scientific reports

Check for updates

OPEN

Advanced music classification using a combination of capsule neural network by upgraded ideal gas molecular movement algorithm

Peiyan Chen[1], Jichi Zhang[2]✉ & Arsam Mashhadi[3,4]✉

Music genres classification has long been a challenging task in the field of Music Information Retrieval (MIR) due to the intricate and diverse nature of musical content. Traditional methods have struggled to accurately capture the complex patterns that differentiate one genre from another. However, recent advancements in deep learning have presented new opportunities to tackle this challenge. One such approach is the use of Capsule Neural Networks (CapsNet), which have shown promise in capturing hierarchical relationships within data. Nevertheless, the performance of CapsNet models heavily depends on the optimal configuration of their parameters, which is a complex task. To address this issue, this research proposes a novel methodology that combines CapsNet with an upgraded version of the Ideal Gas Molecular Movement (UIGMM) optimization algorithm. By utilizing the UIGMM algorithm, the parameters of the CapsNet model can be fine-tuned, thereby enhancing its ability to accurately recognize and classify different music genres. The effectiveness of this proposed model is evaluated using three benchmark datasets: ISMIR2004, GTZAN, and Extended Ballroom. Through comparative analysis against state-of-the-art models, the proposed approach demonstrates superior performance, highlighting its potential as a robust tool for music genre classification.

Keywords Music genre classification, Capsule neural network, Ideal gas molecular movement, UIGMM, Deep learning, Metaheuristic algorithm, ISMIR2004, GTZAN, Extended ballroom

Throughout the course of history, music has experienced numerous transformations, and one of the notable changes has been the emergence of diverse musical styles. This development has significantly expanded the realm of music science for individuals interested in this field[1]. The evolution of different music genres can be attributed to various factors including the type of performance, musical instruments and musicians involved, as well as the historical and geographical context of different regions[2].

For many music enthusiasts, the desire to begin learning a musical instrument often arises, yet they may be unsure about which style to pursue or which styles align with their interests[3]. To address this, it is beneficial to approach the exploration of music styles from different perspectives. This can involve examining the instruments utilized in a particular style, the unique rhythms and steps associated with it, as well as the various arrangements and genres within that style[4]. By gaining a comprehensive understanding of these aspects, one can develop a more informed perspective.

Furthermore, actively listening to different types of music can be highly effective in discovering personal preferences[5]. By immersing oneself in a wide range of musical genres, individuals can identify the styles that resonate with them the most[6]. This exploration can not only aid in finding one's preferred way of listening to songs but also serve as a starting point for embarking on a journey to learn music[7].

There are various ways for different music genres classification[8]. Due to its subjective nature, music primarily depends on personal preference, making these classifications highly contentious. This is because there is a potential for music to overlap across different styles.

The field of Music Information Retrieval (MIR) has long grappled with the challenge of automatically classifying music genres[9]. This task is made difficult by the diverse and complex nature of music, which encompasses a wide range of features and characteristics that can vary greatly between genres. Traditional approaches have typically relied on manual feature extraction and classical machine learning techniques, which, while somewhat effective,

[1]Wenhua College, Wuhan 430074, Hubei, China. [2]Central Saint Martins College of Art and Design, University of the Arts London, London N1C 4AA, UK. [3]Arak Branch, Islamic Azad University, Arak, Iran. [4]College of Technical Engineering, The Islamic University, Najaf, Iraq. ✉email: j.zhang0120211@arts.ac.uk; arsammashhadi@gmail.com

fail to capture the deeper, hierarchical patterns present in music data. However, the emergence of deep learning has provided new tools to address this challenge which show promise in understanding the spatial hierarchies within music data.

For instance, Regarding the present study, Dong[10] suggested an approach that integrated human research data in classifying music genres auditory neurophysiology model. A straightforward CNN (Convolutional Neural Network) was utilized to categorize a small segmentation of the music. After that, the music genres were ascertained via dividing it into small segmentations and integrating forecasts of CNN. Once the training stage was accomplished, the approach could obtain the accuracy of human with the value of 70%.

Chillara et al. discovered a superior machine learning algorithm compared to the ones existed that could forecast various genres of music[11]. Initially, several models of categorization were made and trained utilizing dataset FMA (Free Music Archive). The efficiencies of all of the models were contrasted with each other. Some of the models were trained using the songs' mel-spectrograms in addition to their attributes of audio, while others were trained merely utilizing the songs' spectrograms. It was revealed that CNN could obtain the best accuracy amid all models with the value of 88.54%.

A research was conducted by Ashraf et al.[12] that GLR (Global Layer Regularization) was suggested in accordance with RNN and CNN hybrid approach utilizing Mel-spectrograms for evaluating training and accuracy. The results obtained led to an improvement in efficacy based on the information collected from the FMA (Free Music Achieve) and GTZAN datasets. Furthermore, GTZAN and FMA's accuracies turned out to be 87.79% and 68.87%. The model proposed utilized the attributes of the spatiotemporal ground effectively and employed a GLR to obtain remarkable accuracy, outperforming additional advanced investigations.

Sharma et al.[13] implemented a study and utilized two models of categorization. In the present study, Mel frequency cepstral coefficients were utilized similar to features. Moreover, it carried out some methods, such as DNN (one, two, and three layers), CNN (one, two, and three layers), SVM (Sigmoid, Polynomial & Gaussian Kernel), and RNN-LSTM as a method. A three-channel input was made by integrating some attributes, such as Scalogram, Spectrogram, and MFCC, and employed several approaches, such as ResNet-50, CNN (one, two, and three layers), and VGG-16. RNN-LSTM and a three-layered CNN could perform significantly great compared to other methods.

In the following, Li et al.[14] implemented an investigation and utilized several cutting-edge DNN approaches for music classification. Additionally, spectrograms evaluated the efficacy of the models. Initially, the files of audio got changed to spectrograms through model transformation. Next, the songs got categorized through deep learning. Two models, namely balanced ResNet50_trust and trusted function of loss, were created to alleviate the problem of overfitting during the training process. In the end, in the end, the efficacy of the various Deep Neural Networks was contrasted. The results depicted that the model could gain the accuracy of 71.56% that immensely high compared to other models.

As can be observed, advancements in deep learning have been useful for music genre classification, resulting in exceptional performance levels. Nevertheless, current deep learning approaches, including Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), exhibit certain limitations in their ability to capture hierarchical relationships among features and in optimizing model parameters. Instead, the proposed CapsNet/UIGMM model combines the advantages of capsule networks with upgraded ideal gas molecular movement (UIGMM) optimization to facilitate a more effective understanding of hierarchical relationships and parameter optimization. Unlike traditional methods, CapsNet/UIGMM employs a hybrid strategy that influences the benefits of both capsule networks and UIGMM to improve the accuracy and robustness against the noise. Additionally, CapsNet/UIGMM demonstrates enhanced optimization efficiency by utilizing UIGMM for parameter optimization within the capsule network to surpass the effectiveness of conventional optimization techniques.

However, due to the computational intensity and sensitivity to hyperparameters of CapsNets[15], the research introduces the use of an Upgraded Ideal Gas Molecular Movement (UIGMM) optimizer. This optimizer efficiently optimizes the CapsNet model parameters, aiming to improve performance and practical applicability in music genre classification. By integrating a metaheuristic algorithm with deep learning, this research offers an innovative approach to navigate complex search spaces more efficiently, potentially leading to better-performing models. A metaheuristic algorithm is a technique to quickly solve a problem when classical methods are too slow, or to find an approximate solution when classical methods cannot find an exact solution to the problem. In fact, metaheuristic algorithms change optimality, perfection, precision and accuracy for speed. The goal of these algorithms is to generate a solution in a reasonable time to solve the current problem.

The contributions of this research are manifold, including the advancement in music genre classification techniques by integrating CapsNet with UIGMM, setting new standards for accuracy and efficiency. This methodological innovation, inspired by physical sciences, provides a fresh perspective on optimizing complex neural network models.

Database

This study utilized three separate and widely recognized datasets to assess the efficiency and reliability of our recently developed approach designed specifically for music genre identification. The chosen datasets for this objective were GTZAN, ISMIR2004, and the expanded Ballroom, each presenting distinct features and difficulties that accurately represent the variety encountered in the field of music genre categorization. In the following, a brief description about each dataset has been described.

GTZAN

The first dataset is known as GTZAN. The GTZAN dataset is a highly regarded standard for categorizing music genres. It comprises 1000 audio tracks, each with a duration of 30 s, representing ten different genres: classical,

blues, disco, country, jazz, hip-hop, pop, metal, rock, and reggae. These tracks are encoded as 22,050 Hz Mono 16-bit audio files in wav format[16]. The dataset is accessible from the following link: https://www.kaggle.com/andradaolteanu/gtzan-dataset-music-genre-classification.

ISMIR2004

The ISMIR2004 dataset was compiled from the 5th International Conference on Music Information Retrieval in 2004. The dataset comprised 1000 audio snippets, each lasting 30 s[17]. These excerpts were uniformly divided across 6 genres: classical, blues and jazz, electronic, metal and punk, rock and pop, and world. Furniture piece with a flat top and one or more legs, used for various purposes such as eating, working, or displaying objects. The dataset is accessible via the provided hyperlink: https://ismir2004.ismir.net/genre_contest/index.html#genre.

Extended ballroom

The Extended Ballroom dataset is an enhanced variant of the well-known Ballroom dataset, providing a broader selection of songs and a detailed inventory of track repetitions. The development of the website was prompted by the scarcity of tracks, subpar audio quality, and the accessibility of the original website[18]. The dataset collected audio extracts and metadata from the internet, and different sorts of repetitions were tagged using semi-automated methods. The product provides enhanced audio fidelity, a six-fold increase in track capacity, five more categories of rhythmic patterns, and annotations for various forms of repetitions. The dataset's potential applications in the music industry are broadened. The dataset is accessible via the provided hyperlink: http://anasynth.ircam.fr/home/media/ExtendedBallroom.

Figure 1 displays the various genres and the corresponding amount of samples used in the datasets.

The study intends to thoroughly evaluate the performance of the suggested method across a diverse range of music genres, spanning from broad and generic categories to highly specialized and subtle ones, by utilizing these three datasets. This technique guarantees that the efficacy of the method is evaluated not only based on its capacity to identify diverse genres, but also on its ability to detect small differences within a genre. This comprehensive evaluation ensures a thorough assessment of the method's skills in recognizing music genres.

Audio characteristics and normalization

In our investigation on music genre classification, the audio features that encapsulate the essence of various genres have been extracted by meticulously analyzing the audio signal. The forthcoming classification task will be firmly based on a comprehensive compilation of diverse data that has been painstakingly gathered. In the subsequent section, a comprehensive explanation is provided regarding the features that have been utilized.

Spectral crest

The spectral crest is a measure of how sharp or prominent a peak is in a spectrum. It is calculated by comparing the highest value in the spectrum to the average value of the spectrum. Higher value of spectral crest indicates that there are only a few prominent components in the spectrum, whereas lower value of spectral crest suggests that the spectrum is relatively noisy or uniform. This can be mathematically defined as follows:

$$Crest = \frac{\max\limits_{k\in\,[b_1,b_2]} S_k}{\frac{1}{b_2-b_1}\sum_{k=b_1}^{b_2} S_k} \tag{1}$$

The signal's magnitude spectrum is denoted as S_k, while the frequency range of interest is represented by $[b_1, b_2]$. The spectral crest serves as a commonly used spectral descriptor for audio features extraction. The spectral crest effectively captures certain aspects of the sound's tone, timbre, and musical style.

Spectral centroid

A measuring instrument that may be used to estimate the center of mass of a spectrum and hence the brightness of the sound is the spectral centroid. To get it, we take the signal's frequencies and average them out, using their magnitudes as weights. This can be mathematically defined as follows:

$$C = \frac{\sum_{n=0}^{N-1} f(n)\, x(n)}{\sum_{n=0}^{N-1} x(n)} \tag{2}$$

where, $x(n)$ specifies the magnitude of the n-th frequency bin, and $f(n)$ describes the center frequency of bin.

Spectral entropy

The degree of disorder or instability in a spectrum may be measured by spectral entropy. What we mean by this is the output we get when we take the negative sum of the product of the normalized spectrum and its logarithm. A low spectral entropy indicates a more concentrated or ordered spectrum, while a high entropy indicates a more uniform or random one. This can be mathematically defined as follows:

$$SE = -\sum_{k\in\,[b_1,b_2]} p_k \log p_k \tag{3}$$

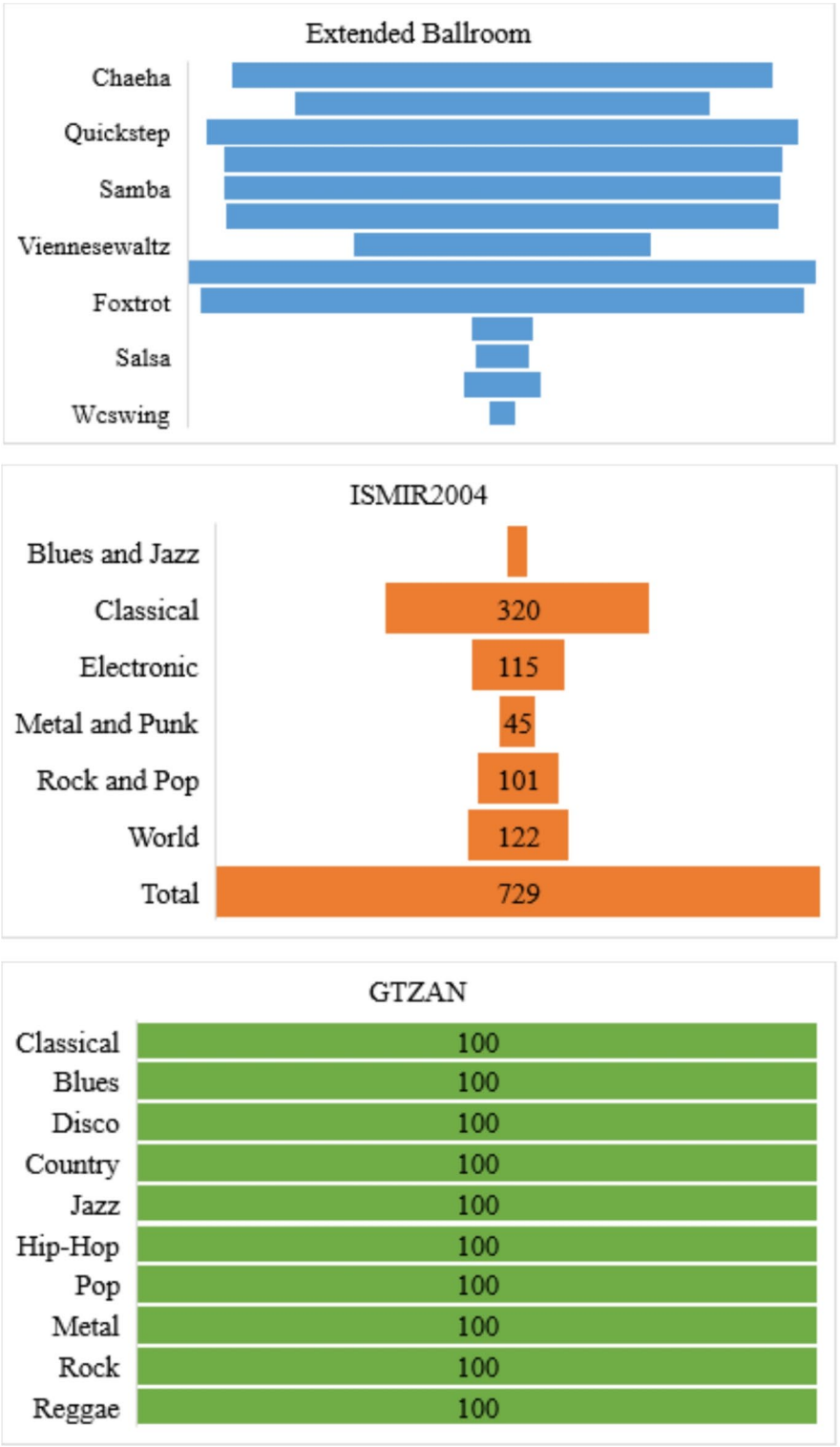

Fig. 1. Various genres and the corresponding amount of samples used in the datasets.

where, $[b_1, b_2]$ denotes the targeted frequency range, and p_k represents the normalized signal spectrum. The spectrum is divided by the sum of its values, representing $\sum_{k\in[b_1,b_2]} p_k = 1$, in order to accomplish normalization.

Spectral flux

The rate of change in a spectrum over time is called its spectral flux. The square of the difference between the normalized spectra of two consecutive frames is used to compute it. When the spectral flux is low, the

spectrum is more stable, and when it's high, the spectrum is changing quickly. Spectral flux may be expressed mathematically as:

$$SF = \sum_{k \in [b_1, b_2]} \left(p_k^{(n)} - p_k^{(n-1)} \right)^2 \tag{4}$$

Where denote the normalized spectra of the most recent and prior frames, respectively, as $p_k^{(n)}$ and $p_k^{(n-1)}$, respectively.

Pitch

Spectral analysis, autocorrelation, and cepstrum are some of the methods used to quantify the perceived fundamental frequency, which in turn determines the pitch of a sound. These techniques analyze the harmonic structure of an audio source to assist extract its pitch. Finding the signal period that minimizes the difference function is the goal of the YIN algorithm, which is a popular method for pitch estimation. Next, we get the inverse of this time to get the pitch, which shows how high or low the sound is.

$$P = \frac{F_s}{\underset{\tau}{\operatorname{argmax}} R_x(\tau)} \tag{5}$$

This is accomplished in the following way: with F_s representing the sample frequency and $R_x(\tau)$ describing the signal autocorrelation function, i.e.

$$R_x(\tau) = \sum_{n=0}^{N-1-\tau} x(n)\, x(n+\tau) \tag{6}$$

where, τ describes the delayed samples, and $x(n)$ and $x(n+\tau)$ determine the signal value at the current time and the next time point, respectively. Figure 2 displays a sample of features that have been extracted using the aforementioned characteristics.

Figure 2 commences with the "original image", a visual representation of the audio waveform that depicts changes in amplitude over time, laying the foundation for further analysis. Next, "Spectral crest" quantifies the ratio of peak spectral magnitude, providing insights into the sound's frequency dominance by highlighting its spectral "peakiness". The "Spectral centroid" then measures the weighted average frequency of the spectrum, indicating the overall brightness of the sound and contributing to the understanding of its pitch and timbre. "Spectral entropy" evaluates the randomness in the spectral distribution, assessing the complexity of the sound.

"Spectral flux" tracks temporal changes in the spectrum, identifying shifts in the audio signal that indicate dynamic variations. Lastly, "Pitch" focuses on the fundamental frequency, which is crucial for analyzing melody and harmony as it directly correlates with the perceived musical note or tone. Together, these features encompass a comprehensive methodology for dissecting and comprehending the intricacies of audio signals, playing a pivotal role in tasks such as music genre classification and sophisticated sound analysis.

Normalization

The audio characteristics are scaled via normalization, which involves translating their values into a range between 0 and 1. To do this, we divide each value by the feature's range after removing the feature's minimum value. In this case, we normalized using the Min-Max approach. In the following, the Min-Max Normalization formula has been defined:

$$X_N = \frac{X - \mathrm{X}}{\overline{\mathrm{X}} - \mathrm{X}} \tag{7}$$

where, X_N specifies the normalized value, and the original value is denoted by X, whereas the lowest and maximum values of the features are represented by $\overline{\mathrm{X}}$ and X, respectively.

Capsule neural network

CapsNet has been found to be a kind of Artificial Neural Network that intends to efficiently seizure ranked associations by the use of capsules rather neurons just like the essential components of calculation. Clusters of neurons that represent the probability and attributes of a specific feature are known as capsules. Various characteristics of an object can be expressed through them, including its presence, direction, form, and location[19,20]. In addition, capsules have the ability to grasp the spatial connections and diverse positions of input characteristics that are usually disregarded in conventional neural networks.

A CapsNet is composed of several layers of capsules, each of which represents a unique level of abstraction. A dynamic routing mechanism connects the higher-level capsules to the lower-level ones. The coupling coefficients between capsules are updated in an iterative manner[21]. This update process is performed in accordance with the alignment of their posture variables. The CapsNet has the ability to understand the connections between different pieces and complete entities, giving more importance to capsules that demonstrate a stronger correlation with capsules that are of higher-level.

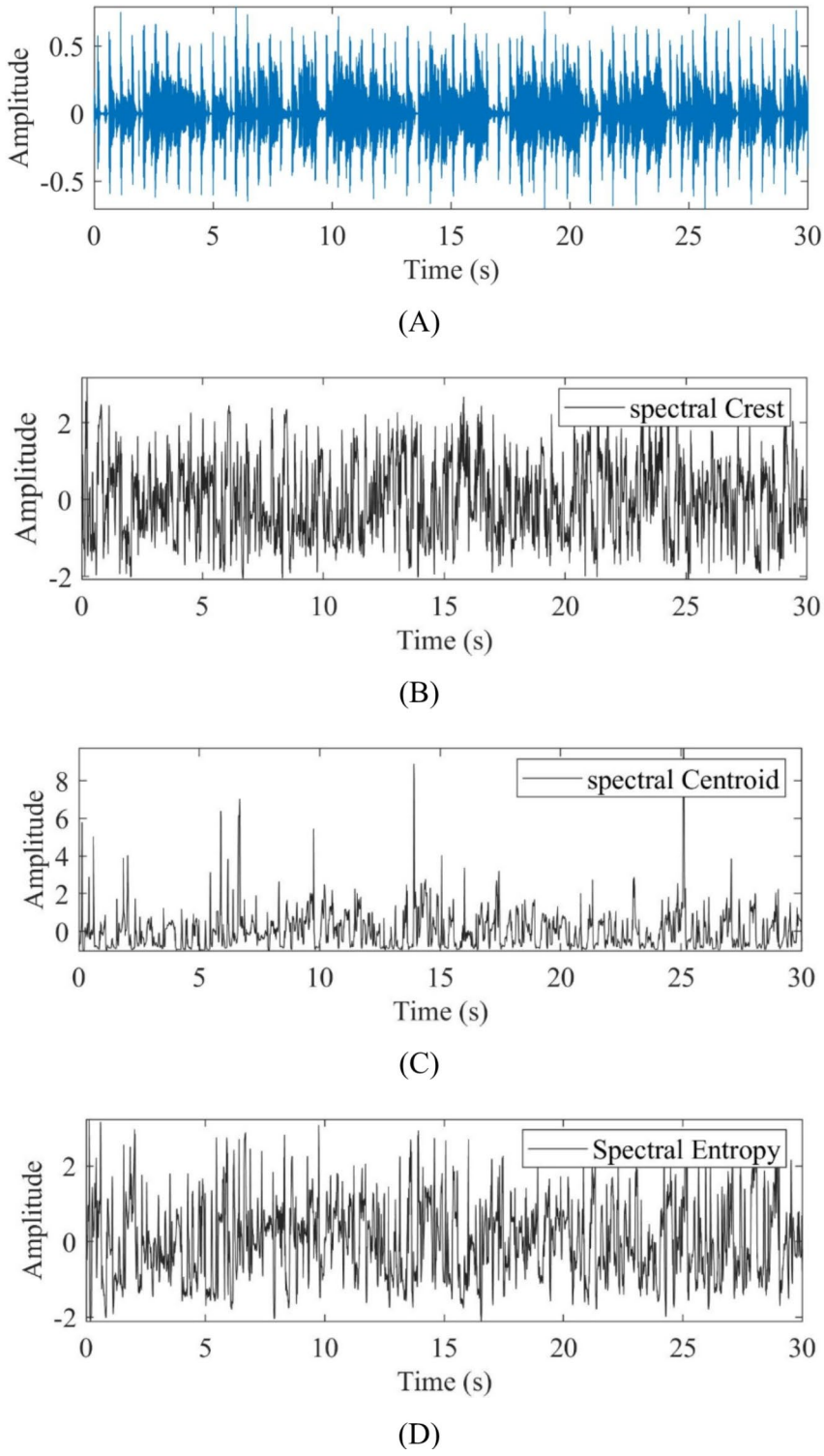

Fig. 2. Sample of features extracted using the aforementioned characteristics: (**A**) original image, (**B**) spectral crest, (**C**) spectral centroid, (**D**) spectral entropy, (**E**) spectral flux, (F) pitch.

The CapsNet's output is composed of a series of vectors that each vector corresponds to a particular category. The intended class is defined by vectors that include both the possibility and pose variables. The probability of the class is indicated by the vector's length, whereas the pose variables are represented by the direction of it. A margin function of loss has been utilized for training the CapsNet[22]. The present function motivates the proper category to possess longer vectors and discourage improper categories no to possess longer vectors. In addition,

the margin function of loss includes a term of regularization that reprimands the present neural network to allocate great possibilities to several categories.

Every one of the capsules existing within the initial layers includes a vector of activity p_j that represents geographic data in the instantiation element of the index of capsule, which has been demonstrated by j. The vector of output p_j of the lower-grade capsule j has been utilized within the next layer $1+1$ that has been used by all capsules. Candidate i obtains the input p_i and computes output of it by the utility of the matrix of weight $Y_{j,i}$ within the layer $1+1$. The ultimate prediction vector $p_{j,i}$ illustrates amount of the initial j that belongs to category i.

$$p_{j,\mathrm{i}} = Y_{j,\mathrm{i}} \times p_j \tag{8}$$

The agreement level between the candidates, including the main factor to predict the candidate j's manner to the candidate i, has been ascertained via multiplying the main forecasting factor of candidate j by coefficient of coupling[23]. Both candidates are linked it has been ensured they are aligned. Hence, the couping coefficient raises once the reverse occurs. For evaluating the effect of the individual' function of squeezing, the entire amount (Yx_i) has been allocated and assigned to the main forecasting of candidate i.

$$Yx_i = \sum_{j=1}^{N} q_{j,i}\widehat{p}_{i,j} \tag{9}$$

$$d_i = \frac{\|Yx_i\|^2}{1+\|Yx_i\|^2}\frac{Zc_j}{\|Ys_i\|} \tag{10}$$

$$q_{j,i} = \frac{\exp\left(b_{j,i}\right)}{\sum_k \exp\left(b_{j,i}\right)} \tag{11}$$

Compression is linked with potentiality and limits an individual's output to values n within the range of 0 to 1. The layer of capsule d_i moves to the subsequent layer which exhibits the same manner to the prior layer. The primary prediction level has been verified by q_j and is generalizable to the layer $1+1$. In each cycle, the element outcome $\widehat{p}_{j,i}$ is obtained. All capsules' vector has been seen as the integration of two values that are numerical.

The main difference lies in the fact that a Capsule compresses features and generates a probability, whereas a set of instantiation units is used to determine the layer's reliability. Once the lower-degree capsule is in agreement with a capsule that is of a higher-degree, the association between whole and an element has been signified by it that highlights the importance of the way. The perception refers to the Dynamic routing-by-agreement.

There exist two main layers within the present network. the layer of input is in charge for managing the input data that is the consequence of procedure regarding pre-training. The following layer refers to the layer of capsule, adopting the discriminating patterns tasks in data and classifying them[24]. The layer of capsule comprises 32 channels of convolution; its kernel size is 9×9; and its stride is two. The instantiation produced a vector that depicts the possible association between presence and function of a capsule in genre margin loss. An infection is indicated by every tear capsule, and each of these capsules has assigned a specific rate of the loss of margin, which is ascertained using the subsequent formula. The vector's maximum value has been indicated by D_h in all genres.

$$G_{D_h} = U_{D_h} \max\left(0, b^+ - D_{Vr}\right)^2 + \alpha\left(1 - U_{Vr}\right)\max(\left(0, D_{Vr} - b^-\right)^2 \tag{12}$$

In accordance with the present research, in the presence of a specific genre, the value of U_{D_h} is one. Moreover, the values of b^+ and b^- are, in turn, 0.9 and 0.1. by declining the effect of all genre capsules, the parameter a can be gained. It should be noted that genre capsules play a role in the procedure of regularization. The properly recognized genres must be divided by the entire quantity of the genres for computing the accuracy.

$$Accuracy = \frac{\sum Correct\ identified\ fractures}{Total\ EquationNumber\ of\ fractures} \tag{13}$$

For proceeding to the subsequent phase, improving the hyper-parameters of the present network is of utmost importance.

The present research has employed optimization for recognizing a hyper-parameter within a D-dimensional context that its main purpose is to diminish the validation function to the least.

The function OB can identify a series of c hyperparameters. The identification of validation problems can be effectively accomplished by configuring the network in a suitable manner. A suggested solution that is according to the optimization of the OB is represented subsequently that allows for the automated mapping of the optimal hyperparameters.

$$\begin{aligned} &\min_{x \in R^G} OB\ (x,\ \theta; Z_{val}) \\ &s.t.\theta = \arg\min OB\left(\theta; S_{train}\right) \end{aligned} \tag{14}$$

It is pretty demanding to address the problem in the prior equation due to the intricacy of the OB index. The dataset of training is indicated by the utility of $s.t.\theta$.; moreover, Y is utilized for the purposes of validation. The purpose of the process of training is minimizing the procedure of learning and handling a set that has been immensely populated with x's values.

The upgraded ideal gas molecular movement is utilized for handling the pricy index of error. Additionally, it has been employed for optimizing the network's hyperparameters.

Upgraded ideal gas molecular movement

Gas molecules might be similar to some possible agents, travailing and exploring quickly total the capacity in the vessel as an optimization area, probing total the possible areas for a global optimum solution. They work together and interchange data with one additional after every crash, as they move the inner side of the area[25]. The interchange of data is discovered as a modification in the molecules' velocity after they strike and go to novel situations in diverse time stages. The IGMM algorithm's core notion is to discover the global optimum throughout these direction-finding and crashes. The optimization process is designated in the next segment.

Initialization

As the 1st specification, molecules' numbers outspread regularly and travel stochastically in diverse routes. This specification supports the notion of utilizing a steady spreading for producing the first population. It is anticipated to speed up the search procedure to the optimum solution. Once the solutions' first population is designated stochastically using a steady spreading from the design variables' permissible span, the first temperature of 273 K is regulated in the velocity equation.

Computation of molecular masses

In the IGMM algorithm, a comparative scale was executed to the fitness of the normalized agent by describing a mass parameter in line with every gas molecule's fitness. Calculation associated with gas molecules' velocity exposes the detail that emphasizes a reverse correlation among the elements' velocity and their masses[26]. Consequently, the molecule that has a bigger mass move with a lower velocity and vice-versa. Therefore, the mass calculation needs to be determined such that proper molecules travel at a lower velocity. This circumstance can be defensible by their inclination to stay in the present area which is more possible to be reasonable. This is obviously opposed to fewer appropriate molecules[27]. As this study is concentrated on minimalizing the objective, on the basis of the rule deliberated, every solution is allocated a mass regarded to its fitness with the subsequent equation:

$$\frac{\frac{1}{fit(i)}}{\frac{1}{\sqrt{\sum\left(fit^2(i)\right)}}} \tag{15}$$

here the $i-th$ molecule's mass is defined by m_i and $fit\,(i)$ determines the $i-th$ molecule's fitness regarding the cost value for the problem. The top molecule would be set up during a comparison of molecular masses in each iteration.

Computation molecule's collision possibility

Gas molecules' Physical prevailing calculations reveal that perfect gas molecules crash with one additional with a definite specific probability. Namely, it surges with a surge in the moving time and the distance that is moved by the molecules. For employing this specific in the optimization procedure, a novel parameter called MCP (molecules collision probability) is specified. In the initial part of the optimization procedure, once the problem area is not recognized and no interchange of data has happened between the agents, the MCP is located at its minimum level[28]. Nevertheless, with more processed during optimization and the optimum area is set up, the crash possibility rises exponentially in line with the next relation and grasps its maximum level.

$$MPC = 1 - e^{(-0.63\times\, iter)} \tag{16}$$

Computing molecules' novel velocity and situation

All 2 molecules either crash with one additional on the basis of their particular crash possibility or continue traveling non-collided on the basis of the velocity relation. Regarding this occurrence, the next stages would continue to compute the novel velocity and situation of every molecule.

Calculating molecular velocity at the collision incidence (MVCI)

The 2nd and 3rd features of perfect gases reveal molecular communications. The molecule that has a larger mass is supposed motionless and the lighter molecule transfers along with the hypotheses about the elastic collision among molecules. Moreover, molecule's motion velocity is calculated with $x = \Delta\, vt$ (for t = 1) by sub-tracting the situation of the 2 molecules.

$$\left(v_1^d\right)' = \frac{(m_1 - Em_2)}{m_1 + m_2} \times\, v_1^d \tag{17}$$

$$\left(v_2^d\right)' = \frac{(1+E)m_1}{m_1 + m_2} \times\, v_1^d \tag{18}$$

here v_1^d indicates the early velocity of the 1st molecule before the influence, whilst supposed $v_2^d = 0$, and consequently, $\left(v_1^d\right)'$ and $\left(v_2^d\right)'$ signify the last velocities after the impact, respectively, d specifies the of the optimization problem's dimension. As already mentioned, in the elastic collisions, the amount of E is one, but

in the former formula, this parameter is specified as a variable to assurance convergence in this algorithm. Consequently, in the 1st limited stages of the optimization procedure, the amount of this variable is about one but with a surging in the optimization cycles' number its amount drops animatedly on the basis of the next linear Eq.

$$E = 1 - \left(\frac{iter}{maxIt}\right) \tag{19}$$

here $iter$ and $maxIt$ designate the existing and the max iterations of the optimization process, in turn. After calculated the novel velocity of every molecule, its novel situation is calculated by the next formulas:

$$\left(z_1^d\right)' = z_2^d + rand\left(v_1^d\right)' \tag{20}$$

$$\left(z_1^d\right)' = z_2^d + rand\left(v_2^d\right)' \tag{21}$$

here z_2^d indicates the situation of the motionless molecule before the influence and consequently, $\left(z_1^d\right)'$ and $\left(z_2^d\right)'$ show the novel situations after the influence, in turn. $rand$ signifies a stochastic standard distributed amount that is from 0 to 1.

Computing nocollision molecular velocity (NCMV)

In the isolated media of ideal gases, if the supposed molecular does not have any crash, the ith molecule's novel velocity is defined as follows:

$$\left(v_i^d\right)' = 1.7\sqrt{\frac{kT_i}{m_i}} \tag{22}$$

Because the Boltzmann coefficient has a reverse relation with the molecules' number, its amount is regulated to $k = 1/nVar$, here $nVar$ is the molecules' number in the optimization procedure. Every molecule's velocity adapts to the molecule's mass and temperature. Therefore, in this stage, it is necessary to calculate the novel temperature of every molecule. To do so, a sub-tractive formulation is determined by the next formula:

$$T_i' = T_i - 1/m_i \tag{23}$$

Having defined every molecule's novel velocity, the novel situation is calculated by the following formula:

$$\left(z_i^d\right)' = z_i^d + rand\left(v_i^d\right)' \tag{24}$$

Convergence principles

In the optimization's final phase, the convergence is checked. The optimization process is regulated totally, if the finest outcome in the existing iteration barely demonstrates any variations in a permissible tolerance for several iterations. Additional adequate convergence principle is a place the distance between the finest solution and the average punished cost value of total designs in an iteration decreases to a wanted amount. Additional normally measured ending principle in meta-heuristic algorithms is someplace the max number of iterations is done while convergence dose not happens. This is wherever the optimization process might dismiss deprived of a normal convergence.

Upgraded ideal gas molecular movement

The Ideal Gas Molecular Movement method is being improved to enhance its effectiveness in complex optimization problems with limited parameters. This method uses chaotic maps, which offer nonlinearity, statistical features, and unpredictability, enabling more thorough investigation and better results. These maps are increasingly used in computer engineering, particularly for replacing pseudo-random numbers with Gaussian distributions and producing random numbers. However, the method is vulnerable to changes in parameter settings and lacks search agent diversity.

Chaotic maps' nonlinearity, statistical properties, and unpredictability make them more appealing for their use in metaheuristic optimization, particularly in producing random numbers and replacing pseudo-random ones with normal distribution ones. Numerous improvements to optimization methods have their roots in the chaotic improved optimization strategy. One well-known method for producing pseudo-random integers is the Kent map, which is defined as follows:

$$x_{n+1} = \begin{cases} \frac{x_n}{m}, & 0 < x_n \leq m \\ \frac{1-x_n}{1-m}, & m < x_n < 1 \end{cases} \tag{25}$$

By considering the Kent map for all determined molecules, the upgraded state may be determined using the following formula:

$$\left(z_i^d\right)' = z_i^d + x_n \tag{26}$$

where, $x_0 = 0.5$.
Chaos in (0, 1) is generated using the Kent map, as seen in Fig. 3.

The "elimination phase" is another mechanism that can be used by metaheuristic algorithms to increase their efficiency and speeds up their convergence to the optimal solution. In this mechanism, all of the potential solutions have been sorted by importance, with the goal of letting the algorithm zero in on the best ones. Eliminating the inadequate results improves the algorithm's efficiency and brings it closer to the optimal answer. As part of the elimination process, we sort all of the solutions by the value of the objective function and choose the ones that scored the lowest.

Algorithm authentication

We tested the upgraded Ideal Gas Molecular Movement algorithm with 23 esteemed classical benchmark functions. The functions were all 30 dimensions each. These selected functions, studied in-depth by academic scholars, cover a wide variety of scenarios. Table 1 depicts key details such as the chosen functions, the search space dimensions, and the best solutions.

The upgraded Ideal Gas Molecular Movement algorithm aims to identify the minimum value of the existing objective functions. To assess its efficacy, it has been compared to well-known and respected algorithms such as the World Cup Optimization (WCO)[29], Butterfly Optimization Algorithm (BOA)[30], Gaining Sharing Knowledge (GSK)[31], Growth Optimizer (GO)[32], and War Strategy Optimization (WSO)[33].

In order to ensure a fair comparison, each algorithm utilizes a group of 50 search candidates and goes through 200 iterations, resulting in a total of 1000 implementations. The detailed settings for each algorithm can be found in the renowned Table 2.

Table 3 presents a magnificent exhibition of the standard deviation (STD) and the majestic average of the objective function, encompassing all test functions and algorithms. Each algorithm, with utmost grace, is executed autonomously on the classical benchmark functions, adorning the stage with a total of 15 remarkable performances.

Upon conducting an analysis of the findings, it becomes apparent that the upgraded UIGMM algorithm excels in various instances. For instance, when examining the average values of the objective function for F1 and F6, UIGMM consistently displays significantly lower mean values in comparison to alternative algorithms. This indicates its effectiveness in identifying the minimum value for these particular functions. Furthermore, in terms of standard deviation (StD), UIGMM consistently showcases competitive or superior performance across multiple functions, thereby highlighting its stability and dependability in discovering optimal solutions.

Furthermore, the comparison reveals that UIGMM achieves noteworthy outcomes for specific functions, such as F10 and F16, where it maintains exceptionally low mean values and minimal standard deviations. This underscores its robustness in optimizing these particular functions. However, it is important to acknowledge that for certain functions, such as F3 and F5, UIGMM's performance in terms of mean and standard deviation is relatively less impressive when compared to other algorithms. The findings demonstrate that the application of this algorithm yields important efficiency in addressing optimization problems, including the specific objectives of this study. Consequently, the proposed UIGMM has been employed to minimize Eq. (14) for the optimal selection of the Capsule Network. This network comprises a convolutional layer dedicated to feature extraction, primary capsules for identifying low-level features, digit capsules for recognizing high-level features, a routing mechanism that directs the output from the primary capsules to the digit capsules based on the dot product of the outputs and their corresponding weights, and an output layer that integrates a fully connected layer with a Softmax activation function to generate the final output. The convolutional layer is equipped with 32 filters, each utilizing a 3×3 kernel size and ReLU activation. The primary and digit capsules consist of 32 and 10 capsules, respectively, with each capsule utilizing 8 and 16 convolutional filters, also with a 3×3 kernel size and ReLU activation, to efficiently extract and process the input data. Figure 4 shows a simple schematic diagram of the Capsule Net architecture used in this research.

Simulation results

The present research identified sound spectra and extracted characteristics using a Hybrid CapsNet/UIGMM method based on An improved version of the gray lag goose optimizer. Features including pitch, spectral

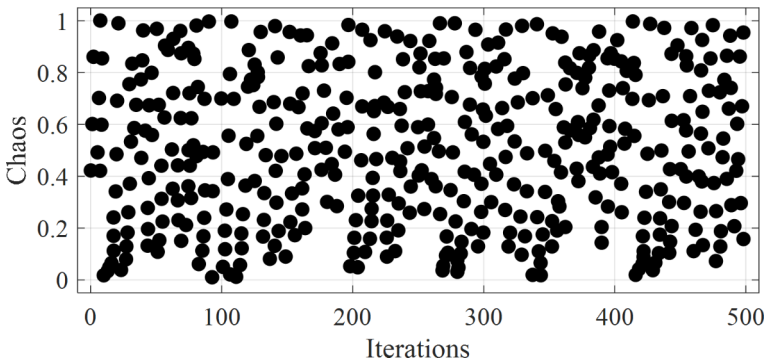

Fig. 3. Chaos made by the Kent map.

Function	Range	Dim	f^*
$f_1(z) = \sum_{i=1}^{n} z_i^2$	[-100, 100]	30	0
$f_2(z) = \sum_{i=1}^{n} \lvert z_i \rvert + \Pi_{i=1}^{n} \lvert z_{ix} \rvert$	[-10, 10]	30	0
$f_3(z) = \sum_{i=1}^{n} \left(\sum_{j=1}^{i} z_j \right)^2$	[-100, 100]	30	0
$f_4(z) = maxi \{ \lvert z_i \rvert, 1 \leq i \leq n \}$	[-100, 100]	30	0
$f_5(z) = \sum_{i=1}^{n-1} \left[100(z_{i+1} - z_i^2)^2 + (z_i - 1)^2 \right]$	[-30, 30]	30	0
$f_6(z) = \sum_{i=1}^{n} ([z_i + 0.5])^2$	[-100, 100]	30	0
$f_7(z) = \sum_{i=1}^{n} i z_i^4 + random\,[0,1]$	[-1.28, 1.28]		
$f_8(z) = \sum_{i=1}^{n} - z_i \sin\left(\sqrt{\lvert z_i \rvert} \right)$	[-500,500]	30	-418.9829 xD
$f_9(z) = \sum_{i=1}^{n} \left[z_i^2 - 10\cos(2\pi z_i) + 10 \right]$	[-5.12, 5.12]	30	0
$f_{10}(z) = -20\exp\left(-0.2\sqrt{\frac{1}{N} \sum_{i=1}^{n} z_i^2} \right) - exp\left(\frac{1}{n} \sum_{i=1}^{n} \cos(2\pi z_i) \right) + 20 + e$	[-32, 32]	30	0
$f_{11}(z) = \frac{1}{4000} \sum_{i=1}^{n} z_i^2 - \Pi_{i=1}^{n} \cos\left(\frac{z_i}{\sqrt{i}} \right) + 1$	[-600, 600]	30	0
$f_{12}(z) = \frac{\pi}{n} \left\{ 10\sin(\pi y1) + \sum_{i=1}^{n-1} (y_i - 1)^2 \left[1 + 10sin^2(\pi y_i + 1) \right] + (y_n - 1)^2 \right\} + \sum_{i=1}^{n} u(z_i, 10,100,4)$ $y_i = 1 + \frac{z_i + 1}{4},\ u(z_i, a, k, m) = \begin{cases} k(z_i - a)^m, & z_{i>a} \\ 0, & -a < z_i < a \\ k(-z_i - a)^m, & z_i < a \end{cases}$	[-50, 50]	30	0
$F_{13}(x) =$ $0.1 \left\{ sin^2(3\pi z_1) + \sum_{i=1}^{n} (z_i - 1)^2 \left[1 + sin^2(3\pi z_i + 1) \right] + (z_n - 1)^2 \left[1 + sin^2(2\pi z_n) \right] \right\} +$ $\sum_{i=1}^{n} u(z_i, 5,100,4)$	[-50, 50]	30	0
$f_{14}(z) = \left(\frac{1}{500} + \sum_{j=1}^{25} \frac{1}{j + \sum_{i=1}^{2} (z_i - a_{ij})^6} \right)^{-1}$	[-65, 65]	2	1
$f_{15}(z) = \sum_{i=1}^{11} \left[a_i - \frac{z_1 (b_i^2 + b_i z)}{b_i^2 + b_i z_3 + z_4} \right]^2$	[-5, 5]	4	0.00030
$f_{16}(z) = 4z_1^2 - 2.1z_1^4 + \frac{1}{3} z_1^6 + z_1 z_2 - 4z_2^2 + 4z_2^4$	[-5, 5]	2	-1.0316
$f_{17}(z) = \left(z_2 - \frac{5.1}{4\pi^2} z_1^2 + \frac{5}{\pi} z_1 - 6 \right)^2 + 10 \left(1 - \frac{1}{8\pi} \right) cosz_1 + 10$	[-5, 5]	2	0.398
$f_{18}(z) = \left[1 + (z_1 + z_2 + 1)^2 \left(19 - 14z_1 + 3z_1^2 - 14z_2 + 6z_1 z_2 + 3z_2^2 \right) \right]$ $\left[30 + (2z_1 - 3z_2)^2 (18 - 32z_1 + 12z_1^2 + 48z_2 - 36z_1 z_2 + 27z_2^2) \right]$	[-2, 2]	2	3
$f_{19}(z) = - \sum_{i=1}^{4} c_i . \exp\left(- \sum_{j=1}^{3} a_{ij} (z_j - p_{ij})^2 \right)$	[1, 3]	3	-3.86
$f_{20}(z) = - \sum_{i=1}^{4} c_i . \exp\left(- \sum_{j=1}^{6} a_{ij} (z_j - p_{ij})^2 \right)$	[0, 1]	6	-3.32

Table 1. Key details of the employed functions for the algorithm validation.

Algorithm	Parameter/value
World Cup Optimization (WCO)[29]	$playoff = 0.04,\ ac = 0.3$
Butterfly Optimization Algorithm (BOA)[30]	$P = 0.8,\ \alpha = 0.8,\ c = 0.03$
Gaining Sharing Knowledge (GSK)[31]	$p = 0.2,\ k_r = 0.8,\ k_f = 0.4$
Growth Optimizer (GO)[32]	$P_1 = 8,\ P_2 = 1e-3,\ P_3 = 0.5$
War Strategy Optimization (WSO)[33]	$w = 0.2,\ a = 0.5,\ d = 0.2,\ s = 0.1$

Table 2. The detailed settings for each algorithm studied in this research.

Function	Mean/StD	WCO	BOA	GSK	GO	WSO	UIGMM
F1	Mean	2.13E-59	2.84E-08	2.05E-17	2.24E-59	4.04E-07	1.98E-59
	StD	4.96E-30	2.89E-04	2.07E-09	6.55E-30	5.18E-04	4.57E-30
F2	Mean	5.84E-35	3.24E-04	2.16E-08	6.02E-35	3.12E+01	5.50E-35
	StD	6.82E-18	3.29E-02	4.98E-05	7.15E-18	5.04E+00	6.33E-18
F3	Mean	1.22E-14	1.36E+01	1.99E+02	1.28E-14	2.26E+02	1.16E-14
	StD	2.42E-07	2.11E+00	7.12E+00	2.91E-07	1.02E+01	2.36E-07
F4	Mean	1.44E-14	5.74E-01	1.07E-03	1.81E-14	7.19E+00	1.44E-14
	StD	1.40E-07	3.00E-01	8.24E-02	1.65E-07	1.34E+00	1.37E-07
F5	Mean	8.56E+00	5.18E+01	2.30E+01	2.17E+01	1.25E+02	7.96E+00
	StD	6.94E-01	5.30E+00	3.84E+00	7.49E-01	1.20E+01	6.84E-01
F6	Mean	5.54E-11	2.64E-08	6.50E-11	4.84E-01	4.50E-07	5.48E-11
	StD	4.80E-10	3.43E-04	2.14E-09	4.58E-01	5.39E-04	4.54E-10
F7	Mean	5.81E-04	5.79E-02	1.64E-02	6.23E-04	4.11E-02	5.75E-04
	StD	2.40E-03	1.06E-01	6.77E-02	1.76E-02	1.16E-01	2.28E-03
F8	Mean	-5.98E+03	-5.01E+03	-1.94E+03	-4.20E+03	-4.97E+03	-5.59E+03
	StD	1.46E+00	3.46E+01	1.50E+01	2.45E+01	2.23E+01	1.35E+00
F9	Mean	5.22E-01	4.07E+01	1.30E+01	5.68E-01	6.00E+01	5.05E-01
	StD	6.02E-01	2.98E+00	1.47E+00	1.21E+00	3.98E+00	5.69E-01
F10	Mean	1.35E-05	5.98E-02	3.29E-04	2.17E+00	1.48E+00	1.24E-05
	StD	6.80E-07	4.86E-01	2.09E-05	7.66E-01	6.45E-01	6.30E-07
F11	Mean	1.06E-03	7.13E-03	2.81E+00	1.79E-03	1.88E-01	9.94E-04
	StD	2.37E-03	7.65E-02	1.06E+00	6.19E-02	2.66E-01	2.30E-03
F12	Mean	1.10E-11	5.31E-03	1.95E-02	3.37E-02	4.18E-01	1.09E-11
	StD	1.39E-12	1.31E-01	1.79E-01	1.27E-01	9.91E-01	1.30E-12
F13	Mean	1.72E-10	2.13E-03	1.16E-03	4.16E-01	1.22E-03	1.68E-10
	StD	2.59E-11	5.16E-02	5.46E-02	3.66E-01	6.66E-02	2.56E-11
F14	Mean	8.65E-01	3.18E+00	3.84E+00	3.31E+00	1.44E+00	8.45E-01
	StD	3.23E-01	1.44E+00	1.31E+00	1.61E+00	9.08E-01	3.13E-01
F15	Mean	9.25E-05	7.31E-04	2.95E-03	3.13E-03	2.00E-03	8.86E-05
	StD	3.03E-03	1.29E-02	3.69E-02	7.11E-02	5.33E-02	2.62E-03
F16	Mean	-7.13E-01	-7.35E-01	-7.51E-01	-7.83E-01	-7.77E-01	-7.80E-01
	StD	5.56E-05	2.77E-04	2.64E-04	1.48E-04	2.14E-04	5.13E-05
F17	Mean	2.53E-01	3.27E-01	3.21E-01	3.07E-01	3.19E-01	2.46E-01
	StD	2.51E-08	2.40E-08	2.44E-08	9.72E-03	2.24E-07	2.34E-08
F18	Mean	2.05E+00	2.39E+00	2.67E+00	2.78E+00	2.10E+00	1.99E+00
	StD	6.32E-14	3.02E-08	4.67E-08	2.15E+00	5.20E-07	5.88E-14
F19	Mean	-3.25E+00	-2.76E+00	-2.65E+00	-2.98E+00	-2.82E+00	-3.44E+00
	StD	2.32E-15	6.77E-08	6.99E-08	3.98E-02	1.59E-07	2.31E-15
F20	Mean	-2.71E+00	-2.45E+00	-2.70E+00	-2.79E+00	-2.15E+00	-2.57E+00
	StD	3.23E-08	1.88E-01	3.42E-08	2.62E-01	1.87E-01	3.01E-08
F21	Mean	-7.49E+00	-5.65E+00	-4.65E+00	-7.47E+00	-4.58E+00	-7.53E+00
	StD	5.35E-01	1.60E+00	1.46E+00	1.23E+00	1.40E+00	5.20E-01
F22	Mean	-7.82E+00	-6.72E+00	-7.88E+00	-7.94E+00	-5.14E+00	-7.50E+00
	StD	3.17E-01	1.24E+00	5.60E-01	5.26E-01	1.31E+00	3.02E-01
F23	Mean	-7.89E+00	-6.56E+00	-8.06E+00	-9.15E+00	-6.09E+00	-9.10E+00
	StD	6.39E-01	1.23E+00	9.12E-01	7.10E-01	1.61E+00	5.92E-01

Table 3. Grand exhibition of the standard deviation (STD) and the majestic average of the objective function, encompassing all test functions and algorithms.

entropy, spectral centroid, and spectral crest were retrieved and identified using the Hybrid CapsNet/UIGMM. Applying the model to three benchmark datasets, including ISMIR2004 (ISMIR), GTZAN, and Extended Ballroom (Ballroom), and comparing its results to those of state-of-the-art approaches allowed us to assess its performance. By analyzing the outcomes, the music's categorization into its respective genres, the model's accuracy has been tested. A Windows 11 PC was used entirely for the assessment in MATLAB-R2020a. The computer made use of 64 GB of RAM and an Intel® CoreTM i7-9700 K CPU running at 2.3 GHz.

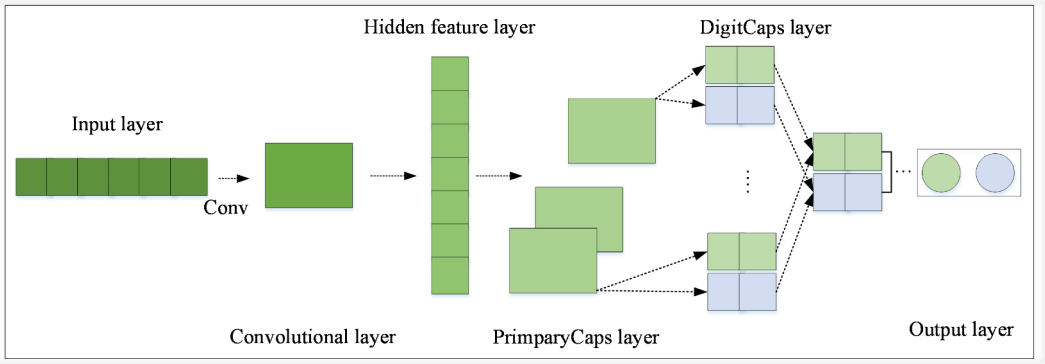

Fig. 4. A simple schematic diagram of the CapsNet used in this research.

Measurement indicators

The classification quantity of the proposed CapsNet/UIGMM is determined using five commonly used measurement indicators. These indicators, namely Precision, Sensitivity, Accuracy, F1-score, and Specificity, are utilized to assess the performance of the model. The mathematical equations representing these indicators are provided below.

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

$$Recall = \frac{TP}{TP + FN} \tag{28}$$

$$F1 = 2 \times \frac{Precision \times Sensitivity}{Precision + Sensitivity} \tag{29}$$

$$\left(z_i^d\right)' = z_i^d + x_n \tag{30}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{31}$$

where, the evaluation metrics involved TP (true positive) cases, FP (false positive) cases, FN (false negative) cases, and TN (true negative) cases.

In this study, the evaluation of CapsNet/UIGMM was conducted in two distinct phases. Firstly, the performance of CapsNet/UIGMM was compared to that of CapsNet and CapsNet/IGMM. Following this, in the second phase, the performance evaluation was carried out using a 5-fold comparison of CapsNet/UIGMM against five other state-of-the-art methods. These methods included Convolutional Neural Network (CNN)[10], CNN2[11], RNN[12], RNN-LSTM[13], balanced ResNet50_trust[14].

Ablation analysis

In this section, an ablation study has been applied to assess the efficacy of each element within the Capsule Net architecture. Here, three state of the proposed model are used. One for the original CapsNet, another one for combination of the CapsNet with the original Ideal Gas Molecular Movement (IGMM) algorithm, and finally, the proposed CapsNet/UIGMM. Table 4 and Fig. 5 present the assessment outcomes for the three methodologies, namely CapsNet, and CapsNet/IGMM, in the context of music genre classification.

Figure 5 illustrates a column chart showing the evaluation results of three distinct approaches, namely CapsNet, CapsNet/IGMM, and CapsNet/UIGMM.

The performance of CapsNet/UIGMM in music genre classification tasks is evaluated against CapsNet and CapsNet/IGMM using the ISMIR, GTZAN, and E-Ballroom datasets. The results demonstrate that CapsNet/UIGMM outperforms the other models, showcasing its superior ability to handle complex audio signal processing and classification scenarios.

Across all datasets, CapsNet/UIGMM consistently achieves higher Precision, Sensitivity, and F1 Scores, indicating its exceptional effectiveness in accurately identifying and classifying music genres. Notably, in the GTZAN dataset, it achieves near-perfect precision and showcases high Sensitivity and F1 scores, highlighting its robustness and adaptability in diverse musical contexts.

Although CapsNet/UIGMM's accuracy rates may be slightly lower than CapsNet in some cases, they remain competitive, suggesting that the model's precision and Sensitivity capabilities are not significantly compromised. Additionally, CapsNet/UIGMM exhibits notably high specificity scores, demonstrating its proficiency in minimizing false positives, which is crucial for reliable music genre classification. The AUC results further emphasize CapsNet/UIGMM's enhanced ability to accurately distinguish between different genres, surpassing the other models in all evaluation metrics. This comprehensive analysis underscores the potential of CapsNet/

Dataset	Method	Precision (%)	Accuracy (%)	Specificity (%)	Sensitivity (%)	Fl Score (%)	AUC (%)
ISMIR	CapsNet	93.248	94.681	87.597	93.667	85.108	83.454
	CapsNet/IGMM	87.061	90.263	82.132	91.972	74.740	82.585
	CapsNet/UIGMM	97.934	93.820	96.256	96.235	88.332	87.290
GTZAN	CapsNet	94.830	95.637	88.847	95.284	86.159	85.051
	CapsNet/IGMM	88.651	92.059	83.526	93.052	75.594	83.416
	CapsNet/UIGMM	99.569	95.145	97.698	97.954	89.780	88.568
E-Ballroom	CapsNet	94.020	94.348	87.931	94.217	85.179	83.484
	CapsNet/IGMM	88.336	91.828	83.342	92.120	74.364	81.889
	CapsNet/UIGMM	98.195	94.323	97.476	96.143	88.761	87.344

Table 4. Evaluation results based on CapsNet, CapsNet/IGMM, and CapsNet/UIGMM.

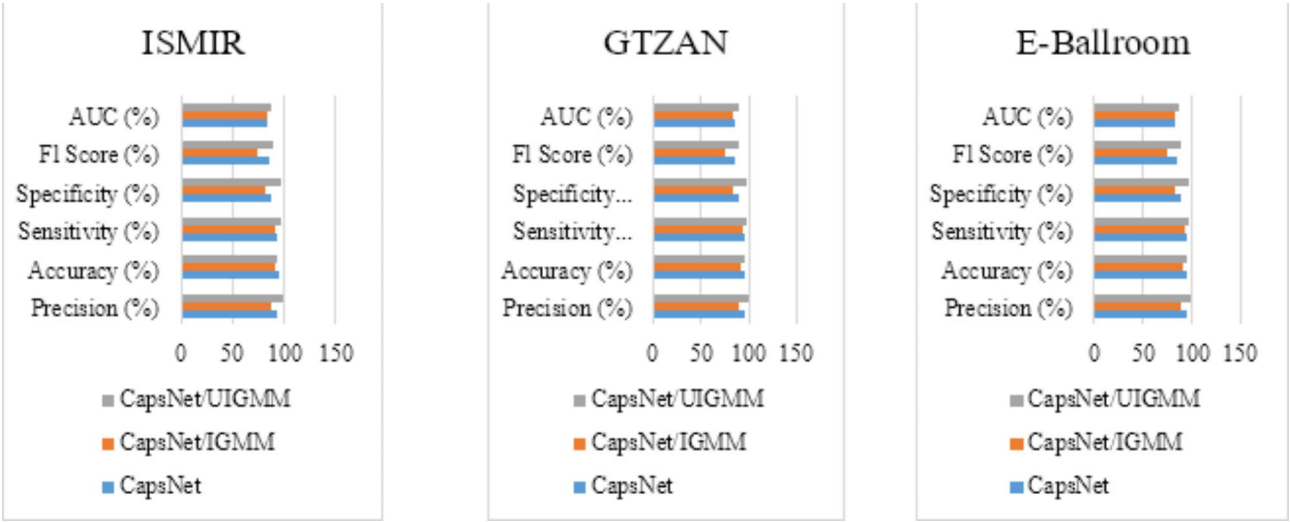

Fig. 5. Column chart of the evaluation results of the three approaches, CapsNet, CapsNet/IGMM, and CapsNet/UIGMM.

Hyperparameter	Selected Value
Number of Primary Capsules	32
Number of Digit Capsules	10
Number of Convolutional Filters	64
Kernel Size	3
Batch Size	64
Learning Rate	0.01

Table 5. Hyperparameter Selection results.

UIGMM as a highly reliable and precise tool for audio signal processing and classification, offering significant improvements over existing methodologies.

The study also investigates the performance of the model on the validation set for different values of the hyperparameters and selected the values that resulted in the best performance [Table 5].

The findings indicate that the chosen hyperparameters achieved optimal performance on the validation set.

Comparative analysis

Following the successful validation of the proposed CapsNet/UIGMM in the diagnosis of foot fracture x-ray images, our next step involved a comprehensive performance comparison with five established and recently developed methods. These methods encompass Convolutional Neural Network (CNN1)[10], Convolutional Neural Network (CNN2)[11], RNN[12], RNN/LSTM[13], balanced ResNet50_trust (ResNet50t)[14]. The statistical results derived from this analysis are presented in Table 6 and Fig. 5.

Figure 6 illustrates a bar chart depicting the performance of the CapsNet/UIGMM model in comparison to other methods.

The performance of CapsNet/UIGMM in diagnosing foot fractures from x-ray images was compared to five established methods, namely CNN1, CNN2, RNN, RNN/LSTM, and ResNet50_trust (ResNet50t), across three datasets (ISMIR, GTZAN, and Ballroom). The comparative analysis revealed that CapsNet/UIGMM exhibited superior performance in terms of precision, accuracy, Sensitivity, specificity, F1 score, and AUC percentages.

In the ISMIR dataset, CapsNet/UIGMM outperformed the other models with a precision of 96.180%, accuracy of 91.896%, and notably higher sensitivity, specificity, F1 score, and AUC percentages. These results indicate the exceptional ability of CapsNet/UIGMM to accurately diagnose foot fractures from x-ray images. This trend of superior performance continued across the GTZAN and Ballroom datasets, where CapsNet/UIGMM consistently achieved the highest scores in all metrics. In the GTZAN dataset, CapsNet/UIGMM achieved a precision of 96.622% and an accuracy of 93.167%.

Similarly, in the Ballroom dataset, CapsNet/UIGMM achieved a precision of 96.679% and an accuracy of 93.234%. Additionally, CapsNet/UIGMM demonstrated outstanding sensitivity, specificity, F1 score, and AUC percentages in both datasets. These results highlight the effectiveness of CapsNet/UIGMM in delivering highly accurate and reliable diagnoses compared to conventional CNN, RNN, and ResNet models. This signifies a significant advancement in the application of deep learning techniques for medical imaging diagnostics.

Conclusions

The field of Music Information Retrieval (MIR) has experienced noteworthy progress in recent decades, driven by the increasing demand to efficiently organize, search, and suggest music in digital format. Within the various tasks in MIR, the classification of music into genres poses a particular challenge due to the intricate and subjective nature of musical genres. Traditional approaches to this task have relied on manually crafted feature extraction combined with classical machine learning algorithms. However, the emergence of deep learning has revolutionized the field, allowing models to automatically learn complex patterns directly from the data. In this study, a novel methodology based on Capsule Neural Networks (CapsNet) was proposed to capture hierarchical relationships in the data, offering a promising architecture for addressing the nuanced task of music genre classification. CapsNet's performance relies on precise parameter configuration, which can hinder its application in music genre classification. However, solving this optimization problem is challenging due to the vast range of possible parameter values and intricate interactions. Traditional optimization techniques often struggle due to computational inefficiency or the inability to find the global optimum in complex environments. The research enhances the accuracy of CapsNet in music genre classification by combining it with an upgraded Ideal Gas Molecular Movement (UIGMM) optimization algorithm. This method not only improves the model's accuracy but also offers a method for optimizing complex neural network architectures, contributing to the broader field of deep learning. This research evaluated three benchmark datasets in the MIR community: ISMIR2004, GTZAN, and Extended Ballroom, assessing various music genres. The UIGMM-optimized CapsNet model showed superior performance in classifying these genres compared to existing models. The integration of UIGMM with CapsNet serves as a blueprint for future research, overcoming optimization challenges in other complex neural network architectures. This approach not only improves accuracy and efficiency in MIR models but also opens new avenues as the future work for exploring the synergy between deep learning and metaheuristic optimization techniques.

Dataset	Method	Precision (%)	Accuracy (%)	Specificity (%)	Sensitivity (%)	Fl Score (%)	AUC (%)
ISMIR	CNN1	85.907	84.500	81.128	87.751	76.597	71.116
	CNN2	89.228	90.594	85.885	93.990	83.049	79.049
	RNN	87.997	88.429	80.643	88.988	72.159	78.804
	RNN/LSTM	91.668	91.398	85.541	93.883	82.129	82.680
	ResNet50t	88.081	86.388	80.688	91.811	75.080	78.156
	CapsNet/UIGMM	96.180	91.896	92.459	95.212	87.873	83.894
GTZAN	CNN1	87.615	84.935	84.057	89.424	77.478	71.566
	CNN2	91.104	90.509	86.269	94.841	80.961	80.445
	RNN	86.521	88.207	81.305	87.498	72.633	77.799
	RNN/LSTM	91.191	91.435	85.540	91.281	81.219	81.671
	ResNet50t	85.677	84.719	80.888	91.288	75.794	78.633
	CapsNet/UIGMM	96.622	93.167	92.975	95.536	86.223	86.650
Ballroom	CNN1	85.646	84.708	83.090	86.876	76.908	70.723
	CNN2	90.494	89.615	87.287	92.945	83.515	80.587
	RNN	88.416	88.793	81.944	88.204	73.161	79.113
	RNN/LSTM	91.889	92.843	85.826	90.407	83.939	81.170
	ResNet50t	85.530	86.253	79.371	90.083	73.723	79.226
	CapsNet/UIGMM	96.679	93.234	93.532	95.688	87.286	86.051

Table 6. Comparison analysis of the proposed CapsNet/ESSO for foot fracture diagnosis.

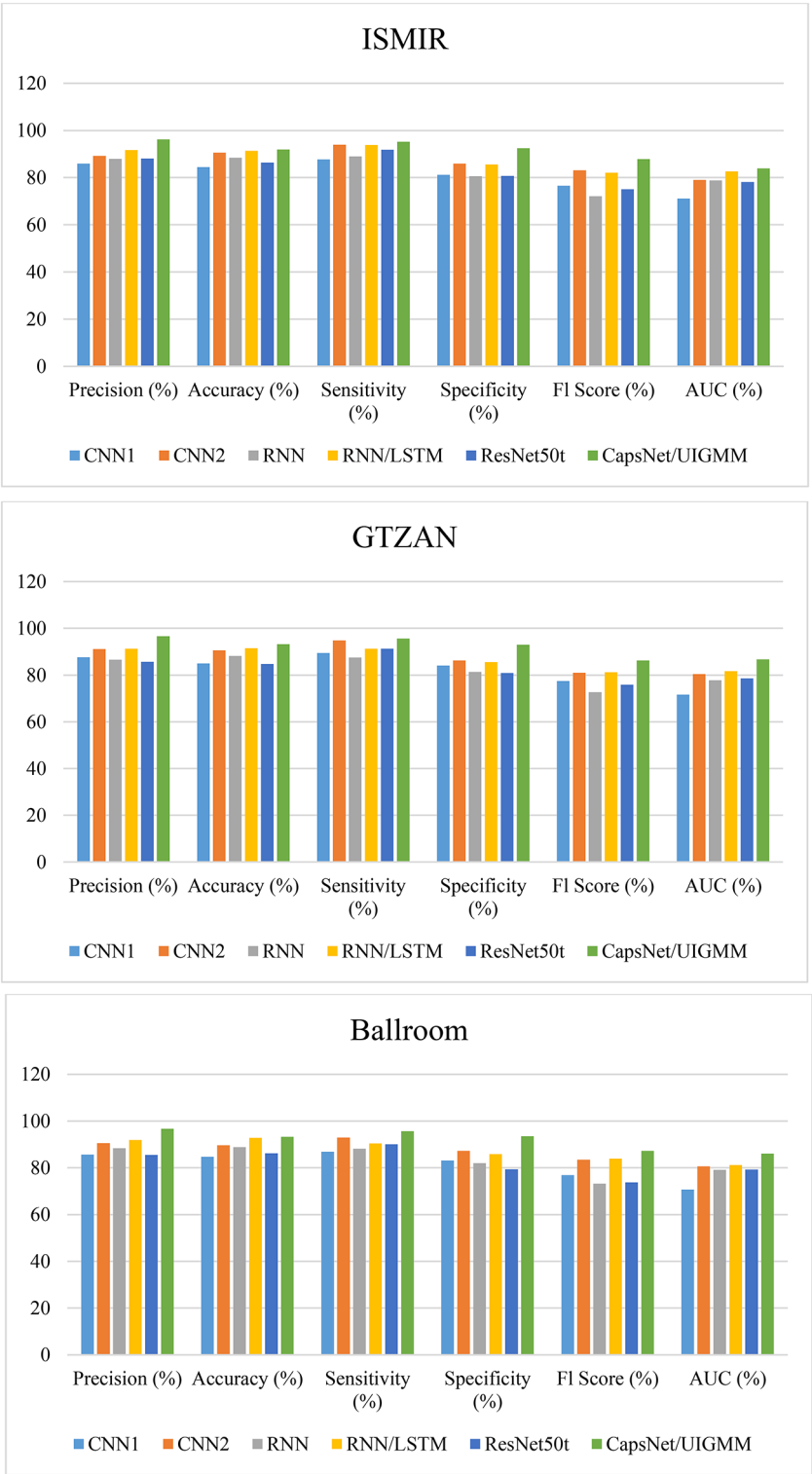

Fig. 6. Bar chart of the CapsNet/UIGMM toward other comparative methods.

Data availability

All data generated or analysed during this study are included in this published article.

Received: 13 October 2024; Accepted: 28 November 2024
Published online: 28 December 2024

References

1. Farajzadeh, N., Sadeghzadeh, N. & Hashemzadeh, M. PMG-Net: Persian music genre classification using deep neural networks. *Entertainment Comput.* **44**, 100518 (2023).
2. Pelchat, N. & Gelowitz, C. M. Neural network music genre classification. *Can. J. Electr. Comput. Eng.* **43**(3), 170–173 (2020).
3. Hung, Y. N. et al. Low-resource music genre classification with cross-modal neural model reprogramming. In *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2023).
4. Liu, C. et al. Bottom-up broadcast neural network for music genre classification. *Multimedia Tools Appl.* **80**, 7313–7331 (2021).
5. Jena, K. K. et al. A hybrid deep learning approach for classification of music genres using wavelet and spectrogram analysis. *Neural Comput. Appl.* 1–26 (2023).
6. Sun, J. et al. *Memristor-Based Neural Network Circuit of Associative Memory with Overshadowing and Emotion Congruent Effect* (IEEE Transactions on Neural Networks and Learning Systems, 2024).
7. Liu, D. et al. Dynamical analysis of high-order Hopfield neural network with application in WBANs. *Phys. Scr.* **99**(8), 085258 (2024).
8. Vishnupriya, S. & Meenakshi, K. Automatic music genre classification using convolution neural network. In *International Conference on Computer Communication and Informatics (ICCCI)* (IEEE, 2018).
9. Lerch, A. *Music Similarity Detection and Music Genre Classification.* (2023).
10. Dong, M. *Convolutional neural network achieves human-level accuracy in music genre classification.* arXiv preprint arXiv:1802.09697 (2018).
11. Chillara, S. et al. Music genre classification using machine learning algorithms: A comparison. *Int. Res. J. Eng. Technol.* **6**(5), 851–858 (2019).
12. Ashraf, M. et al. A globally regularized joint neural architecture for music classification. *IEEE Access.* **8**, 220980–220989 (2020).
13. Sharma, A. K. et al. Classification of Indian classical music with time-series matching deep learning approach. *IEEE Access.* **9**, 102041–102052 (2021).
14. Li, J. et al. An evaluation of deep neural network models for music classification using spectrograms. *Multimedia Tools Appl.* 1–27 (2022).
15. Yu, Y. et al. Music auto-tagging with capsule network. In *Data Science: 6th International Conference of Pioneering Computer Scientists, Engineers and Educators, ICPCSEE 2020, Taiyuan, China, September 18–21, 2020, Proceedings, Part I 6* (Springer, 2020).
16. GTZAN. *GTZAN Dataset - Music Genre Classification* (Kaggle, 2020).
17. Cano, P. et al. *ISMIR 2004 Audio Description Contest* (Music Technology Group of the Universitat Pompeu Fabra, Tech. Rep, 2006).
18. Marchand, U. & Peeters, G. *The extended ballroom dataset.* (2016).
19. Meng, Q. et al. A single-phase transformer-less grid-tied inverter based on switched capacitor for PV application. *J. Control Autom. Electr. Syst.* **31**(1), 257–270 (2020).
20. Yuan, K. et al. Optimal parameters estimation of the proton exchange membrane fuel cell stacks using a combined owl search algorithm. *Energy Sour. Part a Recover. Utilization Environ. Eff.* **45**(4), 11712–11732 (2023).
21. Ahmadova, S. & Erek, M. A review on Ripple, a financial intermediary Coin. *Akademik İzdüşüm Dergisi* **7**(2), 117–130 (2022).
22. Cai, X. et al. Breast cancer diagnosis by convolutional neural network and advanced thermal exchange optimization algorithm. *Computational and Mathematical Methods in Medicine* 2021 (2021).
23. Yang, Z. et al. Robust multi-objective optimal design of islanded hybrid system with renewable and diesel sources/stationary and mobile energy storage systems. *Renew. Sustain. Energy Rev.* **148**, 111295 (2021).
24. Bo, G. et al. Optimum structure of a combined wind/photovoltaic/fuel cell-based on amended Dragon fly optimization algorithm: A case study. *Energy Sour. Part a Recover. Utilization Environ. Eff.* **44**(3), 7109–7131 (2022).
25. Zhang, L. et al. A deep learning outline aimed at prompt skin cancer detection utilizing gated recurrent unit networks and improved orca predation algorithm. *Biomed. Signal Process. Control* **90**, 105858 (2024).
26. Chang, L., Wu, Z. & Ghadimi, N. A new biomass-based hybrid energy system integrated with a flue gas condensation process and energy storage option: An effort to mitigate environmental hazards. *Process Saf. Environ. Prot.* **177**, 959–975 (2023).
27. Ghiasi, M. et al. A comprehensive review of cyber-attacks and defense mechanisms for improving security in smart grid energy systems: Past, present and future. *Electr. Power Syst. Res.* **215**, 108975 (2023).
28. Rezaie, M. et al. Model parameters estimation of the proton exchange membrane fuel cell by a modified Golden Jackal optimization. *Sustain. Energy Technol. Assess.* **53**, 102657 (2022).
29. Li, S. et al. Evaluating the efficiency of CCHP systems in Xinjiang Uygur Autonomous Region: An optimal strategy based on improved mother optimization algorithm. *Case Stud. Therm. Eng.* **54**, 104005 (2024).
30. Arora, S. & Singh, S. Butterfly optimization algorithm: A novel approach for global optimization. *Soft. Comput.* **23**, 715–734 (2019).
31. Mohamed, A. W., Hadi, A. A. & Mohamed, A. K. Gaining-sharing knowledge based algorithm for solving optimization problems: A novel nature-inspired algorithm. *Int. J. Mach. Learn. Cybernet.* **11**(7), 1501–1529 (2020).
32. Zhang, Q. et al. Growth optimizer: A powerful metaheuristic algorithm for solving continuous and discrete global optimization problems. *Knowl. Based Syst.* **261**, 110206 (2023).
33. Ayyarao, T. S. et al. War strategy optimization algorithm: A new effective metaheuristic algorithm for global optimization. *IEEE Access.* **10**, 25073–25105 (2022).

Author contributions

Peiyan Chen, Jichi Zhang and Arsam Mashhadi wrote the main manuscript text and prepared figures. All authors reviewed the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.Z. or A.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.