diagnostics

MDPI

Article

A Novel Framework of Manifold Learning Cascade-Clustering for the Informative Frame Selection

Lei Zhang [1,†,*], Linjie Wu [2,†], Liangzhuang Wei [1], Haitao Wu [2] and Yandan Lin [3,*]

[1] Academy for Engineering and Technology, Fudan University, Handan 220, Shanghai 200433, China
[2] ENT Institute and Otorhinolaryngology Department, Eye & ENT Hospital of Fudan University, Shanghai 200433, China
[3] School of Information Science and Technology, Fudan University, Handan 220, Shanghai 200433, China
* Correspondence: leizhang18@fudan.edu.cn (L.Z.); ydlin@fudan.edu.cn (Y.L.)
† These authors contributed equally to this work.

Abstract: Narrow band imaging is an established non-invasive tool used for the early detection of laryngeal cancer in surveillance examinations. Most images produced from the examination are useless, such as blurred, specular reflection, and underexposed. Removing the uninformative frames is vital to improve detection accuracy and speed up computer-aided diagnosis. It often takes a lot of time for the physician to manually inspect the informative frames. This issue is commonly addressed by a classifier with task-specific categories of the uninformative frames. However, the definition of the uninformative categories is ambiguous, and tedious labeling still cannot be avoided. Here, we show that a novel unsupervised scheme is comparable to the current benchmarks on the dataset of NBI-InfFrames. We extract feature embedding using a vanilla neural network (VGG16) and introduce a new dimensionality reduction method called UMAP that distinguishes the feature embedding in the lower-dimensional space. Along with the proposed automatic cluster labeling algorithm and cost function in Bayesian optimization, the proposed method coupled with UMAP achieves state-of-the-art performance. It outperforms the baseline by 12% absolute. The overall median recall of the proposed method is currently the highest, 96%. Our results demonstrate the effectiveness of the proposed scheme and the robustness of detecting the informative frames. It also suggests the patterns embedded in the data help develop flexible algorithms that do not require manual labeling.

Keywords: unsupervised learning scheme; manifold learning; deep convolutional neural networks; laryngoscopic images; informative frame selection

check for updates

Citation: Zhang, L.; Wu, L.; Wei, L.; Wu, H.; Lin, Y. A Novel Framework of Manifold Learning Cascade-Clustering for the Informative Frame Selection. *Diagnostics* **2023**, *13*, 1151. https://doi.org/10.3390/diagnostics13061151

Academic Editor: Ming-Huwi Horng

Received: 31 January 2023
Revised: 5 March 2023
Accepted: 10 March 2023
Published: 17 March 2023

1. Introduction

Laryngeal cancer (LC), grouped with head and neck squamous cell cancer (HNSCC), is the 7th most common cancer (men 5th and women 13th) [1]. In Europe, it is the second most common malignancy of the head and neck region [2]. A landmark report confirmed the evidence of the association between tobacco smoke and cancer in the 1950s, explicitly in the head and neck tumors, including the larynx [3,4]. Nearly 87% of LC patients are tobacco users (central Europe), while 60–89% of LC is attributed to a combination of tobacco smoking and alcohol drinking (South America). Fortunately, quitting cigarettes lowers the probability of developing laryngeal cancer, especially for those who had smoked for more than ten years [5]. The Surveillance, Epidemiology, and End Results (SEER) also reported new laryngeal cancer cases fell by some 50% from 1975 to 2016, credited to lower smoking rates in younger populations and the evolving tobacco-related marketing [1].

The 5-year survival rate is one of the most concerning indicators in the larynx community for computer-aided diagnosis (CADx) [6,7]. It notes the percentage of people still alive more than five years after confirmed laryngeal cancer [8]. In recent cancer statistics, laryngeal cancer is not one of the leading cancers in the United States [9]. Unfortunately, the 5-year survival rate has dropped from 66 to 63% in the past 40 years [10], while approximately 60% of patients present with advanced stage disease (stage III or IV) at diagnosis [11].

1.1. Early Diagnosis and Narrow-Band Imaging

Nowadays, surgical techniques performed for laryngeal function preservation are feasible. However, the early diagnosis of laryngeal cancer is still the primary means of clinical intervention. Ref. [12] showed that patients with different levels of laryngeal carcinomas (Tis, T1, and T2) have an 80–90% probability of healing in the early stage, during an approximately 60% cure rate for more advanced tumors. Ref. [4] calls on governments and social institutions to get involved in cancer prevention education, early diagnosis, surveillance, and monitoring of public health interventions.

The current diagnosis of LC at the early stage is not out of the scope of an endoscopy. The literature has well summarized the pros and cons of the two popular non-invasive endoscopy techniques, narrow-band imaging (NBI) and white light endoscopy (WLE). NBI is more expensive than WLE from a clinical perspective. However, it is beneficial for adding more definition to the tumor margins and highlighting the features of submucosal vascularization [7,13]. Additionally, ref. [14] reported that NBI is 21% more diagnostically sensitive than conventional WLE.

1.2. Informative Frame Selection for CADx

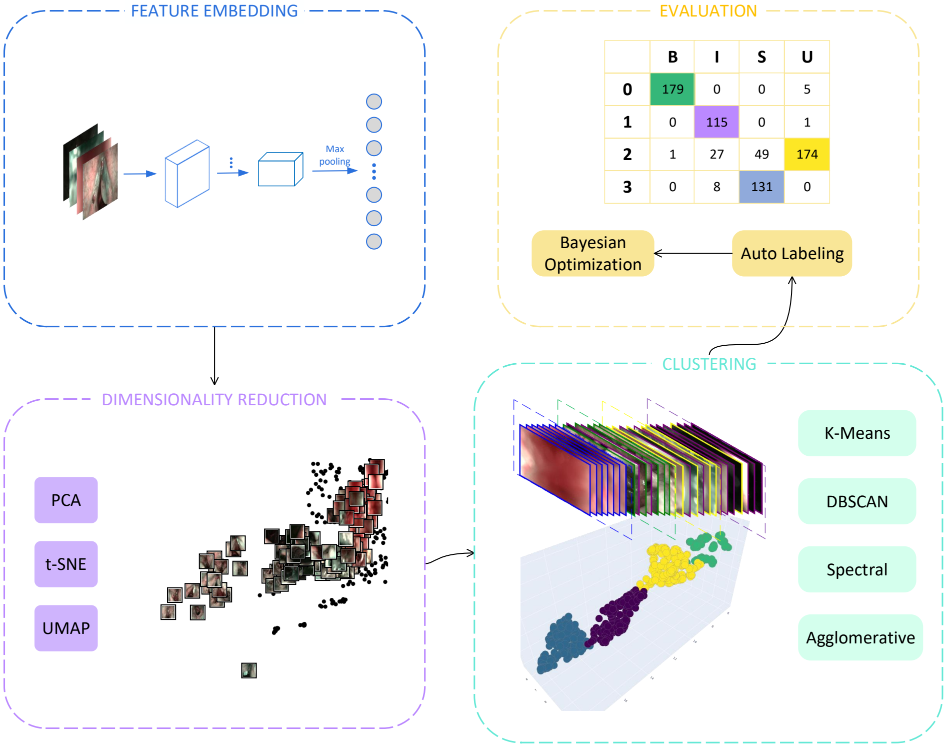
Manually selecting the informative frames from the vast number of candidates available in endoscopy is tedious and time-consuming. It also requires experienced endoscopists. More than 10,000 endoscopy frames are produced per patient during each examination. It may cost the physician a lot of time (over 60 min) to read the images [15,16]. Not all frames are helpful; most are redundant (blurred, bubbles, etc.), and only a tiny proportion are related to lesions or abnormalities. Therefore, selecting the most informative instances (images) is essential for the subsequent automatic diagnosis, such as classification and segmentation of the lesions [17,18]. If the uninformative frames are removed, the accuracy of the automatic detection of the abnormalities will be increased [19].

This issue seemed to be trivial with the emergence of deep learning techniques. However, we examined several recent studies that employed artificial neural networks for the computer-aided diagnosis of laryngeal cancer [7,20–23] and found that although the datasets were labelled and well-structured, the number of images still ranged from 3000 to 25,000. Such samplings applied to clinical applications are still a considerable workload for science, technology, engineering, and math (STEM) researchers willing to cooperate with physicians [24].

On the other hand, collecting high-quality data is a critical research component, ultimately improving laryngeal cancer patient management [1]. Interestingly, the public datasets related to informative frame selection are rare, only 3% (3/97) compared to vibration analysis, lesion recognition, etc. [24]. Our fundamental concern is that existing algorithms are insufficient for extracting specific patterns of high-quality data in enormous data with noise. The solution for informative frame selection should be close to a natural and clinical setting, with unstructured and unlabeled data. Ref. [18] reported a similar scenario in the field of biomedicine and biology.

1.3. Organization

The remaining paper is organized as follows. Section 2 illustrates the related works and the missing part of the puzzle. Section 3 describes the proposed approach to select the informative frames in an unsupervised fashion, a scheme combined with feature embedding, dimensionality reduction methods, and the primary clustering methods. Section 4 presents the experimental details, including the dataset and metrics. It includes the automatic labeling algorithm and a cost function for hyperparameter tuning. Results are presented in Section 5 and discussed in Section 6. Finally, the paper concludes with suggestions in Section 7.



	B	I	S	U
0	179	0	0	5
1	0	115	0	1
2	1	27	49	174
3	0	8	131	0

Figure 1. The flowchart scheme.

2. Related Work

Many medical screenings include informative frame selection, such as image screening for laryngeal disease detection and redundant image elimination. These are not imited to laryngoscopy [17,25], e.g., gastrointestinal endoscopy [26], wireless capsule endoscopy (WCE) [27,28], optical endomicroscopy [29], and colonoscopy [30]. We have grouped these methods into two categories.

2.1. Criterion-Based Feature Extraction

This method's main characteristic is based on observing basic information about the image, such as color, texture, and geometric features. The statement is that the sharp images respond to high-frequency content [31].

Ref. [26] found that the informative and uninformative frames can be distinguished in the frequency domain according to an energy histogram. Later, the k-means algorithm was employed in the endoscopic video manifolds to cluster the informative frames. However, the frequency spectrum contains superfluous information. Ref. [17] selected a set of descriptors followed by a support vector machine to classify the NBI endoscopic frames. Their method overcame the specific threshold. Similarly, ref. [17] observed that informative frames have higher spatial frequencies than blurred ones. Refs. [27,32] extracted the features with a local color histogram in the HSV space and isolated non-informative frames with a support vector machine classifier. Refs. [25,33] employed the Shannon entropy to eliminate the uninformative images in their specific endoscopic applications. Ref. [29] grouped pure noise and motion artifacts into uninformative frames of optical endomicroscopy. They used grey level cooccurrence matrix (GLCM) texture metrics, such as contrast, energy, entropy, etc., to detect the uninformative frame in the video sequences. Ref. [34] proposed four criteria for screening sharp laryngoscopic images, including sum-modulus-differences and the energy Laplacian of the image, gradient magnitude maximization, and variance. In their latest study, ref. [35] analyzed the hue and geometric features of the laryngoscopic images and introduced the peak signal-to-noise ratio (PSNR) to calculate the nearest frames in the video.

2.2. Learning-Based Feature Extraction

The criterion-based methods focus on distinguishing informative frames with criterion functions and are subjective and time-consuming. However, learning-based approaches extract and represent features without any explicit method [16]. Such methods find representations directly from the input frames without a specific feature set or descriptors. Ref. [36] constructed two convolutional neural networks to detect the blurry and water frames from colonoscopy videos.

Recently, several studies have adapted the idea of transfer learning, one of the emerging techniques in deep learning. Ref. [37] employed the pre-trained Inception-v3 network to end-to-end classify the informative and non-informative frames from a colonoscopy. Ref. [38] used the fine-tuned VGG16 networks to classify the different categories of the NBI frames from laryngoscopic videos. Ref. [39] evaluated several state-of-the-art convolutional neural networks in the NBI-InfFrames dataset and achieved state-of-the-art performance. As of this writing, a study proposed a dataset to evaluate the pre-trained ResNet-18 model. The dataset contains 22,000 laryngoscopic frames. It is the second dataset announced for the problem of informative frame selection and the largest one [40].

Therefore, no matter what feature extraction methods are employed, considerable research has been dedicated to employing supervised learning methods as the classifier due to the types of uninformative frames defined. However, we cannot easily employ such methods to detect uninformative frames for unknown types in clinical setting data. A revolutionary study categorized the frames of WCE videos into four representative groups without defining any frame types [28]. They employed non-negative matrix factorization (NMF) and fuzzy C-means (FCM) to eliminate 85% of uninformative frames. However, they did not take the cluster number determination into account.

2.3. Contributions

We try to answer whether informative frame selection can be conducted without supervision or guidance from the labels. We hypothesised that informative frame selection does not rely on data labels. Our proposed unsupervised framework (Figure 1) remains competitive compared to the state-of-the-art methods in the NBI-InfFrames dataset. The main contributions of this paper are summarized as follows:

- A new scheme couples the state-of-the-art dimensionality reduction techniques and clustering methods for solving the issue. The proposed scheme extracts feature faster than the baseline method [17] and require no effort on the labeling data. This ensures the reliability and effectiveness of the proposed scheme in a clinical setting datasets.
- For the frame reduction or video summarization, the future direction for unsupervised learning methods should involve cluster determination [28]. In this work, we introduce a metric under our scheme, the Calinski-Harabasz Index, to automatically determine cluster numbers.
- In this work, we propose an automatic cluster labeling algorithm using bijections mapping for evaluating the classification performance of unsupervised methods. We further propose a Bayesian optimization cost function algorithm to boost classification performance.
- To the best of the authors' knowledge, none of the existing works in the literature on computer-aided diagnosis of laryngoscopy attempt to solve the problem using an unsupervised scheme based on the feature learning method. In addition, our methods achieved comparable performance to state-of-the-art supervised learning methods [17,38,39].

3. Methods

The unsupervised learning algorithm is a typical process to find patterns in data without labels [41]. It is usually used for clustering, feature extraction, or dimensionality reduction [42].

3.1. Feature Embedding

Two categories of feature extraction can be summarized: the methods that rely on conventional machine learning and those that rely on neural networks. Ref. [16] illustrated more specific divisions as spatial domain, frequency domain, and feature learning methods. Considering this work, the dataset we use in Section 4.1 only contains 720 images. Thus, we use a small architecture neural network. Another fact is that ref. [38] employed the fine-tuned VGG16 succeeded on NBI-InfFrames, motivated by transfer learning.

The vanilla VGG16 network without any pre-trained weights is used in this work for feature extraction. Moreover, the last two layers were dropped (Figure 2).

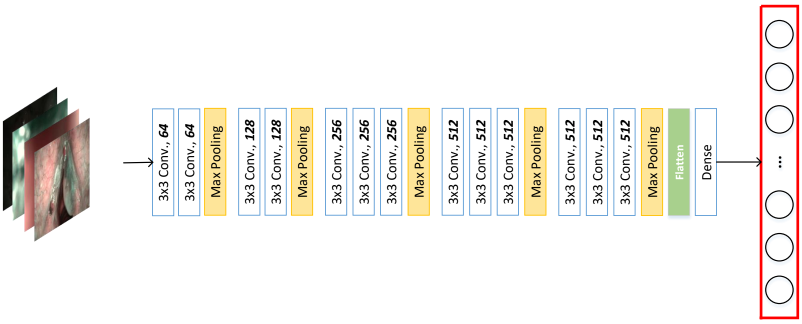


Figure 2. The four images represent the four types of categories in the NBI-InfFrames. The neural network consists of the vanilla VGG16 without the last two layers, the fully connected layer, and the prediction layer. The generated feature embedding dimensionality is 4096.

3.2. Dimensionality Reduction

The dimensionality reduction technique is not rare in the area of medical imaging. It is widely used for feature selection and visualization. In the area related to endoscopic images, ref. [26] proposed a manifold learning method named EVM for projecting the endoscopic video into the local structure of the manifold, and ref. [28] introduced an unsupervised data reduction algorithm for the capsule endoscopy video segments, the non-negative matrix factorization method. This work introduces the current state-of-the-art techniques to compare with a recently proposed algorithm.

3.2.1. PCA

Principle components analysis (PCA) is a popular dimensionality reduction technique that maps the data points from high- to low-dimensional space with the linear transformation. In mathematical terms, the transformation can be denoted as [43]

$$\mathbf{Y} = \mathbf{X}\mathbf{M}, \tag{1}$$

where $\mathbf{X}$ represents the original data or features, $\mathbf{Y}$ is the matrix of the transformed data points y_i, and the mapping relationship is represented by $\mathbf{M}$. PCA aims to find the M that maximizes the cost function trace $tr(\mathbf{M}^T cov(\mathbf{X})\mathbf{M})$, where $cov(X)$ is the sample covariance matrix. Consequently, it is transferred to solve the eigenproblem of the d principle eigenvectors (principal components),

$$cov(\mathbf{X})\mathbf{M} = \lambda\mathbf{M}. \tag{2}$$

In the experiment, d is decided by analysis of the variance. Nearly 80% of the variance can be explained by the 50 principal components. Thus, we choose $d = 50$ in Section 5.1.

Minimizing the cost function forms of the Euclidean distance to find the M is commonly used in multidimensional scaling [43].

$$\phi(\mathbf{Y}) = \sum_{ij}\left(d_{ij}^2 - \|\mathbf{y}_i - \mathbf{y}_j\|^2\right), \tag{3}$$

where d_{ij} represents the Euclidean distance between the x_i and x_j in high-dimensional space, and the $\|\mathbf{y}_i - \mathbf{y}_j\|^2$ is the square Euclidean distance between the low-dimensional space.

3.2.2. t-SNE

A stochastic neighbour embedding technique, t-SNE, was presented by Matten and Hinton [44]. It is widely recommended due to its good visualizations. SNE aims to find the optimal mapping relationship between high and low-dimensional space by minimizing the mismatch between $p_{j|i}$ and $q_{j|i}$. As the derivative version of SNE, t-SNE overcomes the drawbacks of using the non-symmetric Kullback–Leibler divergence to measure the faithfulness of the pairwise distance. The Student-t distribution is used to compute the similarity between two data points in the low-dimensional space instead of a Gaussian, thus named t-SNE.

To minimize the cost function of the conditional probabilities $p_{j|i}$ and $q_{j|i}$, t-SNE turns to minimize a single Kullback–Leibler divergence between the joint probability, P, in high-dimensional space and a joint point probability, Q, in low-dimensional space. Thus, the cost function is denoted as

$$C_{t-SNE} = KL(P\|Q) = \sum_{i \neq j} \sum_{j} p_{ij} \log \frac{p_{ij}}{q_{ij}}, \tag{4}$$

where the low-dimensional pairwise similarity map by q_{ij} and the high-dimensional map p_{ij} are given by

$$p_{ij} = \frac{\exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right)}{\sum_{k \neq l} \exp\left(-\|x_k - x_l\|^2 / 2\sigma^2\right)}, q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_{k \neq l} \left(1 + \|y_k - y_l\|^2\right)^{-1}}.$$

The p_{ij} is a Gaussian distribution to approximate the high-dimensional space, while the q_{ij} is a Student-t distribution. The detailed optimization of the cost function using the stochastic gradient descent is available in Section 3.3.4 of this article [44].

3.2.3. UMAP

Although the t-SNE is good at revealing important global structures, it cannot go beyond locality-preserving limits. Uniform manifold approximation and projection (UMAP) [45] takes a big step in preserving the global structure of the large dataset, as well as the small one. It implies that the distance of the inter-class data points is more distinguishable under the dimensionality reduction techniques with the globality-preserving properties. This is the key to understanding that UMAP outperforms other dimensionality reduction techniques. We will explain this finding further in Section 6.

Our interest in the UMAP came from visualising artworks of the Metropolitan Museum of Art collection [46]. We observed that the artworks are widely distributed according to the brightness and darkness of the average intensity after projection by UMAP. Also, the distinguishable characteristic is embedded in the NBI-InfFrames dataset. The image thumbnails of the informative video frames are well separated from the non-informative frames (Figure 1 DIMENSIONALITY REDUCTION).

Since the mathematical language to understand the UMAP is similar to the t-SNE, we present this part in Appendix B. It is also reported that the UMAP is nine times faster than the t-SNE as evaluated on the MNIST dataset with the scikit-learn toolkit [47].

3.3. Clustering

We divide the clustering methods into four main categories in this work. They are hierarchical methods, partitioning methods, graph-based methods, and density-based methods [18]. Existing clustering methods, such as fuzzy and soft clustering, are not included in the investigation.

3.3.1. Agglomerative

Hierarchical clustering groups the data point into a binary tree called a dendrogram [48]. Agglomerative clustering is one of the two hierarchical clustering methods widely used for its simplicity. Its history dates back at least to the 1950s. Agglomerative clustering builds clusters in a bottom-up fashion, starting with each data point from its cluster. In the subsequent step, the two closest clusters will be merged until all data points are grouped into one cluster. Different from the iterative clustering algorithms, the data points cannot be further merged adjustments or split once the building progress of the agglomerative clustering is made. The property can be helpful for applications with an unknown number of clusters, such as bioinformatics applications [49].

Factors such as similarity measures, criterion functions, and initial conditions, decide the effectiveness of the clustering methods. For agglomerative clustering, there are three definitions of the similarity between two clusters: single-link, complete-link, and average-link [48]. Such linked strategies appear in the form of the hyperparameters (Section 4.4) in our experiments.

3.3.2. K-Means

A partitional clustering method manipulates a single partition on the data points, while the hierarchical clustering method holds a whole structure of the clustering, described as the dendrogram [50]. Therefore, the computation time of the hierarchical clustering performed on the large-size dataset is unbearable. K-means is one of the partitional clustering methods with linear computational time employed in a wide range of applications.

A pattern $\mathbf{x}$ is defined as a singleton (observation, feature vector, or datum) on the clustering algorithm, consisting of d dimensional measurements $\mathbf{x} = (x_1, \dots x_d)$. A pattern set is denoted as $\mathscr{X} = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, which can be viewed as a matrix of samples and features $(n \times d)$. The clustering $\mathscr{L}$ of the k-means can be established from the pattern set $\mathscr{X}$ by employing a squared error criterion as follows

$$e^2(\mathscr{X}, \mathscr{L}) = \sum_{j=1}^{K} \sum_{i=1}^{n_j} \left\| \mathbf{x}_i^{(j)} - \mathbf{c}_j \right\|^2, \tag{5}$$

where the $\mathbf{x}_i^{(j)}$ is the i^{th} pattern of the j^{th} cluster, and $\mathbf{c}_j$ is the centroid calculated by the mean of the member points in the j^{th} cluster.

In summary, the algorithm of k-means starts with the initial position and relocates patterns to the clusters according to the similarity measurement until the convergence criterion is satisfied. This study was motivated by the classification performance of the k-means outperforming the baseline work. In the implementation, the MinibatchKMeans was employed due to the faster convergence and low difference in accuracy compared to k-means.

3.3.3. Spectral Clustering

Until now, we can conclude that the goal of the clustering methods is to minimize the differences in the same cluster and maximize the difference between the clusters. Spectral clustering aims to achieve this by partitioning the similarity graph, consisting of the eigenvalues of the similarity matrix of the data.

Spectral clustering is an algorithm that the k-means performed on the eigenvectors of the graph Laplacian [26]. In addition, it can be viewed as a variant kernel k-means overcoming the drawback of k-means in the non-linear space.

Given the undirected graph $G = (V, E)$, it consists of two elements: the vertex $v_i \in V$ and the edge $e_{ij} \in E$. Each vertex represents a data point x_i. If the edge e_{ij} between two points x_i and x_j is larger than a certain threshold, we denote they are connected and weighted by the similarity s_{ij}, thus $s_{ij} \geq 0$.

Assuming the graph G is weighted, an adjacency matrix $W = (w_{ij})_{i,j=1,\ldots,n.}$ is then generated, which describes the similarity between the vertices v_i and v_j. In addition, $w_{ij} = 0$ in the adjacency matrix W means there is no connection between the two points.

The degree d_i of the vertex v_i is naturally introduced, as $d_i = \sum_{j=1}^{n} w_{ij}$. It counts all the weights starting from the vertex v_i, referred to as v_{ij}, but not including the vertices ending at v_i, referred to as v_{ji}. The degree matrix of D is referred to as the diagonal of the matrix with the degree of vectors $d_1, \ldots, d_n$.

An A is denoted as the subset of the vertices V, $A \subset V$; thus, we have the indicator vector $1_A = (f_1, \ldots, f_n)' \in \mathbb{R}^n$ [51],

$$f_i = \begin{cases} 1, & \text{if } v_i \in A \\ 0, & \text{otherwise} \end{cases}. \tag{6}$$

The unnormalized graph Laplacian is generated from the components of the matrix of D and the matrix of W. The defined matrix can be given as

$$L = D - W. \tag{7}$$

Once the graph is constructed, the spectral clustering algorithm can be interpreted to the k-means. Computing the first k eigenvectors of the L from the equation (Equation (7)), denoted as the $u_1, \ldots, u_k$. Using the terms of the k-means, a pattern $\mathscr{U} \in \mathbb{R}^{n \times k}$, k is the column number of the U. Later, the clustering $\mathscr{Y} = \{y_i | i = 1, \ldots, n\}$, the rows i represent the samples. Different criterion functions can be employed on the $\mathscr{U}$ and $\mathscr{Y}$, such as the Euclidean, Manhattan, cosine, etc. Finally, the algorithm outputs the cluster indices, $A_i = \{j | y_j \in C_i\}$.

3.3.4. DBSCAN

DBSCAN does not hold the common assumption that clusters can be decided by minimizing and maximizing the difference between intra-clusters and inter-clusters. An essential insight is the density of the points; the density of the points in each cluster is higher than the density of the points outside the cluster, and vice versa. For the outliers, the noise (points) density is lower than regular clusters.

Unlike the traditional clustering methods, which hardly assign each point to a cluster, DBSCAN assigns the probability to each point, one of the density-based clustering methods. It only requires a few input parameters (one input parameter used in [52]) and computational efficiency compared to the other three algorithms (Appendix C). The shape of the clusters made by the partitioning methods is convex, while the hierarchical algorithm is prohibitive for the computational time, especially for the large data size. However, it also brings the inconvenience of finding the intended number of clusters for aligning to the number of the ground-truth classes of the dataset. Therefore, we propose Algorithm 1, which helps to find the optimal result for the classification performance.

The algorithm starts from a radius of neighbours; the shape of the radius is determined by the distance measurements. For example, two points, p and q, are neighbours in the Manhattan space. It can be denoted as $dist(p, q)$, and the shape of these two points is rectangular. The definition of the neighborhood of a point p can be denoted as $N_{Eps}(p)$,

$$N_{Eps}(p) = \{q \in D | dist(p, q) \leq Eps\}. \tag{8}$$

Other rules for DBSCAN are maintaining its arbitrary shape to discover the non-convex clusters, such as the density-reachable, density-connected, cluster, and noise [52]. In our experiment, we only specify the minimum number of points ($MinPts$) and the radius of the shape Eps.

Algorithm 1 Cost function for the Bayesian optimization searching

```
Require: $S_{rec}$, recall of the method; $S_{pre}$, precision of the method; $\mathcal{S}_{params}$, parameter space
Ensure: number of clusters K=4
  $\alpha \leftarrow 100$                                  ▷ penalty the label counts deviation
  $\beta \leftarrow 0.01$                                  ▷ weight for the impact of the variance
  repeat
      if method is MinibatchKMeans or Agglomerative or Spectral then
          $p \leftarrow 0.1 \cdot |S_{rec} - S_{pre}|$
          $\mathrm{C} \leftarrow (1 - S_{rec}) + p$, s.t. $\mathcal{S}_{params}$
      else if menthod is DBSCAN then
          count the outlier labels, -1 in the cluster, $N_{out}$
          count the number of kinds of labels except for outliers in the cluster, $N_{in}$
          calculate the variance of the counts of each group deviate from 180, $\sigma$
          $\mathrm{C} \leftarrow N_{out} + (1 - \beta) \cdot (\alpha \cdot |N_{in} - K|) + \beta \cdot \sigma$, s.t. $\mathcal{S}_{params}$
      end if
      $\min_{\mathcal{S}_{params}}$ C.
  until stop-iteration criteria satisfied
  return best parameters in $\mathcal{S}_{params}$
```

4. Evaluation

The experiments are performed on a Linux platform with a 2.00 Hz CPU, 16 GB Tesla P100-PCIE GPU, and 16 GB RAM. The source code for this work is publicly available at https://github.com/portgasray/UL-IFS-LC (accessed on 1 January 2023). In Section 4.1, we describe the dataset and the definition of the categories. Section 4.2 illustrates the evaluation metrics for analysis results. Section 4.3 presents an algorithm for automatically delivering the intent label to the clusters. Finally, the cost function used for finding the optimal result is described in Section 4.4.

4.1. Dataset

NBI-InfFrames is a dataset acquired with the NBI endoscopic system (Olympus Visera Elite S190 video processor and an ENF-VH rhino-laryngo videoscope) with a frame rate of 25 fps and image size of 1920 × 1072 pixels [17]. It is the current known labelled dataset that is access available in the laryngoscopy area.

The dataset contains 720 frames in 4 categories collected from 18 patients affected by laryngeal squamous cell carcinoma (SCC). Each category consists of 180 frames, informative (I), blurred (B), specular reflection (S), and underexposed (U), Table A1. The dataset was manually labelled by three human evaluators and split into three folders according to the following criteria [17]:

- B, frames should show a homogeneous and widespread blur.
- I, frames should have adequate exposure and visible blood vessels; they may also present micro-blur and small portions of specular reflections (up to 10 per cent of the image area).
- S, frames should present bright white/light-green bubbles or blobs, overlapping with at least half of the image area.
- U, frames should present a high percentage of dark pixels, even though small image portions (up to 10 per cent of the image area) with over or regular exposure are allowed.

Sample images for the four classes are shown in Figure 3. The color intensity bar range from 0 to 255 of the whole dataset is shown at the bottom.

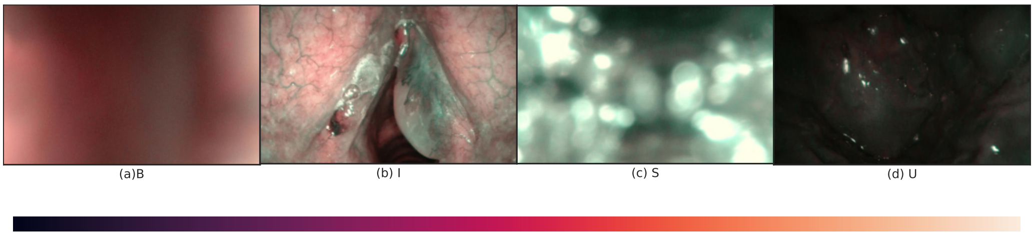


Figure 3. Visualization of sampled examples in the NBI-InfFrames: (**a**) B: blurred frame; (**b**) I: informative frame; (**c**) S: frame with saliva and specular reflections; (**d**) U: underexposed frame. The intensity bar of the dataset is at the bottom.

4.2. Evaluation Metrics

The following metrics will be used to evaluate the classification performance of the state-of-the-art models and our proposed methods.

$$\text{Precision}(\mathbf{Prec}_{class}) = \frac{TP}{TP + FP} \tag{9}$$

$$\text{Sensitivity}(\mathbf{Rec}_{class}) = \frac{TP}{TP + FN} \tag{10}$$

$$\text{F1-score}(\mathbf{F1}_{class}) = \frac{2}{\frac{1}{\mathbf{Rec}_{class}} + \frac{1}{\mathbf{Prec}_{class}}} = \frac{2TP}{2TP + FP + FN} \tag{11}$$

ROC/AUC, $\mathbf{FPR} = \frac{FP}{FP+TN}$ is the x-axis for the false positive rate. $\mathbf{TPR} = \frac{TP}{TP+FN}$ is the y-axis for the true positive rate.

The Calinski–Harabasz Index [53] (Variance Ratio Criterion) is a ratio of the sum of the inter-clusters (between-group) dispersion and the intra-cluster dispersion (within-group) for all clusters.

$$BGSS = \sum_{k=1}^{K} n_k \times \|C_k - C\|^2 \tag{12}$$

The between-group sum of squares (BGSS) is a weighted sum of squared distances between each cluster centroid and the centroid of the whole dataset, where C_k is the centroid of the cluster k, C is the centroid of the whole dataset, and n_k is the number of the data points in the cluster k.

$$WGSS_k = \sum_{i=1}^{n_k} \|X_{ik} - C_k\|^2 \tag{13}$$

The within-group sum of squares (WGSS) calculates the distance between the data points and the centroid of the same cluster. X_{ik} is the data points in the cluster k, and C_k is the centroid of the cluster k.

$$\text{Calinski-Harabasz Index}(CH) = \frac{\frac{BGSS}{K-1}}{\frac{WGSS}{N-K}} = \frac{BGSS}{WGSS} \times \frac{N-K}{K-1} \tag{14}$$

Finally, the Calinski–Harabasz Index is calculated from a ratio sum of BGSS and WGSS, where N is the number of all data points and K is the number of clusters divided by the algorithm, a big score (CH) indicates a well-separated performance.

The silhouette score [54] analyzes the separation distance between clusters, and the number ranges from $[-1, 1]$. A value close to one means the clusters are far away from each, and vice versa. A negative score $[-1, 0]$ indicates that the data points are assigned to the wrong clusters.

$$\text{Silhouette Score}(\mathbf{SC}) = \frac{(b-a)}{max(a,b)}, \tag{15}$$

where a represents the mean distance between a sample point and other points in the same cluster, while b represents the mean distance between a sample point and other points in the nearest cluster.

4.3. Automatic Cluster Labeling

Evaluating the performance of a clustering algorithm is not as trivial as counting the precision and recall of a supervised classification algorithm [55]. In this section, we propose an automatic cluster labeling algorithm to close the gap that compares the two different kinds of algorithms.

Since the cluster labels have no intention of understanding the ground-truth classes, we manually assign the meaningful intent label to the clustering result by viewing the images in the cluster, which is time-consuming and tedious. Before diving deep into the algorithm, we need to recap some sets and map theories.

Given a set of N unlabeled image instances, which can be denoted as $X = \{x_1, x_2, ..., x_N\}$, the N is 720 for the NBI-InfFrames in our experiments. Therefore, the intent classes X_G, are referred to as $X_G = \{X_I, X_B, X_S, X_U\}$. Each class in the NBI-InfFrames can be represented by $X_G = \{X_j = \{x_1, x_2, ..., x_{180}\} | j \in \{I, B, S, U\}\}$, where I, B, S, U are the categories of the dataset. There are 180 image instances for each category.

A collection of labelled clusters X_C is generated from the clustering methods in the scheme, $X_C = \{X_{C_i} | i \in \{0, 1, 2, 3\}\}$, where i refers to the four clusters with the inattentive labels. We can conclude the solution for this problem into a bijections map problem. The key is to find the mapping relationship from the inattentive labels generated from clusters X_C to the intent classes X_G. The mapping function $f : X_C \to X_G$ can be denoted by

$$\forall X_j \in X_G, \exists! X_{C_i} \in X_C \textit{ such that } X_j = f(X_{C_i}), \tag{16}$$

where $\exists! X_{C_i}$ represents exactly one X_{C_i} exists.

The implementation of finding the mapping relationship between cluster pseudo labels and the intent labels is proposed (Algorithm 2). In addition, we visualize one of the possible results of the algorithm (Figure 1 EVALUATION).

Algorithm 2 Automatic cluster labeling

```
Require: $X_{C_i}$,images grouped by cluster labels;$X_G$,images with meaningful intent labels
Ensure: number of clusters is 4
  $i \leftarrow \{0, 1, 2, 3\}$
  $j \leftarrow \{I, B, S, U\}$
  for each cluster $X_{C_i}$ in $X_C$ do
      for each class $X_j$ in $X_G$ do
          calculate intersection number of the $X_{C_i}$ and $X_j$, $n(X_{C_i} \cap X_j)$
      end for
      find the max intersection number of $X_{C_i}$ in $X_G$, max $\{n(X_{C_i} \cap X_G)\}$
      update the mapping relationship, $f : i \to j$
  end for
  return $f$
```

4.4. Cost Function in Hyperparameter Tuning

Hyperparameter tuning contributes to the performance of the proposed methods, which exceed the state-of-the-art supervised learning algorithms. As we introduced 4 clustering methods and 3 kinds of dimensionality reduction methods, the Cartesian product of the two kinds of methods is 12 possible hyperparameter spaces. The mission of the cost function is to find the optimal combination among these spaces (Algorithm 1).

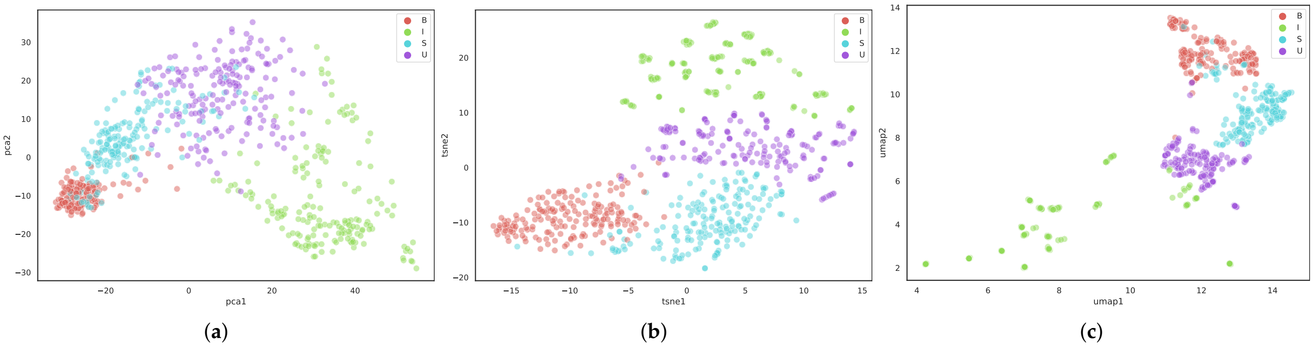
Bayesian optimization (BO) and random research (RS) are designed for this purpose. Evidence shows BO 100× sample efficiency gains more than RS [56]. Our trial on the task finding optimal average sensitivity demonstrated that the BO is faster than RS with 262 s in the 200 steps evaluation and 840 s in 500 steps evaluation. Meanwhile, the difference in average sensitivity between BO and RS is less than 0.3%.

Several strategies are employed to optimize finding the results of the DBSCAN. For the methods of MinibatchKMeans, agglomerative, and spectral clustering, the proposed cost function aims to find the best sensitivity. Meanwhile, it penalizes a deviation. For the DBSCAN, the cost function is used to find the number of clusters, which must be identical to the ground-truth numbers of the classes.

5. Experiments and Results

5.1. Comparison of Dimensionality Reduction Methods

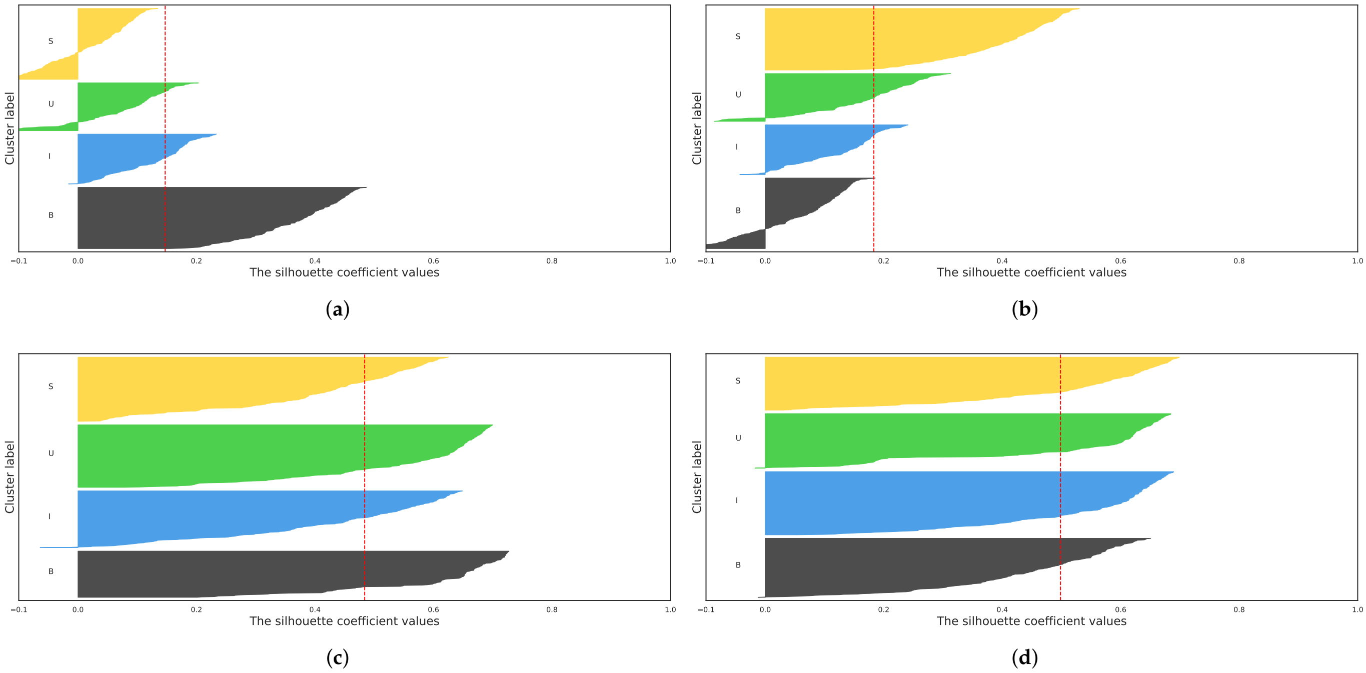
With the support of the ground-truth label from the NBI-InfFrames dataset, we visualize the feature embeddings using three different dimensionality reduction methods. The data points of different classes are distinguishable in (b) and (c). In contrast, the points of class *S* and class *U* are overlapped in (a) (Figure 4). We cannot further infer from the visual inspection that the UMAP is the most favourable. However, the t-SNE and UMAP are better at visualization than PCA.



(**a**) (**b**) (**c**)

Figure 4. The projections of the feature embedding using different dimensionality reduction methods: (**a**) the original feature embeddings projected by PCA; (**b**) the original feature embeddings projected by t-SNE; (**c**) the original feature embeddings projected by UMAP. The four different frame classes classified by the ground-truth labels are reported (B: blurred frames, I: informative frames, S: frames with saliva or specular reflections, U: underexposed frames).

To further evaluate the clustering preference of features projected from PCA, t-SNE, and the UMAP, we introduced the silhouette score (SC) (Equation (15)) for the quantitative analysis of the methods above. We observed that the original feature embeddings of the data points in class *S* and class *U* are negative (Figure 5a). The situation changed while using the PCA projection. However, classes *B*, *I*, and *U* are partly negative. Meanwhile, three classes surpass the average SC, and one is left behind (Figure 5b). Unlike the first two analyses of the silhouette score, the silhouette analysis of the projected features with t-SNE or UMAP is beyond the average silhouette score (Figure 5c,d). Moreover, the negative part of the classes is much less than the first two (Figure 5). Finally, based on the average silhouette score, the clustering performance of the UMAP projected features is a little better than the feature embeddings projected by the t-SNE (0.5 versus 0.48, average silhouette score).



(**a**)

(**b**)

(**c**)

(**d**)

Figure 5. Silhouette analysis for K-means clustering on proposed dimensionality reduction methods. Different frame classes (B: blurred frames, I: informative frames, S: frames with saliva or specular reflections, U: underexposed frames) are in different colors. The red dotted line represents the average silhouette score (avg_sc), and the negative part of the cluster indicates the incorrect clustering. (**a**) silhouette analysis for K-means clustering on vanilla feature embedding (avg_sc = 0.15); (**b**) silhouette analysis for K-means clustering on PCA projected feature embeddings (avg_sc = 0.18); (**c**) silhouette analysis for K-means clustering on t-SNE projected feature embeddings (avg_sc = 0.48); (**d**) Silhouette analysis for K-means clustering on UMAP projected feature embeddings (avg_sc = 0.50).

We can deduce that the cluster performance of the introduced UMAP projected features is most favourable via the visual inspection and silhouette analysis of the introduced dimensionality reduction methods. In the subsequent experiments, we combine the cluster methods and several introduced dimensionality reduction methods and further evaluate their classification performance with precision, recall, and F1-score. Eventually, we want to know whether the UMAP is most promising.

Take a typical clustering method as an illustration. The classification performance of the K-means combined with UMAP is best compared to combine with PCA or t-SNE or without projection (the median precision $\mathbf{Prec}_{class} = 92\%$, recall $\mathbf{Rec}_{class} = 94\%$, F1-score $\mathbf{F1}_{class} = 93\%$ with respective smallest IQR $= 7\%, 7\%, 5\%$ are reported in Table 1). Meanwhile, the detection of class *I* is most robust ($\mathbf{Rec}_{class} = 95\%$), too, from the perspective of the task purpose. Such observation also can be found in other experiments (Tables 2 and 3).

Until now, we can conclude that UMAP is better than PCA or t-SNE or without projection for clustering performance both from qualitative and quantitative analysis perspectives. The classification performance of the clustering methods coupled with UMAP is also better than the alternative dimensionality reduction methods, such as PCA and t-SNE, in terms of precision, recall, and F1-score. Finally, we further compare the different clustering methods combined with UMAP in the next section.

Table 1. Classification performance of the feature embeddings using the state-of-the-art dimensionality reduction approaches (PCA, t-SNE, UMAP). Results are evaluated under the K-means clustering (KM). Precision ($\mathbf{Prec}_{class}$), class-specific recall ($\mathbf{Rec}_{class}$), and F1-score ($\mathbf{F1}_{class}$) are reported for the different frame classes (**B**: blurred frames, **I**: informative frames, **S**: frames with saliva or specular reflections, **U**: underexposed frames) of the NBI-InfFrames dataset. The three metrics' median and interquartile range (IQR) are also reported.

	Vanilla K-Means			**PCA + K-Means**			**t-SNE + K-Means**			**UMAP + K-Means**		
	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$
B	**0.90**	0.97	0.93	**0.90**	0.97	0.93	0.89	0.97	0.93	0.89	**0.98**	**0.94**
I	**1.00**	0.87	0.93	**1.00**	0.87	0.93	**1.00**	0.81	0.89	**1.00**	**0.95**	**0.97**
S	**0.94**	0.78	0.85	**0.94**	0.78	0.85	0.88	**0.87**	0.87	0.93	0.86	**0.89**
U	0.78	**0.97**	0.87	0.79	**0.97**	0.87	0.80	0.89	0.85	**0.90**	0.93	**0.92**
Median	0.92	0.92	0.90	0.92	0.92	0.90	0.89	0.88	0.88	**0.92**	**0.94**	**0.93**
IQR	0.13	0.15	0.07	0.13	0.15	0.07	0.11	0.09	**0.05**	**0.07**	**0.07**	**0.05**

Table 2. Classification performance of the feature embeddings using the state-of-the-art dimensionality reduction approaches. Results are evaluated by Spectral clustering (Spec). Precision ($\mathbf{Prec}_{class}$), class-specific recall ($\mathbf{Rec}_{class}$), and F1-score ($\mathbf{F1}_{class}$) are reported for the different frame classes (**B**: blurred frames, **I**: informative frames, **S**: frames with saliva or specular reflections, **U**: underexposed frames) of the NBI-InfFrames dataset. The three metrics' median and interquartile range (IQR) are also reported. The dash in the cells indicates the failure on the PCA projected features and the t-SNE projected features.

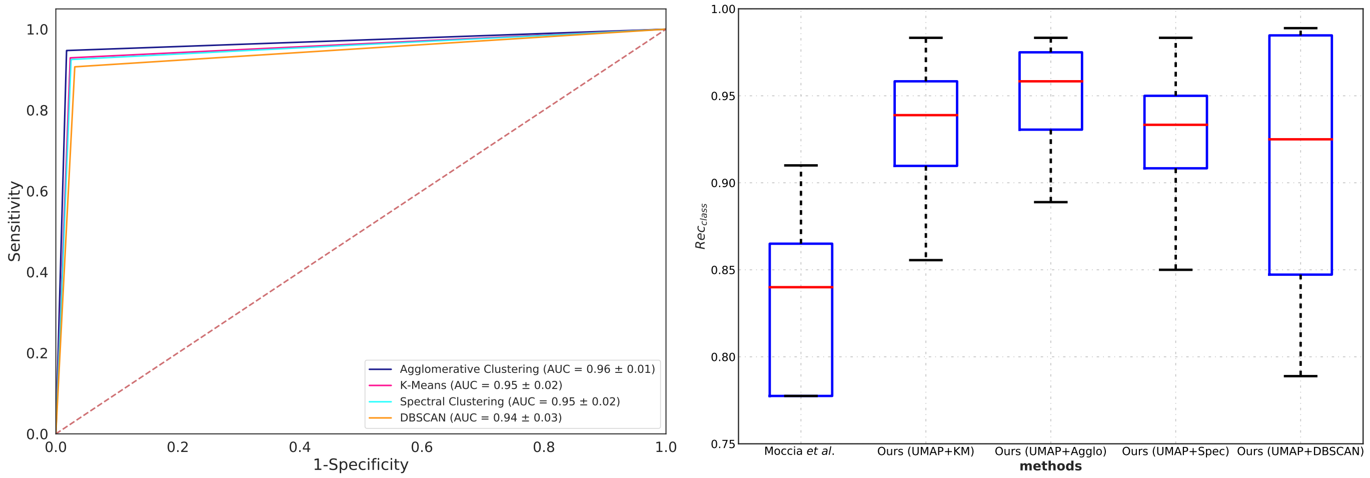
	Spectral Clustering			**PCA + Spectral Clustering**			**t-SNE + Spectral Clustering**			**UMAP + Spectral Clustering**		
	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$
B	0.25	1.00	0.40	0.25	1.00	0.40	0.31	1.00	0.47	0.89	0.98	0.93
I	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.74	0.85	0.99	0.94	0.97
S	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.92	0.85	0.88
U	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.91	0.93	0.92
Median	-	-	-	-	-	-	-	-	-	**0.92**	**0.94**	**0.93**
IQR	-	-	-	-	-	-	-	-	-	**0.06**	**0.07**	**0.05**

Table 3. Classification performance of the feature embeddings using the state-of-the-art dimensionality reduction approaches. Results are evaluated under agglomerative clustering (Agglo). Precision ($\mathbf{Prec}_{class}$), class-specific recall ($\mathbf{Rec}_{class}$), and F1-score ($\mathbf{F1}_{class}$) are reported for the different frame classes (**B**: blurred frames, **I**: informative frames, **S**: frames with saliva or specular reflections, **U**: underexposed frames) of the NBI-InfFrames dataset. The three metrics' median and interquartile range (IQR) are also reported.

	Agglomerative Clustering			**PCA + Agglomerative Clustering**			**t-SNE + Agglomerative Clustering**			**UMAP + Agglomerative Clustering**		
	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$
B	0.89	0.98	**0.93**	**0.91**	0.94	**0.93**	0.89	**0.99**	**0.93**	0.89	0.98	**0.93**
I	**1.00**	0.83	0.91	**1.00**	0.83	0.91	0.95	0.92	0.93	0.99	**0.94**	**0.97**
S	**0.99**	0.88	0.93	0.90	**0.91**	0.91	**0.99**	0.60	0.75	**0.99**	0.89	**0.94**
U	0.84	**0.99**	0.91	0.84	0.95	0.89	0.71	0.93	0.81	**0.94**	0.97	**0.95**
Median	0.94	0.93	0.92	0.91	0.93	0.91	0.92	0.93	0.87	**0.97**	**0.96**	**0.95**
IQR	0.13	0.13	**0.02**	0.09	0.08	**0.02**	0.17	0.20	0.15	**0.08**	**0.06**	0.03

5.2. Classification Performance Comparison

Based on the prior observations, the clustering methods coupled with the UMAP projected feature achieved the best performance in clustering and classification. We compared the different clustering methods (Section 3.3) coupled with UMAP under the ROC/AUC curve. The comparison results suggest that the agglomerative clustering method obtained the highest mean AUC score, followed by K-means, spectral clustering, and DBSCAN (AUC = 96%, 95%, 95%, 94%, respectively, reported in Figure 6a). In addition, the agglomerative clustering obtained the best median recall with relatively the smallest IQR among all clustering methods ($\mathbf{Rec}_{class} = 95\%$, IQR $= 6\%$ in Table 3). Meanwhile, the proposed clustering methods coupled with UMAP are beyond the baseline (Figure 6b).



(**a**) (**b**)

Figure 6. Classification performance comparison of the proposed methods: (**a**) receiver operating characteristic (ROC) curves and area under ROC curve (AUC). The mean area under the ROC curve (±standard deviation) of each method is reported in the legend; (**b**) The boxplot of recall ($\mathbf{Rec}_{class}$) for comparison of the proposed clustering methods. The comparison in terms of $\mathbf{Rec}_{class}$ for the proposed methods and method proposed by [17].

It is worth noting the spectral clustering failed on the PCA and t-SNE projected features while succeeding in the UMAP case (Table 2). We detailed the classification performance of precision, recall, and F1-score of the proposed clustering methods coupled with different dimensionality clustering methods in Tables 1 and 2. An exception is that DBSCAN differs from the other methods since one cannot specify the number of clusters in advance. Fortunately, we proposed a cost function in the Bayesian optimization searching algorithm to approach the exact number of clusters.

In this section, we found that agglomerative clustering coupled with the UMAP achieved the best classification performance in all clustering methods, and the proposed methods achieved comparable performance compared to the baseline from the perspective of statistical significance ($p > 0.1$, Wilcoxon signed-rank test).

5.3. Comparison with Benchmarks

The agglomerative clustering exceeded the baseline by 12% (Table 4). We were inspired to compare the best method in our scheme with several benchmarks on the NBI-InfFrame [17,38,39].

We can infer from the statistical significance that our method (UMAP + Agglo) is comparable to the performance benchmarks (Figure 7). Our method (UMAP + Agglo) achieves identical performance to the current best benchmark in terms of the complete statistics of recall ($p = 1$, Wilcoxon signed-rank test).

So far, we can conclude the best method in our scheme, the agglomerative clustering coupled with UMAP (UMAP + Agglo), obtained a mean AUC = 96% with a standard deviation (±0.01). For detecting the informative frame (class *I*), the method achieved a 97% mean AUC (Figure 8a). In addition, we illustrated the relative confusion matrix of the method for visualization of the classification results (Figure 8).

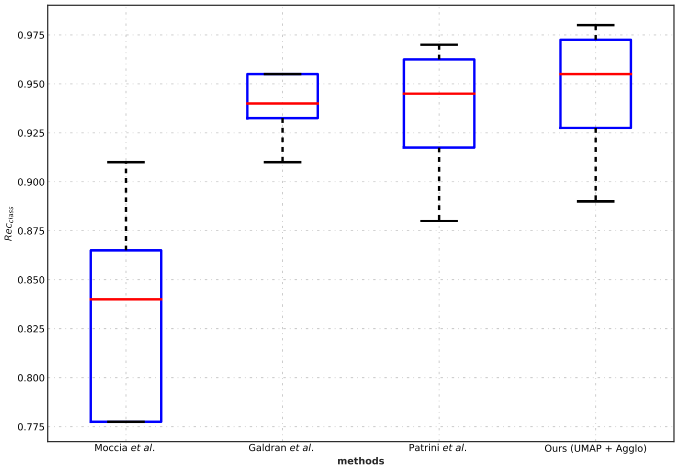
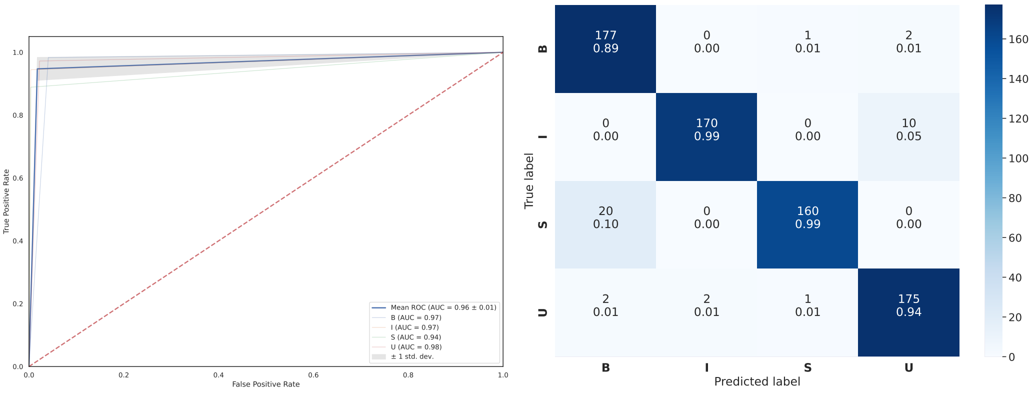


Figure 7. Boxplot of recall ($\mathbf{Rec}_{class}$) for comparing with benchmark studies. We compared our method (UMAP + Agglo) with [17,38,39] quantitatively using the NBI-InfFrame dataset for evaluation. The difference between the class-specific recall from ours and the other three methods is not statistically significant (relative *p*-value is 0.125, 1.000, 0.625, Wilcoxon signed-rank test). The overall median recall of the proposed method (UMAP + Agglo) outperformed Moccia et al. by 12% absolute.

Table 4. Classification performance of the state-of-the-art methods and our proposed method. Precision ($\mathbf{Prec}_{class}$), class-specific recall ($\mathbf{Rec}_{class}$), and F1-score ($\mathbf{F1}_{class}$) are reported for the four frame classes (**B**, **I**, **S**, **U**). Results from Moccia et al., (2018) [17] proposed SVM with the manually selected feature set, Patrini et al., (2020) [38] proposed VGG16 fine-tuned method, and Galdran et al., (2019) [39] proposed SqueezeNet-based method. The three metrics' median and interquartile range (IQR) are also reported.

	SVM [17]			**Fine-tuned SqueezeNet [39]**			**Fine-tuned VGG16 [38]**			**Ours (UMAP + Agglo)**		
	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$	$\mathbf{Prec}_{class}$	$\mathbf{Rec}_{class}$	$\mathbf{F1}_{class}$
B	0.76	0.83	0.79	**0.94**	0.94	0.94	0.92	0.96	0.94	0.89	0.98	0.93
I	0.91	0.91	0.91	0.97	**1.00**	0.98	0.97	0.97	0.97	0.99	0.94	0.97
S	0.78	0.62	0.69	0.93	**0.91**	0.91	0.93	0.88	0.91	**0.99**	0.89	0.94
U	0.76	0.85	0.80	**0.97**	0.94	0.95	0.92	0.93	0.93	0.94	0.97	0.95
Median	0.77	0.84	0.80	0.96	0.94	**0.95**	0.93	0.95	0.94	**0.97**	**0.96**	**0.95**
IQR	0.09	0.16	0.12	0.04	**0.05**	0.04	**0.03**	0.06	0.04	0.08	0.06	**0.03**



(**a**) (**b**)

Figure 8. Classification performance of the proposed method (UMAP + Agglo): (**a**) for quantitative analysis, the receiver operating characteristic (ROC) curves and the area under the ROC curve (AUC); the mean (±standard deviation) area under the ROC curve is reported by the solid blue lines (a grey area) in the legend. The area under the ROC (AUC) for each class is reported, too; (**b**) confusion matrix for the proposed method (UMAP + Agglo); the color bar on the right represents the number of frames in each class (B: blurred frames, I: informative frames, S: frames with saliva or specular reflections, U: underexposed frames).

5.4. Cluster Number Determination

We analyzed the classification performance of the proposed clustering methods, as we knew the types of informative and uninformative frames. The types of frames may not be predictable in actual clinical data. Thus, the definition of the informative frame does not exist. We can hypothesise that the exact number of classes of the NBI-InfFrmae dataset is unknown. In other words, the dataset is not well-classed. We introduced the Calinski-Harabasz Index (Equation (14)) to determine the optimal number of clusters.

We observed existing results matched the exact number of classes of the NBI-InfFrmae dataset. The Calinski–Harabasz Index score in the column of UMAP indicated the optimal number of clusters is four (Table 5). The result suggested that the proposed scheme can still achieve comparable performance in the situation of unknown labels of the NBI-InfFrame.

Table 5. Calinski–Harabasz Index (CH) analysis on the cluster number of the K-means coupled with three state-of-the-art dimensionality reduction approaches. Evaluated methods reported without dimensionality reduction (vanilla), PCA, t-SNE, and UMAP. The tested cluster number ranges from 2 to 6.

n_cluster	Vanilla↑	PCA ↑	t-SNE ↑	UMAP ↑
2	**185.55**	**226.95**	1445.76	1097.70
3	148.47	186.72	1470.55	1285.14
4	122.82	157.26	**1696.34**	**1529.70**
5	104.90	136.15	1564.44	1384.12
6	93.23	122.66	1500.42	1357.81

6. Discussion

The purpose of informative frame selection is to detect the informative frames among all kinds of frames. As the class of the NBI-InfFrame is well-defined in advance, we intentionally hide the labels of the dataset and proposed an unsupervised learning scheme. Finally, we evaluated the classification performance of the proposed methods by unfold the labels.

There are several reasons the agglomerative clustering achieved the best classification performance in our scheme:

- We extracted the features using a suitable scale convolutional neural network.
- We employed a persevering global-structure features method, which keeps a minimal number of features for the subsequent clustering.
- The agglomerative clustering maintained a bottom-up fashion dendrogram, which is efficient for the small dataset.

Since the difference in the classification performance between the clustering methods is tiny, we finally attributed the success of the proposed scheme to the introduced dimensionality reduction method, UMAP. Meanwhile, the proposed cost function for the Bayesian optimization searching algorithm is vital in finding optimal parameters in the vast space.

Still, we cannot infer from the classification results that the learning-based feature exaction is superior to the criterion-based one. Most importantly, we revealed an unsupervised scheme that achieved a comparable classification performance to the supervised learning methods without defining types of frames. However, there are several drawbacks to this work:

- The introduced metric for cluster number determination needs to be further demonstrated in the dataset close to the clinical setting.
- The proposed automatic cluster labeling algorithm is conditioned on the number of clusters, which should be identical to the defined number of the class.
- The cost function in Bayesian optimization aims to find the best average recall of all classes; thus, the time consumption of the searching algorithm is enormous (Appendix C).

7. Conclusions

In this work, we developed a novel unsupervised framework, integrated with the neural network and dimensionality reduction methods coupled with clustering methods, that can distinguish the informative and uninformative frames from laryngoscopic videos. An automatic cluster labeling algorithm and a cost function for the Bayesian optimization are proposed to manifest the classification performance of the framework.

Several experiments were conducted. The comparison result of the different dimensionality reduction methods showed that the t-SNE and UMAP are more suitable than PCA in our scheme. Furthermore, the UMAP best fits the 4 clusters with an average silhouette score of 0.5, while the t-SNE is 0.48. The four clustering methods coupled with UMAP all obtained comparable performance to the baseline. The comparison among the four clustering methods coupled with UMAP indicated that agglomerative clustering achieved the best classification performance. An overall median classification recall of 96% among four frame classes was achieved with 12% over the baseline. Informative video frames were classified with a recall of 94%. Such performance is comparable to the current optimal benchmark from the statistical significance perspective. Moreover, the Calinski-Harabasz Index in the t-SNE and UMAP separately indicates the optimal number of the cluster is the same as the class number. This evidence motivates the application of the proposed scheme to vanilla clinical data.

This work stands on the experiments of the NBI-InfFrame; further evaluation experiments on the different datasets can be conducted for the community. In other words, the effort of the methods to improve the clinical data quality is worthwhile but not only devoted to machine learning algorithms with manual labeling.

Author Contributions: Conceptualization, L.Z. and L.W. (Linjie Wu); methodology, L.Z.; software, L.Z.; validation, L.Z.; visualization, L.Z.; writing—original draft preparation, L.Z.; writing—review and editing, L.Z., L.W. (Liangzhuang Wei) and Y.L.; supervision, L.W. (Linjie Wu) and Y.L.; project administration, H.W. and Y.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research is supported by the Science and Technology Commission of Shanghai Municipality of China (Grant NO 20Y11901900) and the Health and Family Planning Commission of Shanghai Municipality of China (Grant NO 2019SY059).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data sharing is applicable to this article.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. NBI-InfFrmaes Dataset

Table A1. NBI-InfFrames dataset. The dataset is split into three folders, and each folder contains six videos (video ID) and 60 images for each class (I, B, S, U). I: informative frame; B: blurred frame; S: frame with saliva or specular reflections; U: underexposed frame.

	Video ID	**I**	**B**	**S**	**U**
Fold 1	1	10	10	20	11
	2	10	0	6	9
	3	10	0	0	2
	4	10	40	23	20
	5	10	10	11	3
	6	10	0	0	15
	total	60	60	60	60

Table A1. *Cont.*

	Video ID	I	B	S	U
Fold 2	7	10	28	19	0
	8	10	8	21	5
	9	10	3	10	10
	10	10	21	10	16
	11	10	0	0	14
	12	10	0	0	15
	total	60	60	60	60
Fold 3	13	10	17	0	10
	14	10	21	34	22
	15	10	0	11	10
	16	10	0	9	5
	17	10	12	0	2
	18	10	10	6	11
	total	60	60	60	60

Appendix B. Understanding UMAP from Cost Function

Uniform Manifold Approximation and Projection (UMAP) is a recently proposed manifold learning technique for dimensionality reduction. Understanding it can be based on the t-SNE (Equation (4)). we can rewrite the cost function of t-SNE as

$$C_{t-SNE} = \sum_{i \neq j} p_{ij} \log p_{ij} - p_{ij} \log q_{ij}. \tag{A1}$$

Instead of minimizing the original KullbackLeibler divergence, LargeVis [57] maximizes a likelihood function,

$$C_{LV} = \sum_{i \neq j} p_{ij} \log w_{ij} + \gamma \sum_{i \neq j} \log\left(1 - w_{ij}\right). \tag{A2}$$

where the w_{ij} is derived from low-dimensional space approximated by the Student t-distribution $q_{ij} = \frac{w_{ij}}{\sum_{k \neq l} w_{kl}}$, thus $w_{ij} = \left(1 + \|y_i - y_j\|_2^2\right)^{-1}$.

Eventually, we can write the cost function of UMAP as

$$C_{UMAP} = \sum_{i \neq j} v_{ij} \log\left(\frac{v_{ij}}{w_{ij}}\right) + \left(1 - v_{ij}\right) \log\left(\frac{1 - v_{ij}}{1 - w_{ij}}\right), \tag{A3}$$

where v_{ij} is derived from the Gaussian distribution $p_{j|i} = \frac{v_{j|i}}{\sum_{k \neq i} v_{k|i}}$ in the high-dimensional space, $v_{j|i} = \exp\left(-\|x_i - x_j\|_2^2 / 2\sigma_i^2\right)$,

$$v_{ij} = \left(v_{j|i} + v_{i|j}\right) - v_{j|i} v_{i|j}. \tag{A4}$$

Put all together, the cost function of UMAP can be presented in the non-constant form

$$C_{UMAP} = \sum_{i \neq j} v_{ij} \log v_{ij} + \left(1 - v_{ij}\right) \log\left(1 - v_{ij}\right) - v_{ij} \log w_{ij} - \left(1 - v_{ij}\right) \log\left(1 - w_{ij}\right). \tag{A5}$$

Appendix C. Time Consumption Analysis

Appendix C.1. Feature Extraction

The computational time of the proposed feature extraction method (Section 3.1) is ∼0.03 s for 720 images, while the one image computational time is ∼0.03 s (Moccia et al. (2018) [17]).

Appendix C.2. Hyperparameter Tuning

The proposed cost function used for Bayesian Optimization (BO) searching (Section 4.4).

Table A2. The parameters and the CPU time of the clustering methods coupled with UMAP.

	K-Means	Agglomerative	Spectral	DBSCAN
Parameters (num)	4320	35,840	26,880	54,000
CPU time (sec)	13,078	37,523	37,039	19,940
BO search (step)	1000	5000	3000	2000
CPU time (sec) /per step	13.08	7.50	12.35	9.97

References

1. Bradley, P.J.; Piazza, C.; Paderno, A. A Roadmap of Six Different Pathways to Improve Survival in Laryngeal Cancer Patients. *Curr. Opin. Otolaryngol. Head Neck Surg.* **2021**, *29*, 65–78. [CrossRef] [PubMed]
2. Lauwerends, L.J.; Galema, H.A.; Hardillo, J.A.U.; Sewnaik, A.; Monserez, D.; van Driel, P.B.A.A.; Verhoef, C.; Baatenburg de Jong, R.J.; Hilling, D.E.; Keereweer, S. Current Intraoperative Imaging Techniques to Improve Surgical Resection of Laryngeal Cancer: A Systematic Review. *Cancers* **2021**, *13*, 1895. [CrossRef]
3. Sasco, A.; Secretan, M.; Straif, K. Tobacco Smoking and Cancer: A Brief Review of Recent Epidemiological Evidence. *Lung Cancer* **2004**, *45*, S3–S9. [CrossRef] [PubMed]
4. Brawley, O.W. The Role of Government and Regulation in Cancer Prevention. *Lancet Oncol.* **2017**, *18*, e483–e493. [CrossRef] [PubMed]
5. Zuo, J.J.; Tao, Z.Z.; Chen, C.; Hu, Z.W.; Xu, Y.X.; Zheng, A.Y.; Guo, Y. Characteristics of Cigarette Smoking without Alcohol Consumption and Laryngeal Cancer: Overall and Time-Risk Relation. A Meta-Analysis of Observational Studies. *Eur. Arch. Oto-Rhino-Laryngol.* **2017**, *274*, 1617–1631. [CrossRef]
6. Zhou, X.; Tang, C.; Huang, P.; Mercaldo, F.; Santone, A.; Shao, Y. LPCANet: Classification of Laryngeal Cancer Histopathological Images Using a CNN with Position Attention and Channel Attention Mechanisms. *Interdiscip. Sci. Comput. Life Sci.* **2021**, *13*,666–682.
7. Xiong, H.; Lin, P.; Yu, J.G.; Ye, J.; Xiao, L.; Tao, Y.; Jiang, Z.; Lin, W.; Liu, M.; Xu, J.; et al. Computer-Aided Diagnosis of Laryngeal Cancer via Deep Learning Based on Laryngoscopic Images. *EBioMedicine* **2019**, *48*, 92–99. [CrossRef]
8. Cancer.Net. Laryngeal and Hypopharyngeal Cancer: Statistics. Available online: https://www.cancer.net/cancer-types/laryngeal-and-hypopharyngeal-cancer/statistics (accessed on 30 April 2022).
9. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer Statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30. [CrossRef]
10. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer Statistics, 2016. *CA Cancer J. Clin.* **2016**, *66*, 7–30. [CrossRef]
11. Steuer, C.E.; El-Deiry, M.; Parks, J.R.; Higgins, K.A.; Saba, N.F. An Update on Larynx Cancer. *CA: Cancer J. Clin.* **2017**, *67*, 31–50. [CrossRef]
12. Marioni, G.; Marchese-Ragona, R.; Cartei, G.; Marchese, F.; Staffieri, A. Current Opinion in Diagnosis and Treatment of Laryngeal Carcinoma. *Cancer Treat. Rev.* **2006**, *32*, 504–515. [CrossRef]
13. Paderno, A.; Holsinger, F.C.; Piazza, C. Videomics: Bringing Deep Learning to Diagnostic Endoscopy. *Curr. Opin. Otolaryngol. Head Neck Surg.* **2021**, *29*, 143–148. [CrossRef]
14. Ni, X.G.; Zhang, Q.Q.; Wang, G.Q. Narrow Band Imaging versus Autofluorescence Imaging for Head and Neck Squamous Cell Carcinoma Detection: A Prospective Study. *J. Laryngol. Otol.* **2016**, *130*, 1001–1006. [CrossRef]
15. Qureshi, W.A. Current and Future Applications of the Capsule Camera. *Nat. Rev. Drug Discov.* **2004**, *3*, 447–450. [CrossRef]
16. Ali, H.; Sharif, M.; Yasmin, M.; Rehmani, M.H.; Riaz, F. A Survey of Feature Extraction and Fusion of Deep Learning for Detection of Abnormalities in Video Endoscopy of Gastrointestinal-Tract. *Artif. Intell. Rev.* **2020**, *53*, 2635–2707. [CrossRef]
17. Moccia, S.; Vanone, G.O.; Momi, E.D.; Laborai, A.; Guastini, L.; Peretti, G.; Mattos, L.S. Learning-Based Classification of Informative Laryngoscopic Frames. *Comput. Methods Programs Biomed.* **2018**, *158*, 21–30. [CrossRef]
18. Paolanti, M.; Frontoni, E. Multidisciplinary Pattern Recognition Applications: A Review. *Comput. Sci. Rev.* **2020**, *37*, 100276. [CrossRef]
19. Maghsoudi, O.H.; Talebpour, A.; Soltanian-Zadeh, H.; Alizadeh, M.; Soleimani, H.A. Informative and Uninformative Regions Detection in WCE Frames. *J. Adv. Comput.* **2014**, *3*, 12–34. [CrossRef]
20. Ren, J.; Jing, X.; Wang, J.; Ren, X.; Xu, Y.; Yang, Q.; Ma, L.; Sun, Y.; Xu, W.; Yang, N.; et al. Automatic Recognition of Laryngoscopic Images Using a Deep-Learning Technique. *Laryngoscope* **2020**, *130*, E686–E693. [CrossRef]
21. Cho, W.K. Comparison of Convolutional Neural Network Models for Determination of Vocal Fold Normality in Laryngoscopic Images. *J. Voice* **2020**, *36*, 590–598. [CrossRef] [PubMed]
22. Cho, W.K.; Lee, Y.J.; Joo, H.A.; Jeong, I.S.; Choi, Y.; Nam, S.Y.; Kim, S.Y.; Choi, S.H. Diagnostic Accuracies of Laryngeal Diseases Using a Convolutional Neural Network-Based Image Classification System. *Laryngoscope* **2021**, *131*, 2558–2566. [CrossRef] [PubMed]
23. Yin, L.; Liuy, Y.; Pei, M.; Li, J.; Wu, M.; Jia, Y. Laryngoscope8: Laryngeal Image Dataset and Classification of Laryngeal Disease Based on Attention Mechanism. *Pattern Recognit. Lett.* **2021**, *150*, 207–213. [CrossRef]

24. Yao, P.; Usman, M.; Chen, Y.H.; German, A.; Andreadis, K.; Mages, K.; Rameau, A. Applications of Artificial Intelligence to Office Laryngoscopy: A Scoping Review. *Laryngoscope* **2021**, *132*, 1993–2016. [CrossRef]
25. Kuo, C.F.J.; Chu, Y.H.; Wang, P.C.; Lai, C.Y.; Chu, W.L.; Leu, Y.S.; Wang, H.W. Using Image Processing Technology and Mathematical Algorithm in the Automatic Selection of Vocal Cord Opening and Closing Images from the Larynx Endoscopy Video. *Comput. Methods Programs Biomed.* **2013**, *112*, 455–465. [CrossRef]
26. Atasoy, S.; Mateus, D.; Meining, A.; Yang, G.Z.; Navab, N. Endoscopic Video Manifolds for Targeted Optical Biopsy. *IEEE Trans. Med. Imaging* **2012**, *31*, 637–653. [CrossRef]
27. Bashar, M.K.; Mori, K.; Suenaga, Y.; Kitasaka, T.; Mekada, Y. Detecting Informative Frames from Wireless Capsule Endoscopic Video Using Color and Texture Features. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention—MICCAI, New York, NY, USA, 6–10 September 2008*; Metaxas, D., Axel, L., Fichtinger, G., Székely, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 603–610. ._72. [CrossRef]
28. Iakovidis, D.; Tsevas, S.; Polydorou, A. Reduction of Capsule Endoscopy Reading Times by Unsupervised Image Mining. *Comput. Med. Imaging Graph.* **2010**, *34*, 471–478. [CrossRef]
29. Perperidis, A.; Akram, A.; Altmann, Y.; McCool, P.; Westerfeld, J.; Wilson, D.; Dhaliwal, K.; McLaughlin, S. Automated Detection of Uninformative Frames in Pulmonary Optical Endomicroscopy. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 87–98. [CrossRef]
30. Yao, H.; Zhang, X.; Zhou, X.; Liu, S. Parallel Structure Deep Neural Network Using CNN and RNN with an Attention Mechanism for Breast Cancer Histology Image Classification. *Cancers* **2019**, *11*, 1901. [CrossRef]
31. Krotkov, E. Focusing. *Int. J. Comput. Vis.* **1988**, *1*, 223–237. [CrossRef]
32. Bashar, M.; Kitasaka, T.; Suenaga, Y.; Mekada, Y.; Mori, K. Automatic Detection of Informative Frames from Wireless Capsule Endoscopy Images. *Med. Image Anal.* **2010**, *14*, 449–470. [CrossRef]
33. Park, S.; Sargent, D.; Spofford, I.; Vosburgh, K.; A-Rahim, Y. A Colon Video Analysis Framework for Polyp Detection. *IEEE Trans. Biomed. Eng.* **2012**, *59*, 1408–1418. [CrossRef] [PubMed]
34. Kuo, C.F.J.; Kao, C.H.; Dlamini, S.; Liu, S.C. Laryngopharyngeal Reflux Image Quantization and Analysis of Its Severity. *Sci. Rep.* **2020**, *10*, 10975. [CrossRef]
35. Kuo, C.F.J.; Lai, W.S.; Barman, J.; Liu, S.C. Quantitative Laryngoscopy with Computer-Aided Diagnostic System for Laryngeal Lesions. *Sci. Rep.* **2021**, *11*, 10147. [CrossRef]
36. Islam, A.B.M.R.; Alammari, A.; Oh, J.; Tavanapong, W.; Wong, J.; de Groen, P.C. Non-Informative Frame Classification in Colonoscopy Videos Using CNNs. In Proceedings of the 2018 3rd International Conference on Biomedical Imaging, Signal Processing. Association for Computing Machinery, Bari, Italy, 11–12 October 2018; pp. 53–60. [CrossRef]
37. Yao, H.; Stidham, R.W.; Soroushmehr, R.; Gryak, J.; Najarian, K. Automated Detection of Non-Informative Frames for Colonoscopy through a Combination of Deep Learning and Feature Extraction. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 2402–2406. [CrossRef]
38. Patrini, I.; Ruperti, M.; Moccia, S.; Mattos, L.S.; Frontoni, E.; De Momi, E. Transfer Learning for Informative-Frame Selection in Laryngoscopic Videos through Learned Features. *Med. Biol. Eng. Comput.* **2020**, *58*, 1225–1238. [CrossRef]
39. Galdran, A.; Costa, P.; Campilho, A. Real-Time Informative Laryngoscopic Frame Classification with Pre-Trained Convolutional Neural Networks. In Proceedings of theIEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019 pp. 87–90. [CrossRef]
40. Yao, P.; Witte, D.; Gimonet, H.; German, A.; Andreadis, K.; Cheng, M.; Sulica, L.; Elemento, O.; Barnes, J.; Rameau, A. Automatic Classification of Informative Laryngoscopic Images Using Deep Learning. *Laryngoscope Investig. Otolaryngol.* **2022**, *7*, 460–466. [CrossRef]
41. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.; van Ginneken, B.; Sánchez, C.I. A Survey on Deep Learning in Medical Image Analysis. *Med. Image Anal.* **2017**, *42*, 60–88. . [CrossRef]
42. Ravì, D.; Wong, C.; Deligianni, F.; Berthelot, M.; Andreu-Perez, J.; Lo, B.; Yang, G.Z. Deep Learning for Health Informatics. *IEEE J. Biomed. Health Inform.* **2017**, *21*, 4–21. [CrossRef]
43. Van Der Maaten, L.; Postma, E.; Van den Herik, J. Dimensionality reduction: A comparative review. *J. Mach. Learn. Res.* **2009**, *10*, 13.
44. Van Der Maaten, L.; Hinton, G. Visualizing Data Using T-SNE. *J. Mach. Learn. Res.* **2008**, *9*, 2579–2605.
45. McInnes, L.; Healy, J.; Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* **2020**, arXiv:1802.03426.
46. Conlen, M.; Hohman, F. The Beginner's Guide to Dimensionality Reduction. In Proceedings of the Workshop on Visualization for AI Explainability (VISxAI) at IEEE VIS, Berlin, Germany, 21–26 October 2018.
47. Coenen, A.; Pearce, A. Understanding UMAP. Available online: https://pair-code.github.io/understanding-umap/ (accessed on 29 March 2022).
48. Kotsiantis, S.; Pintelas, P. Recent Advances in Clustering: A Brief Survey. *WSEAS Trans. Inf. Sci. Appl.* **2004**, *1*, 73–81.
49. Ackermann, M.R.; Blömer, J.; Kuntze, D.; Sohler, C. Analysis of Agglomerative Clustering. *Algorithmica* **2014**, *69*, 184–215. [CrossRef]
50. Jain, A.K.; Murty, M.N.; Flynn, P.J. Data Clustering: A Review. *ACM Comput. Surv.* **1999**, *31*, 264–323. . 331499.331504. [CrossRef]
51. Von Luxburg, U. A Tutorial on Spectral Clustering. *Stat. Comput.* **2007**, *17*, 395–416. [CrossRef]

52. Ester, M.; Kriegel, H.P.; Sander, J.; Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In Proceedings Second International Conference on Knowledge Discovery and Data Mining, Portland, OR, USA, 2–4 August 1996; Volume 96, pp. 226–231.
53. Caliński, T.; Harabasz, J. A Dendrite Method for Cluster Analysis. *Commun. Stat.* **1974**, *3*, 1–27. . 03610927408827101. [CrossRef]
54. Rousseeuw, P.J. Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis. *J. Comput. Appl. Math.* **1987**, *20*, 53–65. [CrossRef]
55. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
56. Turner, R.; Eriksson, D.; McCourt, M.; Kiili, J.; Laaksonen, E.; Xu, Z.; Guyon, I. Bayesian Optimization Is Superior to Random Search for Machine Learning Hyperparameter Tuning: Analysis of the Black-Box Optimization Challenge 2020. *arXiv* **2021**, arXiv:2104.10201. https://doi.org/10.48550/arXiv.2104.10201.
57. Tang, J.; Liu, J.; Zhang, M.; Mei, Q. Visualizing Large-scale and High-dimensional Data. In Proceedings of the 25th International Conference on World Wide Web, Montréal, Québec, QC, Canada, 11–15 April 2016; pp. 287–297. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.