	90.56	91.45	4.32
HOA	98.47	95.99	98.76	91.99	93.95	1.53

Table 2. Evaluation of classifier with actual features for GTZAN dataset

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	F-measure (%)	Error rate
Autoencoder	82.57	80.25	81.23	83.56	81.87	17.4
RNN	89.15	87.56	83.56	86.25	86.90	10.8
DNN	79.54	75.68	75.56	79.56	77.57	20.4
BiLSTM	95.64	93.56	95.58	86.24	89.75	4.35
GRU	90.25	87.56	90.26	88.57	88.06	9.75
LSTM	95.68	92.35	96.56	90.25	91.29	4.31

Received: November 6, 2024. Revised: November 30, 2024.

Table 3. Evaluation of classifier with optimized features for GTZAN dataset

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	F-measure (%)	Error rate
Autoencoder	84.57	82.35	83.56	85.45	83.87	15.4
NN	90.15	89.25	85.65	88.56	88.90	9.85
DNN	81.24	79.25	78.56	80.24	79.74	18.7
BiLSTM	98.21	95.68	97.58	89.77	92.63	1.79
GRU	92.35	89.45	91.25	90.45	89.95	7.64
LSTM	98.47	95.99	98.76	91.99	93.95	1.53

Table 4. K-fold validation of proposed LSTM for GTZAN dataset

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	F-measure (%)	Error rate
K=3	97.15	92.35	95.25	90.25	91.29	2.85
K=5	98.47	95.99	98.76	91.99	93.95	1.53
K=7	96.56	90.45	94.56	89.26	89.85	3.44
K=9	95.24	88.45	93.65	87.57	88.00	4.76

Table 3 shows the evaluation of various classifier with optimized features to compare performance with proposed LSTM for GTZAN dataset. The conventional techniques such as autoencoder, RNN, DNN, BiLSTM and GRU performance are taken and compared with proposed HOA.

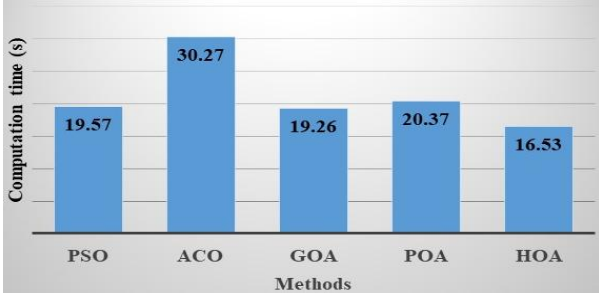
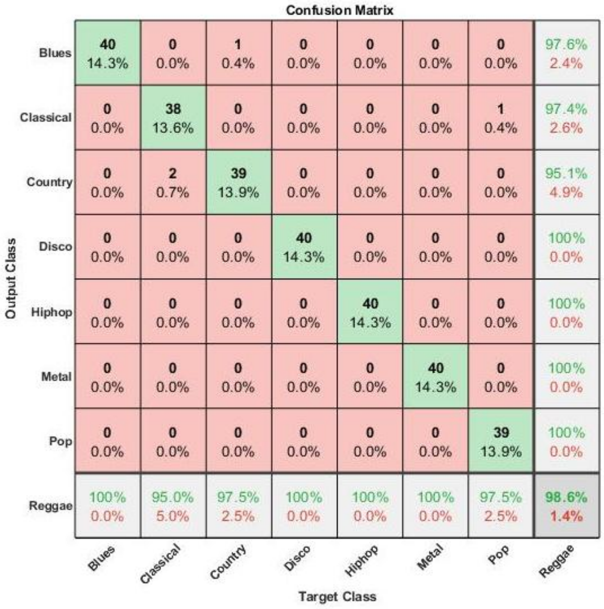


Figure. 2 Evaluation of feature selection in terms of computation time for GTZAN dataset



	Blues	Classical	Country	Disco	Hiphop	Metal	Pop	Reggae
Blues	40 14.3%	0 0.0%	1 0.4%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	97.6% 2.4%
Classical	0 0.0%	38 13.6%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	1 0.4%	97.4% 2.6%
Country	0 0.0%	2 0.7%	39 13.9%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	95.1% 4.9%
Disco	0 0.0%	0 0.0%	0 0.0%	40 14.3%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
Hiphop	0 0.0%	0 0.0%	0 0.0%	0 0.0%	40 14.3%	0 0.0%	0 0.0%	100% 0.0%
Metal	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	40 14.3%	0 0.0%	100% 0.0%
Pop	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	39 13.9%	100% 0.0%
Reggae	100% 0.0%	95.0% 5.0%	97.5% 2.5%	100% 0.0%	100% 0.0%	100% 0.0%	97.5% 2.5%	98.6% 1.4%

Figure. 3 Confusion matrix for GTZAN dataset

The LSTM with optimized features achieves 98.47%, 95.99%, 98.76%, 91.86%, 93.95% and 1.53 of accuracy, sensitivity, specificity, PPV, f-measure and error rate respectively.

Table 4 shows the performance of k-fold validation of the proposed LSTM for GTZAN dataset. The different K-values such as K=3,5,7 and 9 are considered with metrics of accuracy, sensitivity, specificity, PPV, f-measure and error rate for the GTZAN dataset. The LSTM achieves better performance at K=5 in terms of 98.47%, 95.99%, 98.76%, 91.86%, 93.95% and 1.53 of accuracy, sensitivity, specificity, PPV, f-measure and error rate respectively. Fig. 2 displays the computation time of the proposed HOA with various conventional techniques such as PSO, ACO, GOA and POA. The HOA achieves less computation time of 16.53s due to its efficiency in balancing exploration and exploitation through its adaptive search mechanism. It minimizes the number of iterations and quickens the convergence to the best solutions.

4.1 Comparative analysis

The comparison of the proposed HOA-LSTM with existing techniques for the GTZAN, ISMIR2004 and MagnaTagATune datasets is provided in this subsection. The existing methods such as DL BAG [20], RGLUformer [21], RF [22], SVM [23] and BiLSTM-VGG-16 Net [24] are taken to compare the proposed HOA-LSTM performance. The different metrics such as accuracy, sensitivity, specificity, PPV and f-measure are considered to evaluate the performance. The HOA-LSTM achieves 98.47%, 95.99%, 98.76%, 91.86% and 93.95% of accuracy, sensitivity, specificity, PPV and f-measure respectively for GTZAN dataset. The HOA-LSTM achieves 97.58%, 97.75%, 95.68% and 97.26% of accuracy, sensitivity, specificity and f-measure respectively for ISMIR2004 dataset.

Received: November 6, 2024. Revised: November 30, 2024.

Table 5. Comparison of proposed HOA-LSTM for GTZAN dataset

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	F-measure (%)
DL BAG [20]	93.51	93.25	93.78	NA	92.99
RGLUformer [21]	96.8	91.3	NA	89.4	90.3
RF [22]	90	91.62	NA	90.78	91.20
SVM [23]	91.8	NA	NA	NA	NA
BiLSTM-VGG-16 Net [24]	97.8	93.8	NA	87.9	77.8
Proposed HOA-LSTM	98.47	95.99	98.76	91.99	93.95

Table 6. Comparison of proposed HOA-LSTM for ISMIR2004 dataset

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-measure (%)
DL BAG [20]	92.49	92.61	92.38	92.03
SVM [23]	90.9	NA	NA	NA
BiLSTM-VGG-16 Net [24]	96.5	94	NA	77
Proposed HOA-LSTM	97.58	97.75	95.68	97.26

Table 7. Comparison of proposed HOA-LSTM for MagnaTagATune dataset

Method	Accuracy (%)	Sensitivity (%)	Specificity (%)	F-measure (%)
DL BAG [20]	92.13	92.44	91.82	91.75
Proposed HOA-LSTM	96.74	95.87	93.48	95.82

The HOA-LSTM achieves 96.74%, 95.87%, 93.48% and 95.82% of accuracy, sensitivity, specificity and f-measure respectively for MagnaTagATune dataset. Table 5, 6 and 7 shows the comparison of proposed HOA-LSTM for GTZAN, ISMIR2004 and MagnaTagATune dataset respectively.

4.2 Discussion

The performance of HOA and LSTM is compared against various optimization algorithms and classifiers. Evaluation includes different feature selection approaches, classification using both actual and optimal features, and k-fold validation. In prior works, DL BAG [20] faced overfitting issues due to irrelevant features, decreasing model performance. RGLUformer [21] experienced gradient vanishing during training, leading to misclassification and reduced accuracy. The RF model [22] failed to classify music genres optimally due to poor boundary recognition, while SVM [23] struggled with feature dimensionality, which hurt classification performance. The BiLSTM-VGG-16 Net [24] was hindered by noise signals, reducing classification efficiency. To address these shortcomings, this research proposes the HOA-LSTM approach. The HOA-based feature selection method effectively eliminates irrelevant features, improving model performance, while LSTM handles genre classification by addressing the vanishing gradient issue, leading to higher classification accuracy.

5. Conclusion

The HOA-LSTM method proposed in this research enhances classification accuracy by leveraging different feature extraction techniques. Temporal and structural features are derived from raw audio signals to differentiate genres. HOA-based feature selection removes inappropriate features, improving model performance through better convergence and reduced computation time by efficiently exploring and exploiting the search space. LSTM classifies the genre classes, mitigating the vanishing gradient problem and capturing temporal dependencies and patterns in audio signals, thus enhancing classification accuracy. The proposed HOA-LSTM method achieves an accuracy of 98.47% and an error rate of 1.53% on the GTZAN dataset, outperforming conventional methods. Future work aims to further reduce the error rate and improve music genre classification performance.

Notation

Notation	Description
X_i	Location of ith solution
r	Random number in range of $[0, 1]$
lb_j and ub_j	Lower and upper bounds of jth variable
N	Population size
m	Number of decision variables
X_i^{Mhippo}	Location of male hippopotamus
$Dhippo$	Dominant hippopotamus location
I_1 and I_2	Integer among 1 and 2

MG_i	Mean score of randomly selected hippopotamus
ϱ_1 and ϱ_2	Integer random numbers in 0 or 1
h_1 and h_2	Randomly selected vectors
X_i^{HippoR}	Hippopotamus location
f, D, g and σ	Uniform random numbers
$\overrightarrow{RL}$	Random vector with Levy distribution
ϑ	Constant
Γ	Gamma function
t and T	Present and maximum iteration
$s1$	Random vector
X_t	Input sequence
h_t	Hidden layer output
f_t, i_t and O_t	Forgetting, input and output gates
σ	Sigmoid function
$tanh$	Activation function
W	Weight matrix
$\times$	Point pair product
W_f and b_f	Forget gate weight and bias
W_i and b_i	Input gate weight and bias
W_o and b_o	Output gate weight and bias
C_t	Present cell state value
TP	True positive
TN	True negative
FP	False positive
FN	False negative

Conflicts of Interest

The authors declare no conflict of interest.

Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision and project administration, have been done by 2nd author.

References

[1] N. Farajzadeh, N. Sadeghzadeh, and M. Hashemzadeh, "PMG-Net: Persian music genre classification using deep neural networks," *Entertainment Computing*, Vol. 44, p. 100518, 2023.

[2] Z. Xu, "Construction of intelligent recognition and learning education platform of national music genre under deep learning", *Frontiers in Psychology*, Vol. 13, p. 843427, 2022.

[3] Y. Li, Z. Zhang, H. Ding, and L. Chang, "Music genre classification based on fusing audio and lyric information", *Multimedia Tools and Applications*, Vol. 82, No. 13, pp. 20157-20176, 2023.

[4] Y. Singh, and A. Biswas, "Robustness of musical features on deep learning models for music genre classification", *Expert Systems with Applications*, Vol. 199, p. 116879, 2022.

[5] G. R. Sumanth, U. Priyadarshini, "Music genre classification using linear regression compared with extreme gradient boost algorithm with improved accuracy", *Journal of Pharmaceutical Negative Results*, Vol. 13, N0. 4, pp. 1659-1665, 2022.

[6] Y. Mao, G. Zhong, H. Wang, and K. Huang, "Music-CRN: An efficient content-based music classification and recommendation network", *Cognitive Computation*, Vol. 14, No. 6, pp. 2306-2316, 2022.

[7] N. N. Wijaya, and A. R. Muslikh, "Music-genre classification using Bidirectional long short-term memory and mel-frequency cepstral coefficients", *Journal of Computing Theories and Applications*, Vol. 1, No. 3, pp. 243-256, 2024.

[8] J. Liu, "An automatic classification method for multiple music genres by integrating emotions and intelligent algorithms", *Applied Artificial Intelligence*, Vol. 37, No. 1, p. 2211458, 2023.

[9] B. Kumaraswamy, "Optimized deep learning for genre classification via improved moth flame algorithm", *Multimedia Tools and Applications*, Vol. 81, No. 12, pp. 17071-17093, 2022.

[10] Z. Liu, T. Bian, and M. Yang, "Locally activated gated neural network for automatic music genre classification", *Applied Sciences*, Vol. 13, No. 8, p. 5010, 2023.

[11] A. E. C. Salazar, "Hierarchical mining with complex networks for music genre classification", *Digital Signal Processing*, Vol. 127, p. 103559, 2022.

[12] Q. He, "A Music Genre Classification Method Based on Deep Learning", *Mathematical Problems in Engineering*, Vol. 2022, No. 1, p. 9668018, 2022.

[13] K. M. Hasib, A. Tanzim, J. Shin, K. O. Faruk, J. Al Mahmud, and M. F. Mridha, "Bmnet-5: A novel approach of neural network to classify the genre of bengali music based on audio features", *IEEE Access*, Vol. 10, pp. 108545-108563, 2022.

[14] Y. Singh, and A. Biswas, "Lightweight convolutional neural network architecture design for music genre classification using evolutionary stochastic hyperparameter selection", *Expert Systems*, Vol. 40, No. 6, p. e13241, 2023.

[15] P. S. Yadav, S. Khan, Y. V. Singh, P. Garg, and R. S. Singh, "A Lightweight Deep Learning-Based Approach for Jazz Music Generation in MIDI Format", *Computational Intelligence and*

Neuroscience, Vol. 2022, No. 1, p. 2140895, 2022.

[16] B. Jaishankar, R. Anitha, F. D. Shadrach, M. Sivarathinabala, and V. Balamurugan, "Music Genre Classification Using African Buffalo Optimization", *Computer Systems Science & Engineering*, Vol. 44, No. 2, pp. 1823-1836, 2023.

[17] Y. H. Cheng, and C. N. Kuo, "Machine learning for music genre classification using visual mel spectrum", *Mathematics*, Vol. 10, No. 23, p. 4427, 2022.

[18] Z. Wen, A. Chen, G. Zhou, J. Yi, and W. Peng, "Parallel attention of representation global time–frequency correlation for music genre classification", *Multimedia Tools and Applications*, Vol. 83, No. 4, pp. 10211-10231, 2024.

[19] M. Faizan, I. Intzes, I. Cretu, and H. Meng, "Implementation of deep learning models on an SoC-FPGA device for real-time music genre classification", *Technologies*, Vol. 11, No. 4, p. 91, 2023.

[20] S. K. Prabhakar, and S. W. Lee, "Holistic approaches to music genre classification using efficient transfer and deep learning techniques", *Expert Systems with Applications*, Vol. 211, p. 118636, 2023.

[21] C. Xie, H. Song, H. Zhu, K. Mi, Z. Li, Y. Zhang, J. Cheng, H. Zhou, R. Li, and H. Cai, "Music genre classification based on res-gated CNN and attention mechanism", *Multimedia Tools and Applications*, Vol. 83, No. 5, pp. 13527-13542, 2024.

[22] M. Chaudhury, A. Karami, and M.A. Ghazanfar, "Large-scale music genre analysis and classification using machine learning with apache spark", *Electronics*, Vol. 11, No. 16, p. 2567, 2022.

[23] X. Cai, and H. Zhang, "Music genre classification based on auditory image, spectral and acoustic features", *Multimedia Systems*, Vol. 28, No. 3, pp. 779-791, 2022.

[24] W. Hongdan, S. SalmiJamali, C. Zhengping, S. Qiaojuan, and R. Le, "An intelligent music genre analysis using feature extraction and classification using deep learning techniques", *Computers and Electrical Engineering*, Vol. 100, p. 107978, 2022.

[25] GTZAN dataset link: https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification (accessed on September 2024).

[26] M.H. Amiri, N. Mehrabi Hashjin, M. Montazeri, S. Mirjalili, and N. Khodadadi, "Hippopotamus optimization algorithm: a novel nature-inspired optimization algorithm", *Scientific Reports*, Vol. 14, No, 1, p.5032, 2024.

[27] S. Alomari, K. Kaabneh, I. AbuFalahah, S. Gochhait, I. Leonova, Z. Montazeri, M. Dehghani, and K. Eguchi, "Carpet Weaver Optimization: A Novel Simple and Effective Human-Inspired Metaheuristic Algorithm", *International Journal of Intelligent Engineering & Systems*, Vol. 17, No. 4, pp. 230-242, 2024, doi: 10.22266/ijies2024.0831.18.

[28] T. Hamadneh, K. Kaabneh, O. AlSayed, G. Bektemyssova, Z. Montazeri, M. Dehghani, and K. Eguchi, "Sculptor Optimization Algorithm: A New Human-Inspired Metaheuristic Algorithm for Solving Optimization Problems", *International Journal of Intelligent Engineering & Systems*, Vol. 17, No. 4, pp. 564-575,2024, doi: 10.22266/ijies2024.0831.43.

[29] P.D. Kusuma, and A. Dinimaharawati, "Swarm Bipolar Algorithm: A Metaheuristic Based on Polarization of Two Equal Size Sub Swarms", *International Journal of Intelligent Engineering & Systems*, Vol. 17, No. 2, pp. 377-389, 2024, doi: 10.22266/ijies2024.0430.31.

[30] P.D. Kusuma, and M. Kallista, "Migration-Crossover Algorithm: A Swarm-based Metaheuristic Enriched with Crossover Technique and Unbalanced Neighbourhood Search", *International Journal of Intelligent Engineering & Systems*, Vol. 17, No. 1, pp. 698-710, 2024, doi: 10.22266/ijies2024.0229.59.

[31] T. Hamadneh, B. Batiha, O. Alsayyed, G. Bektemyssova, Z. Montazeri, M. Dehghani, and K. Eguchi, "On the Application of Potter Optimization Algorithm for Solving Supply Chain Management Application", *International Journal of Intelligent Engineering & Systems*, Vol. 17, No. 5, pp. 88-99, 2024, doi: 10.22266/ijies2024.1031.09.