XLVI CILAMCE
VITÓRIA - ES - BRAZIL - 2025
XLVI IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGENEERING

Content-based macroscopic microbial image retrieval: Preliminary results

Angela P. M. Muñante, Antonio R. S. Parmezan

Institute of Mathematics and Computer Science, University of São Paulo
Av. Trabalhador São-carlense 400, 13566-590, São Carlos, SP, Brazil
angimunantem@usp.br, parmezan@usp.br

Abstract. As image databases grow and evolve, the need for effective search engines becomes increasingly critical. While Content-Based Image Retrieval (CBIR) has advanced considerably across various domains, its application in microbiology remains limited—particularly from a macroscopic perspective, where pure culture images often appear visually similar and are difficult to distinguish. This study pioneers a preliminary investigation into the development of a CBIR framework for accessing images of microorganisms grown on solid media. Our proposal leverages deep neural networks to extract image features and applies Uniform Manifold Approximation and Projection (UMAP) to reduce the dimensionality of feature vectors without sacrificing performance. To enable fast and efficient retrieval, we implement a K-dimensional tree structure in combination with the *k*-nearest neighbors algorithm. We evaluate the framework's performance using a leave-one-out strategy, considering precision and recall, quantified by mean average precision, along with execution time. Experiments conducted on two original microbial image sets, as well as on their aggregated version, yielded promising results, with vision transformers outperforming other pre-trained models. Our findings highlight the potential of the developed CBIR system for accurately retrieving culture images and demonstrate its applicability in both clinical and microbiological research.

Keywords: Image search, Feature extraction, Dimensionality reduction, Multidimensional indexing, Information retrieval

1 Introduction

Multimedia content analysis plays a crucial role in various computer vision applications, with digital images representing a significant portion of this data. It is well known that the complexity of multimedia content, especially images, has increased significantly in recent years. In particular, millions of images are stored daily in databases across different domains [1–3].

Retrieving one or more relevant images in a database presents a challenging research problem for the artificial intelligence community [4]. The most common way to access information in these collections is through keywords, much like users routinely do in Internet search engines [5]. However, this approach presupposes the existence of consistent and semantically rich textual annotations associated with each image—a condition that is not always feasible or guaranteed. Moreover, keywords often fail to capture subtle and distinctive pictorial aspects, compromising the effectiveness of the query in contexts where visual similarity is essential.

To overcome the aforementioned limitations, Content-Based Image Retrieval (CBIR) emerges as a solution that directly exploits the visual characteristics of images, without relying on metadata [2, 4]. In this task, images—which may be associated with textual annotations—are sampled from a collection and described by pictorial features such as color, texture and shape, reflecting their inherent semantics. The similarity between two images is quantified by a distance function applied to their corresponding feature vectors. This measure allows similar images to cluster in nearby regions of the representation space, while visually distinct images are mapped to more distant areas. Indexing structures benefit from this arrangement to store similar images in physically close locations on the computer's disk, optimizing query efficiency [6].

Even so, the user's perception of relevance may not align with the clustering achieved by combining feature extractors and the similarity function, leading to a problem known as the semantic gap [5]. This discrepancy is further exacerbated by the fact that different users may judge the relevance of the same image differently for a particular query. Therefore, user participation in the retrieval process—e.g., through relevance feedback techniques—is essential to align the system with human perception [1, 7]. CBIR engines have been successfully developed in a

wide range of domains, including medicine, remote sensing, public security, and the textile industry, supporting tasks such as medical image analysis, video surveillance, criminal activity detection, and automated visual pattern inspection, thereby contributing to expert decision-making [2, 4].

Within microbiology, CBIR systems have great potential for supporting the identification of microorganisms. Although approximately 10 million species have already been cataloged, this denotes only 0.001% of the estimated number on the planet [8]. Many of these strains produce chemical substances with biotechnological and pharmacological applications [9]. Retrieving morphologically similar images can quickly provide information about the taxa of unknown microorganisms, supporting microbial biodiscovery and, consequently, aiding the inference of the compounds these organisms produce. Although some initial CBIR studies have explored microscopic images [10–14], the use of macroscopic records—such as those obtained through photographs of cultures grown in Petri dishes—offers additional advantages, such as reduced time, lower operational costs, and a decrease in the need for highly skilled labor, both during system development and in its deployment [9, 15, 16].

In this work, we conduct a pioneering investigation into the development of a CBIR framework for accessing images of microorganisms cultivated on solid media. Our proposal employs deep neural networks to extract image features and applies Uniform Manifold Approximation and Projection (UMAP) to reduce the dimensionality of the feature vectors. To enable fast and efficient retrieval, we implement a K-Dimensional Tree (KD-Tree) structure in combination with the k-Nearest Neighbors (kNN) algorithm. We evaluate the framework's performance using a leave-one-out strategy, considering precision and recall, measured by Mean Average Precision (MAP), along with execution time. Tests on two original microbial image sets, as well as on their aggregated version, yielded promising results, with vision transformers outperforming other pre-trained models.

Our research marks a first step toward building a prototype computational module for CBIR tailored to a little-explored domain, with the potential to accelerate the morphological screening of microorganisms and pave the way for more accessible and scalable solutions in microbial biodiscovery.

The remainder of this paper is structured as follows: Section 2 describes the content-based macroscopic microbial image retrieval framework; Section 3 reports the empirical evaluation of our proposal, detailing the experimental setup and discussing the results; and Section 4 summarizes our contributions and outlines directions for future research.

2 Proposed Framework

We organized the CBIR framework into six steps for didactic purposes, as illustrated in Figure 1. It aims to retrieve the k most similar images of cultures in Petri dishes, guided by the visual representation of a given query, thus allowing life science professionals to compare microorganism morphology and support their recognition.

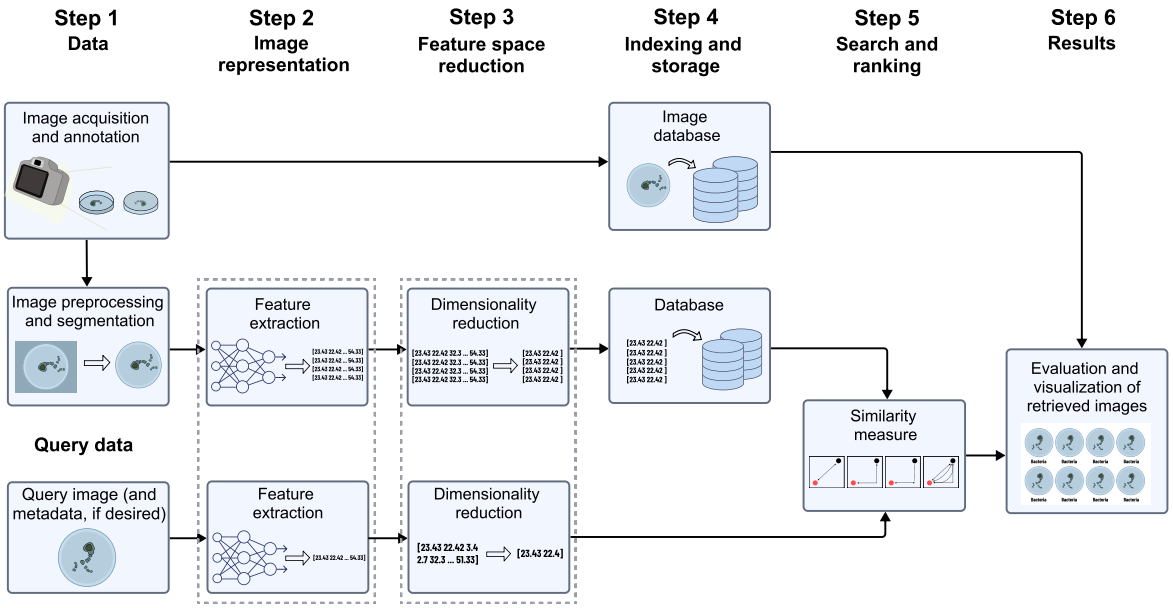
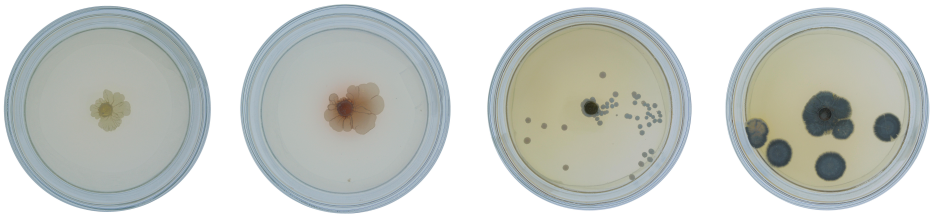


Figure 1. Proposed framework for content-based macroscopic microbial image retrieval.

In Step 1, we prepare the database and the queries of interest. Our research group has been developing the first image collection of microorganisms grown on solid media, with records captured and taxonomically annotated

according to guidelines designed to boost computer vision tasks [9]. Figure 2 shows front-view images of glass Petri dishes containing pure cultures of actinobacteria taken from the aforementioned database, which also includes back-view images for each sample.



(a) A18 (b) A18 (c) A54a (d) AM76

Figure 2. Pure cultures of actinobacteria grown on different solid media: A18 (*Amycolatopsis albidoflavus*), A54a (*Streptomyces lunalinharesii*), and AM76 (*Streptomyces venezuelae*). Front-view images of Petri dishes.

Parmezan et al. [9] details the protocol for cultivating microorganisms in triplicate, acquiring images with a customized low-cost device at equidistant time intervals throughout the microbial growth curve, performing taxonomic labeling following the Linnaean system, and applying preprocessing and segmentation procedures for distinct macroscopic image analysis purposes. In this work, we employ a processing routine that comprises global white balance correction based on the gray world assumption with a strength control value of 0.7, followed by background removal to preserve only the region corresponding to the Petri dish, as depicted in Figure 2.

In Step 2, we transform each image into a visual representation vector using handcrafted descriptors and/or attributes extracted from pre-trained deep learning models (off-the-shelf features). Defining an optimal set of descriptors for macroscopic images of microorganisms remains a challenging task. Therefore, reusing deep neural networks pre-trained on large-scale datasets under the transfer learning paradigm is a viable and effective strategy, even in specialized domains such as microbiology.

In Step 3, we apply a dimensionality reduction technique to compress the extracted representation vectors, improving the CBIR framework's computational efficiency while preserving the topological structure of the original feature space. We recommend UMAP [17] due to its superior ability to capture nonlinear relationships, which are often present in features derived from pre-trained deep neural networks. Although UMAP incurs higher computational costs and offers lower reproducibility compared with linear methods such as Principal Component Analysis (PCA), it proves more effective in scenarios that require robustness to morphological variation and the preservation of semantic structure.

In Step 4, we organize the reduced representation vectors into a data structure suitable for multidimensional indexing, which enables efficient search in the feature space. Several alternatives have been proposed in the literature, including KD-Trees, ball trees, and hashing-based methods [1, 4, 6]. The choice among these depends on factors such as representation vector dimensionality, dataset size, and the selected similarity function. Additional considerations contemplate query efficiency, index construction time, support for dynamic updates, and storage management, since each feature vector must be efficiently linked to its respective image as metadata for retrieval.

In Step 5, upon receiving a query image, the CBIR system replicates the procedures from Steps 2 and 3 to extract its representation vector. The resulting embedding is then compared against the indexed data through a similarity measure. While the optimal choice of metric may vary depending on the application domain [18], Euclidean and cosine distances have demonstrated consistent effectiveness in visual retrieval tasks.

In Step 6, our framework returns the k most similar records to one or more query images. We estimate the effectiveness of the CBIR system by employing data resampling techniques and specific performance measures, the latter typically based on taxonomic annotations, to quantify retrieval quality. When necessary, domain experts may be involved in evaluating the results to assess the degree of correspondence between the retrieved images and the queries, thereby helping to interpret the remaining semantic gap.

3 Empirical Evaluation

We assessed the effectiveness and efficiency of the proposed CBIR framework through a series of experiments using an image collection of actinobacteria cultures. Our goal was to investigate how different off-the-shelf features, representation vector dimensionality, and indexing configurations influence retrieval performance. We structured the evaluation around three research questions: (i) How do distinct pre-trained deep learning models behave as macroscopic microbial image descriptors? (ii) How does dimensionality affect retrieval accuracy and

query time? (iii) How consistent is the search engine across different datasets and microorganism classes? The following sections outline the empirical protocol and present the results.

3.1 Experimental Setup

The empirical protocol followed the steps portrayed in Figure 1. In accordance with Parmezan et al. [9], we cultivated 24 actinobacterial strains in ISP2 medium. Starting on the third day post-inoculation, we photographed each Petri dish every 24 hours over eight consecutive days—from the emergence of the first colony to its decline. We captured the Petri dishes' front and back sides, for a total of 48 images per species.

We obtained two datasets, ISP2-A and ISP2-B, each containing 1,152 images acquired in 2024 and 2025, respectively. We also considered a combined version, ISP2-A+ISP2-B, created by merging both sets. Image preprocessing and segmentation were carried out as specified in Section 2.

For feature extraction, we employed two convolutional neural networks—(i) ResNet50 and (ii) EfficientNet—and two vision transformers—(i) Vision Transformer (ViT) and (ii) Shifted WINdow transformer (SWIN)—all pre-trained on ImageNet (Table 1). RGB images were resized to 224×224 pixels to meet the input requirements of these architectures, normalized to the range $[0, 1]$, and standardized using ImageNet's mean values ([0.485, 0.456, 0.406]) and standard deviations ([0.229, 0.224, 0.225]). To ensure the networks functioned solely as feature extractors, we froze their fully connected layers. This procedure yielded 2,048-dimensional embeddings with ResNet50, 1,280 with EfficientNet, and 768 with both ViT and SWIN.

Table 1. Details of the pre-trained models adopted to derive representation vectors.

Model	Input resolution (pixels)	Feature dimensionality	Reference
ResNet50	224×224	2,048	[19]
EfficientNet	224×224	1,280	[20]
ViT	224×224	768	[21]
SWIN	224×224	768	[22]

Given the high dimensionality of the feature vectors, both storage and direct querying would be computationally expensive. To mitigate this issue and assess the impact of dimensionality on the CBIR system's performance, we applied the UMAP algorithm, projecting the representation vectors onto 2, 8, and 32-dimensional spaces. We set the distance metric to Euclidean, fixed the random state parameter at 42 to guarantee reproducibility, and kept all other input arguments at their default values.

For multidimensional indexing, we employed the KD-Tree data structure and computed similarity between embeddings using Euclidean distance. We adopted the *k*NN method to retrieve the k most similar images, with $k = 15$. This choice is justified by the design of the ISP2-A and ISP2-B datasets: each species is represented by 16 images, so when one serves as a query, 15 same-class records remain for comparison.

Due to the relatively limited number of images, we considered the leave-one-out validation technique because of its greater robustness against overfitting. In each iteration, we applied z-normalization separately to the training and test sets before running UMAP. We evaluated retrieval performance in terms of precision and recall, both incorporated into the MAP measure, as well as query time, estimated in seconds. MAP provides a global indicator of retrieval quality by computing the mean of the Average Precision (AP_q) obtained for each query q. Let S be the total number of queries; MAP is defined in Equation 1.

$$\text{MAP} = \frac{1}{S} \sum_{q=1}^{S} AP_q \tag{1}$$

For a single query, the Average Precision is expressed as in Equation 2, where P_k denotes the precision at the k^{th} retrieved image, R_k is the corresponding recall value, and N indicates the number of relevant records considered in the calculation.

$$AP = \frac{1}{N} \sum_{k=1}^{N} P_k R_k \tag{2}$$

The above formulation approximates the area under the precision–recall curve by averaging precision values

across recall levels. This approach rewards CBIR systems that rank relevant images higher in the retrieval list, thereby providing a reliable indicator of both ranking quality and retrieval effectiveness.

We conducted the experiments on a virtual machine equipped with two Intel Xeon E5-2690 v2 processors and 125 GB of RAM. All implementations were developed in Python employing the `TorchVision`[1], `scikit-learn`[2], `Pillow`[3], and `umap-learn`[4] libraries. The processed data supporting the findings of this study are available from the first author upon reasonable request.

3.2 Results and Discussion

Table 2 summarizes the main results of this work, listing the macro-averaged MAP over 24 actinobacterial strains, obtained from distinct datasets, pre-trained deep learning models, and UMAP-reduced dimensions. We report standard deviations in parentheses and highlight in bold the best-performing configuration for each dataset.

Table 2. Macro-averaged MAP scores (%) computed across 24 classes, considering different datasets, image representations, and UMAP dimensions.

Model	#Components	ISP2-A	ISP2-B	ISP2-A+ISP2-B
ResNet50	2	58.13 (12.68)	71.45 (16.97)	59.38 (12.47)
	8	69.65 (10.27)	77.92 (14.24)	70.62 (09.82)
	32	69.40 (10.39)	77.70 (14.47)	70.29 (10.08)
EfficientNet	2	66.43 (10.52)	73.48 (15.28)	63.24 (11.79)
	8	76.44 (09.01)	80.60 (11.26)	75.77 (09.43)
	32	75.92 (08.35)	80.24 (12.26)	75.73 (09.20)
ViT	2	73.87 (14.46)	77.75 (14.29)	70.27 (10.39)
	8	**80.57** (**11.25**)	81.56 (11.92)	**78.98** (**08.43**)
	32	79.83 (11.92)	**82.11** (**11.18**)	78.57 (07.92)
SWIN	2	67.66 (12.06)	73.30 (15.28)	64.81 (09.90)
	8	72.98 (11.54)	79.98 (10.77)	73.72 (07.86)
	32	73.01 (10.94)	79.96 (10.83)	73.58 (07.42)

The transformer-based models outperformed the convolutional neural networks as feature extractors. ViT achieved the best results on all three datasets, reaching 80.57% MAP on ISP2-A (with eight dimensions) and 82.11% on ISP2-B (with 32 dimensions). In contrast, ResNet50 yielded the lowest MAP scores, indicating greater difficulty in capturing visual patterns relevant to the task.

When assessing dimensionality reduction, we observed that compressing the representation vectors to only two components significantly degraded retrieval quality, likely due to the loss of discriminative information during projection. Increasing the dimensionality to eight or 32 components led to substantial gains, although the distinction between these settings was marginal, with the eight-dimensional configuration generally producing slightly better results. Finally, our analysis of MAP standard deviations revealed moderate variability among the bacterial species, ranging from 9% to 14%, suggesting that the CBIR framework consistently retrieved relevant records across the 24 evaluated classes.

We also examined the conceived CBIR system with respect to query execution time. Figure 3 displays the average search time for each representation derived from the pre-trained models under different dimensionality reduction settings. Representations obtained from convolutional neural networks resulted in slower retrieval responses, particularly at 32 dimensions, showing that the KD-Tree indexing structure operated more efficiently when paired with visual transformer features. Moreover, in all scenarios, increasing the number of components

[1] `https://docs.pytorch.org/vision/stable/`.
[2] `https://scikit-learn.org/stable/`.
[3] `https://pypi.org/project/pillow/`.
[4] `https://umap-learn.readthedocs.io/en/latest/`.

led to longer query times, reinforcing the importance of dimensionality reduction for maintaining the CBIR framework's efficiency and scalability.

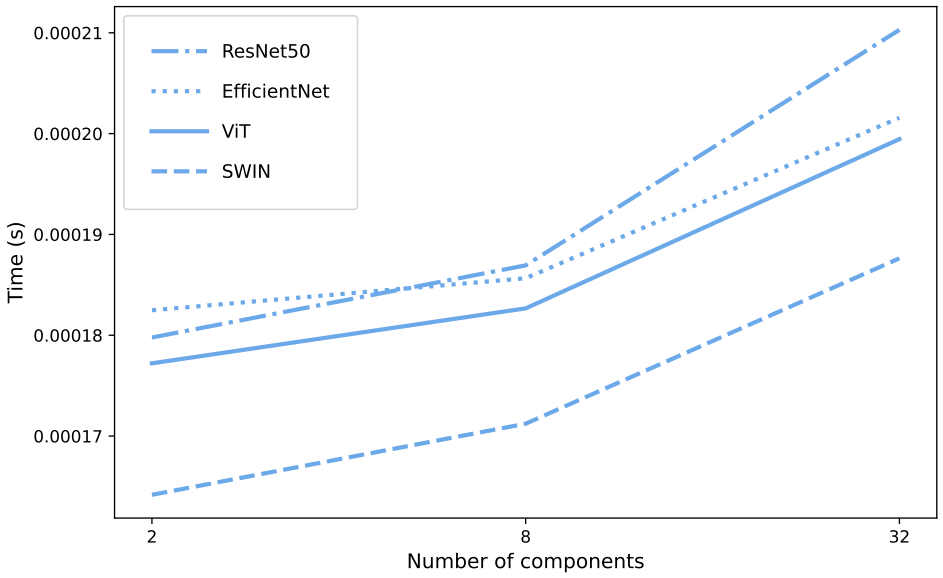


Figure 3. Average query execution time (in seconds) for each off-the-shelf feature set under UMAP output dimensions of 2, 8, and 32 components.

4 Conclusion

This paper proposed and assessed a CBIR system for macroscopic microbial images. The framework integrates visual representations extracted by pre-trained deep neural networks, dimensionality reduction via UMAP, KD-Tree indexing, and similarity search using *k*NN with Euclidean distance.

Our results demonstrated that the ViT-derived image features achieved the best overall performance, yielding the highest MAP scores across all three investigated datasets. Reducing the feature space to eight dimensions proved effective in preserving retrieval quality while considerably reducing query time. In contrast, convolutional neural networks, particularly ResNet50, exhibited lower retrieval accuracy and longer average search times.

We found that combining ViT, UMAP, an efficient multidimensional indexing structure, and *k*NN with Euclidean distance is a viable approach for retrieving actinobacteria images to support the microbial biodiscovery process. This synergy helps bridge the semantic gap between human visual perception and the computational representation and comparison of images, paving the way for more accessible and practical content-based information retrieval systems in the microbiological domain.

In forthcoming work, we plan to adapt our CBIR system to adopt the Ball Tree data structure for multidimensional indexing and to employ cosine similarity as the matching criterion in the retrieval engine.

Acknowledgements. This study was financed, in part, by the São Paulo Research Foundation (FAPESP), Brasil. Process Numbers #2019/17721-9, #2022/02176-8, and #2024/07589-4.

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] X. Li, J. Yang, and J. Ma. Recent developments of content-based image retrieval (cbir). *Neurocomputing*, vol. 452, pp. 675–689, 2021.

[2] A. Latif, A. Rasheed, U. Sajid, J. Ahmed, N. Ali, N. I. Ratyal, B. Zafar, S. H. Dar, M. Sajid, and T. Khalil. Content-based image retrieval and feature extraction: A comprehensive review. *Math. Probl. Eng.*, vol. 2019, n. 1, pp. 9658350, 2019.
[3] M. S. Lew, N. Sebe, C. Djeraba, and R. Jain. Content-based multimedia information retrieval: State of the art and challenges. *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 2, n. 1, pp. 1–19, 2006.
[4] I. M. Hameed, S. H. Abdulhussain, and B. M. Mahmmod. Content-based image retrieval: A review of recent trends. *Cogent Eng.*, vol. 8, n. 1, pp. 1927469, 2021.
[5] Y. Liu, D. Zhang, G. Lu, and W.-Y. Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognit.*, vol. 40, n. 1, pp. 262–282, 2007.
[6] A. W. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, n. 12, pp. 1349–1380, 2000.
[7] J.-H. Su, W.-J. Huang, S. Y. Philip, and V. S. Tseng. Efficient relevance feedback for content-based image retrieval by mining user navigation patterns. *IEEE Trans. Knowl. Data Eng.*, vol. 23, n. 3, pp. 360–372, 2010.
[8] K. J. Locey and J. T. Lennon. Scaling laws predict global microbial diversity. *Proc. Natl. Acad. Sci.*, vol. 113, n. 21, pp. 5970–5975, 2016.
[9] A. R. S. Parmezan, D. R. Gonçalves, and S. O. Rezende. A unified framework for Petri dish image acquisition and processing to promote consistency in microorganism identification. In *arXiv*, 2025.
[10] H. Yu, B. Lu, X. Ouyang, Y. Yang, Y. Zhang, H. Meng, M. Grzegorzek, X. Zhao, C. Li, and H. Lei. Texture features based microbiological image retrieval. In *ICFEICT*, pp. 470–481, Beijing, China. Springer, 2024.
[11] Y. Zou, C. Li, K. Shirahama, T. Jiang, and M. Grzegorzek. Content-based image retrieval of environmental microorganisms using double-stage optimisation-based fusion. *Inf. Eng. Express*, vol. 3, n. 4, pp. 43–53, 2017.
[12] Y. Zou, C. Li, K. Shirahama, T. Jiang, and M. Grzegorzek. Environmental microorganism image retrieval using multiple colour channels fusion and particle swarm optimisation. In *ICIP*, pp. 2475–2479, Phoenix, USA. IEEE, 2016a.
[13] Y. Zou, C. Li, K. Shirahama, F. Schmidt, T. Jiang, and M. Grzegorzek. Content-based microscopic image retrieval of environmental microorganisms using multiple colour channels fusion. In *Computer and Information Science*, volume 605 of *Studies in Computational Intelligence*, pp. 119–130. Springer, Okayama, Japan, 2016b.
[14] Y. L. Zou, C. Li, Z. Boukhers, K. Shirahama, T. Jiang, and M. Grzegorzek. Environmental microbiological content-based image retrieval system using internal structure histogram. In *CORES*, pp. 543–552, Wrocław, Poland. Springer, 2016c.
[15] da J. P. R. Silva and A. R. S. Parmezan. Assessing the feasibility of a spatio-temporal approach for recognizing microorganisms in image sequences. In *COTB*, volume 16, pp. 602–604, Itajaí, Brazil. UNIVALI, 2025.
[16] D. R. Gonçalves, A. R. S. Parmezan, L. F. P. Oliveira, S. P. Lira, R. G. S. Berlinck, and S. O. Rezende. Petri dish image-capturing guidelines for artificial intelligence-based microorganism recognition. In *CBM*, pp. 615–1, Foz do Iguaçu, Brazil. SBM, 2023.
[17] J. Healy and L. McInnes. Uniform manifold approximation and projection. *Nat. Rev. Methods Primers*, vol. 4, n. 1, pp. 82, 2024.
[18] A. R. S. Parmezan, V. M. A. Souza, and G. E. A. P. A. Batista. Time series prediction via similarity search: Exploring invariances, distance measures and ensemble functions. *IEEE Access*, vol. 10, pp. 78022–78043, 2022.
[19] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pp. 770–778, Las Vegas, USA. IEEE, 2016.
[20] M. Tan and Q. Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In *ICML*, pp. 6105–6114, Long Beach, USA. PMLR, 2019.
[21] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *ICLR*, pp. 1–21, Vienna, Austria. OpenReview.net, 2021.
[22] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, pp. 10012–10022, Montreal, Canada. IEEE, 2021.