A new design of multimedia big data retrieval enabled by deep feature learning and Adaptive Semantic Similarity Function

D. Sujatha¹ · M. Subramaniam² · Chinnanadar Ramachandran Rene Robin³

Received: 16 August 2021 / Accepted: 15 January 2022 / Published online: 5 February 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

Nowadays, multimedia big data have grown exponentially in diverse applications like social networks, transportation, health, and e-commerce, etc. Accessing preferred data in large-scale datasets needs efficient and sophisticated retrieval approaches. Multimedia big data consists of the most significant features with different types of data. Even though the multimedia supports various data formats with corresponding storage frameworks, similar semantic information is expressed by the multimedia. The overlap of semantic features is most efficient for theory and research related to semantic memory. Correspondingly, in recent years, deep multimodal hashing gets more attention owing to the efficient performance of huge-scale multimedia retrieval applications. On the other hand, the deep multimodal hashing has limited efforts for exploring the complex multilevel semantic structure. The main intention of this proposal is to develop enhanced deep multimedia big data retrieval with the Adaptive Semantic Similarity Function (A-SSF). The proposed model of this research covers several phases “(a) Data collection, (b) deep feature extraction, (c) semantic feature selection and (d) adaptive similarity function for retrieval. The two main processes of multimedia big data retrieval are training and testing. Once after collecting the dataset involved with video, text, images, and audio, the training phase starts. Here, the deep semantic feature extraction is performed by the Convolutional Neural Network (CNN), which is again subjected to the semantic feature selection process by the new hybrid algorithm termed Spider Monkey-Deer Hunting Optimization Algorithm (SM-DHOA). The final optimal semantic features are stored in the feature library. During testing, selected semantic features are added to the map-reduce framework in the Hadoop environment for handling the big data, thus ensuring the proper big data distribution. Here, the main contribution termed A-SSF is introduced to compute the correlation between the multimedia semantics of the testing data and training data, thus retrieving the data with minimum similarity. Extensive experiments on benchmark multimodal datasets demonstrate that the proposed method can outperform the state-of-the-art performance for all types of data.

Keywords Multimedia big data retrieval · Adaptive Semantic Similarity Function · CNN-based semantic feature extraction · Optimal feature selection · Spider Monkey-Deer Hunting Optimization Algorithm · Map-reduce framework

Communicated by A. Liu.

✉ D. Sujatha
sujatharaghhu@gmail.com

M. Subramaniam
subbu.21074@gmail.com

Chinnanadar Ramachandran Rene Robin
renerobin@sairam.edu.in

¹ Information Technology, St. Peter's College of Engineering and Technology, Chennai, India

² Dept. of Information Technology, SreeVidyanikethan Engineering College, Tirupati 517 102, India

³ Dept. of Computer Science and Engineering, Sri Sairam Engineering College, Chennai 602109, India

1 Introduction

In recent years, the emerging growth in multimedia data has been focused on different applications such as social networks, transportation, health, e-commerce, etc. Data diversity has become one of the major components of multimedia big data [9], where multimedia documents suffer from variations in the storage structures and data formats to show the equivalent semantic information [10]. Thus, retrieving and managing multimedia documents reflect the intent of users in heterogeneous big data platforms, which is considered a major problem. The efficient and sophisticated retrieval approaches need accessing of specific data in huge scale datasets [11]. Moreover, the major component for showing

the large-scale data produced in a faster manner is big data owing to the advancements in cloud computing algorithms and technologies, storage, communication, and sensing [12]. However, this gigantic data has more number of problems in industries and businesses. Consequently, it offers more possibilities for eminent future growth by considering the efficient use of data for evaluation [13]. Effective and reliable access is offered for suitable data in huge scale image repositories due to the complexities in their contents that have been analyzed in recent years [18].

Multimedia big data retrieval has more benefits inclusive of switching a distributed pattern for storing and processing the enormous multimedia contents [14]. Even though multimedia computing solves the computational burden and maintenance issues, it suffers from the storage and processing of multimedia big data. In a big data environment, a huge amount of commodity computers requires more storage capacity and enormous computation power while generating multimedia content [2]. As more applications and services are provided, processed, edited, and retrieved some multimedia documents, the significant multimedia contents with their storage structures and data formats demonstrate similar semantic information [15]. Hence, multimedia big data storage and retrieval are affected by the most common limitation called heterogeneity. Another complexity in the multimedia retrieval process is users' intention. In most of the existing methods, the retrieval process is often restricted to a similar category of multimedia content like images [16]. It minimizes the Quality of Service (QoS) of multimedia retrieval due to the failure in identifying the intention of users' search owing to the lack of type diversity [17].

The semantic feature can be overlapped that plays a crucial role in the research field in terms of semantic memory [24]. In recent years, more research works is carried out in the similarity measurement domain [19]. Different applications employ Euclidean distance [20] as a significant measure. Some of the specific scenarios use Mahalanobis distance [21] as a similarity measurement metric. Though, retrieval systems are impacted by the query conditions through inappropriate expressions. These problems are solved by adopting some deep learning methods, in which fuzzy matching techniques [22] are employed for reflecting the user's intent. On the other hand, these retrieval models do not precisely reflect the users' intentions concerning the incomplete information [23]. Hence, this model proposes a new multimedia big data retrieval model along with the semantic similarity function.

The major contributions of the developed model are given here as follows:

- To present a new multimedia big data retrieval model with deep semantic feature extraction technique, optimal semantic feature selection, and A-SSF-based retrieval

under the map-reduce framework with the aid of a new hybrid meta-heuristic-based algorithm.

- To choose the most significant semantic features of multimedia big data using a new hybrid algorithm called SM-DHOA with the objective of correlation maximization among the features. Therefore, this process gets unique and most representative features for further processes.
- To develop a new A-SSF-based retrieval with SM-DHOA under the map-reduce framework for the efficient retrieval process. It aims to maximize the correlation and F1-score among the features for retrieving the most suitable data according to the user query.
- To introduce a new algorithm called SM-DHOA with the integration of Deer Hunting Optimization Algorithm (DHOA) and Spider Monkey Optimization (SMO) algorithms for selecting the optimal semantic features and to optimize the weight function of features in map-reduce framework to maximize the retrieval performance with high precision and F1-score.
- To validate the efficiency of the suggested A-SSF-based retrieval model with different existing approaches in terms of standard performance measures along with statistical analysis.

The remaining section of the developed model is given here. The existing works are reviewed in Sect. 2. The proposed multimedia data retrieval using adaptive semantic similarity and deep learning is depicted in Sect. 3. The optimal semantic feature selection and map reducer-based multimedia data retrieval are explained in Sect. 4. The proposed SM-DHOA for enhanced multimedia data retrieval is discussed in Sect. 5. The results are evaluated in Sect. 6. Finally, the concluding statements are given in Sect. 7.

2 Literature survey

2.1 Related work

In 2015, Guo et al. [1] have developed an economical and effective structure called "Semantic-based Heterogeneous Multimedia Retrieval (SHMR)" to use the low cost for retrieving and storing the semantic information. Initially, this model has addressed individuality in big data environments with heterogeneous multimedia retrieval. Then, a new method was proposed for extracting and representing the semantic information. Moreover, this model has offered the multimedia data in semantic storage using a NoSQL-based technique, which was processed in a parallel way among the distributed nodes. Last, the developed model has achieved good user experience and high retrieval precision through a user feedback-driven approach and MapReduce-based retrieval technique. The suggested model has shown better

efficiency in terms of economic efficiency and retrieval performance.

In 2018, Ahmad et al. [2] have presented a more effective approach named Bi-Directional Fast Fourier Transform (BD-FFT) to get the condensed binary codes from the high dimensional deep features. It was used for saving the memory and simplified the computation for efficient retrieval. The transformed codes were considered as the hash codes to access the images in the huge scale datasets through Approximate Nearest Neighbor (ANN) search methods. It has achieved better retrieval accuracy based on the shorter length codes, which were experimented on Convolutional Neural Network (CNN) features. The experimentations were carried out on seven different huge scale datasets to demonstrate the efficiency of the suggested retrieval model.

In 2020, Xia et al. [3] have developed a deep correlation mining method for developing the multimedia retrieval model using “Levinberg-Marquard algorithm-based Deep Typical Correlation Analysis method (LM-DCCA)”. This model has intended to train the diverse medial features based on the suggested algorithm. Then, the correlation among the trained features was attained to compare the features of diverse multimedia data. This method has addressed the problem of local optima problems with the Levenberg–Marquart method. The experimentation was conducted on different datasets to show the efficiency of cross-media retrieval. This model has shown superior retrieval approaches while comparing with other progressed multimedia retrieval approaches.

In 2020, Sun et al. [4] have implemented a new graph classification-based semantic similarity model by introducing the feature reduction technique. Here, the suggested technique has initially learned the representations of vectors through neural language systems with subtree patterns. Further, similar subtree patterns were merged semantically to get the novel features. This model has offered a novel feature discrimination score for selecting the most discriminative features. The experimental analysis was conducted on real datasets to show the efficiency of the suggested retrieval method using the FRS-KELM graph classifier.

In 2018, Beltran et al. [5] have developed a new multimedia retrieval model using probabilistic Latent Semantic Analysis (pLSA) for solving the problems in big data along with data reduction strategy. This model has aimed to enhance the extraction procedure for Content-Based Multimedia Retrieval (CBMR). The developed system has focused on discussing the result of document reduction than other existing approaches, which were performed using pLSA and Latent Dirichlet Allocation (LDA) in CBMR. The suggested model has used the retrieval performance in CBMR.

In 2014, Yilmaz et al. [6] have implemented a “RELIEF-based modality weighting approach termed RELIEF-MM”. This method was the modified version of challenges like

handling unbalanced datasets, challenges with multi-labeled data and noise, class-specific feature selection, and utilizing the technique through machine learning identifications. It has focused on enhanced weight estimation function using RELIEF-MM to show the reliability and representation abilities of modalities with their discrimination capacity at lesser computational complexity. The experimental analysis was carried out on “TRECVID 2007, TRECVID 2008, and CCV datasets” that have shown the robustness and accuracy of the modality weighting on multimedia data retrieval model.

In 2018, Guo et al. [7] have suggested a new heterogeneous multimedia data retrieval model using Semantic Ontology Retrieval (SOR) for storing and retrieving the ontologies by processing the big data tools. Initially, the environment of ontology representation and semantic extraction were solved on multimedia big data. Secondly, SOR was used as the retrieval technique for model definition. Third, MapReduce was used for parallel processing of SOR to propose a new retrieval framework based on distributed nodes. Last, a user feedback method was presented for achieving a better user experience and high retrieval precision. The numerical results have shown that the designed SOR was appropriate for heterogeneous multimedia big data and semantic-based retrieval.

In 2019, Liu et al. [8] have implemented the “Deep Hashing with Multilevel Similarity Learning (DHMSL)” to learn the discriminative and compact hash codes to explore the “multilevel semantic similarity correlations of multimedia data.” Initially, the unified binary hash codes were learned for exploring the multilevel similarity correlation through the exploitation of semantic label information and the local structure. Consequently, the unified hash codes were made compact by taking the constraints like quantization and bit balance. This model has combined the Deep Neural Networks (DNN) for learning the feature representations using learned unified binary codes. The developed model has minimized the prediction errors among the outputs of the networks and unified binary codes. The simulation results were applied on two extensively used multimodal datasets to show the performance on both text and image-query-image tasks.

In 2018, Alghamdi et al. [30] have employed two types of transfer learning techniques to retrain the pre-trained VGG-Net (Fine-tuning and VGG-Net as fixed feature extractor) and obtained two new networks VGG-MI1 and VGG-MI2. In the VGG-MI1 model, the last layer of the VGG-Net model was replaced with a specific layer according to our requirements, and various functions were optimized to reduce overfitting.

In 2018, Sedik et al. [31] have suggested two data-augmentation models to enhance learnability of the Convolutional Neural Network (CNN) and the Convolutional Long Short-Term Memory (ConvLSTM)-based deep learning

models (DADLMs) and, by doing so, boost the accuracy of COVID-19 detection.

In 2018, Sedik et al. [32] have implemented a deep learning model based on CNN and a hybrid model that combined CNN with ConvLSTM to compute a three-tier probability that a biometric has been tempered. Simulation-based experiments have indicated that the alteration detection accuracy matches those recorded in advanced methods with superior performance in terms of detecting central rotation alteration to fingerprints.

2.2 Problem statement

Multimedia big data are dependable to retrieve semantic information concerning video, text, image, or audio from data sources. The extraction of such information is a major and challenging problem owing to their growth in media. Thus, there is a huge demand for fast multimedia retrieval models with the smallest utility gap. These challenges have occurred with the size and diversity of media in the corresponding field. Numerous approaches are developed in multimedia retrieval that has diverse features and challenges, as given in Table 1. MapReduce [1] offers better economic efficiency and also reduces investments in hardware, though this model was not able to adapt to a real-time Internet environment and also suffers from slow retrieval speed. BD-FFT

[2] increases the retrieval performance and overall efficiency and also improves the retrieval accuracy with deep features. However, this model does not perform on sparse features. LM-DCCA [3] has enhanced the training with a better convergence effect and also improves the overall performance. This model does not permit users for selecting the counter samples based on the user's feedback strategy. FRS-KELM graph classifier [4] attains better classification accuracy and selected the more discriminative features. On the other hand, it is not suitable for applying the developed feature reduction approach in Convolutional Neural Networks (CNN). Prior-based probabilistic Semantic Analysis (PpLSA) [5] improves the efficiency of information reduction and also increases the number of parameters required for extracting the information. Conversely, the developed approach does not support an incremental retrieval. RELIEF-MM [6] guarantees a higher accuracy rate with better performance and also a computationally efficient one. However, this model will not use some caching strategies that affect efficiency. MapReduce [7] achieves a good user experience and high retrieval precision and it cannot be applied in real-time applications. Multimodal hashing [8] is efficiently performed on multimedia retrieval. If the indiscriminative hash bits control the hash code, then the retrieval performance can be decreased. These challenges can be further used for developing new

Table 1 Features and challenges of existing multimedia big data retrieval

Author [citation]	Methodology	Features	Challenges
Guo et al. [1]	MapReduce	<ul style="list-style-type: none"> • It offers better economic efficiency • It also reduces the investments in hardware 	<ul style="list-style-type: none"> • This model was not able to adapt to a real-time Internet environment and also suffers from slow retrieval speed
Ahmad et al. [2]	BD-FFT	<ul style="list-style-type: none"> • This model increases the retrieval performance and overall efficiency • It also improves retrieval accuracy with deep features 	<ul style="list-style-type: none"> • This model does not perform on sparse features
Xia et al. [3]	LM-DCCA	<ul style="list-style-type: none"> • It enhances the training with better convergence effects • It also improves the overall performance 	<ul style="list-style-type: none"> • This model does not permit users for selecting the counter samples based on the user's feedback strategy
Sun et al. [4]	FRS-KELM graph classifier	<ul style="list-style-type: none"> • It attains better classification accuracy and selected the more discriminative features 	<ul style="list-style-type: none"> • It does not suitable for applying the developed feature reduction approach in Convolutional Neural Networks (CNN)
Beltran et al. [5]	PpLSA	<ul style="list-style-type: none"> • It improves the efficiency of information reduction • It also increases the number of parameters required for extracting the information 	<ul style="list-style-type: none"> • The developed approach does not support an incremental retrieval
Yilmaz et al. [6]	RELIEF-MM	<ul style="list-style-type: none"> • It guarantees a higher accuracy rate with better performance • It is also a computationally efficient one 	<ul style="list-style-type: none"> • It would not use some caching strategies that affect efficiency
Guo et al. [7]	MapReduce	<ul style="list-style-type: none"> • It achieves a good user experience and high retrieval precision 	<ul style="list-style-type: none"> • It cannot be applied in real-time applications
Liu et al. [8]	multimodal hashing	<ul style="list-style-type: none"> • It efficiently performs multimedia retrieval 	<ul style="list-style-type: none"> • The retrieval performance can be degraded while the indiscriminative hash bits dominate the Hase code

Fig. 1 Multimedia data retrieval using adaptive semantic similarity and deep learning

multimedia big data retrieval along with semantic feature similarity.

3 Proposed multimedia data retrieval using adaptive semantic similarity and deep learning

3.1 Architectural representation

Multimedia big data have emerged in recent years, which offers different mobile technologies and online services. Hence, more researches have been focusing on the multimedia era to target diverse constraints of big data analytics like capturing, storing, mining, indexing, and retrieval of multimedia big data. In recent days, the widespread and fast utilization of multimedia data consists of text, video, audio, and image with easy access and availability for managing the multimedia systems. The major problem in big data analytics is how to minimize the storage capacity and computational

time when maintaining accurate outcomes even for small datasets. Therefore, to get semantic information about the text, video, audio, and image from data sources, the multimedia retrieval system is necessary. However, the extraction of this multimedia information is more complex owing to the growth in media. Moreover, there is a high demand for faster multimedia retrieval with the least utility gap. Conversely, it also suffers from the size of media and diversity. As the major requirement of multimedia is to match the content and other information based on the user's necessity, it has also offered best-matched data retrieval. On the other hand, there are some utility gaps among the delivered data and expected data, which happen due to the different conditions like failures in retrieval models for reaching each media data, false content description and poor query generation. These challenges must be considered while implementing a new retrieval model that assists in searching for entire items saved on the internet by any other platform through social networking sites or direct uploads. Moreover, a good multimedia retrieval model must have completeness and

soundness. However, achieving the complete goodness in the retrieval model is challenging due to the retrieved irrelevant media. Likewise, completeness refers to the retrieval of entire significant multimedia data according to the user query, which is not properly attained on the internet. The multimedia big data suffer from the challenges like transmission, storage, additional compression, and analysis problems related to ensuring computing and scalability efficiency, dealing with understanding and cognition complexity, organization of heterogeneous and unstructured data, and solving the QoS and real-time requirements. Here, the technical and corresponding scientific issues promote new multimedia big data retrieval models with intelligent approaches, as depicted in Fig. 1.

The proposed multimedia data retrieval model consists of two stages: training and testing. Initially, the multimedia data including text, video, audio, and image as multimedia queries are collected from benchmark datasets, which are first processed to the feature extraction stage for getting the significant information. This suggested model has used CNN for getting the semantic features for describing the semantic contents from multimedia data. The CNN can extract the most useful features through the pooling layer, which extracts the most important features for enhancing the retrieval performance by reducing the redundant data or dimensionality of data. The main advantages for the convolutional layer consist of many filters which apply convolution operation to the input to capture some special features and pass the result to the next layer. Due to these advantages, CNN is properly trained even using different images from the same class. Optimal feature selection reduces the falsely selected features. Removing the irrelevant data improves learning accuracy and reduces the computation time. The extracted semantic features are given to the feature selection process for getting the optimal semantic features by applying a newly proposed SM-DHOA technique. This process is essential for the multimedia retrieval model to get the necessary information. When the huge set of features is processed, the retrieval model suffers from computation and time complexity. Therefore, feature selection is necessary for getting the most representative features. Weighted feature extraction trial-and-error for determining the appropriate number of extracted features can be avoided. A weight-based feature extraction approach to reduce the number of features for text classification. In the training stage, the weighted features are stored in the feature library, which is used for the retrieval process in a testing stage. Moreover, in the testing stage, all the above-mentioned processes are carried out for multimedia queries. The optimally selected semantic features using the SM-DHOA are given to the map-reduce framework. Here, it consists of two functions like mapper and reducer for retrieving the appropriate multimedia data. The optimal features are mapped into the mapper and generate

Fig. 2 CNN-based semantic feature extraction for multimedia retrieval process

the intermediate data, which is further given to the reducer. The reducer performs the retrieval performance using the A-SSF. For computing this similarity function, the optimal features of both trained data and query data are multiplied by a weight function. This weight function is optimized by the SM-DHOA and using the weighted features, the semantic similarity function is computed, and hence it is termed as A-SSF. The DHO algorithm can highly balance the exploration and exploitation phases and improve the accuracy of the proposed model. SM is a new type of swarm intelligence-based algorithm. It is used for finding the best-fit solution and also applied to solve complex optimization problems. This aims to maximize the precision and F1-score for attaining an efficient retrieval process.

3.2 Dataset description

This developed multimedia big data retrieval model uses two datasets for experimentation, as follows.

Dataset 1: It is collected from “<http://press.liacs.nl/mirfl/ickr/mirdownload.html>: Access Date: 2021-05-20”. It consists of 25,000 files like multimedia data like images, video, text, and audio files with different directories.

Dataset 2: It is gathered from the “https://github.com/satishrdd/Fast_image_retrieval: Access Date: 2021-05-20”. It consists of images from both query and database with different lists, where the database includes 1,000 files and the query consists of 54 files.

The input multimedia queries collected from both datasets are termed as Q_n , where $n = 1, 2, \dots, N$ and the total number of multimedia data in the dataset is indicated as N .

3.3 CNN-based semantic feature extraction

It is the initial stage of the proposed multimedia big data retrieval model, which is carried out by applying CNN. The feature extraction is the essential step for developing a new

retrieval model. It is the process of extracting the necessary features from multimedia data. It is required for extracting every feature of multimedia queries, which helps in retrieving the relevant multimedia data according to the queries. This stage gets the high dimensional features from the pooling layer of CNN.

CNN [27] is used in this developed multimedia big data retrieval model because of its automatic and efficient feature extraction capabilities without human intervention, which is also a computationally efficient approach. Moreover, this architecture is less dependent on pre-processing, and thus, CNN is more efficient in this model. Moreover, it is simple and easy for processing to offer higher accuracy in the retrieval process. CNN is a feedforward network, which consists of different layers like “convolutional and pooling (or subsampling)” layers that are grouped into modules. The multimedia query is given as the input for CNN, which is processed into different stages “convolutional and pooling” for attaining the specific representations, and finally, the output classes are attained from the output layer. This paper extracts the necessary semantic features from the pooling layers.

The input multimedia queries Q_n are given to the input layer of CNN, which is served as the feature extractors [30] for learning the feature representations from the Q_n . The convolution layer consisted of different neurons that are organized into feature maps, in which each neuron has its receptive field. Each neuron is correlated with a neighborhood of neurons in the earlier layer through a set of trainable weights, namely filter bank. Here, a new feature map is computed by convolving the inputs through the learned weights. Moreover, the convolved results are forwarded with a non-linear activation function. Here, the m th output feature map FE_m is computed in Eq. (1).

$$FE_m = \zeta(R_m * Q_n). \quad (1)$$

Fig. 3 Optimal semantic feature selection model

Here, the term R_m denotes the convolutional filter concerning with m th feature map, ζ represents the non-linear activation function, and the 2D convolutional operator is denoted as ‘ $\zeta(*)$ ’ that is responsible for extracting the non-linear features from the input Q_n . The pooling layer of CNN [31, 32] is used for minimizing the spatial resolution of the feature maps to get the spatial invariance for input translations and distortions. Most of the model employs average pooling aggregation layers for propagating the average of entire input values. The max-pooling layers choose the largest element in the corresponding receptive field as formulated in Eq. (2).

$$FE_{mps} = \max_{(u,v) \in \delta_{ps}} Q_{nmuv}. \quad (2)$$

In Eq. (2), the element at location (u, v) included through the pooling region δ_{ps} is termed as Q_{nmuv} that is used for embodying a receptive field about the position (p, s) , and the output of the pooling operation is denoted as FE_{mps} that is correlated through the m th feature map. Therefore, the final semantic features FE_m are attained using the pooling layer of CNN, where $m = 1, 2, \dots, M$ and M indicate the total number of semantic features and it is obtained as 784 features, which is further given to the retrieval process. This CNN-based semantic feature extraction is diagrammatically represented in Fig. 2.

4 Optimal semantic feature selection and map reducer-based multimedia data retrieval

4.1 Optimal semantic feature selection

The major contribution of this developed multimedia data retrieval model is to select the optimal semantic features from the extracted CNN-based semantic features FE_m . The feature selection is the process of reducing the noisy, excess, and extraneous multimedia feature subset to select the optimal features. Finally, the most significant and optimal features are selected using the SM-DHOA from the extracted semantic feature subsets. The optimal feature selection is performed using the SM-DHOA technique to reduce the retrieval time. The optimally selected semantic features using SM-DHOA are termed as FS_{m*} , where $m* = 1, 2, \dots, OM$ and OM represents the total number of optimal semantic features. The objective model here is minimizing the correlation between features.

The correlation CT among the optimal semantic features FS_{m*} is found for attaining the minimum correlated features, which is the primary objective of the developed model as formulated in Eq. (3).

Fig. 4 Adaptive semantic similarity in map reduce framework

$$FF_1 = \arg \max_{\{FS_{ms}\}} \left(\frac{1}{CT} \right), \quad (3)$$

$$CT = \frac{OM \sum fsfh - \sum fs \sum fh}{\sqrt{\left(OM \sum fs^2 - (\sum fs)^2 \right) - \left(OM \sum fh^2 - (\sum fh)^2 \right)}}. \quad (4)$$

Here, the two features in the optimal semantic feature set are represented as fs and fh , and the total number of extracted optimal semantic features is indicated as OM . It is attained as 10. Hence, the proposed multimedia retrieval model aims to get unique semantic features with minimum correlation by the SM-DHOA. This optimal semantic feature selection enhances multimedia retrieval efficiency. The optimal semantic feature selection is shown in Fig. 3.

4.2 Adaptive semantic similarity in map-reduce framework

This phase is used for retrieving the multimedia data using the map-reduce framework for handling the big data. The input to the map reduce-based retrieval process uses the optimal semantic features FS_{m*} . The map reduce model consist of different phases such as mapper function and reducer function. Here, the mapper function is utilized for mapping the optimal features to provide a transitional values and the reducer function is performed for processing the transitional values corresponding with the equal transitional features. In the map-reduce model, the input and output are considered as the query id and returned list, respectively. The optimally selected semantic features are given to the mapper, where these features are mapped together for getting the intermediate results. Further, the reducer finds the similarity function for efficient retrieval, where the weight function is multiplied with the features of the reducer and feature library. This

weight is optimized using the SM-DHOA for developing the A-SSF. The weighted feature extraction of the feature library FE_{tr}^{TR} is formulated based on Eq. (5).

$$FE_{tr}^{TR(new)} = FE_{tr}^{TR} \times wg_{tr}^{TR}. \quad (5)$$

Here, the new semantic feature set in the feature library after multiplying the weight function is represented as $FE_{tr}^{TR(new)}$ and the weight function used for the feature library is termed as wg_{tr}^{TR} , where FE_{tr}^{TR} denotes the features present in the feature library. Similarly, the features attained from the reducer are considered as FS_{tr}^{Rd} , which is multiplied with the weight function for getting the better retrieval results that are derived in Eq. (6).

$$FS_{tr}^{Rd(new)} = FS_{tr}^{Rd} \times wg_{tr}^{Rd}. \quad (6)$$

Here, the new semantic weighted feature set is denoted as $FS_{tr}^{Rd(new)}$, and the weight function employed for the reducer is represented as wg_{tr}^{Rd} . Therefore, the weighted features of both feature library and reducer are checked with adaptive similarity function, where the weights wg_{tr}^{Rd} and wg_{tr}^{TR} are optimized using the SM-DHOA. Therefore, the adaptive similarity function is formulated in the reducer as given in Eq. (7).

$$SF\left(FE_{tr}^{TR(new)}, FS_{tr}^{Rd(new)}\right) = \frac{1}{E_{FE_{tr}^{TR(new)}} + E_{FS_{tr}^{Rd(new)}}} \left| \left(FE_{tr}^{TR(new)} \right) \cap \left(FS_{tr}^{Rd(new)} \right) \right|. \quad (7)$$

Here, the total number of weighted features of $FE_{tr}^{TR(new)}$ and $FS_{tr}^{Rd(new)}$ are represented as $E_{FE_{tr}^{TR(new)}}$ and $E_{FS_{tr}^{Rd(new)}}$, respectively.

To summarize, the mapper function considers the pairs of record values and attained the intermediate results. Then, each pair is processed to get the multimedia retrieval process using the adaptive similarity function by the reducer. Finally, the returned list is attained as the multimedia retrieval content. The adaptive semantic similarity in the map-reduce framework is shown in Fig. 4.

5 Proposed SM-DHOA for enhanced multimedia data retrieval

5.1 Developed SM-DHOA

This proposed multimedia data retrieval model uses SM-DHOA for selecting the optimal features for getting the significant features. The suggested model has offered a new adaptive semantic similarity in map-reduce framework using the SM-DHOA for optimizing the weight function in the

reducer model. This algorithm aims to enhance the multimedia data retrieval for big data.

DHOA [25] and SMO [26] algorithms are selected because of their features, in which DHOA offers efficient balancing among exploitation and exploration stages with superior searching capabilities, which has determined the optimal position effectively. It provides excellent sensitivity and solves real-time optimization problems owing to the robustness and ability of faster convergence, though DHOA suffers from premature convergence. Thus, there is a need of adopting a new SM-DHOA with the SMO algorithm. It offers convergence more quickly and requires fewer resources. Hence, SMO is adopted with DHOA for improving multi-media data retrieval.

In this proposed SM-DHOA, the propagation through the position angle of DHOA is replaced by the global leader phase of the SMO algorithm. Hence, if ($f < 1$), then the solutions are updated based on the DHOA with leader and successor positions, or else the solutions are updated based on SMO algorithm with global leader phase.

The DHOA is motivated by considering the nature of the humans to hunt the deer with the mobility of hunters to get the best location with “leader and successor” solutions. The deer positions are updated until finding the deer. First, the deer is enclosed by the hunters with some constraints like “deer position and wind angle”. This algorithm has special characteristics like efficient teamwork among hunters for finding the deer, and better-attacking procedure to achieve the optimal location of prey through the position of leader and successor. Moreover, the prey or deer has diverse special characters as given in [25], which are studied for an efficient hunting process. DHOA is derived by initializing the population as given in Eq. (8).

$$X = \{X_1, X_2, \dots, X_p\}; \quad 1 < k \leq p. \quad (8)$$

Here, the number of hunters in the X th population is indicated as p . Moreover, the constraints like deer position and wind angle are derived in Eqs. (9) and (10), respectively.

$$\phi_k = 2\pi r, \quad (9)$$

$$\omega_k = \phi + \pi. \quad (10)$$

In the aforementioned equation, wind angle ω is necessary due to the circular search space. Here, the wind angle is termed as ϕ , the recent iteration is represented as k , ϕ indicates the deer’s position angle and r specifies the random number in the range of [0, 1]. The positions are propagated based on diverse constraints. Initially, the position of the optimal search area is anonymous and so, the solutions are randomly considered that is near to the optimal space through formulating the fitness function. As mentioned above, the solutions are considered as leader position and

successor position, which are represented as X^{ldr} and X^{Sur} . These are considered as the first best position and successor position of the hunters, respectively.

(i) “*Propagation through Leader’s position*”: At the initial stage, the best positions are formulated, and then every individual in the population is tried to get the best position. The enclosing behavior to find deer is formulated in Eq. (11).

$$X_{k+1} = X^{ldr} - Y \cdot f \cdot |B \times X^{ldr} - X_k|. \quad (11)$$

In Eq. (24), the position at the next iteration is denoted as X_{k+1} , the random number of wind speed is represented as f that varies among $[0, 2]$, the position of the current iteration is termed as X_k and coefficient vectors are represented as Y and B as formulated in Eqs. (12) and (13), respectively.

$$Y = \frac{1}{4} \log \left(k + \frac{1}{k_{\max}} \right) p, \quad (12)$$

$$B = 2 \times rnd. \quad (13)$$

Here, the term p is a parameter with the range of $[-1, 1]$, the maximum iteration is considered as $j_{\max}$, and rnd specifies the random number in the bounding range of $[-1, 1]$.

(ii) “*Propagation through position angle*” based on global leader position of SMO algorithm: SMO is a population-based technique inspired by the social interactions among the spider monkeys, which has been considered the fission–fusion social strategy for intelligent foraging behavior. The monkeys have different groups of leaders like local and global leaders and corresponding group members. This hierarchy is used for maintaining defensive boundaries and social bonds. It has different phases like “local leader phase, global leader phase, local leader learning phase, global

leader learning phase, local leader decision phase, and global leader decision phase”. In this developed SM-DHOA algorithm global leader phase is considered for updating the position. Once, the local leader phase is performed for finding the food, the entire spider monkeys reevaluate their positions based on the experience of members of the local group and global leader. This process helps in identifying the most suitable positions through Eq. (14).

$$X_{newkj} = X_{kj} + rn[0, 1](GL_j - X_{kj}) + rn[-1, 1](X_{sj} - X_{kj}). \quad (14)$$

In Eq. (14), the term X_{newkj} expresses the new position update based on the global leader, the k th spider monkey at j th dimension is formulated as X_{kj} , the random number is denoted as rn , the global leader position at j th dimension is formulated as GL_j and the s^{th} spider monkey at j th dimension is indicated as X_{sj} , where an arbitrarily chosen index is given as $j \in \{1, 2, \dots, J\}$.

(iii) “*Propagation through Successor’s position*”: The exploration process of DHOA is formulated by altering the vector B with the enclosing behavior. When $B < 1$ the random search is initialized. Thus, the position is updated with the position of the successor by considering the obtained first-best solution, which permits a global search that is derived in Eq. (15).

$$X_{k+1} = X^{Sur} - Y \cdot f \cdot |B \times X^{Sur} - X_j|. \quad (15)$$

Here, the successor position of the search agent is represented as X^{Sur} in the recent population.

Hence, the designed SM-DHOA is formulated based on the random parameter f , which decides the position updating through the SMO and DHOA. The pseudo-code of the proposed algorithm is given in Algorithm 1.

Algorithm 1: Proposed SM-DHOA

```

Initialization of population  $X$ 
Obtain first and second-best solutions like  $X^{ldr}$  and  $X^{Sur}$ , respectively.
Start
  While ( $k < k_{max}$ )
    for every solution in  $X$ 
      Compute the fitness function of every solution
      Determine the random parameters
      if ( $f < 1$ )
        if ( $|B| \geq 1$ )
          Position update based on DHOA
            Update the search agents based on leader position by Eq. (12)
          Else
            Update the search agents based on successor position by Eq. (15)
          End if
        Else
          Position update based on SMO
            Update the search agents based on the global leader phase of SMO using Eq. (14)
          End if
      End for
    Compute fitness function of every solution
    Update  $X^{ldr}$  and  $X^{Sur}$ 
     $k = k + 1$ 
  End while
Return  $X^{ldr}$ 
Terminate

```

This developed SM-DHOA is inspired by the hybridization concept, which offers efficient convergence behavior for the proposed multi-media data retrieval model. The flow-chart of the designed algorithm is given in Fig. 5.

5.2 Objective model for a-ssf in data retrieval

The proposed multimedia data retrieval model focuses on efficient retrieval using a new SM-DHOA with optimal semantic feature selection with A-SSF. The weighted feature extraction is performed to implement the A-SSF

computation and the optimization [33] strategy is used to tune the weight function, thus calling the similarity function A-SSF. Here, the map-reduce framework is used for retrieval by optimizing the weight function of features in the reducer for increasing the performance in terms of precision and F1-score. The major multi-objective function FF_2 for A-SSF using SM-DHOA is derived in Eq. (16).

$$FF_2 = \arg \min_{\{w_{gr}^{Rd}, w_{gr}^{TR}\}} \left(\frac{1}{prn + F1 - score} \right). \quad (16)$$

Fig. 5 Flowchart of the designed SM-DHOA algorithm

Here, precision is “the ratio of positive observations that are predicted exactly to the total number of observations that are positively predicted” as formulated in Eq. (17).

$$prn = \frac{te^{ps}}{te^{ps} + fe^{ps}}. \quad (17)$$

The F1-score is the “harmonic mean between precision and recall. It is used as a statistical measure to rate performance” as derived in Eq. (18).

$$F1 - \text{score} = \frac{2te^{ps}}{2te^{ps} + fe^{ps} + fe^{ng}}. \quad (18)$$

Here, terms te^{ps} , te^{ng} , fe^{ps} and fe^{ng} refer to the “true positives, true negatives, false positives, and false negatives,” respectively. The multi-objective function focuses on improving the retrieval performance through adaptive semantic similarity in the map-reduce framework using SM-DHOA.

6 Results and discussions

6.1 Experimental setup

The proposed multimedia big data retrieval was developed in MATLAB 2020a, and the experimental analysis was carried out. From the datasets, video to video and audio to an image is analyzed using dataset 1 and also text to image and image to image is analyzed using dataset 2. Here, the performance of the proposed model was compared over the conventional methods like PSO [28], GWO [29], DHOA [25], and SMO [26] and existing models like BD-FFT [2], FRS-KELM [4], and RELIEF-MM [6]. The experimental analysis was performed by considering the number of iteration as 10 and the number of populations as 10.

6.2 Performance measures

Various performance metrics are used for evaluating the performance, which is described below.

- (a) MCC: “correlation coefficient computed by four values”.

$$\text{MCC} = \frac{te^{ps} \times te^{ng} - fe^{ps} \times fe^{ng}}{\sqrt{(te^{ps} + fe^{ps})(te^{ps} + fe^{ng})(te^{ng} + fe^{ps})(te^{ng} + fe^{ng})}}. \quad (19)$$

- (b) NPV: “probability that subjects with a negative screening test truly don't have the disease”.

$$\text{NPV} = \frac{te^{ng}}{fe^{ng} + te^{ng}}. \quad (20)$$

- (c) FDR: “the number of false positives in all of the rejected hypotheses”.

$$\text{FDR} = \frac{fe^{ps}}{fe^{ps} + te^{ps}}. \quad (21)$$

- (d) FPR: “the ratio of the count of false-positive predictions to the entire count of negative predictions”.

$$\text{FPR} = \frac{fe^{ps}}{fe^{ps} + te^{ng}}. \quad (22)$$

- (e) FNR: “the proportion of positives which yield negative test outcomes with the test”.

$$\text{FNR} = \frac{fe^{ng}}{te^{ng} + te^{ps}}. \quad (23)$$

- (f) Sensitivity: “the number of true positives, which are recognized exactly”.

$$\text{Se} = \frac{te^{ps}}{te^{ps} + fe^{ng}}. \quad (24)$$

- (g) Specificity: “the number of true negatives, which are determined precisely”.

$$\text{Sp} = \frac{te^{ng}}{te^{ng} + fe^{ps}}. \quad (25)$$

- (h) Accuracy: It is a “ratio of the observation of exactly predicted to the whole observations”.

$$\text{Ac} = \frac{(te^{ps} + te^{ng})}{(te^{ps} + te^{ng} + fe^{ps} + fe^{ng})}. \quad (26)$$

Fig. 6 Performance analysis on the precision of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for **a** text to image retrieval, **b** image to image retrieval, **c** video to video retrieval, and **d** audio to image retrieval

Fig. 7 Performance analysis on the precision of the proposed Multimedia Big Data Retrieval model with different machine learning-based algorithms for **a** text to image retrieval, **b** image to image retrieval, **c** video to video retrieval, and **d** audio to image retrieval

- (i) Recall: It is measured “as the number of true positives divided by the total number of true positives and false negatives.”

$$Re = \frac{te^{ps}}{te^{ps} + fe^{ng}}. \quad (27)$$

6.3 Analysis of precision

The performance of the designed multimedia big data retrieval model using SM-DHOA with A-SSF is analyzed in terms of precision by varying the number of multi-media data retrieved, which is depicted in Figs. 6 and 7 for meta-heuristic-based algorithms and conventional models, respectively. When considering the number of the multi-media data retrieved as 20 for text to image retrieval, the suggested SM-DHOA gets 7.9%, 4.3%, 6.7%, 6%, 4.4%, 6.7%, and 5.5% higher precision rates than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively. For an image to image retrieval, the proposed SM-DHOA attains

5.7%, 4.5%, 6.97%, 5%, 6.9%, 6.9%, and 6.9% maximum precision rate than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, while taking the number of multi-media data retrieved as 40. Similarly, the performance of the developed SM-DHOA is analyzed in terms of precision with the number of multimedia data retrieved as 60, which is 3.7%, 5%, 2.4%, 7.6%, 2.4%, 1.2%, and 9% superior to PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for a video to video retrieval. Likewise, the efficiency of the recommended SM-DHOA for audio to image retrieval is 11%, 1.28%, 2.5%, 6.7%, 8%, 6.75%, and 5.3% progressed than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, while taking the number of multi-media data retrieved as 80. Finally, the proposed model shows better performance by comparing with the existing approaches.

6.4 Analysis of F1-score

For the statistical analysis, the number of the multi-media data retrieval was taken as R@100 for image retrieval. In

Fig. 8 Performance analysis on F1-score of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for **a** text to image retrieval, **b** image to image retrieval, **c** video to video retrieval, and **d** audio to image retrieval

our proposed model, image 1 of the actual and predicted images are compared in terms of recall which is regarded as $R@1$, further the set of 2 of the images are compared in terms of recall which is regarded as $R@2$. Similarly, the comparison is made for the $R@100$ images. The proposed multimedia big data retrieval is evaluated in terms of F1-score with different optimization-based algorithms and classifiers as given in Figs. 8 and 9, respectively. The F1-score of the suggested SM-DHOA shows the better performance, which is 8%, 9.4%, 10.7%, 6.8%, 8%, 6.8%, and 9.4% advanced than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, by considering the number of multi-media data retrieved as 40 for text to image retrieval. While taking the image to image retrieval, the performance of the recommended SM-DHOA is 8.75% advanced than PSO and GWO, 8.75% advanced than DHOA, 7.4% advanced than SMO, 4.8% advanced than BD-FFT, 6% advanced than FRS-KELM, and 6% advanced than RELIEF-MM when evaluating the number of multimedia data retrieved as 60. Likewise, when considering the video to video retrieval with the number of multi-media

data retrieved as 80, the developed SM-DHOA is 6.75%, 3.9%, 8.2%, 9.7%, 9.4%, 10.9%, and 8.7% better than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively. Finally, when taking the number of multimedia data retrieved as 100, the implemented SM-DHOA is 11.7%, 13.4%, 8.5%, 16.9%, 5.6%, 8.69%, and 15.3% enhanced than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for audio to image retrieval. Further, the designed multimedia big data retrieval model using SM-DHOA with A-SSF establishes superior performance in terms of F1-score when compared to other approaches.

6.5 Overall performance analysis on the accuracy

The overall performance of the designed multimedia big data retrieval model is analyzed in terms of accuracy with different optimization-based algorithms and classifiers as given in Fig. 10. While considering the text to image retrieval, the performance of the designed SM-DHOA is 9.3%, 6.8%, 4.44%, and 4.44% superior to PSO, GWO,

Fig. 9 Performance analysis on F1-score of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for **a** text to image retrieval, **b** image to image retrieval, **c** video to video retrieval, and **d** audio to image retrieval

Fig. 10 Performance analysis on Accuracy for the proposed multimedia big data retrieval model for four datasets with **a** different optimization-based algorithms and **b** classifiers

DHOA, and SMO, respectively. The performance of the designed SM-DHOA is 4.3%, 3.2%, and 1% enhanced than BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for the video to video retrieval. Hence, the performance of the

designed model with A-SSF using SM-DHOA gets a more accurate retrieval rate than other conventional models.

Table 2 Statistical analysis of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for four datasets in terms of precision

Algorithms	Best	Worst	Mean	Median	Standard deviation
Text to image retrieval					
PSO [28]	88	68	81.11	83	6.5426
GWO [29]	91	65	80.72	80	8.6528
DHOA [25]	89	67	81.09	83	7.3788
SMO [26]	89	65	80.63	82	7.6866
SM-DHOA	95	75	87.49	89	6.2869
Image to image retrieval					
PSO [28]	90	67	81.19	83	7.37
GWO [29]	88	70	82.17	84	6.1694
DHOA [25]	89	70	81.45	84	6.3268
SMO [26]	90	69	82.05	84	6.5757
SM-DHOA	95	75	88.29	90	6.0658
Video to video retrieval					
PSO [28]	91	65	80.02	82	9.2321
GWO [29]	90	70	82.3	83	6.5017
DHOA [25]	90	67	80.63	83	7.5954
SMO [26]	90	69	82.31	84	6.5052
SM-DHOA	95	75	86.72	89.5	6.8329
Audio to image retrieval					
PSO [28]	88	65	79.88	82.5	8.1182
GWO [29]	91	70	83.73	86	6.0233
DHOA [25]	90	70	82.53	83.5	6.1913
SMO [26]	91	69	81.48	83	6.9391
SM-DHOA	95	75	86.79	88	6.7664

6.6 Statistical analysis on precision

The statistical analysis on precision for the developed multimedia data retrieval model is given in Tables 2 and 3, respectively, by considering different query data. “The mean is the average value of the best and worst values and the median is referred to as the center point of the best and worst values, whereas the standard deviation is represented as the degree of deviation between each execution”. While considering the text to image retrieval, the accuracy of the SM-DHOA is 7.8%, 8.3%, 7.8%, 8.5%, 9.2%, 9.6%, and 7.69% higher than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively. Likewise, the statistical analysis on precision for all the datasets shows superior performance while comparing with the other existing approaches.

6.7 Statistical analysis on F1-score

The statistical analysis on F1-score of the proposed multimedia data retrieval model with A-SSF is given in Tables 4 and 5. The performance of the SM-DHOA is 7.5%, 8%, 6.7%, 7.6%, 4.4%, 5.5%, and 9.11% advanced than PSO, GWO,

Table 3 Statistical analysis of the proposed Multimedia Big Data Retrieval model with conventional models for four datasets in terms of precision

Algorithms	Best	Worst	Mean	Median	Standard deviation
Text to image retrieval					
BD-FFT [2]	91	66	80.11	81	7.9923
FRS-KELM [4]	89	67	79.67	82.5	7.5866
RELIEF-MM [6]	90	65	81.12	83	8.4331
SM-DHOA	95	75	87.36	89	6.2998
Image to image retrieval					
BD-FFT [2]	88	65	77.27	77	0.34441
FRS-KELM [4]	91	65	77.952	77	0.35267
RELIEF-MM [6]	91	65	77.47	77	0.4413
SM-DHOA	89	66	76.988	77	0.3814
Video to video retrieval					
BD-FFT [2]	88	67	81.33	83	6.9892
FRS-KELM [4]	90	69	81.95	83.5	6.3044
RELIEF-MM [6]	90	69	81.93	84.5	6.6564
SM-DHOA	95	75	86.99	88	6.381
Audio to image retrieval					
BD-FFT [2]	91	67	81.3	83.5	7.8515
FRS-KELM [4]	88	69	80.16	81	6.3576
RELIEF-MM [6]	88	67	81.82	84.5	6.8466
SM-DHOA	95	75	87.15	89.5	6.6002

DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, while considering the audio to image retrieval. The median of the suggested SM-DHOA is 5.9%, 7.22%, 5.9%, 5.3%, 3.4%, 5.2%, and 7.8% higher than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for audio to image retrieval. Similarly, superior performance is attained by the proposed model than existing approaches.

6.8 Statistical analysis on recall

The performance of the implemented SM-DHOA for the multimedia data retrieval is validated in Tables 6 and 7. The efficiency of the recommended SM-DHOA is 7%, 6%, 7%, 7.3%, 7.7%, 6.9%, and 8.3% higher than PSO, GWO, DHOA, SMO, BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for an image to image retrieval. The median of the proposed SM-DHOA is 7.3%, 2.3%, 3.5%, 3.5%, 7.8%, 7%, and 7.8% enhanced than PSO, GWO, DHOA, SMO,

Table 4 Statistical analysis of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for four datasets in terms of F1-score

Algorithms	Best	Worst	Mean	Median	Standard deviation
Text to image retrieval					
PSO [28]	91	66	80.84	82.5	7.9184
GWO [29]	90	69	80.87	83	6.8515
DHOA [25]	89	65	80.93	84	7.1085
SMO [26]	91	69	82	83.5	6.849
SM-DHOA	95	75	87.7	89	6.2418
Image to image retrieval					
PSO [28]	90	65	80.25	82	8.2846
GWO [29]	91	65	82.04	85	8.262
DHOA [25]	88	67	81.13	82.5	7.0132
SMO [26]	88	66	80.06	82	7.4952
SM-DHOA	95	76	87.26	88.5	6.3891
Video to video retrieval					
PSO [28]	91	70	82.25	82	7.1157
GWO [29]	88	69	80.8	82	6.1496
DHOA [25]	88	66	80.74	83.5	7.3356
SMO [26]	88	68	80.37	83	6.6858
SM-DHOA	95	75	87.23	90	6.865
Audio to image retrieval					
PSO [28]	89	68	81.61	84	6.5473
GWO [29]	89	67	81.25	83	6.8878
DHOA [25]	91	70	82.23	84	6.1952
SMO [26]	91	65	81.58	84.5	7.9419
SM-DHOA	95	76	87.81	89	6.1013

BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for the image to image retrieval. Therefore, the performance of the designed model is validated in terms of recall while compared with the other approaches.

7 Conclusion

This paper has developed a new deep multimedia big data retrieval with A-SSF based on hybrid SM-DHOA. The gathered multimedia data were given to the deep CNN-based semantic feature extraction for getting the significant features. Then, the optimal semantic features were selected using the SM-DHOA to minimize the correlation among

Table 5 Statistical analysis of the proposed Multimedia Big Data Retrieval model with different conventional models for four datasets in terms of F1-score

Algorithms	Best	Worst	Mean	Median	Standard deviation
Text to image retrieval					
BD-FFT [2]	91	66	82.03	85	7.744
FRS-KELM [4]	88	66	80.64	84	7.2995
RELIEF-MM [6]	89	69	81.23	83	6.1511
SM-DHOA	95	75	87.36	89	6.2998
Image to image retrieval					
BD-FFT [2]	89	68	82.27	83.5	6.7806
FRS-KELM [4]	88	65	79.96	83.5	7.3566
RELIEF-MM [6]	89	69	82.3	84	6.4424
SM-DHOA	95	75	87.24	89.5	6.4934
Video to video retrieval					
BD-FFT [2]	91	65	81.2	85	7.8341
FRS-KELM [4]	88	67	81.71	84.5	7.1426
RELIEF-MM [6]	88	68	81.12	85	7.1042
SM-DHOA	95	75	86.99	88	6.381
Audio to image retrieval					
BD-FFT [2]	90	71	83.43	86.5	6.0574
FRS-KELM [4]	91	69	82.58	85	6.4841
RELIEF-MM [6]	89	65	79.87	83	8.4467
SM-DHOA	95	75	87.15	89.5	6.6002

the optimal features. Furthermore, the map-reduce framework with A-SSF was developed using SM-DHOA for an efficient retrieval process. Through the performance analysis, while considering dataset 1, the accuracy of the developed SM-DHOA was 4.3%, 3.2%, and 1% enhanced than BD-FFT, FRS-KELM, and RELIEF-MM, respectively, for the video to video retrieval. Hence, superior performance was observed by the proposed model while compared to the conventional algorithms. In our proposed SM-DHOA model, results were found only for text-to-image, image-to-image, video-to-video, and audio-to-image. In future works, the proposed model extends towards the utilization of audio-to-video or audio-to-image-to-image processing by utilizing intelligent approaches like the ensemble learning method.

Table 6 Statistical analysis of the proposed Multimedia Big Data Retrieval model with different meta-heuristic-based algorithms for four datasets in terms of recall

Algorithms	Best	Worst	Mean	Median	Standard deviation
Text to image retrieval					
PSO [28]	91	70	81.69	83.5	6.5547
GWO [29]	89	66	79.45	79	7.3751
DHOA [25]	91	69	81.44	81.5	7.3955
SMO [26]	88	69	81.1	83.5	6.274
SM-DHOA	95	75	87.36	89	6.1242
Image to image retrieval					
PSO [28]	91	67	81.82	82	7.5617
GWO [29]	91	68	82.59	86	7.0268
DHOA [25]	90	66	81.72	85	7.3045
SMO [26]	88	67	81.61	85	6.2018
SM-DHOA	95	76	87.57	88	5.8968
Video to video retrieval					
PSO [28]	91	65	80.02	82	9.2321
GWO [29]	90	70	82.3	83	6.5017
DHOA [25]	90	67	80.63	83	7.5954
SMO [26]	90	69	82.31	84	6.5052
SM-DHOA	95	75	86.72	89.5	6.8329
Audio to image retrieval					
PSO [28]	88	70	82.03	84	6.274
GWO [29]	91	67	81.94	84	7.0307
DHOA [25]	88	68	81.63	84	6.7475
SMO [26]	91	68	81.93	84	7.6081
SM-DHOA	95	75	88.01	90	6.2997

Table 7 Statistical analysis of the proposed Multimedia Big Data Retrieval model with different conventional models for four datasets in terms of recall

Algorithms	Best	Worst	Mean	Median	Standard deviation
Text to image retrieval					
BD-FFT [2]	91	68	82.41	84.5	6.7091
FRS-KELM [4]	90	68	82.62	84.5	6.5886
RELIEF-MM [6]	89	67	81.45	84	6.4829
SM-DHOA	95	75	87.36	89	6.2998
Image to image retrieval					
BD-FFT [2]	88	68	80.99	83	7.0875
FRS-KELM [4]	89	67	81.54	83.5	7.2925
RELIEF-MM [6]	89	67	80.53	83	6.7934
SM-DHOA	95	75	87.24	89.5	6.4934
Video to video retrieval					
BD-FFT [2]	88	67	81.33	83	6.9892
FRS-KELM [4]	90	69	81.95	83.5	6.3044
RELIEF-MM [6]	90	69	81.93	84.5	6.6564
SM-DHOA	95	75	86.99	88	6.381
Audio to image retrieval					
BD-FFT [2]	89	66	81.55	83.5	7.1243
FRS-KELM [4]	90	69	81.95	85	6.8819
RELIEF-MM [6]	88	66	80.2	82	7.1619
SM-DHOA	95	75	87.15	89.5	6.6002

References

- Guo, K., Pan, W., Mingming, Lu., Zhou, X., Ma, J.: An effective and economical architecture for semantic-based heterogeneous multimedia big data retrieval. *J. Syst. Softw.* **102**, 207–216 (2015)
- Ahmad, J., Muhammad, K., Lloret, J., Baik, S.W.: Efficient conversion of deep features to compact binary codes using Fourier decomposition for multimedia big data. *IEEE Trans. Industr. Inf.* **14**(7), 3205–3215 (2018)
- Xia, D., Miao, L., Fan, A.: A cross-modal multimedia retrieval method using depth correlation mining in big data environment. *Multimedia Tools Appl.* **79**, 1339–1354 (2020)
- Sun, Z., Huo, H., Huan, J., Vitter, J.S.: Feature reduction based on semantic similarity for graph classification. *Neurocomputing* **397**, 114–126 (2020)
- Fernandez-Beltran, R., Pla, F.: Prior-based probabilistic latent semantic analysis for multimedia retrieval. *Multimedia Tools Appl.* **77**, 16771–16793 (2018)
- Yilmaz, T., Yazici, A., Kitsuregawa, M.: RELIEF-MM: effective modality weighting for multimedia information retrieval. *Multimedia Syst.* **20**, 389–413 (2014)
- Guo, K., Liang, Z., Tang, Y., Chi, T.: SOR: an optimized semantic ontology retrieval algorithm for heterogeneous multimedia big data. *J. Comput. Sci.* **28**, 455–465 (2018)
- Liu, Q., Jin, L., Li, Z., Tang, J.: Multimedia retrieval by deep hashing with multilevel similarity learning. *J. Vis. Commun. Image Represent.* **59**, 150–158 (2019)
- Uma, R., Muneeswaran, K.: OMIR: ontology-based multimedia information retrieval system for web usage mining. *Cybernet. Syst.* **48**(4), 393–414 (2017)
- Giangreco, I., Schuldt, H.: ADAM pro: database support for big multimedia retrieval. *Datenbank Spektrum* **16**, 17–26 (2016)
- Zhou, P., Wang, K., Xu, J., Wu, D.: Differentially-private and trustworthy online social multimedia big data retrieval in edge computing. *IEEE Trans. Multimedia* **21**(3), 539–554 (2019)
- Wason, R., Jain, V., Narula, G.S., Balyan, A.: Deep understanding of 3-D multimedia information retrieval on social media: implications and challenges. *Iran J. Comput. Sci.* **2**, 101–111 (2019)
- Lee, C., Min, G., Chen, W.: Data mining and machine learning technologies for multimedia information retrieval and recommendation. *Multimedia Tools Appl.* **75**, 4845–4849 (2016)
- Cheruiyot, W., Tan, G.-Z., Musau, F., Mushi, J.C.: Query quality refinement in singular value decomposition to improve genetic algorithms for multimedia data retrieval. *Multimedia Syst.* **17**, 507–521 (2011)
- Maki, W.S., Krinsky, M., Muñoz, S.: An efficient method for estimating semantic similarity based on feature overlap: reliability and validity of semantic feature ratings. *Behav. Res. Methods* **38**, 153–157 (2006)
- Baâzaoui, A., Abderrahim, M., Barhoumi, W.: Dynamic distance learning for joint assessment of visual and semantic similarities

- within the framework of medical image retrieval. *Comput. Biol. Med.* **122**, 103833 (2020)
17. Sun, Z., et al.: Feature reduction based on semantic similarity for graph classification. *Neurocomputing* **397**, 114–126 (2020)
 18. Ashraf, R., Ahmed, M., Ahmad, U., et al.: MDCBIR-MF: multimedia data for content-based image retrieval by using multiple features. *Multimedia Tools Appl.* **79**, 8553–8579 (2020)
 19. Hu, C., Xu, Z., Liu, Y., Mei, L., Chen, L., Luo, X.: Semantic link network-based model for organizing multimedia big data. *IEEE Trans. Emerg. Top. Comput.* **2**(3), 376–387 (2014)
 20. Jayasena, K.P.N., Li, L., Xie, Q.: Multi-modal multimedia big data analyzing architecture and resource allocation on cloud platform. *Neurocomputing* **253**, 135–143 (2017)
 21. Zhou, Z., Zhao, L.: Cloud computing model for big data processing and performance optimization of multimedia communication. *Comput. Commun.* **160**, 326–332 (2020)
 22. Tang, J., Lin, J., Li, Z., Yang, J.: Discriminative deep quantization hashing for face image retrieval. *IEEE Trans. Neural Netw. Learn. Syst.* **29**, 6154–6162 (2018)
 23. Jin, L., Li, K., Hu, H., Qi, G.-J., Tang, J.: Semantic neighbor graph hashing for multimodal retrieval. *IEEE Trans. Image Process.* **27**, 1405–1417 (2018)
 24. Jin, L., Li, K., Li, Z., Xiao, F., Qi, G.-J., Tang, J.: Deep semantic-preserving ordinal hashing for cross-modal similarity search. *IEEE Trans. Neural Netw. Learn. Syst.* **30**(5), 1429–1440 (2019)
 25. Brammy, G., Praveena, S., Ninu Preetha, N.S., Ramya, R., Rajakumar, B. R. and Binu, D.: Deer Hunting Optimization Algorithm: a new nature-inspired meta-heuristic paradigm. *Comput. J.* (2019)
 26. Agrawal, V., Rastogi, R., Tiwari, D.C.: Spider Monkey Optimization: a survey. *Int. J. Syst. Assur. Eng. Manage.* **9**(4), 929–941 (2018)
 27. Putzu, L., Piras, L., Giacinto, G.: Convolutional neural networks for relevance feedback in content based image retrieval. *Multimedia Tools Appl.* **79**, 26995–27021 (2020)
 28. Alloui, T., Boussebough, I., Chaoui, A.: A Particle Swarm Optimization Algorithm for web information retrieval: a novel approach. *Int. J. Intell. Inf. Technol.* **11**(3), 1200–1216 (2015)
 29. Benyl Renita, D., Seldev Christopher, C.: Novel real time content based medical image retrieval scheme with GWO-SVM. *Multimedia Tools Appl.* **79**, 17227–17243 (2020)
 30. Alghamdi, A., Hammad, M., Ugail, H., et al.: Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. *Multimed Tools Appl.* (2020). <https://doi.org/10.1007/s11042-020-08769-x>
 31. Sedik, A., Iliyasu, A.M., El-Rahiem, B.A., Samea, M.E.A., Abdel-Raheem, A., Hammad, M., Peng, J., El-Samie, F.E.A., El-Latif, A.A.A.: Deploying machine and deep learning models for efficient data-augmented detection of COVID-19 infections. *Viruses* **12**(7), 769 (2020)
 32. Sedik, A., Tawalbeh, L., Hammad, M., El-Latif, A.A.A., El-Banby, G.M., Khalaf, A.A.M., El-Samie, F.E.A.: Deep learning modalities for biometric alteration detection in 5G networks-based secure smart cities. *IEEE Access* **9**, 94780–94788 (2021)
 33. Malipatil, S., Maheshwari, V., and Chandra, M.B.: Area optimization of CMOS full adder design using 3T XOR. In 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), pp. 192–194 (2020)

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