Research article

A novel approach for identification of zoonotic trypanosome utilizing deep metric learning and vector database-based image retrieval system

Veerayuth Kittichai [a], Weerachat Sompong [a], Morakot Kaewthamasorn [b], Thanyathep Sasisaowapak [c], Kaung Myat Naing [c], Teerawat Tongloy [c], Santhad Chuwongin [c], Suchansa Thanee [b], Siridech Boonsang [d,*]

[a] Faculty of Medicine, King Mongkut's Institute of Technology Ladkrabang, Thailand
[b] Veterinary Parasitology Research Unit, Faculty of Veterinary Science, Chulalongkorn University, Bangkok, Thailand
[c] College of Advanced Manufacturing Innovation, King Mongkut's Institute of Technology Ladkrabang, Thailand
[d] Department of Electrical Engineering, Faculty of Engineering, King Mongkut's Institute of Technology Ladkrabang, Thailand

A B S T R A C T

Trypanosomiasis, a significant health concern in South America, South Asia, and Southeast Asia, requires active surveys to effectively control the disease. To address this, we have developed a hybrid model that combines deep metric learning (DML) and image retrieval. This model is proficient at identifying Trypanosoma species in microscopic images of thin-blood film examinations. Utilizing the ResNet50 backbone neural network, a trained-model has demonstrated outstanding performance, achieving an accuracy exceeding 99.71 % and up to 96 % in recall. Acknowledging the necessity for automated tools in field scenarios, we demonstrated the potential of our model as an autonomous screening approach. This was achieved by using prevailing convolutional neural network (CNN) applications, and vector database based-images returned by the KNN algorithm. This achievement is primarily attributed to the implementation of the Triplet Margin Loss function as 98 % of precision. The robustness of the model demonstrated in five-fold cross-validation highlights the ResNet50 neural network, based on DML, as a state-of-the-art CNN model as AUC >98 %. The adoption of DML significantly improves the performance of the model, remaining unaffected by variations in the dataset and rendering it a useful tool for fieldwork studies. DML offers several advantages over conventional classification model to manage large-scale datasets with a high volume of classes, enhancing scalability. The model has the capacity to generalize to novel classes that were not encountered during training, proving particularly advantageous in scenarios where new classes may consistently emerge. It is also well suited for applications requiring precise recognition, especially in discriminating between closely related classes. Furthermore, the DML exhibits greater resilience to issues related to class imbalance, as it concentrates on learning distances or similarities, which are more tolerant to such imbalances. These contributions significantly make the effectiveness and practicality of DML model, particularly in in fieldwork research.

* Corresponding author.
   E-mail address: siridech.bo@kmitl.ac.th (S. Boonsang).

1. Introduction

Trypanosomiasis is a major public health concern globally, encompassing regions including South America, South Asia, and Southeast Asia [1]. The transmission occurs through the bites of blood-sucking insects like Tsetse flies and tabanids. Notably, the parasite has the capability to infect both animals, particularly livestock, and humans, resulting in zoonosis. Approximately 300,000 individuals are at risk, especially those with acquired immune deficiency [2]. Trypanosomiasis, also known as sleeping sickness, prevalent in Sub-Saharan Africa, is attributed by *Trypanosoma brucei.* In Latin America, Chagas disease is caused by *T. cruzi.* In other regions, various trypanosome parasite species like *T. lewisi*, *T. congolense*, and *T. evansi* have the potential to be lethal to both humans and animals. To complete their transmission life cycle, these blood-parasitic protozoa of the Kinetoplastida class necessitate different hosts, with a human host being a crucial component require different hosts, including a human host [3,4]. Recently, an individual with a deficiency in the apolipoprotein L1 (APOL1) gene was verified to have the infection, a confirmation established through polymerase chain reaction (PCR) and immunological techniques [5]. Healthy Vietnamese patients without a gene deficiency have also been identified with *T. evansi* infection [1]. The molecular technique, the gold standard method for detection, necessarily requires the involvement of skilled and experienced personnel. Furthermore, this technique typically involves multiple steps, leading to an extended turnaround time and requiring the acquisition of costly equipment.

Epidemiological surveillance serves as a useful approach for identifying reservoir animals with capacity to transmit the disease to humans residing in close proximity to infected animals. Regularly conducting active surveys in is essential to control the transmission of the disease. Microscopic examination is a commonly and widely acknowledged diagnostic method, but it requires an expert technician and exhibits low detection sensitivity. Furthermore, individual technicians' interpretations of microscopic examinations vary. Nevertheless, there are variations in the interpretations of microscopic examination. Nonetheless, microscopic observation remains early commonly used, straightforward, and a rapid screening method, particularly in remote areas lacking costly laboratory equipment. To assist stakeholders living in resource-limited settings, an innovative diagnostic device that can function as an automatic tool is required.

The integration of artificial intelligence with diagnostic devices, specifically computer-aided diagnostic (CADs), represents a promising approach that can serve as a simple tool to detect microscopic images derived from blood films. The observation of pathological changes at the microscopic level can significantly contribute to clinical decision-making. At present, machine learning (ML) and its subfield of algorithms, deep learning (DL), are employed for the diagnosis of malaria infection through the analysis of microscopic images [6–8]. Several neural network models, including the Single-Shot MultiBox Detector (SSD) model [6], adaptive color segmentation and tree-based decision classification [9], segmentation, feature extraction, and SVM classifier [10,11] have been implemented in the medical and veterinary fields. Deep convolutional neural networks (DCNN) and Faster Region-based Convolutional Neural Network (Faster R–CNN) have undergone successfully testing [12].

Current studies into the semantic distance of data points, employing deep metric learning (DML) within the context of the content-based image retrieval (CBIR) analysis, provide an alternative approach utilized to solve an object classification task [13,14]. Previous studies have demonstrated that employing dimensionality reduction and transfer learning techniques for classifying the age of malaria mosquitoes can achieve a high accuracy of 98 %. This implies the efficacy of understanding the population dynamics of the mosquito vector [13]. Utilization of the similar techniques as described earlier, Merchan and colleges (2023) deployed Siamese Neural Networks (SNNs) and Triplet Neural Networks (TNNs) backbones of DML models to differentiate between species within malaria mosquitoes from tick species [14]. The trained model proposed showed an accuracy exceeding 93 %, affirming the capability of the state-of-the-art-algorithm to learn extracted features and generalize the type of DML for application with our trypanosome datasets. As a result of the multidimensional feature vectors being extracted and learned by the DML models, data points with similar features and minimized distance were embedded in a neighboring space, resulting in distinct and compact clustering of positive samples [15]. In contrast to object classification that prediction result relies on class probabilities, the DML measures the true positive based the similarities between the query image and well-trained database. Conceptually, this technique allows deep learning approaches to surpass the classification by enhancing better result and quantifying similarities between two data points, thereby achieving improved results.

The CBIR, the embedding loss serves to map high-dimensional feature vectors onto low-dimensional manifolds based Euclidean distance loss. This approach gives a rapid and accurate retrieval system that effectively compares a query image to a well-trained database [16]. The essential phase in training the DML model relates to embedding feature vectors representing the class signature and mapping them into space. This process is a crucial step on optimizing model parameters using various losses, including *Triplet Loss*, *N-pair-mc*, *Lifted Struct*, and *Proxy-NCA*, respectively [17]. Although the slow convergence observed in pairwise or triplet-wise losses when dealing with a significant amount of data triplets, the PyTorch Triplet-Margin loss have demonstrated to be an appropriate alternative in studies on the DML [18,19]. The CBIR does not solely rely on the embedding space; the feature labels and optimization are also the main components. The CIBR has been applied in in diverse medical field studies, including the current task of analyzing COVID-19 [16], and it has the potential to be highly beneficial in identifying the distribution of Trypanosoma species in Thailand when compared to traditional (non-deep) metric learning methods [20]. In recent medical research, there is an intention to use the DML technique to assist technicians in visualizing and screening content-based histopathological images. This aims to assist in diagnosis patients' specimens by comparing them to the trained database [21,22]. The advantage of the CBIR lies in effectively reducing the time spent searching for taxonomic keys for guide the corrected answer, giving to a high recall yield of 84 % in top-1 (*Recall*@1) recommendations for interpreting histological tasks [23]. When faced with challenges in discriminating ambiguous data, CBIR techniques have the capabilities to give a considered return image for any query image. The image embedding plays a crucial role in CBIR, converting images from Euclidean to multidimensional representations within lower-dimensional manifolds. This transformation

enhances the accuracy and speed of potential retrieval systems [24]. Optimizing image features and labeling of images appear to be key strategies for image retrieval tasks. Although traditional-DML has been explored in the veterinary studies, there is a lack of documented progress in the image retrieval system, especially in applications involving microscopic images. Rarely has a DML-based CBIR system been developed for the study of Trypanosoma species.

Vector databases have found application in several domains, exemplified by diverse use cases. This includes their utilization in large language models (LLM) such as chatGPT, where they generate coherent text-based AI plugins for storing and retrieving information from the database to match given content [25]. Other domains encompass natural language processing (NLP), recommendation systems (RS), and computer vision (CV) [26]. Within the healthcare sector, the vector database assumes an essential role in analysis of genomic data, facilitating personalized and more tailored medical therapies that align with individual genetic makeup. This database also supports medical solutions in this context. Databases encapsulate the complexities of high-dimensional data using mathematical representations refers to as vector databases. Attribute data is obtained by applying an embedding function in feature extraction algorithms on raw microscopic images, thereby transforming unstructured data. The vectors, composed of numerical arrays that represent structured data, assist the model in understanding relationships by transforming them into a low-dimensional space. This capability enables the model to discern connections in the data with greater precision. The significance of the vector database lies in its capacity to facilitate a retrieval system with accurate similarity location and a unique search method based on the distances between two vectors.

The study purpose is to develop a state-of-the-art algorithm based the DML technique. The inference output is conducted through the image retrieval process, aiming to discriminate medically and veterinary important three species of Trypanosome parasites. Our research contributions are as follows.

i.) developed an outstanding DML model for discriminating zoonotic Trypanosoma species; specifically, *T. cruzi*, *T. brucei* and *T. evansi*, utilizing microscopic images obtained from thin-blood film examinations obtaining for archived blood samples.

ii.) with the proposed models, incorporating optimized neural network backbones and metric loss during training conditions, it becomes feasible to correctly identify trypanosome amidst other parasites; like *Leishmania* spp., *Toxoplasma* spp., *Trichomonas* spp., respectively. The proposed model is capable of simultaneously learning the environment and background (including erythrocyte and leukocyte) within an image frame, enabling the discrimination of the parasite of interest from these objects.

iii.) our proposed model showed superior performance in identifying all studied classes and appears to be effective in preventing overfitting. to the use of the five-fold cross-validation method yielded comparable performance to the testing dataset. It can be inferred that the best selected DML neural network model as the exhibit's robustness, showing no significant impact from dataset variation and giving generalization across various datasets, even those collected from public resources.

In the future, this study may contribute the potential to assist public health professionals working with human and animal populations in easily identifying infections caused by unicellular parasitic flagellates. Furthermore, it could prove beneficial in designing strategies for the prevention and treatment of such disease.

2. Materials and methods

2.1. Ethical statement

Archived blood films from both human and animal blood samples, were collected and prepared. The research design received approval from the Institutional Biosafety Committee and the Institutional Animal Care and Use Committee of Faculty of Veterinary Science, Chulalongkorn University (IBC No. 2031033 and IACUC No. 1931027). Also, the research was endorsed by the Human Research Ethic Committee of King Mongkut's Institute of Technology Ladkrabang (EC-KMITL_66_014).

2.2. Dataset

Twelve slides, positively confirming blood-parasite infections, including *T. brucei*, *T. cruzi* and *T. evansi*, were verified by parasitologist experts based on a standard key to characteristics and morphology. The Giemsa stained-thin blood films, mentioned earlier, were photographed under oil-immersion field using a light microscope with digital camera (Olympus + DP21) at Faculty of Medicine, King Mongkut's Institute of Technology Ladkrabang. These images resulted in a total of 2904 pictures containing multiple objects of the parasite stages. Late/mature stage images of both *T. brucei* and *T. cruzi* was obtained through microscopic examination, characterized by a slender shape, long-tails with posterior ends, undulating membranes, and kinetoplasts at anterior end. Nevertheless, various forms of *T. evansi* were collected in comparison to a previous report [27]. Consequently, we categorized the blood stages of the parasite into three stages: a round form, stumpy form, and late stages, respectively. The late stages consisted of intermediate forms (19–24 μm) and slender forms (24–37 μm), exhibiting observed monomorphic forms containing nucleus and kinetoplasts, small flagella and undulating membranes. The stumpy form was smaller in size than the late stage but lacked a flagellum (15–18 μm). The round stage, comprising amastigote and sphaeromastigote stages, typically appeared a circular shape with 1–2 dense nuclei and vacuoles.

With the exception of above data, all the remaining five public datasets were acquired. These include datasets for *Leishmania* spp., *Toxoplasma* spp., *Trichomonas* spp., as well as datasets for Erythrocytes, and Leukocytes, respectively (Table 1; Fig. 1a) [28–30]. These data were sourced from three locations, including url: https://github.com/senli2018/DCTL, https://data.mendeley.com/datasets/

38jtn4nzs6/3, https://github.com/jiangdat/GFS-ExtremeNet. Altogether, 32,276 images containing a single object per an individual image were gathered. The export-crop module was utilized to ensure each image represented a single-parasite, employing the object detection function of the YOLOv4-tiny neural network model within the in-house CIRA CORE platform (https://git.cira-lab.com/cira/cira-core). The application of YOLOv4-tiny-based object detection is proposed for cropping a single blood stage parasite from those present in a microscopic image. Among various YOLO algorithms, the YOLOv4-tiny model outperforms others in terms of accurate localization and classification [8,31], as demonstrated by its high balance accuracy [32–34]. Specifically, the inference capability of this model reaches up to 155 frames per second during processing. The model achieves faster inference speeds compared to its larger counterparts like YOLOv4. In term of accuracy, the "tiny" models sacrifice some level of accuracy in order to achieve faster inference speeds [35]. The trade-off here is accepting a modest decrease in precision in exchange for quicker processing. Additionally, the YOLO model version is capable of learning up to 9000 multi-object classes across the entire potential area with minimal false positive [36]. In this study, all 10 classes were selected and divided into training/validation (26808 images, 83 %) and testing (5468 images, 17 %) sets.

The YOLO models mentioned above have been made more efficient by employing data augmentation techniques. These techniques aim to enhance the model's robustness against the unforeseen diversity present in the image dataset, which might otherwise affect the model's performance. The introduced data augmentation in this study involves rotational angles, incremented by 20-degrees from a minimum of $-180°$ to a maximum of $180°$. Additionally, contrast and brightness adjustments were applied at every 0.2 step, ranging from ±25 % increment between 0.4 and 1.2. Gaussian noise injection was introduced in this dataset, using four different standard deviation values ($\sigma = 0, 10, 20, 30$). Furthermore, a 10-step blur condition was implemented. Augmentation techniques play a crucial role in enhancing the diversity and robustness of microscopic image datasets, which frequently encounter the limited variations in imaging conditions. In this study, various transformations are operated to the original images to address the mentioned challenges. These transformations are carefully chosen based on specific criteria as part of augmentation techniques utilized in the study.

(i) **Increased variability**: The introduction of variations through rotational angles and adjustments in brightness and contrast adds diversity to the dataset. This diversity aids the model's ability to generalize effectively across a spectrum of imaging conditions observed in microscopic samples.

(ii) **Robustness to Gaussian noise**: Annotation can be operated to simulate real-world noise and distortions that may be found in microscopic images. This approach is valuable as it strengthens the model's robust to variations introduced by imperfect imaging conditions, improving its effectiveness on diverse datasets.

(iii) **Reducing of overfitting**: by augmenting the dataset, the model is forced to learn more generalized features, reducing the likelihood of memorizing specific data. This technique aids in mitigating overfitting, thereby enhancing the model's performance on a diverse range of unseen testing image.

The augmentation of data has been shown to enhance the YOLO model's generalization capabilities and reduce overfitting, resulting in improved performance metrics across various stages of parasites [35,37]. The utilization of augmented data contributes to improve precision by providing a more diverse training dataset. This diversity enables the model to better discrimination between different object classes, consequently reducing false positives. Moreover, augmented data boosts recall by exposing the model to a broader spectrum of object variations, enabling it to detect more accurately and reducing instances of false negatives. Lastly, the positive impact on accuracy is evident as augmented data enables the model to recognize objects under diverse conditions, thereby encouragement more robust and reliable predictions. In summary, augmented data plays a role in enhancing generalization, curbing overfitting, and facilitating superior object detection and classification in the YOLO model.

For the training and validation of the models, they were executed on a deep learning framework using an Nvidia RTX2070 GPU. The batch size was set at 32 with a subdivision of 8. The batch size of 32 was chosen to achieve a compromise between computing

Table 1
Sources of dataset.

Class	Train & validation	Test	Total	Source	URL link
Erythrocytes	4400	1497	5897	Public	https://github.com/senli2018/DCTL & https://data.mendeley.com/datasets/38jtn4nzs6/3
Late stage of *T. brucei*	4000	895	4895	Ours	–
Late stage of *T. cruzi*	248	28	276	Ours	–
Late stage of *T. evansi*	4500	818	5318	Ours	–
Leishmania spp.	2150	239	2389	Public	https://github.com/jiangdat/GFS-ExtremeNet & https://data.mendeley.com/datasets/38jtn4nzs6/3
Leukocytes	410	46	456	Public	https://data.mendeley.com/datasets/38jtn4nzs6/3
Round shape of *T. evansi*	2300	352	2652	Ours	–
Stumpy form of *T. evansi*	300	33	333	Ours	–
Toxoplasma spp.	4000	547	4547	Public	https://github.com/senli2018/DCTL & https://data.mendeley.com/datasets/38jtn4nzs6/3 & https://github.com/jiangdat/GFS-ExtremeNet
Trichomonas spp.	4500	1013	5513	Public	https://data.mendeley.com/datasets/38jtn4nzs6/3
Total (83 %: 17 %)	**26808**	**5468**	**32276**	–	–
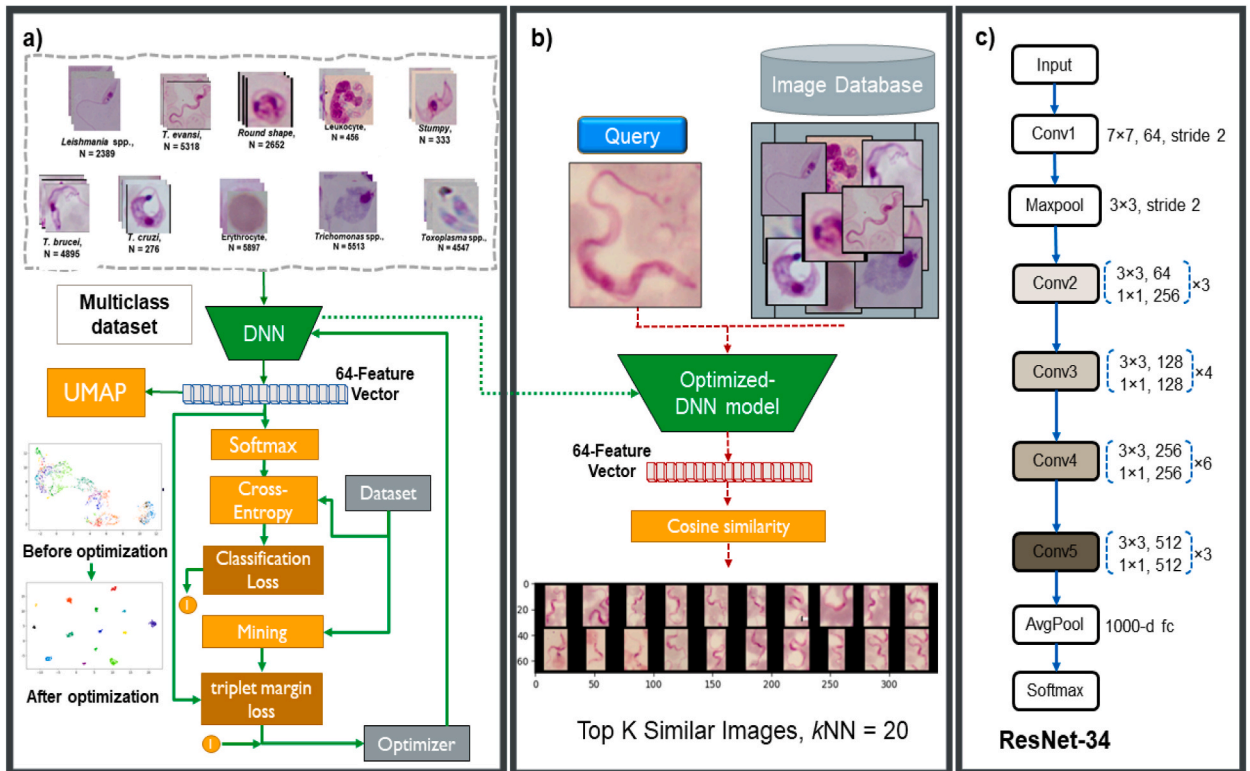
Fig. 1. Architecture. a) Ten-classes of unstructured image data were employed into the DML algorithm. The transformed data, represented as vectors, were observed by embedding them into a vector space. The clustering analysis of these vectors were observed by using UMAP. b) The optimized deep convolutional neural network (DCNN) model was used for inference in an unseen testing process. This involved using a batch of twenty image-retrieval based on a single query image.

efficiency and accuracy of the models. The learning rate was adjusted to 0.001 based on the trained weight reaching optimal accuracy. The training process involved 1000 burn-ins allowed the model to move past initial fluctuations and settle into a learning regime. The number of training steps of 400000, 450000 was set to ensure that the models have seen the data and refined their parameters, as well as scale adjustments of 0.1 and 0.1. The model operated with a momentum of 0.9 and decay of 0.0005. Momentum helps the model navigate through shallow valleys in the loss landscape, while decay prevents the weights from growing excessively large. These values are common choices for these parameters. Subsequently, the performance of the model was evaluated to assist in selecting the most suitable model for further applications.

2.3. Architecture of deep metric learning (DML) approach

We employed the DML technique to learn a specifically extracted feature, namely the vector representation, aiming to improve the performance of our model, which was previously hindered by the pixel-wise classification task. A well-trained neural network model can be visualized, grouping all data-point clusters based on their feature similarities, utilizing an embedded space loss function. In this clustering analysis, positive samples will belong to the same cluster, while negative samples will be in a separate cluster. The vectors embedded into the two-dimensional space were visualized using a general non-linear dimensional reduction with Uniform Manifold Approximation and Projection (UMAP) for Dimension Reduction representation (Figs. 1a and 3). The quality performance of the trained model mentioned above relies on the content-based image retrieval (CBIR) process (Figs. 1b and 2a). The CBIR was implemented to search for similar images to the query one from a pre-built vector database. The inference output of the CBIR would return the twenty most similar images to the mentioned query image based on the K-nearest neighbor (KNN) function at 20 (Kittichai et al., 2022, submitted and see the plot in supplementary materials). Among these, all returned images described above have the most similar and closer distance, optimized by Euclidean distance. The interpretation of true positive results was evaluated by counting the proportion of correctly predicted returned images (true positive) from the twenty-images.

2.4. CNN algorithms and compared model training

CNNs are a class of deep neural networks, most commonly applied to analyzing visual imagery. They are designed to automatically and adaptively learn spatial hierarchies of features from input images. Implementing CNN in DML would typically involve using a deep
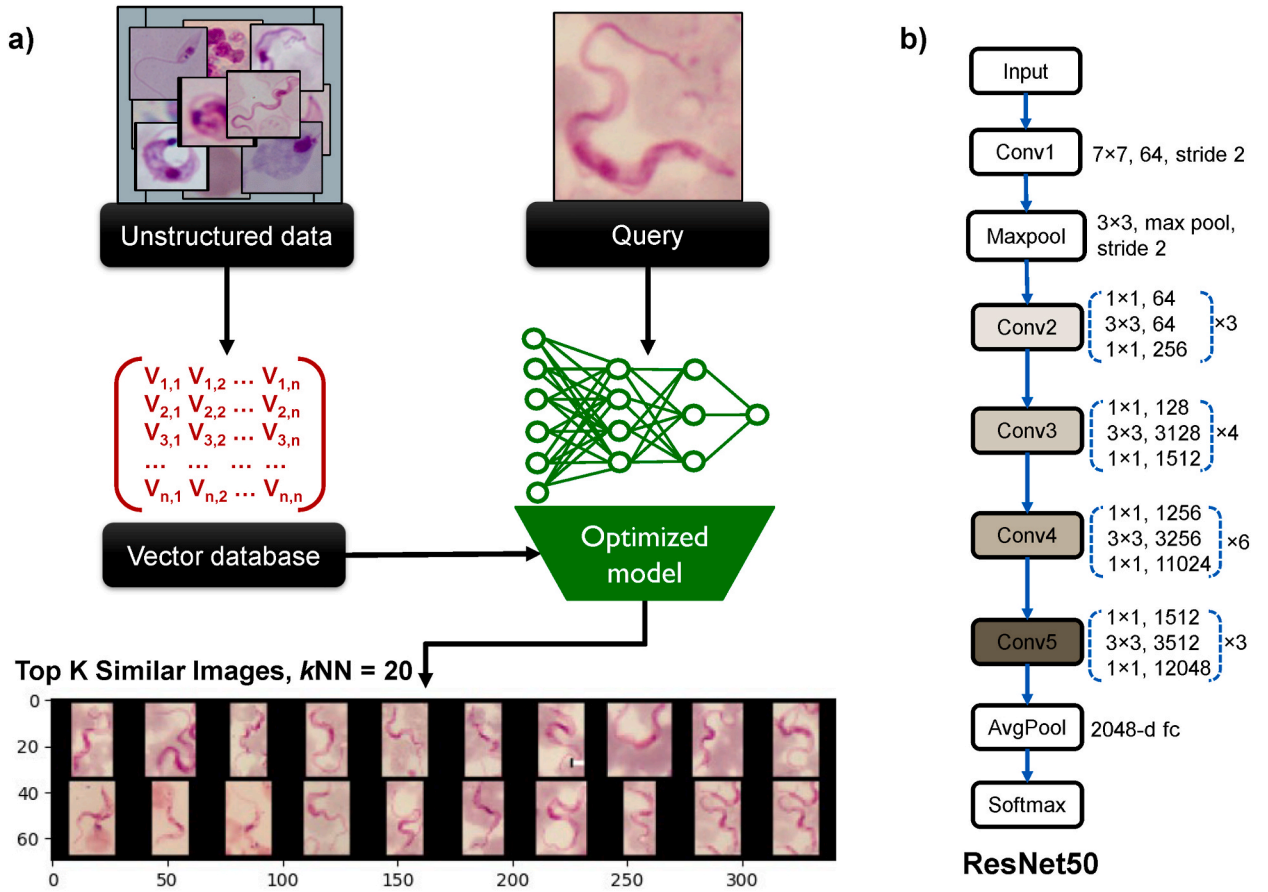
Fig. 2. Vector database and CNN backbones. a) a search and retrieval method from stored vector database, and b) the utilization of the DCNN algorithm along with the optimization of AI model based on the best deep neural network algorithm, ResNet50. Resnet50, as outlined by Farag et al. (2021), incorporating network architectures with 50 parameter layers and 3.8 billion FLOPS, utilizing residual connections to increase dimensions within the architecture [38].

learning framework like TensorFlow or PyTorch [39], which provide high-level APIs for building neural networks.

CNNs consist of multiple layers, including convolutional layers, pooling layers, and fully connected layers. Convolutional layers apply convolution operations to input images, extracting local features using filters or kernels. Pooling layers downsample the feature maps obtained from convolutional layers, reducing the computational complexity. Finally, fully connected layers process the high-level features extracted by previous layers to make predictions.

The performance of different DCNN backbones, including ResNet18, ResNet34, ResNet50 and ResNeXt50 models, was compared in this study (Fig. 3a-d). The selection of ResNet was motivated by its ability to overcome the vanishing gradients problem in many visual recognition tasks [40]. Although various architectures and the number of residual neural network layers appear to be correlated with training accuracy, they also lead to larger memory consumption during model training. Consequently, this study aimed to find the most suitable model that can effectively learn from our dataset while achieving optimal performance. The ResNet50 was ultimately chosen as the best-performing neural network, providing the highest accuracy (Fig. 2b).

2.5. Metric losses and classification loss

The selection of an appropriate metric loss function is crucial for achieving the classification task with the metric learning approach. Previous study has proposed models that achieve this by utilizing distances, regularizes, and reducers [24], similar to what is employed used in this study. The metric loss is employed to measure the vector-based distance between the training dataset plotted into embedded space. Two types of loss support our training phase to attain the optimized leaning model, and these include the metric loss and classification loss as outlined below. Five metric losses were used for embedding losses such as Contrastive loss, Triplet Margin loss, ArcFace loss, Proxy NCA loss, and MultiSimilarity loss. We applied the Cross-Entropy loss for classification loss and the best selected embedding loss of Triplet Margin loss as follows:

Cross-Entropy Loss: This loss function aids in measuring the error of datapoints and is commonly applied in classification tasks. The formula for Cross-Entropy loss in this work, specifically for a multi-class classification problem, is defined in equation (1) as follows:
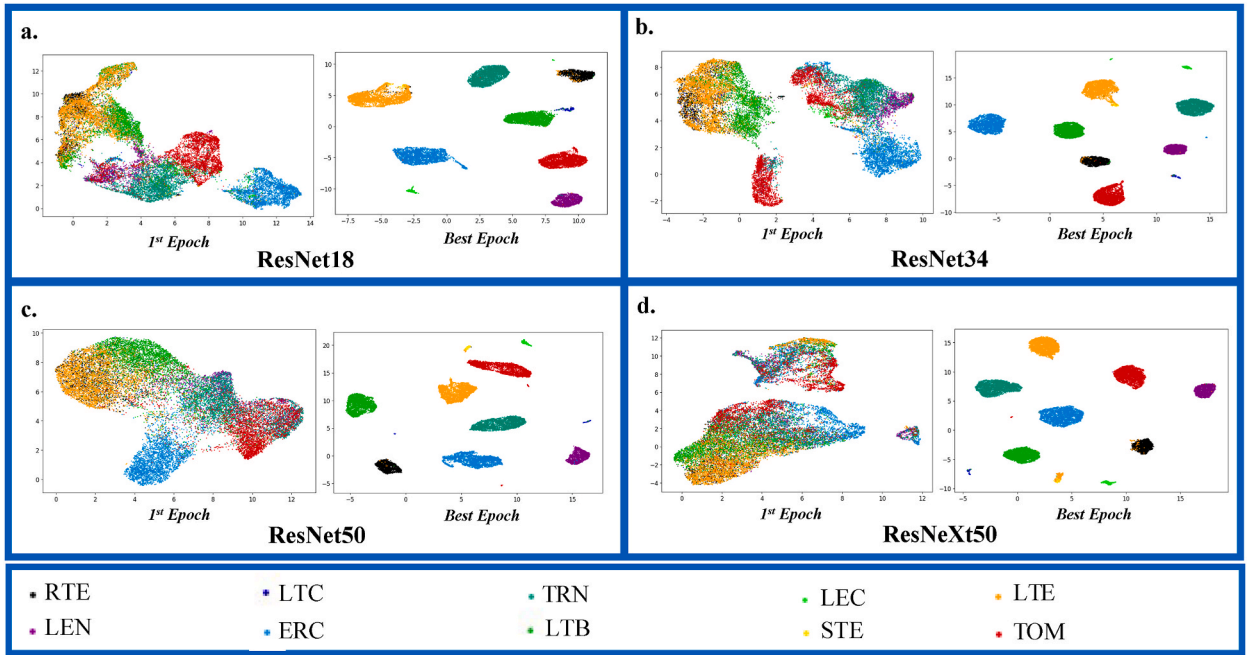
Fig. 3. UMAP. The plots assigned A to D illustrated the clustering analysis of each individual datapoint at best epoch, contrasting with the 1st epoch where no clustering was observed. The comparison encompassed all four DCNN backbones, namely ResNet18, ResNet34, ResNet50 and ResNeXt50. The class name abbreviation include ERC (Erythrocytes), LTB (late stage of *T. brucei*), LTC (late stage for *T. cruzi*), LTE (late stage of *T.evansi*), LEN (Leishmania spp.), LEC (Leukocytes), RTE (round shape of *T. evansi*), STE (stumpy of *T. evansi*), TOM (Toxoplasma spp.), and TRN (Trichomonas spp.).

$$L_{CE} = -\sum_{i=1}^{n} t_i \log p_i \qquad (1)$$

where n is number of classes, t_i is true probability of class i and p_i is predicted probability of that class by well-trained neural network models.

Triplet Margin Loss: The loss was utilized to compute datapoints involving a positive-, a negative and an anchor sample [41]. This loss has enforced the distance of anchor-positive (d_{ap}) is smaller than the distance of anchor-positive pair (d_{an}) by a specified margin. The computation of the Triplet Margin loss can be formulated in equation (2) as follows:

$$L_{triplet} = [d_{ap} - d_{an} + m]_+ \qquad (2)$$

In this equation, the desired difference (dap) and (dan) is controlled by margin parameter (*m*), which was set to 0.1 as default value in this study.

2.6. Vector database

In this study, the vector database serves the purpose of identifying images similar to a given image/vector by assessing their proximity in terms of features, visual content (Fig. 2a). The query vector can be generated either from the same type of data as the stored vectors, such as using an image as a query for an image database, or from different types of data, like using a query image for a text database. The distance between two vectors is measured based on the cosine similarity metric, indicating how closely the vectors are positioned in the vector space. Consequently, the outcome of the similarity search and retrieval is a commonly ranked list of vectors based on k-nearest neighborhood (KNN), presenting the vectors with the lowest distance and highest similarity values to the query vector. This process allows for access to corresponding raw data related to each vector from the original source.

The capabilities of the vector database have the potential to bring about a revolution in numerous domains when integrated with advanced deep learning models. This integration enhances scalability and adaptability, provides multi-user support and ensures data privacy, improves the speed and precision of the search method, and offers user-friendly interfaces for applications.

Cosine Similarity: Cosine similarity is a measure of similarity between two non-zero vectors of an embedding space that measures the cosine of the angle between them [24]. It measures the cosine of the angle between two vectors projected in a multi-dimensional space, as described in equation (3).

$$cosine\_similarity(a, b) = \frac{a.b}{\|a\|\|b\|} \tag{3}$$

Where a·b denotes the dot product of vectors a and b. ‖a‖ denotes the Euclidean norm (magnitude) of vector a. ‖b‖ denotes the Euclidean norm (magnitude) of vector b. This formula calculates the cosine of the angle between the two vectors, normalized by their magnitudes, thus providing a measure of similarity between the vectors. The cosine similarity ranges from -1 (perfectly dissimilar) to 1 (perfectly similar), with 0 indicating no similarity.

K-Nearest Neighbors (KNN): KNN is a simple and intuitive classification algorithm that stores all available database and query image based on a similarity measure (e.g., distance functions) to the k nearest neighbors [42]. The formula for distance calculation depends on the distance metric chosen, commonly Euclidean distance or Manhattan distance. Here's the Euclidean distance formula between two data points x_i and x_j, as described in equation (4):

$$EuclideanDistance(x_i, x_j) = \sqrt{\sum_{l=1}^{n}(x_{i,l} - x_{j,l})^2} \tag{4}$$

Where n is the number of features (dimensions) in the data points. $x_{i,l}$ and $x_{j,l}$ are the l-th feature values of data points x_i and x_j, respectively. To classify a new data point, KNN calculates the distance (typically Euclidean distance) between that point and all other points in the database. It then selects the k nearest points (neighbors) and assigns the majority class among them to the new data point.

2.7. Inference

The evaluation of the proposed model's capabilities included known-image retrieval and clustering against query images. In line within this, the inference process utilized the metric parameters learned by the model during the training phase to propose the model's performance. This involved employing an assigned batch of returned images with K-nearest neighbor (KNN) function from a prebuilt vector database, featuring an optimized model. The vector database served as the image source, incorporating the loaded-trained model and the match finder function to identify similar images to the query image. This was achieved by computing the pairwise distances based on the Cosine Similarity loss function, with a specified threshold set at 0.7 during the testing phase. The inference, based KNN, proved advantageous for similarity searches using the chosen distance metric and was established on the PyTorch library.

2.8. Statistical metrics analysis

The performance of all proposed models will be assessed using various statistical metrics, including a confusion matrix table, precision, recall, accuracy, F1-Score, and specificity [8]. These evaluation metrics were obtained and analyzed from the confusion matrix tables, and the formulas for these metrics are as described in equations (5)–(9) as follows:

$$Reacall = \frac{TP}{TP + FN} \tag{5}$$

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

$$F1\ score = 2 \times \frac{Recall \times Precision}{Recall + Precision} \tag{7}$$

$$Specificity = \frac{TN}{TN + FP} \tag{8}$$

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{9}$$

where TP represents the total number of true positives, TN is the total number of true negatives, FP denotes the total number of false positives, and FN is for the total number of false negatives classifications.

The additional statistics could address the overfitting issue when using distinct independent datasets by minimizing the cross-validated error of the optimal model. The comparisons were studied using the Receiver Operating Characteristic (ROC) curve. The area under the curve (AUC) involved measuring data-folds to determine if a higher AUC value indicates superior model performances.

3. Results

3.1. Model-wise comparison

Primary objective is to identify the best proposed model, utilizing deep metric learning combined with the predictive inference through the CBIR process, as illustrated in Fig. 1. The evaluation of the predictive results encompasses five-evaluation metrics: accuracy, recall, precision, F1 score and specificity. On average, the trained model of using the ResNet50 neural network showed

outstanding performance with an accuracy of 0.997, recall of 0.966, precision of 0.935, F1 score of 0.949 and specificity of 0.998, respectively as shown in Table 2. The finding of the result was supported by a prior publication [43], which employed ResNet-based DML and achieve noteworthy performance. When considering the classification based on the relative performance scores, the ResNet50 network exhibited predominantly scores, indicating higher evaluated metrics. Although the number of the true positive (TP) values of all four neural networks were comparable, the ResNet50 model showed a greater number of TPs in most multiclass scenarios, such as Erythrocytes (TP_{1495}/$total_{1497}$), *T. cruzi* (TP_{27}/$total_{28}$), Round shape (TP_{338}/$total_{352}$) and Stumpy stages (TP_{28}/$total_{33}$) of *T. evansi*, respectively (Suppl. Tables S1–S4). Interestingly, the general accuracy of the trained DML model based the ResNet50 neural network surpassed 99.71 % (Suppl. Table S3).

Overall, the area under the ROC curve exhibited comparable values for all four models, ranging from 0.990 to 0.992, suggesting no significant difference in their performance. However, when evaluating the Stumpy class of *T. evansi*, the AUC values varied, with 0.786 for ResNet18, 0.848 for ResNet34, 0.908 for ResNeXt50 and 0.924 for ResNet50, and support the DML-ResNet50 model as out-performing the others (Fig. 4.; Suppl. Table S5).

3.2. Metric loss comparison

This section aimed to study the optimal metric loss, an important role in minimizing the distant similarity between datapoints during the dimensional reduction process. The predictive result based on five-metric losses are presented in Fig. 4 and Supplementary Table S5. Upon comparing the predictive result, the model that outperformed with the Triplet Margin Loss exhibited an accuracy of 0.997, recall of 0.985, precision of 0.985, F1 score of 0.985 and specificity of 0.998, respectively (Fig. 5, Suppl. Fig. S1).

Consistently, the evaluation metric values per class revealed that the similar pattern for accuracy and specificity. However, two classes, late stage and round shape of *T. evansi*, exhibited varying evaluation metric values. In addition, another pattern set of patterns emerged for recall, precision and F1 score, particularly in the evaluation values of late stage, and stumpy of *T. evansi*. These two patterns suggest less predictive results of true positive values; however, they have no impact on the overall accuracy (Suppl. Figs. S2–6).

### 3.3. K-fold cross validation

In this section we aimed to measure the robustness of the best-selected DML neural network model using the five-fold cross-validation technique. The result showed inconsistent variation in the AUC by class, particularly for the stumpy stage of *T. evansi*: K1: AUC = 0.878, K2: AUC = 0.924, K3: AUC = 0.893, K4: AUC = 0.878 and K5: AUC = 0.916, respectively (Fig. 6a-e). Nevertheless, the area under the ROC curve showed similar values across all experiments: K1:

AUC = 0.989, K2: AUC = 0.990, K3: AUC = 0.991, K4: AUC = 0.990 and K5: AUC = 0.990, respectively (Fig. 6a-e; Suppl. Table S6). This suggests that our proposed DML-based ResNet50 neural network is robustness state-of-art CNN and is not affected by dataset variation, as designed by the cross-validation technique. The research results are supported by previous publication, indicating that employing the five-fold cross validation helps verify whether the chosen model consistently provides general accuracy and minimizes

Table 2
Quality performance of trained DML approaches. All four models such as ResNet18, ResNet34, ResNet50 and ResNeXt50, were studied to find their performance. Five metrics, namely accuracy, recall, precision, F1 score, specificity, were employed to evaluate model's performances. The class name abbreviation include ERC (Erythrocytes), LTB (late stage of *T. brucei*), LTC (late stage for *T. cruzi*), LTE (late stage of *T.evansi*), LEN (Leishmania spp.), LEC (Leukocytes), RTE (round shape of *T. evansi*), STE (stumpy of *T. evansi*), TOM (Toxoplasma spp.), and TRN (Trichomonas spp.). Bold numbers indicate the highest averaged scores for each metric, aiding in the determination of an outstanding model.

Models	Metric	ERC	LTB	LTC	LTE	LEN	LEC	RTE	STE	TOM	TRN	Average
ResNet18	*Accuracy*	1.000	0.998	0.999	0.986	0.997	0.999	0.991	0.994	0.999	0.998	0.996
	Recall	0.999	0.993	0.893	0.943	0.975	0.978	0.943	0.576	0.993	0.996	0.929
	Precision	1.000	0.994	0.862	0.966	0.967	0.957	0.917	0.487	0.993	0.993	0.914
	F1 score	0.999	0.994	0.877	0.954	0.971	0.968	0.930	0.528	0.993	0.995	0.921
	Specificity	1.000	0.999	0.999	0.994	0.998	1.000	0.994	0.996	0.999	0.998	**0.998**
ResNet34	*Accuracy*	0.999	0.998	0.998	0.988	0.998	0.999	0.992	0.995	0.998	0.999	0.996
	Recall	0.997	0.994	0.929	0.954	0.979	0.957	0.935	0.697	0.991	0.997	0.943
	Precision	1.000	0.993	0.722	0.968	0.967	0.978	0.940	0.605	0.991	0.995	0.916
	F1 score	0.999	0.994	0.813	0.961	0.973	0.967	0.937	0.648	0.991	0.996	0.928
	Specificity	1.000	0.999	0.998	0.994	0.998	1.000	0.996	0.997	0.999	0.999	**0.998**
ResNet50	*Accuracy*	1.000	0.999	0.999	0.989	0.998	1.000	0.993	0.996	0.999	0.999	**0.997**
	Recall	0.999	0.993	0.964	0.950	0.975	0.978	0.960	0.848	0.997	0.997	**0.966**
	Precision	1.000	0.998	0.844	0.977	0.975	1.000	0.926	0.636	0.995	0.997	**0.935**
	F1 score	0.999	0.996	0.900	0.963	0.975	0.989	0.943	0.727	0.996	0.997	**0.949**
	Specificity	1.000	1.000	0.999	0.996	0.999	1.000	0.995	0.997	0.999	0.999	**0.998**
ResNeXt50	*Accuracy*	0.999	0.998	0.998	0.988	0.999	1.000	0.992	0.996	0.999	1.000	0.997
	Recall	0.997	0.993	0.893	0.954	0.992	0.978	0.946	0.818	0.996	1.000	0.957
	Precision	0.999	0.997	0.806	0.968	0.996	0.978	0.935	0.643	0.993	0.999	0.931
	F1 score	0.998	0.995	0.847	0.961	0.994	0.978	0.941	0.720	0.995	1.000	0.943
	Specificity	1.000	0.999	0.999	0.994	1.000	1.000	0.996	0.997	0.999	1.000	**0.998**
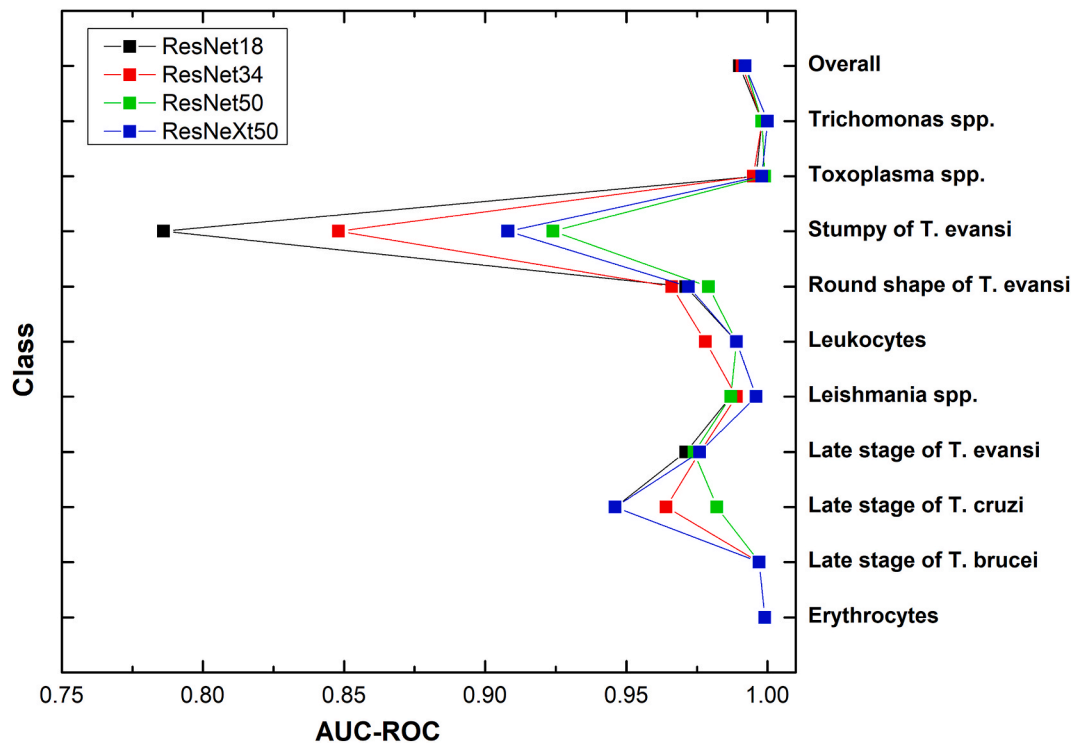
Fig. 4. AUC values for neural network model. This study was conducted to examine the comparison of models by assessing averaged values of the area under the ROC curve (AUC-ROC) for each individual class labels. Models are comprised of ResNet18, ResNet34, ResNet50, and ResNeXt50.

predictive bias [31].

4. Discussion

A novel hybrid approach of deep metric learning and inference-based image-retrieval process has been proposed for automatic screening of microscopic images containing medically and veterinary important blood parasites. This method aims to efficiently identify parasite features in Giemsa staining-blood films, distinguishing them from both a single-red blood cell and white blood cell images. A well-trained model exhibiting superior performance could be employed to discriminate parasite characteristics. The utilization of DML in automatically identifying Trypanosoma species yields findings line with prior research. One such study employed DML to distinguish between mosquito vector species in Thailand [44]. Another investigation focused on classifying lung nodules, employing a two-step content-based image retrieval approach that utilized texture features and a learned distance metric [20]. The findings suggest that employing the suitable AI technique has the potential to assist clinicians in decision-making. Traditional classification methods, relying on pixel identification, are limit in addressing classification challenges, especially in scenarios involving high variation such as medical and biological data. In the light of this, a new deep learning technique could be developed to fix the mentioned classification problem.

A remarkable proposed DML model demonstrates exceptional capability in distinguishing zoonotic Trypanosoma species from other infectious parasites and blood cells, achieving an accuracy surpassing 99.71 %. This outperforms previous studies that employed hybrid deep learning models to identify these three species of Trypanosome parasites [31]. The reason for this discrepancy is the utilization of a larger dataset compared to those studies. Nevertheless, the accuracy achieved in the proposed research is similar to that reported previously, which involved the development of a random forest algorithm for the automatic detection and enumeration of *T. cruzi* trypomastigotes in blood smear samples using mobile phone images [45,46]. The study findings indicate that the model achieved a high recall of 96 % and precision of 98 %, surpassing those of a previous publication which reported a recall of 90.5 % and precision of 87.5 % [45]. Despite Moras et al. (2022) developing the model using mobile phone images captured with a specific camera, the study does not address the applicability of the proposed method to images obtained with various mobile phone cameras or lower resolutions. A recent study employed pre-trained models such as MobileNet-v2, Xception, EfficientNet-b0, ResNet-101, and DenseNet-201. Evaluation of these five deep-learning models demonstrated a high degree of Matthews Correlation Coefficient (MCC) (>0.980) and accuracy (>0.990) for the classification of cutaneous leishmaniasis [47]. Despite the outstanding performance exhibited by these models, the sample size studied remains limited.

Previously, the CBIR system based on DML model was proposed, presenting a robust learning approach to support classical classification task [22–24,48,49]. To address classification challenges, fine-tuning depends on the kNN method, enhancing predictions
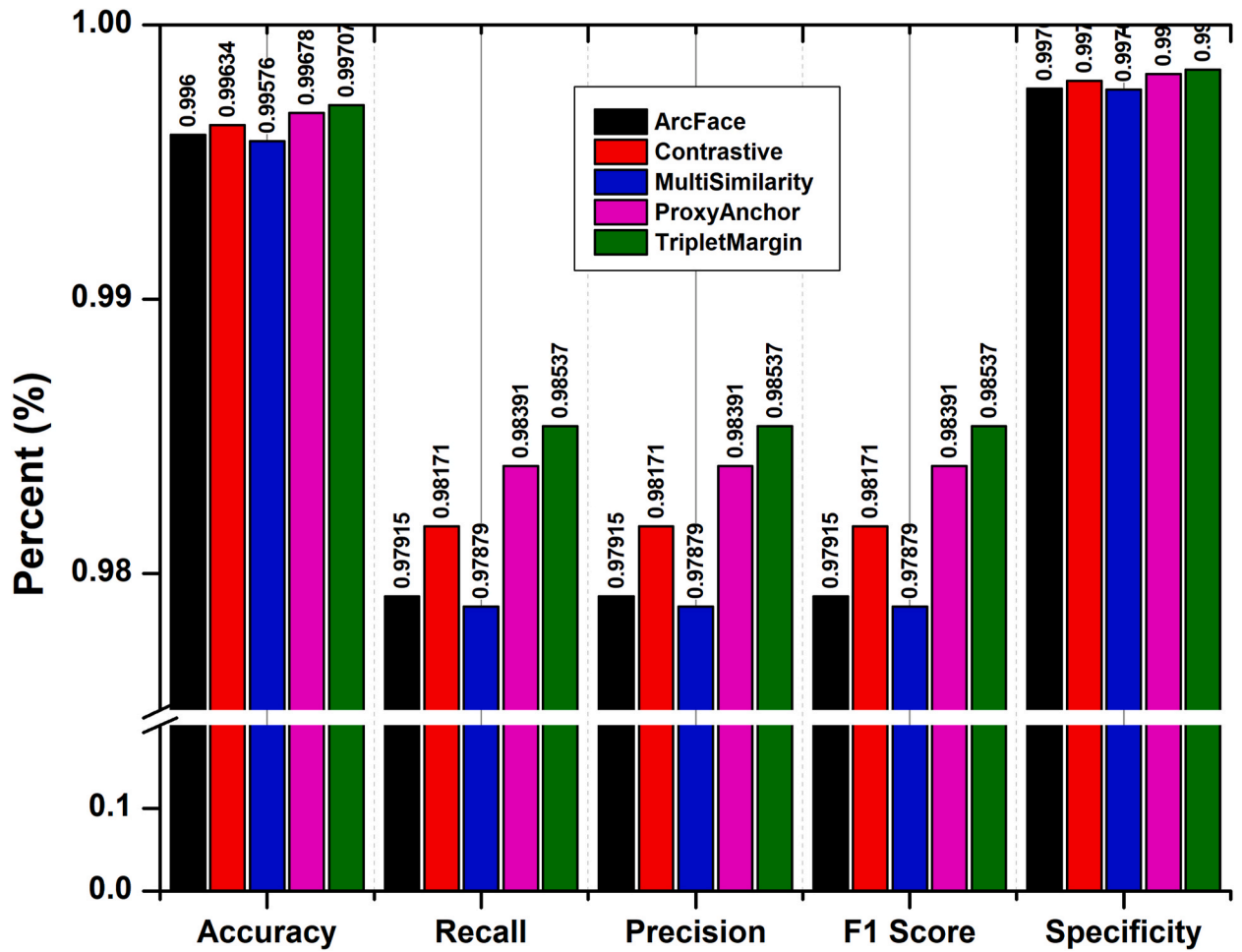
Fig. 5. Evaluation metrics for comparison of metric losses. There are including ArcFace, Contrastive, Mullti Similarity, Proxy Anchor and Triplet Margin, respectively.

based on distant similarity in feature vector information for each testing data, even in conditions of high variation. In contrast to our findings, the earlier publication does not address any limitations concerning the performance or generalizability of the K-Nearest Neighbor (KNN) classifier employed in their framework [43]. Additionally, the requirement for fine-tuning in the CBIR protocol may arise, considering that the number of returned images influences the evaluation metrics [43,50]. Although the kNN value of 10 resulted in the average AUC reaching its peak when comparing to the other values [20], it is noteworthy that employing kNN = 20 in our study yielded comparable results to those reported by Kittichai et al. [44] and represent a soft-voting approach. Nevertheless, it is important to overcome a significant limitation in the CBIR work must be overcome. Determining a suitable kNN is contingent on the specific environment and variation within the biological dataset, leading to instances where approaches fail to search the nearest image of the actual class from the prebuilt vector database as novelty and advantage of the study.

The utilization of deep learning method provides an efficient strategy to enhance analytical processes for extracting semantic and content-discriminated biological features obtained from unseen testing data. Hence, the CBIR system could be highly beneficial for medical and veterinary professionals engaged in the challenging task of screening blood parasites under microscopic observation, especially when dealing with large and complex datasets. As a result, the primary contribution of this study lies advocating the robustness of the state-of-art neural network models for the classification of various classes of blood parasites, encompassing zoonotic species. This study demonstrates a remarkable DCNN model, which may help support in preventing overfitting to biological data sources. The effectiveness of the model was validated through a five-fold cross-validation method. We applied the suggested model to accurately identify an individual label for ten classes, which include six parasites and two types of blood cells. This was achieved by employing the triplet-margin loss, implemented to minimize the errors while reducing the dimensions of the feature vectors through the contrastive learning challenge. Considering research's strength of the study, the developed model is capable of reliably predicting and discriminating multiple-class labels within testing sets, even in the presence of a high degree of variation in biological blood cells. The trustworthy in computer algorithms, which extract feature vectors, can remarkably assist human observers when confronted with semantic disease-related characteristics, such as local lesions in blood film samples. Moreover, limitations in human visual perception
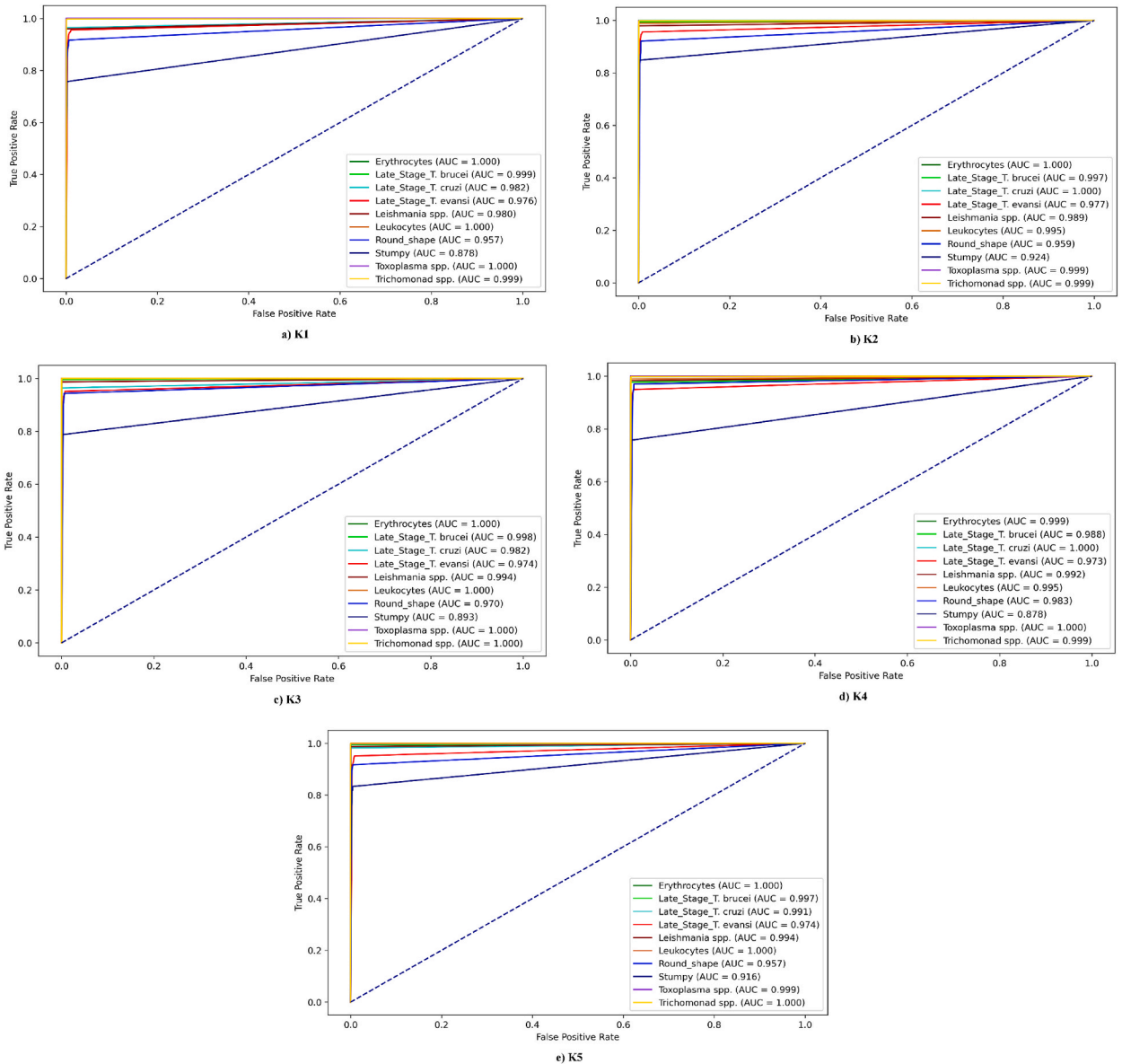
Fig. 6. ROC curve by data-folds. The comparison of area under the ROC curve (AUC) obtained from five-fold cross-validation using the ResNet50 neural network is illustrated in panels a) to e) corresponding to K1 – K5, respectively.

may be observed, particularly in relation to the global shape and position of the parasite in the image. Although addressing the challenge of interpreting the classification task involved balancing the classes, the CBIR system appears to have no impact on prediction. This is because the system solely requires an augmentation condition to regulate the similarities between the visual and semantic gap. It is crucial to acknowledge the limitation concerning the diversity and representativeness of the dataset, particularly in relation to the three classes of Trypanosome parasites, in order to ensure the generalizability of the proposed methodology.

Previous publications showed a high level of performance in identifying datasets related to the one under study (Supplementary Table S8). However, numerous challenges remain to be addressed and improved upon.

1. Limited dataset size and absence of diversity can impact the quality of data, levels of expertise, and imbalance among classes [31, 47]. Morais et al. (2022) concentrated on identifying and counting *T. cruzi* trypomastigotes, neglecting the detection of other parasite stages or different types of parasites that could be present in blood smear samples [45]. This challenge has the potential to impact the model's performance during the training process, potentially resulting in problems such as overfitting and underfitting.

2. The research employed thin blood film images, potentially constraining the generalization of the results to alternative diagnostic techniques or sample diversities [31]. Moreover, the significant color variability within the images diminished the network's performance [46].

3. Furthermore, the study utilized mobile phone images captured using a particular camera with a resolution of 12 megapixels. The extent to which the proposed method can be generalized to images obtained with alternative mobile phone cameras or lower resolutions is not discussed [45].

Addressing these research gaps will require a concerted effort to curate diverse and representative datasets, validate the proposed methods across a range of diagnostic scenarios, and develop robust models that can handle variations in imaging devices and resolutions. The initial two challenges can be solved through augmentation techniques aimed at enlarging the sample size, enhancing data quality, and diversifying color variations, thereby reducing the risks of underfitting and overfitting. This approach facilitates generalization with data obtained from various sources. To address the challenge of probability-based classification, particularly when subtle differences exist between images, a Deep Metric Learning (DML) model based on image retrieval processes can be employed. DML aims to learn a compact and discriminative feature representation that captures the similarity between images. By training the model to minimize the distance between features of similar images and maximize the distance between features of dissimilar images, DML enables the model to distinguish between subtle differences and handle probability-based classification more effectively. In an image retrieval-based approach, the model is trained to map images to a low-dimensional embedding space where similar images are clustered together, and dissimilar images are separated. During inference, a query image is mapped to the embedding space, and its nearest neighbors are retrieved based on the learned metric. This approach allows for more flexible and interpretable classification, as the model can provide a ranked list of similar images rather than a single probability score. Furthermore, the integration of multiple data sources, such as stereomicroscope and mobile phone images, has shown improvements in the performance of the trained DML model. Sensitivity and precision levels surpassing 90 % and 95 %, respectively, were achieved. This incorporation also enhances clustering analysis and provides a framework for predicting region-specific mosquito species [44].

This study introduced a practical CBIR system along with a deep contrastive learning algorithm to aid the classification of zoonotic parasite infections, a technology that has not been previously explored in the realms of medical and veterinary fields. We demonstrated state-of-the-art hybrid models through the combination of the YOLOv4-tiny object detection with deep metric learning algorithms and the inference process using the k-nearest neighbor function. This approach was to find the most proportionate class as the true positive. While the prediction results give a high degree of false negatives for three actual classes, specifically *T. evansi*, addressing this concern could be clarified through further optimization. This may involve increasing the quantity of feature vectors to incorporate additional feature attribute for each class label. Consequently, the positive matching verification process significantly increases the number of TP values through the proposed image retrieval method. The potential for interpreting FN arises when they are obtained with low kNN values, leading to the identification of FP values through confusion matrix tables. In a prior publication, a novel False Negative Elimination (FNE) strategy for image-text matching was introduced, effectively addressing false negatives in triplet loss optimization. This involved a cut-down strategy focusing on hard negatives by reducing the sampling probability of easy negatives [51]. Nevertheless, in this study, the FN issue was resolved by increasing the kNN value up to 20, thereby increasing the TP value (Suppl. Fig. S7). The aforementioned evaluation metrics mentioned as above reveal a higher value for the predictive result of the proposed neural network, suggesting it support for cytopathologists in decision-making tasks. Nonetheless, our research has demonstrated the impact of increasing the number of feature vectors, ranging from the initially examined 64 to higher values such as 128, 256, and 512. ResNet50 algorithms were employed to determine the optimal conditions (Suppl. Fig. S8). The outcomes revealed comparable average accuracy and specificity values (Suppl. Table S7). Notably, with 256 feature vectors, the model exhibited enhanced performance, with an average recall of 96.74 %, precision of 96.74 %, and an F1 score of 96.64 %. Despite these findings indicating that our model tends to be optimized with 256 feature vectors, all averaged metric values show no significant difference when compared to using 64 feature vectors.

5. Conclusions

A highly state-of-the-art algorithm utilizing the DML technique has been methodically developed. The inference output is achieved by employing the image retrieval process, relying on the vector database as novel as an innovative approach to concept learning. The effectiveness of the proposed model was evaluated for its ability to discriminate three medically and veterinary important species of Trypanosome parasites. In addressing the current medical and veterinary challenges posed by trypanosomiasis, an illness caused by a trypanosome parasite, there is a crucial requirement for an automated and innovative tool. This tool is essential to overcome the limitations associated with conventional microscopic methods, including issues such as inter- and intra-rater variation, particularly in facilitating active surveillance. Suggestions from this study underscore the efficacy of employing contemporary applications of convolutional neural network (CNN) and K-nearest neighbors (KNN) for the automated screening process in field settings.

Data availability

Data will be made available on request.

CRediT authorship contribution statement

Veerayuth Kittichai: Writing – original draft, Supervision, Project administration, Methodology, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Weerachat Sompong:** Data curation. **Morakot Kaewthamasorn:** Writing – review & editing, Resources, Investigation. **Thanyathep Sasisaowapak:** Validation, Project administration, Formal analysis, Data curation. **Kaung Myat Naing:** Validation, Formal analysis, Data curation. **Teerawat Tongloy:** Software. **Santhad Chuwongin:** Validation, Software. **Suchansa Thanee:** Resources, Investigation. **Siridech Boonsang:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2024.e30643.

References

[1] N. Van Vinh Chau, L. Buu Chau, M. Desquesnes, S. Herder, N. Phu Huong Lan, J.I. Campbell, et al., A clinical and Epidemiological investigation of the first reported human infection with the zoonotic parasite trypanosoma evansi in Southeast Asia, Clin. Infect. Dis. 62 (8) (2016) 1002–1008, https://doi.org/10.1093/cid/ciw052.

[2] S.J. Black, J.R. Seed, N.B. Murphy, Innate and acquired resistance to African trypanosomiasis, J. Parasitol. 87 (2001) 1–9, https://doi.org/10.1645/0022-3395(2001)087[0001:IAARTA]2.0.CO;2.

[3] M.P. Barrett, R.J. Burchmore, A. Stich, J.O. Lazzari, A.C. Frasch, J.J. Cazzulo, et al., The trypanosomiases, Lancet 362 (9394) (2003) 1469–1480, https://doi.org/10.1016/S0140-6736(03)14694-6.

[4] O. World Health, A new form of human trypanosomiasis in India. Description of the first human case in the world caused by Trypanosoma evansi, Wkly. Epidemiol. Rec. 80 (7) (2005) 62–63.

[5] P.P. Joshi, V.R. Shegokar, R.M. Powar, S. Herder, R. Katti, H.R. Salkar, et al., Human trypanosomiasis caused by Trypanosoma evansi in India: the first case report, Am. J. Trop. Med. Hyg. 73 (3) (2005) 491–495.

[6] C. Marzahl, M. Aubreville, C.A. Bertram, J. Stayt, A.K. Jasensky, F. Bartenschlager, et al., Deep learning-based quantification of pulmonary hemosiderophages in cytology slides, Sci. Rep. 10 (1) (2020) 9795, https://doi.org/10.1038/s41598-020-65958-2.

[7] M. Aubreville, C.A. Bertram, C. Marzahl, C. Gurtner, M. Dettwiler, A. Schmidt, et al., Deep learning algorithms out-perform veterinary pathologists in detecting the mitotically most active tumor region, Sci. Rep. 10 (1) (2020) 16447, https://doi.org/10.1038/s41598-020-73246-2.

[8] V. Kittichai, M. Kaewthamasorn, S. Thanee, R. Jomtarak, K. Klanboot, K.M. Naing, et al., Classification for avian malaria parasite Plasmodium gallinaceum blood stages by using deep convolutional neural networks, Sci. Rep. 11 (1) (2021) 16919, https://doi.org/10.1038/s41598-021-96475-5.

[9] Adi Pamungkas, Adi Kusworo, R. Gernowo, Identification of plasmodium falciparum development phase in malaria infected red blood cells using adaptive color segmentation and decision tree based classification *international Journal of applied engineering research* 10 (2) (2015) 4043–4055.

[10] M.P. Wongsakorn Preedanan, Wongwit Senavongse, Suchada Tantisatirapong, Automated detection of plasmodium falciparum from giemsa-stained thin blood films, IEEE (2016) 215–218.

[11] G. Gopakumar, M. Swetha, G.S. Siva, G.R.K.S. Subrahmanyam, Automatic detection of Malaria infected RBCs from a focus stack of bright field microscope slide images. ICVGIP '16, in: Proceedings of the Tenth Indian Conference on Computer Vision, Graphics and Image Processing, 2016, pp. 1–7.

[12] A.C. Jane Hung, Applying faster R-CNN for object detection on malaria images, arXiv:180409548v2 (2019) 56–61.

[13] E.P. Mwanga, D.J. Siria, J. Mitton, I.H. Mshani, M. Gonzalez-Jimenez, P. Selvaraj, et al., Using transfer learning and dimensionality reduction techniques to improve generalisability of machine-learning predictions of mosquito ages from mid-infrared spectra, BMC Bioinf. 24 (1) (2023) 11, https://doi.org/10.1186/s12859-022-05128-5.

[14] Fernando Merchan, Kenji Contreras, Rolando A. Gittens, Jose R. Loaiza, J.E. Sanchez-Galan, Deep metric learning for the classification of MALDI-TOF spectral signatures from multiple species of neotropical disease vectors, Artificial Intelligence in the Life Sciences 3 (2023) 100071.

[15] H. Yang, E. Jo, H.J. Kim, I.H. Cha, Y.S. Jung, W. Nam, et al., Deep learning for automated detection of cyst and tumors of the jaw in panoramic radiographs, J. Clin. Med. 9 (6) (2020), https://doi.org/10.3390/jcm9061839.

[16] A.L.X. Zhong, D. Wu, H. Ren, K. Kim, Y. Kim, V. Buch, N. Neumark, B. Bizzo, W.Y. Tak, S.Y. Park, Y.R. Lee, M.K. Kang, J.G. Park, B.S. Kim, W.J. Chung, N. Guo, I. Dayan, M.K. Kalra, Q. Li, Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19, Med. Image Anal. (2021), https://doi.org/10.1016/j.media.2021.101993.

[17] X.H.Y. Wang, E. Kodirov, N.M. Robertson, Ranked list loss for deep metric learning, IEEE Trans. Pattern Anal. Mach. Intell. 44 (9) (2022), https://doi.org/10.1109/TPAMI.2021.3068449.

[18] Yingying Zhang, Qiaoyong Zhong, Liang Ma, S.P. di Xie, Learning incremental triplet margin for person Re-identification, in: Proceedings of the AAAI Conference on Artificial Intelligence, 2019, pp. 9243–9250.

[19] Z. Wang, T. L, Two-stage method based on triplet margin loss for pig face recognition, Comput. Electron. Agric. 194 (2022) 106737.

[20] G. Wei, H. Cao, H. Ma, S. Qi, W. Qian, Z. Ma, Content-based image retrieval for lung nodule classification using texture features and learned distance metric, J. Med. Syst. 42 (1) (2018) 13, https://doi.org/10.1007/s10916-017-0874-5.

[21] M.R. Reena, P.M. Ameer, A content-based image retrieval system for the diagnosis of lymphoma using blood micrographs: an incorporation of deep learning with a traditional learning approach, Comput. Biol. Med. 145 (2022) 105463, https://doi.org/10.1016/j.compbiomed.2022.105463.

[22] J. Fang, H. Fu, J. Liu, Deep triplet hashing network for case-based medical image retrieval, Med. Image Anal. 69 (2021) 101981, https://doi.org/10.1016/j.media.2021.101981.

[23] Zheng ZJ. Yushan, Haopeng Zhang, Fengying Xie Member, Yibing Ma, Huaqiang Shi, Yu Zhao, Histopathological whole slide image analysis using context-based CBIR, IEEE Trans. Med. Imag. 37 (7) (2018) 1641–1652, https://doi.org/10.1109/TMI.2018.2796130.

[24] A. Zhong, X. Li, D. Wu, H. Ren, K. Kim, Y. Kim, et al., Deep metric learning-based image retrieval system for chest radiograph and its clinical applications in COVID-19, Med. Image Anal. 70 (2021) 101993, https://doi.org/10.1016/j.media.2021.101993.

[25] L. De Angelis, F. Baglivo, G. Arzilli, G.P. Privitera, P. Ferragina, A.E. Tozzi, et al., ChatGPT and the rise of large language models: the new AI-driven infodemic threat in public health, Front. Public Health 11 (2023) 1166120, https://doi.org/10.3389/fpubh.2023.1166120.

[26] Yikun Han, Chunjiang Liu, P. Wang, A comprehensive survey on vector database: storage and retrieval technique, arXiv:231011703v1 [csDB], Challenge (2023) 1–13.

[27] K.K. Misra, S. Roy, C. A, Biology of Trypanosoma (Trypanozoon) evansi in experimental heterologous mammalian hosts, J. Parasit. Dis. 40 (3) (2016) 1047–1061, https://doi.org/10.1007/s12639-014-0633-1.

[28] S. Li, Q. Yang, H. Jiang, J.A. Cortes-Vecino, Y. Zhang, Parasitologist-level classification of apicomplexan parasites and host cell with deep cycle transfer learning (DCTL), Bioinformatics 36 (16) (2020) 4498–4505, https://doi.org/10.1093/bioinformatics/btaa513.

[29] S.Z. Li, Yang, Microscopic images of parasites species, Mendeley Data (2020), https://doi.org/10.17632/38jtn4nzs6.3.

[30] Hao Jiang, Parasite dataset, Medndeley Data (2021), https://doi.org/10.17632/m3jxgb54c9.4.

[31] V. Kittichai, M. Kaewthamasorn, S. Thanee, T. Sasisaowapak, K.M. Naing, R. Jomtarak, et al., Superior auto-identification of trypanosome parasites by using a hybrid deep-learning model, J. Vis. Exp. 200 (2023), https://doi.org/10.3791/65557.

[32] Nhat-Duy Nguyen, Tien Do, Thanh Duc Ngo, D.-D. Le, An evaluation of deep learning methods for small object detection, Journal of Electrical and Computer Engineering (2020), https://doi.org/10.1155/2020/3189691.

[33] A. Kumar, A. Kalia, A. Sharma, M. Kaushal, A hybrid tiny YOLO v4-SPP module based improved face mask detection vision system, J. Ambient Intell. Hum. Comput. (2021) 1–14, https://doi.org/10.1007/s12652-021-03541-x.

[34] C. Guo, X.L. Lv, Y. Zhang, M.L. Zhang, Improved YOLOv4-tiny network for real-time electronic component detection, Sci. Rep. 11 (1) (2021) 22744, https://doi.org/10.1038/s41598-021-02225-y.

[35] Zicong Jiang, Liquan Zhao, Shuaiyang Li, Y. Jia, Real-time object detection method based on improved YOLOv4-tiny, arXiv preprint arXiv:201104244 (2020), https://doi.org/10.48550/arXiv.2011.04244.

[36] Markov WD. Ilia, Improving cross-domain hate speech detection by reducing the false positive rate, in: Fourth Workshop on NLP for Internet Freedom: Censorship, Disinformation, and Propaganda (NLP4IF 2021), 2021, pp. 17–22.

[37] A. Bochkovskiy, C.-Y. Wang, H.-Y.M. Liao, Yolov4: optimal speed and accuracy of object detection, arXiv preprint arXiv:200410934 (2020).

[38] Hania H. Farag, Lamiaa A.A. Said, Mohamed R.M. Rizk, M.A.E. Ahmed, Hyperparameters optimization for ResNet and Xception in the purpose of diagnosing COVID-19, J. Intell. Fuzzy Syst. 41 (2021) 3555–3571, https://doi.org/10.3233/JIFS-210925.

[39] Musgrave Kevin, Serge Belongie, S.-N. Lim, PyTorch metric learning, arXiv:200809164 (2020).

[40] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.

[41] Kilian Q. Weinberger, John Blitzer, L.K. Saul, Distance metric learning for large margin nearest neighbor classification (NIPS 2005), Adv. Neural Inf. Process. Syst. 18 (2005).

[42] Kilian Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest neighbor classification, J. Mach. Learn. Res. 10 (2009) 207–244.

[43] A. Pal, Z. Xue, B. Befano, A.C. Rodriguez, L.R. Long, M. Schiffman, et al., Deep metric learning for cervical image classification, IEEE Access 9 (2021) 53266–53275, https://doi.org/10.1109/access.2021.3069346.

[44] V. Kittichai, M. Kaewthamasorn, Y. Samung, R. Jomtarak, K.M. Naing, T. Tongloy, et al., Automatic identification of medically important mosquitoes using embedded learning approach-based image-retrieval system, Sci. Rep. 13 (1) (2023) 10609, https://doi.org/10.1038/s41598-023-37574-3.

[45] M.C.C. Morais, D. Silva, M.M. Milagre, M.T. de Oliveira, T. Pereira, J.S. Silva, et al., Automatic detection of the parasite Trypanosoma cruzi in blood smears using a machine learning approach applied to mobile phone images, PeerJ 10 (2022) e13470, https://doi.org/10.7717/peerj.13470.

[46] N. Sanchez-Patino, A. Toriz-Vazquez, N. Hevia-Montiel, J. Perez-Gonzalez, Convolutional neural networks for Chagas' parasite detection in histopathological images, Annu Int Conf IEEE Eng Med Biol Soc 2021 (2021) 2732–2735, https://doi.org/10.1109/EMBC46164.2021.9629563.

[47] A.M. Abdelmula, O. Mirzaei, E. Guler, K. Suer, Assessment of deep learning models for cutaneous Leishmania parasite diagnosis using microscopic images, Diagnostics 14 (1) (2024), https://doi.org/10.3390/diagnostics14010012.

[48] M.R. Reena, A. Pm, A content-based image retrieval system for the diagnosis of lymphoma using blood micrographs: an incorporation of deep learning with a traditional learning approach, Comput. Biol. Med. 145 (2022) 105463.

[49] Y.Z. Pengshuai Yang, Li Lin, Hairong Lv, Jigang Wang, Chengzhan Zhu, Rui Jiang, Deep metric learning approach for histopathological image retrieval, Methods 179 (2020) 14–25, https://doi.org/10.1016/j.ymeth.2020.05.015.

[50] J.V. Sundgaard, J. Harte, P. Bray, S. Laugesen, Y. Kamide, C. Tanaka, et al., Deep metric learning for otitis media classification, Med. Image Anal. 71 (2021) 102034, https://doi.org/10.1016/j.media.2021.102034.

[51] H. Li, Y. Bin, J. Liao, Y. Yang, H.T. Shen, Your negative may not Be true negative: boosting image-text matching with false negative elimination, in: Proceedings of the 31st ACM International Conference on Multimedia, 2023, pp. 924–934.