

Computer-Aided-Diagnosis for Laryngeal Lesion Assessment: A Feature Extraction and Machine Learning Approach Applied on Enhanced Contact Endoscopy Images

Nazila Esmaeili

Vollständiger Abdruck der von der TUM School of Computation, Information and Technology der Technischen Universität München zur Erlangung einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz: Prof. Dr. Rüdiger Westermann

Prüfende der Dissertation:

1. Prof. Dr. Nassir Navab
2. Prof. Dr. Michael Friebe
3. Prof. Dr. Felix Nensa

Die Dissertation wurde am 21.06.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Computation, Information and Technology am 02.09.2024 angenommen.

Abstract

Over the last decade, advances in medical imaging technologies have led to the development of several image-based diagnostic techniques in clinical practice. These techniques aim to provide clinicians with high-quality, accurate, and detailed views of the targeted tissue or organ and provide patients with efficient care. However, it is found that there are new challenges related to analyzing and interpreting the presented information in the images, especially for less-experienced clinicians. Recently, numerous Computer-Aided-Diagnosis (CAD) systems have been introduced to tackle such issues and support clinicians in various diagnostic fields.

When it comes to laryngeal lesion diagnosis, Otolaryngologists utilize endoscopic imaging modalities to examine the status of the vocal fold before performing any surgical biopsy. This examination can provide them with noticeable information, such as the changes in morphology and distribution of sub-epithelial blood vessels in the vocal fold associated with the evolution of benign or malignant pathologies. Otolaryngologists can use magnified and enhanced endoscopic imaging techniques like Contact Endoscopy (CE) combined with Narrow Band Imaging (NBI) to get more detailed visualization of these vascular architectures. Such a modality allows them to perform a detailed examination of vascular changes that can indicate various laryngeal pathologies and reduce the chance of applying surgical biopsy. However, in the visual assessment of CE-NBI images, the similarity and complexity between the vascular patterns of benign and malignant pathologies raise the issue of subjective interpretation during diagnosis, which requires extensive learning from Otolaryngologists.

This thesis presents the development and validation of different feature extraction and Machine Learning (ML) based techniques for laryngeal lesion assessment using CE-NBI images. The methods are designed into two pipelines according to the requirements studied throughout a clinical evaluation with Otolaryngologists, where the subjective analysis of CE-NBI images was assessed and confirmed. In pipeline 1, handcrafted features are combined with ML classifiers to evaluate the visually detectable information on this uncharted imaging modality, including the correlation between laryngeal lesions and geometrical characteristics of vascular patterns, as well as textural attributes. The strategy in pipeline 2 involves the development of Deep Learning (DL) based solutions where the entire image is taken as the source of information. These approaches incorporate different architectures and categorize laryngeal lesions on CE-NBI images.

As data plays a crucial role in the procedure, clinical data acquisition and collection were conducted continuously throughout the development process to generate a data set of CE-NBI images. Following several classification scenarios on this data set, all methods demonstrated the importance and value of the investigated source of information in CE-NBI images for diagnosing laryngeal lesions. Therefore, applying these methods in a CAD system reduces the chance of subjective interpretation that causes the necessity of performing the surgical biopsy in clinical practice. With that, this thesis is positioned in the research area of CAD systems that offers a series of steps toward applying magnified imaging and Optical Biopsy for a more accurate and reliable minimally invasive larynx lesion assessment.

Zusammenfassung

Fortschritte in der medizinischen Bildgebung haben im vergangenen Jahrzehnt zur Entwicklung mehrerer bildgebungsbasierter Diagnosetechniken beigetragen. Diese Techniken zielen darauf ab, Ärzten hochqualitative, hochaufgelöste und detaillierte Sicht auf das betroffene Gewebe oder Organ bereitzustellen um somit eine effektive Patientenversorgung zu erzielen. Insbesondere weniger Erfahrene Ärzte sehen sich dennoch Herausforderungen bei der Analyse und Interpretation der dargestellten Information in den Bildern gegenüber. Zahlreiche computergestützte Diagnosesysteme (CAD, engl.: Computer-Aided-Diagnosis) wurden jüngst vorgestellt um diese Problematik anzugehen und die Ärzte in unterschiedlichen diagnostischen Feldern zu unterstützen.

Im Zusammenhang mit der Diagnose von Kehlkopfläsionen, nutzen Otolaryngologen endoskopische Bildgebungsverfahren um den Zustand der Stimmlippen zu untersuchen, bevor eine chirurgische Biopsie unternommen wird. Diese Untersuchung liefert wertvolle Informationen über Morphologische Änderungen und die Anordnung von subepithelialen Blutgefäßen in den Stimmlippen, die mit der Entstehung gut- oder bösartiger Pathologien einhergehen. Otolaryngologen können von Bildgebungstechniken, wie Kontaktendoskopie (KE) kombiniert mit Narrow Band Imaging (NBI, dt. Schmalband-Bildgebung) Gebrauch machen, die durch eine vergrößerte und verbesserte Darstellung eine detailliertere Visualisierung dieser vaskulären Architekturen ermöglichen. Dieses Verfahren erlaubt es eine eingehende Untersuchung vaskulärer Veränderungen durchzuführen, die unterschiedliche Pathologien der Stimmlippen indizieren können und so den Bedarf nach einer chirurgischen Biopsie reduzieren. Dennoch besteht bei der visuellen Bewertung, durch die Ähnlichkeit und Komplexität der vaskulären Muster von KE-NBI Abbildungen, die Problematik einer subjektiven Interpretation während der Diagnose, was umfangreiches Lernen von den Otolaryngologists erfordert.

In dieser Arbeit wird die Entwicklung und Validierung unterschiedlicher, auf Merkmalsextraktion und maschinellem Lernen basierender, Techniken zur Bewertung von Kehlkopfläsionen in KE-NBI vorgestellt. Die Methoden sind in zwei Vorgehenssträngen entsprechend der vorab eruierten Anforderungen konzipiert. Diese gehen aus einer klinischen Auswertung mit Otolaryngologen hervor, in der die subjektive Analyse von KE-NBI-Bildern ausgewertet und bestätigt wurde. Im Vorgehensstrang 1, werden manuell konstruierte Merkmale mit auf maschinellem Lernen (ML) basierenden Klassifikatoren kombiniert um die visuell detektierbaren Informationen, einschließlich der Korrelation zwischen Kehlkopfläsionen und geometrischen Charakteristika der vaskulären Muster, sowie Textureigenschaften, zu evaluieren. Die Strategie im Vorgehensstrang 2 beinhaltet die Entwicklung einer Deep Learning (DL) basierten Lösung bei der das Bild ganzheitlich als Informationsquelle eingesetzt wird. Dieser Ansatz umfasst unterschiedliche Architekturen und kategorisiert Kehlkopfläsionen auf KE-NBI Bildern.

Da Daten eine essentielle Rolle bei diesem Diagnoseverfahren einnehmen, wurden im Entwicklungsprozess kontinuierlich klinische Daten akquiriert und gesammelt um ein Datenset der KE-NBI-Bilder zusammenzustellen. In mehreren Klassifikationsszenarien, die auf dieses

Datenset angewandt wurden, zeigten alle Methoden die Relevanz und den Mehrwert der untersuchten Informationsquellen in KE-NBI-Bildern für die Diagnose von Kehlkopfläsionen auf. Daraus geht hervor, dass durch die Anwendung dieser Methoden in einem CAD-System die Wahrscheinlichkeit einer subjektiven Interpretation und die damit einhergehende Notwendigkeit einer chirurgischen Biopsie gesenkt werden. Damit ist diese Arbeit im Forschungsbereich von CAD-Systemen angesiedelt und stellt eine Reihe von Maßnahmen für eine erweiterte Bildgebung und optische Biopsie bereit um ein akkurateres und verlässliches minimalinvasives Bewerten von Kehlkopfläsionen zu ermöglichen.

*To my father, mother,
and brother*

With love and eternal appreciation

Acknowledgement

I would like to express my deepest gratitude to my two supervisors, Prof. Dr. Nassir Navab and Prof. Dr. Michael Friebe, for their invaluable guidance, support, and expertise throughout the course of this research project. Their dedication and insightful feedback have played a significant role in shaping the direction and quality of this thesis.

I would also like to extend my heartfelt appreciation to our clinical partners, Prof. Dr. Christoph Arens and Dr. Nikolaos Davaris, for their unwavering collaboration and support during the development of this research project. Their cooperation and willingness to share their expertise have greatly enhanced the value and relevance of this work.

Additionally, I would like to acknowledge the invaluable contributions of my two exceptional advisors, Dr. Alfredo Illanes and Dr. Axel Boese. Their wisdom, mentorship, and thoughtful guidance have been valuable in shaping my academic and professional growth. I am genuinely grateful for their continuous support and encouragement.

Last but not least, I would like to express my deepest gratitude to my father, mother and brother. Their unwavering love, motivation, and belief in my abilities have been a constant source of strength and inspiration throughout this journey. I am truly fortunate to have such incredible individuals in my life, and I am deeply thankful for their unwavering support.

Contents

- CONTENTS..... VIII**
- LIST OF FIGURES AND TABLES X**
- LIST OF ACRONYMS..... XI**
- CHAPTER 1 – INTRODUCTION..... 14**
 - 1.1 LARYNGEAL LESIONS – FROM ORIGIN TO CLINICAL PAINS..... 14
 - 1.1.1 *What does refer to laryngeal lesion?..... 15*
 - 1.1.2 *What can cause laryngeal lesion development?..... 16*
 - 1.1.3 *Laryngeal lesion diagnosis: how does it work?..... 16*
 - 1.1.4 *What option is available for laryngeal lesion treatment?..... 18*
 - 1.1.5 *What are the potential challenges in laryngeal lesion assessment?..... 19*
 - 1.2 MEDICAL IMAGING – BEGINNING OF AN ERA IN LARYNGEAL LESION ASSESSMENT..... 21
 - 1.2.1 *Endoscopy and clinical examination of laryngeal lesions..... 21*
 - 1.2.2 *Rigid versus flexible endoscopes 22*
 - 1.2.3 *Enhanced endoscopy and laryngeal lesion mucosal vascularization..... 23*
 - 1.2.4 *Medical imaging in advanced larynx cancer..... 24*
 - 1.2.5 *What can go wrong with medical imaging?..... 25*
 - 1.3 COMPUTER-BASED TECHNOLOGY – GAME CHANGER IN LARYNGEAL LESION ASSESSMENT..... 28
 - 1.3.1 *Computer-aided-diagnosis systems for endoscopic image analysis..... 28*
 - 1.3.2 *Laryngeal endoscopic image data sets..... 30*
 - 1.3.3 *Pre-processing strategies on laryngeal endoscopic images..... 32*
 - 1.3.4 *Conventional CAD systems for laryngeal endoscopic image..... 33*
 - 1.3.5 *DL-based CAD systems for laryngeal endoscopic image..... 36*
 - 1.3.6 *Laryngeal endoscopic image and hybrid CAD systems..... 41*
 - 1.3.7 *What is needed to improve the performance of CAD systems..... 41*
 - 1.4 MAGNIFYING ENDOSCOPY – A NEW ADVANCEMENT IN LARYNGEAL LESION ASSESSMENT..... 43
 - 1.4.1 *ME with contact endoscopy..... 44*
 - 1.4.2 *Enhanced CE and vocal fold mucosal vascularization..... 45*

| | |
|---|------------|
| 1.4.3 Challenges to integrate ECE into the clinical setting?..... | 47 |
| CHAPTER 2 – CONTRIBUTIONS..... | 49 |
| 2.1 MOTIVATION AND CONTRIBUTIONS..... | 49 |
| 2.2 CONTRIBUTION 1: NOVEL AUTOMATED VESSEL PATTERN CHARACTERIZATION OF LARYNX CONTACT ENDOSCOPIC VIDEO IMAGES..... | 54 |
| 2.2.1 Summary..... | 54 |
| 2.2.2 Contribution..... | 54 |
| 2.2.3 Novel Automated Vessel Pattern Characterization of Larynx Contact Endoscopic Video Images..... | 55 |
| 2.3 CONTRIBUTION 2: LARYNGEAL LESION CLASSIFICATION BASED ON VASCULAR PATTERNS IN CONTACT ENDOSCOPY AND NARROW BAND IMAGING: MANUAL VERSUS AUTOMATIC APPROACH..... | 67 |
| 2.3.1 Summary..... | 67 |
| 2.3.2 Contribution..... | 67 |
| 2.3.3 Laryngeal Lesion Classification Based on Vascular Patterns in Contact Endoscopy and Narrow Band Imaging: Manual versus Automatic Approach..... | 68 |
| 2.4 CONTRIBUTION 3: CYCLIST EFFORT FEATURES: A NOVEL TECHNIQUE FOR IMAGE TEXTURE CHARACTERIZATION APPLIED TO LARYNX CANCER CLASSIFICATION IN CONTACT ENDOSCOPY—NARROW BAND IMAGING..... | 81 |
| 2.4.1 Summary..... | 81 |
| 2.4.2 Contribution..... | 81 |
| 2.4.3 Cyclist Effort Features: A Novel Technique for Image Texture Characterization Applied to Larynx Cancer Classification in Contact Endoscopy—Narrow Band Imaging..... | 82 |
| 2.5 CONTRIBUTION 4: DEEP CONVOLUTION NEURAL NETWORK FOR LARYNGEAL CANCER CLASSIFICATION ON CONTACT ENDOSCOPY-NARROW BAND IMAGING... | 95 |
| 2.5.1 Summary..... | 95 |
| 2.5.2 Contribution..... | 95 |
| 2.5.3 Deep Convolution Neural Network for Laryngeal Cancer Classification on Contact Endoscopy-Narrow Band Imaging..... | 96 |
| CHAPTER 3 – CONCLUSION AND FUTURE WORK..... | 108 |
| CHAPTER 4 – REFERENCES | 115 |
| APPENDIX A – ABSTRACT OF PUBLICATIONS NOT DISCUSSED IN THE DISSERTATION..... | 125 |
| APPENDIX B – COMPARISON OF RELATED WORKS..... | 135 |

List of figures and tables

| | |
|---|-----|
| Figure 1.1. Anatomical structure of larynx in the head and neck region..... | 14 |
| Figure 1.2. The selection of common laryngeal histopathologies according to WHO classification. The level of severity increases from left to right | 15 |
| Figure 1.3. Clinical examination. (a): Indirect mirror laryngoscopy, and (b): Direct laryngoscopy..... | 17 |
| Figure 1.4. White light versus enhanced endoscopy. (a): WLE [26], (b): NBI [26], (c): SPIES Imaging [32], (d): Autofluorescence Imaging [30], (e): i-SCAN Imaging [33], and (f): Hyperspectral Imaging [34]..... | 24 |
| Figure 1.5. Laryngeal lesion assessment process in current clinical settings..... | 27 |
| Figure 1.6. The architecture of conventional versus DL-based CAD systems for endoscopic image analysis [49]..... | 29 |
| Figure 1.7. Examples of DL-based network architecture. (a): ResNets with 34 parameter layers, and (b): VGGNet with 19 convolutional layers [100]..... | 39 |
| Figure 1.8. Model scaling in EfficientNet. (a): Baseline network, (b) to (d): Conventional scaling that only increases one dimension of network width, depth, or resolution, and (e): Compound scaling method [102]..... | 40 |
| Figure 1.9. The contact endoscope with a forward-oblique telescope at 30° viewing angle..... | 45 |
| Figure 1.10. ECE imaging using NBI for different histopathologies of vocal fold. (a): Cyst, (b): Reinke’s edema, (c): Hyperkeratosis, (d): Papillomatosis, (e): Low-Grade Dysplasia, (f): High-Grade Dysplasia, (g): Carcinoma in Situ, and (h): SCC [122]..... | 46 |
| Figure 1.11. Application of CE-NBI during clinical examination of laryngeal lesion..... | 47 |
| Table B.1. Comparison of related works in the context of laryngeal endoscopic image classification..... | 135 |
| Figure B.1. Training and validation results of Method 4 on final CE-NBI image data set with 11144 images - accuracy graph of final models. (a): DenseNet121, (b): ResNet50V2, and (c): EfficientNetB0V2..... | 139 |
| Figure B.2. Testing results of Method 4 on final CE-NBI image data set with 11144 images - confusion matrix. (a): DenseNet121, (b): EfficientNetB0V2, (c): ResNet50V2, and (d): Ensemble model..... | 140 |

List of acronyms

| | |
|-------|---|
| LC | Larynx Cancer |
| WHO | World Health Organization |
| SCC | Squamous Cell Carcinoma |
| ENT | Ear, Nose, and Throat |
| MIS | Minimally Invasive Surgery |
| CT | Computed Tomography |
| MRI | Magnetic Resonance Imaging |
| US | Ultrasound |
| RE | Rigid Endoscope |
| TRL | Transoral Rigid Laryngoscopy |
| FE | Flexible Endoscope |
| TFL | Transnasal Flexible Laryngoscopy |
| WLE | Wight Light Endoscopy |
| IPCL | Intraepithelial Papillary Capillary Loops |
| NBI | Narrow Band Imaging |
| SPIES | Storz Professional Image Enhancement System |
| nm | Nanometer |
| LED | Light Emitting Diode |
| ELS | European Laryngological Society |
| LVC | Longitudinal Vascular Changes |
| PVC | Perpendicular Vascular Changes |
| MIA | Medical Image Analysis |
| MIC | Medical Image Computing |
| AI | Artificial Intelligence |
| ML | Machine Learning |
| DL | Deep Learning |
| ANN | Artificial Neural Network |
| CAD | Computer Aided Diagnosis |
| GI | Gastrointestinal |
| ROI | Region of Interest |
| CNN | Convolutional Neural Network |
| SR | Specular Reflection |
| CLAHE | Contrast Limited Adaptive Histogram Equalization |
| HOG | Histogram of Oriented Gradient |
| FODG | First-Order Derivatives of Gaussian |

| | |
|--------|---------------------------------|
| MF | Matched Filter |
| LBP | Local Binary Patterns |
| GLCM | Gray Level Cooccurrence Matrix |
| SVM | Support Vector Machine |
| RBF | Radial Basis Function |
| kNN | k-Nearest Neighbor |
| RF | Random Forests |
| NB | Naive Bayes |
| LDA | Linear Discriminant Analysis |
| MP | Multilayer Perceptron |
| SGD | Stochastic Gradient Descent |
| VGGNet | VGG Network |
| ResNet | Residual Network |
| HD | High Definition |
| ME | Magnifying Endoscopy |
| CE | Contact Endoscopy |
| ECE | Enhanced Contact Endoscopy |
| GF | Geometrical Features |
| CyEfF | Cyclist Effort Features |
| HGD | Histogram of Gradient Direction |
| RIA | Rotational Image Averaging |
| ANG | Angle |
| DIS | Distance |
| CUR | Curvature |

Chapter 1 – Introduction

1.1 Laryngeal lesions – from origin to clinical pains

The larynx - commonly referred to as the voice box - is a cartilaginous segment in the respiratory tract placed at the anterior aspect of the neck. The internal space of the larynx is divided into three main anatomical parts, as shown in Figure 1.1:

1. The supraglottic region forms an oval cavity and comprises five separate subsites, including the suprahyoid epiglottis, infrahyoid epiglottis, false vocal folds, arytenoids, and aryepiglottic folds.
2. The Glottis includes the true vocal folds themselves and the space between them known as the Rima Glottidis. The true vocal folds are made up of a layer of stratified squamous epithelium overlying the lamina propria, a gel-filled space that is comprised of a superficial, middle, and deep layer.
3. The subglottic region is the space below the glottis and extends from a horizontal plane to the end of the cricoid cartilage [1, 2].

The larynx is responsible for three primary physiological functions, including breathing through the Rima Glottidis, vibration to allow for speech, and airway protection during swallowing. However, this multi-function organ can be affected by a wide range of conditions that influence its performance. Among all the causes of laryngeal malfunction, such as infection, airway obstruction, neurologic disorders, and surgery, the appearance of a lesion turns the red light on and requires more investigation [3].

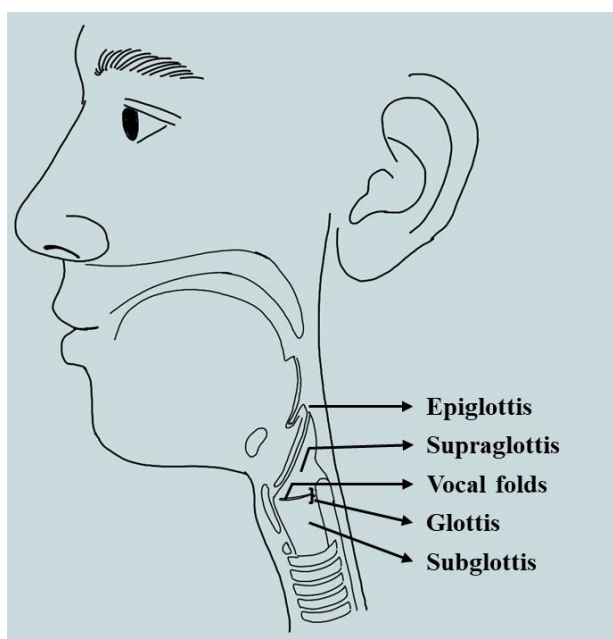


Figure 1.1. Anatomical structure of larynx in the head and neck region.

1.1.1 What does refer to laryngeal lesion?

Laryngeal lesion indicates a form of abnormal tissue growth in any epithelial and nonepithelial structures of the larynx. Laryngeal tumors include a broad spectrum of histopathologies localized in different regions in the larynx that can result into a certain form of organ malfunction. Several classifications based on the geometrical and histopathological attributes of the laryngeal lesions were proposed to structure this large scope of definitions. Nevertheless, the classification of World Health Organization (WHO) turns out to be the most comprehensive division that is mainly recommended for clinical practice. According to this classification, laryngeal lesions are divided into two main categories known as benign and malignant. Figure 1.2 represents an overview of this classification which also indicates the behavior of each histopathology towards malignancy [4, 5].

In this classification, the benign lesion refers to any mass of tissue in the larynx that does not present characteristics of malignancies. Many laryngeal histopathologies are classified into benign categories, including Cyst, Polyp, Reinke’s edema, Hemangioma, Nodule, Granuloma, Amyloidosis, Papillomatosis, Hyperplasia, Hyperkeratosis, and Low-Grade Dysplasia. On the other hand, the malignant laryngeal lesions include a form of malignancy originating from the larynx, where the cellular changes can lead to laryngeal precursor lesions, such as High-Grade Dysplasia and Carcinoma in Situ. If these conditions are not treated, they can eventually lead to an invasive stage known as Larynx Cancer (LC). Squamous Cell Carcinoma (SCC) is the most common histopathology variant of LC derived from the mucosal epithelium and accounts for 85-95% of all malignant tumors of the larynx. According to the studies, LC is the second most common cancer in the head and neck region, where almost two-thirds (75-80%) of these lesions are confined to the Glottic area. In addition, Low-Grade Dysplasia is usually categorized as benign lesions, while High-Grade Dysplasia and Carcinoma in Situ are considered malignant cases [6–8].

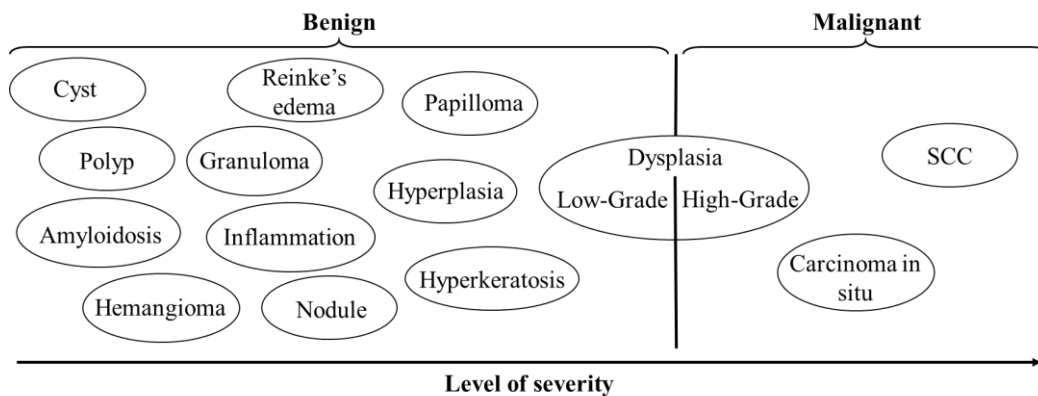


Figure 1.2. The selection of common laryngeal histopathologies according to WHO classification. The level of severity increases from left to right.

1.1.2 What can cause laryngeal lesion development?

LC is the most invasive type of lesion in larynx that requires immediate attention. There are several well-known risk factors related to the advancement of LC; however, all these risks cannot be delineated as the factors leading to the development of other benign laryngeal lesions. The reason behind this fact relies on the classification of risk factors of LC:

1.1.2.1 Epidemiological risk factors: The risk factors listed in this category can be leading causes for the occurrence of all types of laryngeal lesions. Chronic tobacco use and excessive alcohol ingestion are the more critical risks that can increase the chance of cancer by up to 6-fold to 30-fold. On the other hand, this category also includes occupational causes as other risk factors, including asbestos exposure, ionizing radiation, and ingestion of aerosols containing sulfuric acid [4, 8].

1.1.2.2 Histological precursor lesions risk factors: The risk factors in this category are defined explicitly as one of the critical causes for malignant laryngeal lesions, including LC. Accordingly, precursor lesions refer to a range of laryngeal pathologies with a high and variable chance of transforming into a malignancy from 1% to 40%. It is reported that around 90% of malignant tumors in the larynx are raised from this type of lesions. From the clinical point of view, although these pathologies can appear with benign characteristics, they should be considered as a group of risk factors for LC and must be monitored closely. Laryngeal Leukoplakia, Papillomatosis, and laryngeal Dysplasia, mostly found on the vocal fold in the Glottic region, are the primary concerns. Leukoplakia is a descriptive term that refers to a white plaque on the vocal fold and can correspond to various histopathological diagnoses from benign Keratosis to Dysplasia or LC. The spectrum of these cellular changes, from Low- to High-Grade Dysplasia, can expand to Carcinoma in Situ and finally develop into invasive LC. Therefore, the close observation of precursor lesions plays an essential role in the early detection of laryngeal malignancies [9, 10].

1.1.3 Laryngeal lesion diagnosis: how does it work?

All forms of laryngeal lesions can manifest with similar symptoms; however, hoarseness and dysphagia (difficulty in swallowing) are the most frequent symptoms in the presence of laryngeal tumors. According to the guidelines, all patients with any suspicious signs of these symptoms lasting for 3 to 4 weeks need to visit an Otolaryngologist and undergo two types of examination to achieve a precise diagnosis. The ideal is to combine these examinations with the patient's medical history to reduce the time of the diagnosis process and follow a treatment plan [4].

1.1.3.1 Clinical examination: In this stage, the patient will go through a full Ear, Nose, and Throat (ENT) examination. As the laryngeal mucosa is not accessible for a direct assessment, visible tissue changes in the larynx cannot be directly examined by Otolaryngologists. This problem is solved in two ways in the standard and conventional clinical examination:

1. The first solution is the indirect examination of the larynx called “indirect laryngoscopy,” which is performed by placing a small mirror in the back of the throat and angling it down towards the larynx (Figure 1.3-a). Indirect laryngoscopy usually is done in the doctor’s office to evaluate mucosal changes in the larynx, observe the Glottis with particular attention to involvement and mobility of vocal folds, and assess the possible invasions [11].
2. The second solution is called “direct laryngoscopy,” where Otolaryngologist inserts a blade shape instrument called a laryngoscope via the mouth and looks through it to examine the larynx (Figure 1.3-b). This procedure is usually conducted without anesthesia; however, the gag reflex can cause difficulties during the examination. Therefore, this procedure can also be done in the operating room under general anesthesia in the inpatient or outpatient centers. This way, the overall anatomical structures and mucosal changes can be examined with a direct view of the larynx [4].

The application of medical imaging in laryngeal lesion diagnosis is the complementary approach to these two solutions. Nowadays, medical imaging is the standard element of the clinical examination process of laryngeal lesions and is powerfully integrated into indirect and direct laryngoscopy procedures. A more detailed discussion of this topic is provided in Section 1.2.

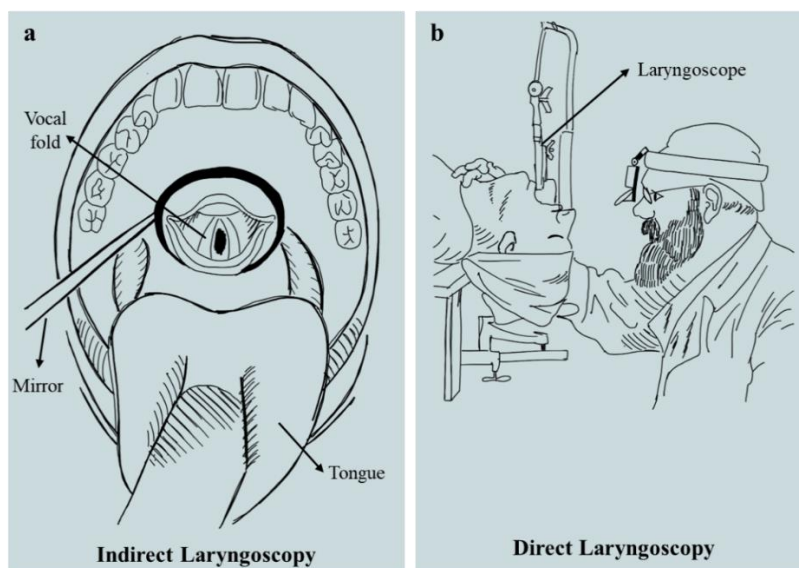


Figure 1.3. Clinical examination. (a): Indirect mirror laryngoscopy, and (b): Direct laryngoscopy.

1.1.3.2 Histopathological examination: The clinical examination offers a complete set of information regarding the larynx's anatomical characteristics and functional attributes; however, it cannot provide any information on the level of cellular changes of the suspicious tissue. As a rule, any mucosal changes in the larynx persisting for longer than 3 to 4 weeks are suspicious of malignancy and must be clarified. Hence, surgical biopsy followed by histopathological examination is the golden standard for a definitive diagnosis of laryngeal lesions. Nowadays, the procedure is combined with direct laryngoscopy (direct laryngoscopy biopsy), meaning that the immediate investigation is carried out before the tissue sample is taken. The surgical biopsy is usually performed under general anesthesia in the operation room of the hospitals [3].

1.1.4 What option is available for laryngeal lesion treatment?

A particular treatment method is often determined by lesion anatomy and resectability, perceived functional outcomes, patient preference, and departmental experience. Additionally, the type of lesion and level of invasiveness of the tumor can lead to different standard treatment strategies. In this regard, surgical techniques used to remove the lesion fall into abroad class, where all the guidelines suggest conducting the operations in the form of Minimally Invasive Surgery (MIS) such as transoral laryngoscopy as much as possible [12–14].

1.1.4.1 Benign laryngeal lesions: The benign lesions can be managed with one or a combination of several strategies, including conservative medical treatment, speech therapy, and surgical procedures. However, the symptom-free outcome may not be reachable in all the cases. It is recommended to set up some follow-up sessions to observe the patient's conditions and perform the surgery when necessary [6].

1.1.4.2 Precursor laryngeal lesions: Surgical excision combined with the follow-up sessions to observe the possible recurrence of the tumor is the primary treatment strategy to manage precursor laryngeal lesions. Laryngeal Dysplasia creates the most challenges for Otolaryngologists in this category because this histopathology shows a wide range of cellular changes that can be classified as benign or malignant. Low- to High-Grade Dysplasia treatment starts usually with surgical tumor removal. After the surgery, patients with Low-Grade Dysplasia have a long time between follow-up appointments, whereas patients with High-Grade Dysplasia have more frequent follow-up appointments during a more extended period [15].

1.1.4.3 Larynx cancer: In the case of LC, cancer staging is the preliminary step to prepare treatment planning. Early LC (T1-T2a, Stage I and II) can be managed via single modality treatment. The application of surgery as an introductory treatment approach gives the chance to reserve radiotherapy as a second-line option if the

tumor recurs. The management of moderately advanced LC (T2b-T3, stage II and IV M0) focuses on laryngeal preservation therapy as guidelines recommend it. This treatment includes the application of radiotherapy and chemotherapy to avoid significant lifestyle change and long-term morbidity associated with a laryngectomy. For the advanced stage LC, many centers are moving from conventional surgical treatment to therapeutic approaches that hopefully preserve the organ's function and survival with better quality of life for the patient. However, the high invasion rate in this stage means many patients will not be suitable for laryngeal preservation treatment and will require surgery in the form of a total laryngectomy followed by radiotherapy [4, 16].

1.1.5 What are the potential challenges in laryngeal lesion assessment?

For the past three decades, LC's incidence and prevalence have increased by 12%-24%. Meanwhile, the LC 5-year survival rate did not improve and faced a decrease of 3%. In 2018, Germany reported 3,310 new LC cases, including 1400 deaths, with the median age of 66 for women and 67 for men, which is earlier than usual for cancer overall. One of the main reasons behind this issue could be the late diagnosis of laryngeal lesions that happens in advanced stages. Although the treatment in the early stages is favorable, over 75% of LC cases are diagnosed at stage III or stage IV. On the other hand, the treatment in advanced stages is too aggressive and reduces the patient's social abilities, leading to decreased quality of life in various parts [7, 8, 17].

A significant portion of malignant laryngeal cases develops from the precursor lesions, classifying LC in a group of cancers suitable for early detection. Recent studies show that the close monitoring of these lesions and scheduled follow-up sessions after their surgical treatment can increase the chance of early detection of malignant changes and improve their prognosis. Furthermore, it is recommended to include repeated histopathological examination during the monitoring process for a more precise assessment of laryngeal tissue.

Screening programs in high-risk populations can be another way to reach the early detection of laryngeal lesions. Nevertheless, in many countries, including Germany, the health insurance screening program does not involve examining LC's risk groups. Additionally, there is a considerable lack of evidence regarding this kind of program's efficiency in reducing laryngeal lesions' incidence and mortality, making it challenging to convince health insurances to include it in their routine programs.

The key message behind all these evaluations and investigations points out the significant role of prompt diagnosis of laryngeal lesions in reaching optimum organ preservation. Now a question may arise: what are the main concerns during the conventional diagnostic procedure that can be an obstacle to early detection?

The first concern derives from the clinical examination procedure. The conventional indirect and direct laryngoscopy without the application of medical imaging could not be accurate enough in diverse conditions. Mirror laryngoscopy could be challenging for both the examiner and the patient and could provide only a limited view of the larynx and proximal trachea. Additionally, the direct laryngoscopy suffered from a lack of magnified view of the larynx, as the examiner looks through the laryngoscope with the naked eye.

The second concern is related to the histopathological examination. Although surgical biopsy remains the gold standard to find the final diagnosis of laryngeal lesions, it is an aggressive procedure for the patient and may cause laryngeal dysfunction. One issue with this procedure can involve the difficult differentiation of cancerous lesions after treatment due to changes in the characteristics of the tissue. Moreover, there is a chance of an uncertain diagnosis on a single biopsy that requires a second operation to collect more specimens.

Therefore, there is an essential need to instruct a less invasive diagnostic method in clinical practice that can provide sensitivity and specificity close to histopathology examination. The application of medical imaging for diagnosing and treating laryngeal lesions extended the opportunities to develop a minimally invasive technique for a more practical examination of the larynx.

1.2 Medical imaging – beginning of an era in laryngeal lesion assessment

Throughout the past century, medical imaging has faced significant advances. This development has allowed clinicians to see a range of body structures and functions to diagnose and manage different diseases and pathological conditions.

The use of medical imaging in diagnosing and treating laryngeal lesions has been integrated into several national guidelines, including in Germany. Due to this organ's particular location and anatomy, only specific imaging modalities have been introduced into the standard process of laryngeal lesion assessment. The surgical microscope was the first imaging tool used during the laryngoscopy procedure in 1960. A few years later, endoscopy imaging started to be integrated into the clinical and histopathological examination of the larynx. The application of this technique resulted in the generation of multiple endoscopy-based procedures, with the primary objective of providing better visualization of the examined region for the clinicians, along with more optimum care for the patient. From the 1980s, other imaging modalities such as Computed Tomography (CT), Magnetic Resonance Imaging (MRI), and Ultrasound (US) emerged to detect pathological and metastatic malignancies of advanced LC stages in the laryngeal region [18–20].

All these advances resulted in micro-laryngoscopy, endoscopic laryngoscopy, and laryngeal CT/MRI imaging to assess laryngeal mucosa and vocal fold's function more accurately and improve the clinical and histopathological examination of the laryngeal lesions.

1.2.1 Endoscopy and clinical examination of laryngeal lesions

Among all the available imaging modalities, endoscopy is the most popular and applicable technique to examine the characteristics of laryngeal pathologies. Endoscopy is an optical imaging technique that provides an inspection of the inner cavities of the human body. Several endoscopy systems are designed and built to provide digital imaging from the examined region, but the general structure of all systems includes four main components: an endoscope, a camera, a video processing unit, and a light source [21].

Nowadays, the clinical examination for the laryngeal lesion assessment should include endoscopic evaluation of the larynx in phonation and respiration position to evaluate discrete mucosa changes and functional aspects, such as vocal fold morphology. The great advantage of this application goes with precursor and cancerous lesions assessment, as the tissue changes can be visualized more accurately in the framework of endoscopic clinical examination. In the case of suspicious findings, more investigation, such as histopathological evaluation,

should be performed, where the specimen can be taken under microscopic guidance, known as micro-laryngoscopy [4].

Different endoscopy-based methods were introduced based on the components of the endoscopy system to assess laryngeal structure and function. It is conveyed that the combination of these options arrives in better performance that meets the clinicians' need. However, the application of them mainly relies on the facilities that the doctor's office and hospitals can offer [22].

1.2.2 Rigid versus flexible endoscopes

The rigid endoscope (RE) was one of the first tools integrated into the clinical examination of the laryngeal lesion. Transoral Rigid Laryngoscopy (TRL) is a minimally invasive procedure that operates along with the same principles as indirect laryngoscopy, where a 70° or 90° rigid endoscope is inserted into the larynx cavity via the mouth. Nowadays, the rigid endoscope application is also integrated into direct laryngoscopy usually performed under general anesthesia to provide a more detailed visualization of the examined area in the larynx. These types of rigid endoscope are longer and have various diameters and viewing angles [23].

Many patients with malignancies in the laryngeal region experience gag reflex, which creates challenges during TRL clinical examination. In this condition, the transnasal use of flexible endoscope (FE) can offer better visualization for laryngeal lesion assessment. Transnasal Flexible Laryngoscopy (TFL) is also a minimally invasive procedure where the flexible endoscope with a diameter of less than 4 mm is inserted through one nostril and is guided to the larynx via the lower nasal passage and pharynx [24].

In clinical examinations, TFL is widely available for the primary assessment of the larynx as it provides more tolerability for patients than TRL. However, this story may change when we focus on evaluating precursor and cancerous laryngeal lesions. Although TFL is a fast and valuable tool for screening and follow-up examination of laryngeal lesions, controlling the position of the flexible endoscope tip and maintaining the symmetry and orientation of the endoscope's image is more complex than the rigid endoscope. This issue can result in low-quality visualization of mucosa change, wrong tissue differentiation, and an inaccurate assessment. Moreover, the FEs are more sensitive to heat and chemical materials of the typical sterilization process, making the sterilization procedure time-consuming. Nevertheless, the new set of flexible endoscopes could solve this issue by providing immediate sterilization at the examination site, reducing the number of endoscopes required for routine TFL in a day [24–26]. The type of endoscope is one of many criteria of the endoscopy system that should be considered in the diagnosing process, as the light source plays a more critical role.

1.2.3 Enhanced endoscopy and laryngeal lesion mucosal vascularization

The endoscopic laryngoscopy can be performed under different light source conditions. White light is the conventional light source in endoscopy systems, providing sufficient illumination for the inspected path and cavity. However, the actual White Light Endoscopy (WLE) imaging has certain limitations, as it cannot precisely visualize the target region to distinguish between epithelial differences. In this case, the pre-operative clinical examination is not always in agreement with the result of histopathological examination and precursor and cancerous lesions might be overlooked [27, 28].

Considering these issues, a new area of endoscopy imaging complementary to WLE emerged, known as Biologic Endoscopy techniques, to provide a deeper insight into the structure of a target lesion and improve the visualization of the tumors' characteristics that are not visible in normal WLE. Enhanced endoscopy imaging is one of these techniques that became the most advanced modalities for better visualization of the laryngeal mucosal changes, tumor margins, and specifically vascularization networks. When laryngeal tissue begins with pathological changes, sub-epithelial vessels of the mucosa lose their typical architecture and tend to show more irregularities. This type of change may form intraepithelial papillary capillary loops (IPCL) around or inside the lesion, which is a critical indicator of the development of precursor and malignant tumors. Enhanced-endoscopy techniques can visualize this variation of vascular structures in real-time as a piece of complementary information during the clinical examination.

Narrow Band Imaging (NBI) [29], Autofluorescence Imaging [30], Hyperspectral Imaging [31], Storz Professional Image Enhancement System (SPIES) [32], and i-SCAN System [33] are the enhancements tools used in endoscopic laryngoscopy. However, the first place usually goes to NBI application due to its availability and efficiency in diagnosing. Figure 1.4 represents an example of each enhancement technique.

NBI is the most well-known and applicable type of enhanced-endoscopy technique in assessing overall head and neck lesions, especially laryngeal pathologies. The technology is provided by Olympus Corporation, Tokyo, Japan, and is recommended by all the guidelines to be used with rigid or flexible endoscopes during the clinical examination of laryngeal lesions, in the case of availability [4]. NBI mode is designed based on the principal correlation between the depth of light penetration and light wavelength, where the longer wavelength results in deeper penetration. With a broadband white light from a xenon or light emitting diode (LED) lamp, an optical NBI filter can allow the passage of a narrowband blue light centered at a range of 400–430 nm, parallel to a narrowband green light centered at a range of 525–555 nm [29].

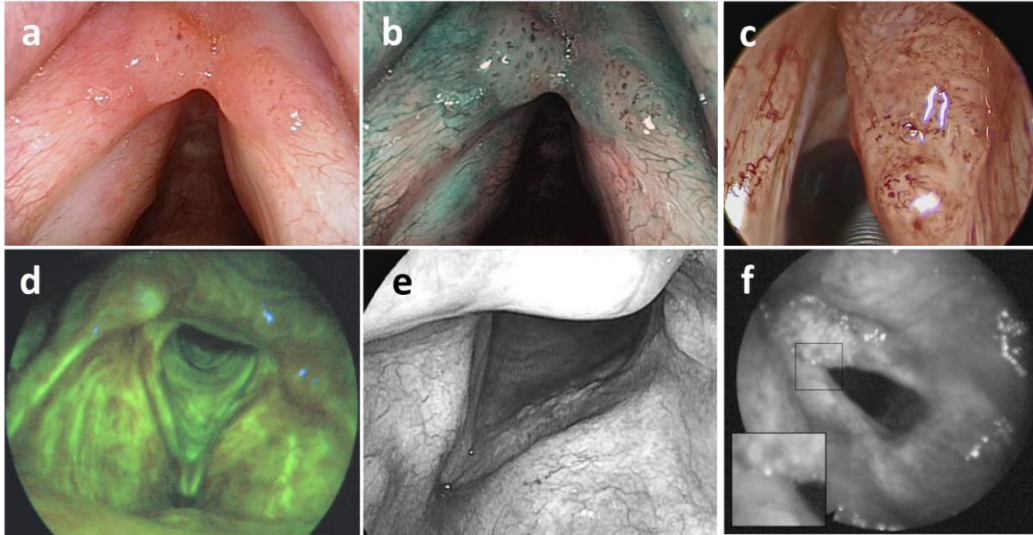


Figure 1.4. White light versus enhanced endoscopy. (a): WLE [26], (b): NBI [26], (c): SPIES Imaging [32], (d): Autofluorescence Imaging [30], (e): i-SCAN Imaging [33], and (f): Hyperspectral Imaging [34].

The narrowband blue light can penetrate through the normal mucosa and the sub-epithelial tissue. In contrast, narrowband green light can penetrate deeper into the tissue and depict vessels in the submucosa [29]. With the use of NBI filter, the evaluation of mucosal vascularization has become an added step into the clinical examination to move toward Biologic Endoscopy. Multiple classification systems were proposed to guide the examiner, where the European Laryngological Society (ELS) guideline is more straightforward and applicable to correlate vascular changes to type of laryngeal lesions. This classification describes benign vascularization as longitudinal vascular changes (LVC) with increase in the number and density of blood vessels along with changes in direction, while malignant lesions are characterized by their newly formed perpendicular vascular changes (PVC), identified as dot-like vessel or IPCLs. However, the vascular variations of some precursor lesions, such as Papillomatosis can create significant challenges on visual differentiation of lesion as they usually do not follow the pattern classification in the guidelines [35–37].

1.2.4 Medical imaging in advanced larynx cancer

Endoscopy is the first imaging option for performing the clinical examination aiming to provide a better perception of laryngeal lesions characteristics. However, due to the necessity of further investigation, such as histopathological examination and metastasis assessments, other imaging modalities have to be involved in laryngeal lesions assessment to overcome the limitation of endoscopy.

For decades, micro-laryngoscopy has been the leading approach to guide the process of surgical biopsy by providing deepened and stable visualization of the target lesion. Moreover, surgeons have used micro-laryngoscopy as the primary

visualization tool during minimally invasive surgical treatments of laryngeal lesions to resect the tumor with the optimum margin [20].

In the case of benign and malignant lesions as well as early-stage LC, the tumor mainly has a small size and is located at the superficial layer of the mucosa. Therefore, endoscopic imaging can provide a good visualization of the lesion and surrounding tissue. Nevertheless, the application of CT or MRI is crucial in patients with advanced stage LC. Both imaging techniques aim to evaluate the involvement of deep structures such as cartilages or lymph nodes. Although MRI can provide better visualization of the soft tissue, CT is the most available tool to evaluate the status of metastasis. CT is not affected by the motion artifacts, and its cross-sectional imaging allows evaluation of the intrinsic and deep soft tissues of the larynx as well as the cartilaginous skeleton [38–40].

1.2.5 What can go wrong with medical imaging?

Figure 1.5 illustrates the current laryngeal lesions assessment process in doctor's office and clinical settings. In the first step, a patient with specific symptoms visits an Otolaryngologist. As most doctors' offices are not equipped with high-tech endoscopy systems or cannot afford one, indirect mirror laryngoscopy is the best examination option. However, in the case of any suspicious finding, the Otolaryngologist must refer the patient to a more equipped center for more detailed investigations. In this step, the patient will undergo clinical examination supported by different medical imaging techniques, where surgical biopsy for histopathological evaluation of suspected lesions remains the primary diagnostic tool. During this step, most Otolaryngologists have a primary diagnosis of the type of lesions and usually decide to proceed with surgical tumor resection to save treatment time. This strategy mainly works for benign and early-stage LC cases as the lesions are minor and superficial. In this case, the biopsy outcome will guide the Otolaryngologists in planning the subsequent follow-up session and treatment procedures.

Now, one question may arise: why imaging, particularly endoscopy, has not improved the early detection of laryngeal lesions, leading to better preservation of the larynx and a reduction in the number of biopsies?

The application of imaging techniques in each phase of clinical examination may deviate according to the facilities of the doctor's offices or hospitals as well as the preferences of the examiner. Nevertheless, endoscopy is a critical element of everyday clinical examination due to its potential to provide a better visualization of the examined region. The evidence in the literature indicates the notable improvements offered by endoscopy in tissue differentiation that was missing in the conventional indirect and direct laryngoscopy. The main proof for this fact is the actual use of this imaging modality in the current diagnosis process [26, 29].

Besides the advantages, introducing a new imaging technique into a routine and standard clinical procedure may raise some challenges. The first concern is related to the learning process. The Otolaryngologist must learn how to operate the imaging system and, more importantly, interpret the newly acquired data with existing information. This issue can be a significant concern for younger doctors due to their level of experience. Therefore, they usually demand more extended supervision and learning [41, 42].

The next issue is associated explicitly with Biologic Endoscopy, a new vision that has been studied during the last decade. Among all the proposed methods in this field, only enhanced endoscopy is integrated into routine clinical procedures and is used by multiple hospitals in Europe. Several studies pointed out the prominent advantage of the NBI over WLE in better differentiation of mucosa changes in the larynx, that can be essential in variation of laryngeal lesions. NBI shows more precise performance in distinguishing mucosal changes of laryngeal precursor lesions, especially for Dysplasia, and early-stage malignant lesions. However, despite the improved accuracy and sensitivity, the specificity of using NBI versus WLE is not significantly higher in the differentiation of benign and malignant lesions [28, 37, 43–45]. One reason behind this concern is associated with the laryngeal tissue changes that human eyes may not recognize. These critical transitions occur in the epithelial layer via the presence of leukoplakia condition and in mucosal vascular structures. Although several guidance tools were proposed to help Otolaryngologists identify these indicators in enhanced endoscopy and interpret them for the diagnosis process, these mechanisms are considered complicated and time-consuming for clinical practice use and again raise the issue of the learning process [35, 46, 47].

All these factors work together to make surgical biopsy followed by histopathological examination the safest diagnosis option, where endoscopic laryngoscopy with the assessment of vocal fold function, micro-laryngoscopy, and CT/ MRI remain the cornerstones of the diagnostic workup. At this moment, computer-based solutions started entering clinical research settings to assist Otolaryngologists during the diagnosis process.

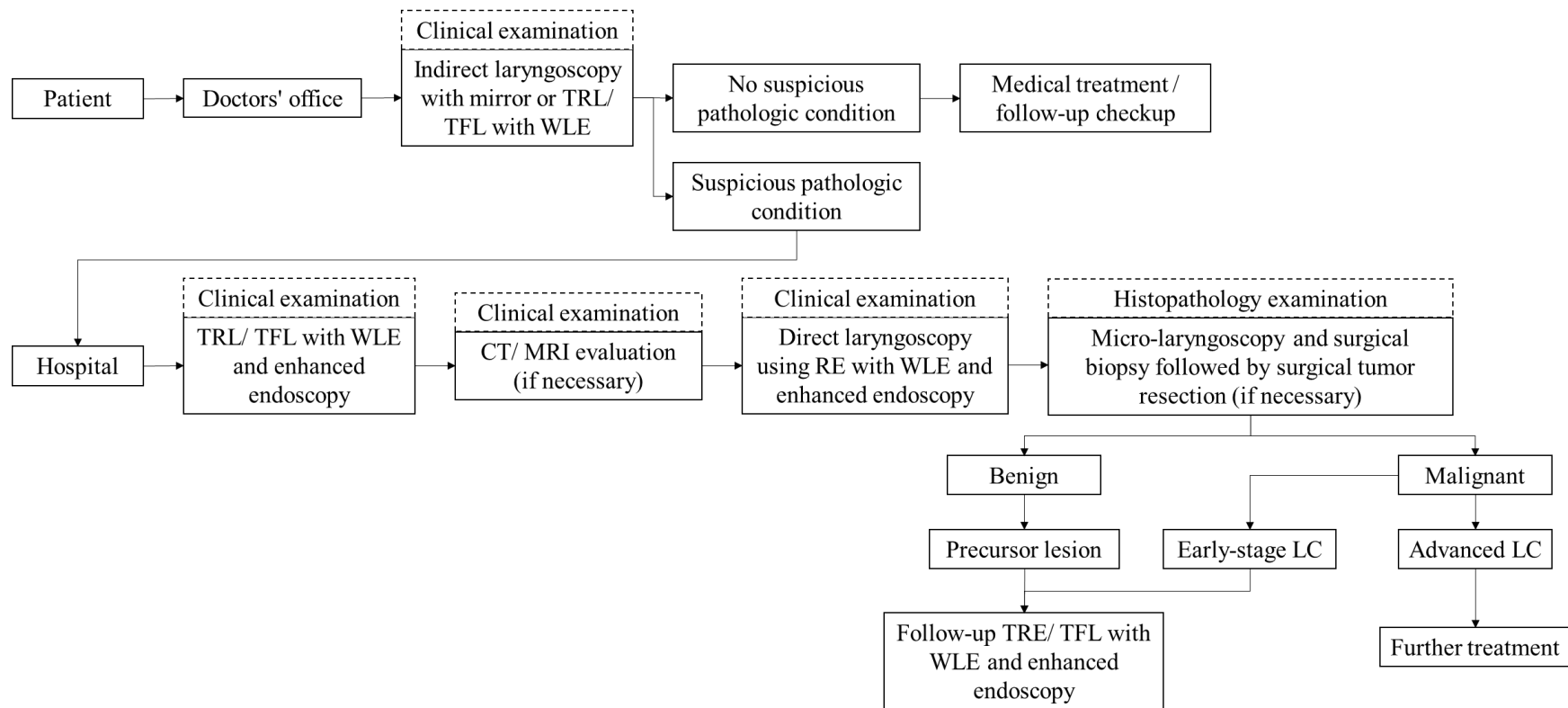


Figure 1.5. Laryngeal lesion assessment process in current clinical settings.

1.3 Computer-based technology – game changer in laryngeal lesion assessment

Along with the advantages of digitalization of the medical imaging modalities, imaging data complexity has introduced several challenges to the diagnostic and treatment process. Nevertheless, the availability of medical imaging big data has formed new opportunities for Medical Image Analysis (MIA). In this regard, Medical Image Computing (MIC) – an intradisciplinary branch of scientific computing – has emerged to develop (semi-)automated evaluation strategies for MIA and cope with the complexity of imaging data. Recently, the tremendous advances in MIC have been dramatically impacted by a field of data science – known as Machine Learning (ML) - that uses computers to perform tasks via building data-driven mathematical model-based strategies that optimize task performance.

ML, the most evolved form of Artificial Intelligence (AI) in medicine, has presented different types of these strategies for various MIA tasks, including image segmentation, image registration, image visualization and image classification. The supervised, unsupervised, semi-supervised, reinforcement, and evolutionary learning methods are the main categories of ML algorithms, while Deep Learning (DL) is the most advanced form of ML techniques that leads the third AI boom. DL has been developed by layering conventional Artificial Neural Networks (ANN) and has become a replacement for traditional image processing techniques in MIC.

1.3.1 Computer-aided-diagnosis systems for endoscopic image analysis

The rapid engagement of technology in MIA leads to the creation of Computer-Aided-Diagnosis (CAD) systems – as a subset of the MIC field – to put all the technology elements together and form a system that can provide real-time assistance for physicians in a variety of everyday clinical tasks. Nowadays, CAD systems are one of the most critical areas of research and development in MIA, where ML-based technologies could be potentially used to enhance human judgment capability or promote physicians' learning process.

From the early research activities in the 1960s, one spotlight of CAD systems was lesion detection and cancer diagnosis by comprehensive evaluation of medical images in a short time [48]. Endoscopic image analysis was no exception to this trend, where the primary effort was started on the development of CAD systems based on endoscopic images of the Gastrointestinal (GI) tract. Depending on the endoscopic technique, different sorts of machine-vision-based systems were developed during the last decade; nevertheless, the application of image processing and ML technologies derives specific architectures divided into two main categories: conventional CAD and DL-based CAD systems.

The general workflow of a CAD system in endoscopic image analysis can be divided into three main steps: image pre-processing, features extraction, and classification [49]. As presented in Figure 1.6, the main architectural difference between conventional and DL-based CAD systems is in the feature extraction and classification steps.

Endoscopic images commonly suffer from noise, such as lens distortions, illumination invariance, scale invariance, rotation invariance, and specular highlights. Therefore, the endoscopic images are pre-processed using different methods according to the image acquisition environment and noise condition. This step may include image normalization, contrast enhancement, image compression, image scaling, image rotation and color space transformation. In some systems, the image pre-processing also involves the region of interest (ROI) segmentation and identification [50].

In conventional endoscopic CAD development, features are manually engineered by a human data scientist using traditional image processing techniques – known as handcrafted features – that can be divided into two main groups. First are the frequency domain features, where the image is transformed into the frequency domain using a frequency transformation technique, and the features are extracted from the processed image. Second are the spatial domain features that refer to direct manipulation and extraction of information from the pixel values in a digital image. Color, texture, and morphological features are the main spatial and frequency feature sets in conventional endoscopic CAD systems. Sometimes, a feature selection step is added to the development workflow to choose suitable and relevant features from the high-level feature sets. Then, supervised classifiers, as the most popular ML strategy in the development of conventional endoscopic CAD systems, are used. They build the final set of features and their relationships – known as the predictive model – through the guidance of the predictive performance in the set of labeled data. This predictive model is the decision-making part of a conventional CAD systems [49, 51, 52].

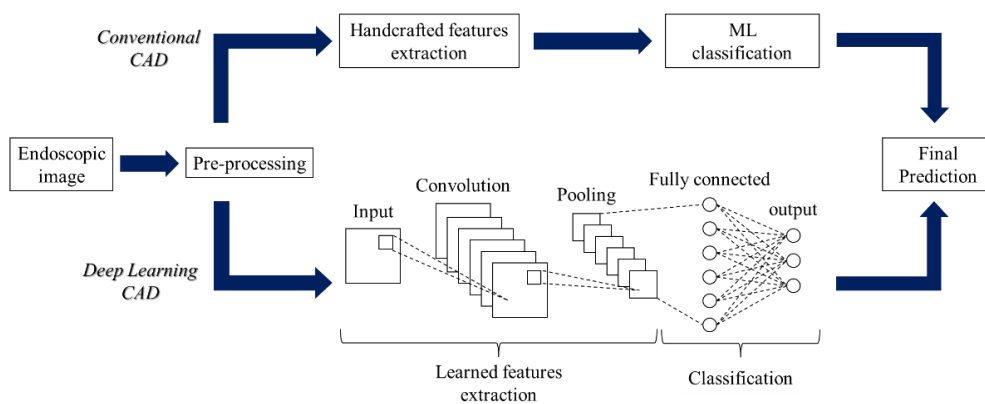


Figure 1.6. The architecture of conventional versus DL-based CAD systems for endoscopic image analysis [49].

In DL-based CAD systems, features – known as learned features – are automatically extracted from the endoscopic images without explicitly defining attributes. In this way, the image representation does not require hard coding for manual feature design. Moreover, DL-based methods have the ability to solve problems from end-to-end rather than breaking them down as in the case of traditional ML algorithms. They use the power of multi-layer ANN to learn compact information from the images and transform it into an output. The number and type of layers, their connection form, and output layer units are adopted according to the application of the DL method which resulted into a wide variety of DL architectures. Convolutional Neural Network (CNN) is the most successful deep model for image and time series data classification. The general architecture of a CNN has three main layers as the combination of convolutional and pooling layers followed by the fully connected layers. These layers are responsible for feature extraction, size reduction of the feature representation, as well as connecting the neurons between two different layers and classification. The application of CNN combined with supervised strategy is the most functional approach in developing endoscopic DL-based CAD systems. In these systems the final trained CNN model plays the role of decision-making section by providing objective guidance [49, 51, 53, 54].

In laryngeal endoscopic image analysis, the first problem is raised from similar visual characteristics in laryngeal tissue that human eyes may not differentiate. Consequently, another issue appeared due to the user-dependent nature of endoscopic image evaluation and its influence on the relatively long learning curve. These reasons resulted in several attempts at developing CAD systems to assess laryngeal lesions, established for laryngeal tissue characterization, lesions classification, and more advanced histopathology classification. All the proposed methods and architectures followed the footsteps of research paths on CAD system development in endoscopic images of the GI tract and fell into the general scheme of conventional or DL-based CAD systems. Apart from architectural differences, the conventional CAD and DL-based systems mainly differ on the amount of data required for their implementation and the number of needed pre-processing steps [55–58].

1.3.2 Laryngeal endoscopic image data sets

Data is the first and most critical element in creating any robust ML-based algorithm. The data set should be large enough to represent the sample data from the target population. In this way, the algorithm can adequately learn the properties of the population to generate a model that can evaluate any new unknown subject from the same population.

In the available investigations, the WL or NBI endoscopic images utilized for forming laryngeal CAD architectures are collected by single or multiple clinical

centers and are submitted mainly as private data sets. These images are acquired directly during the operation or post-extracted from the endoscopic videos captured throughout the procedure. The data annotation involves labeling image data based on the histopathological examination results or the visual assessment of one or more Otolaryngologists. It can also include the manual segmentation of a target region in the image performed by Otolaryngologists. The number of images on these evaluations varies from a minimum of 300 to a maximum of 20000 [59–73].

Publicly available data sets with a limited number of endoscopic images are also used for some of the evaluation phases. The first data set was introduced in 2017 and included 1320 patches with a dimension of 100*100 pixels of 330 NBI images of 33 patients affected by SCC. The data was labeled into four groups according to the tissue characterization of the vocal fold and are named tissue with IPCL-like vessels, leukoplakia, tissue with hypertrophic vessels, and healthy tissue [61]. The second data set was published in 2019 as part of an automated segmentation study. It comprises 535 color images of two patients captured with the stereo endoscope. Each data with a resolution of 512x512 pixels is manually segmented into seven different classes void, vocal folds, other tissue, glottal space, pathology, surgical tool, and intubation [63]. The third data set – known as Laryngoscope8 – is a laryngoscope image data set and aims to move toward an automatic diagnosis of laryngeal disease. It contains 1950 cases with 3057 images that professional Otolaryngologists categorized into eight classes: seven correspond to laryngeal disorders, and one category represents normal tissue [70].

Apart from the data availability, the portion of data in every class of data set should ideally be balanced to reach a robust predictive model. Sometimes, extra effort is required to collect data, especially in rare cases. MIC offers data augmentation as an easy-to-implement solution to save time and effort during data collection [74]. With this method, native data is transformed randomly to create new images. Typical interpretations include rotations, translations, cropping, and image contrast and brightness alterations [75]. However, the proper method selection is crucial to avoid any threat of producing redundancy in data.

Data separation is required in every evaluation scenario in ML development to assess the predictive model accurately. Therefore, the data set is typically divided into training and testing sets. In this way, the training and validation steps are performed on the training set, where two subsets of data are generated. After arriving at the optimum model, the testing set is introduced to the model as unseen data to test the model's performance. An alternative way is to split the data set directly into training, validation, and testing sets. In this configuration, the model is trained on the training set and then is validated and tested separately on the validation and testing sets, respectively [57].

There are three main techniques to create the training and testing sets out of the data set. The k-fold cross-validation technique divides the data set into k folds or groups, selects k-1 groups for the training set to perform the training, and tests the model with the remaining fold. This process is iterated k times, in which a different fold is reserved as the testing set. However, in the case of imbalanced data distribution in the data set, stratified k-fold cross-validation is a better technique than k-fold cross-validation. This method also splits the data into k folds and follows the same strategy for the training and testing process; however, it ensures that each data fold includes the same proportion of samples with a given label. Finally, hold-out cross-validation is a simple way to split the data set into two training and testing sets. Each set's portion depends on the overall size of the data set, but 80% for training and 20% for testing is the most common split using this method [76].

1.3.3 Pre-processing strategies on laryngeal endoscopic images

Among all the known pre-processing methodologies, the studies on laryngeal endoscopic images applied techniques that have already shown reliable performance on GI tract endoscopic image analysis. Furthermore, the level of complexity of the pre-processing methods is usually correlated to the size of the data set. For example, several reports showed that the studies with a limited number of images involved more pre-processing steps than those with larger data sets [58].

In conventional CAD systems, the pre-processing strategies aimed to either reduce the noise – especially specular reflection – or enhance some essential characteristics in the image or both. In this case, bilateral filter and anisotropic diffusion filtering followed by specular reflection (SR) masking are the two proposed approaches [59, 61]. The bilateral filter [77] is a non-linear filter that uses tonal weights beside the spatial weights to replace a pixel value in the image. The anisotropic diffusion [78] is a non-linear and space-variant transformation of the image and, similar to the bilateral filter, aims to reduce the noise and smooth the appearance while preserving the sharp edges. The studies in laryngeal lesion assessment dealt with the specular reflections issue, using adaptive thresholding in the HSV color space approach [79]. It automatically identified SR by exploiting intense brightness and low saturation regions and then masked it.

In DL-based CAD systems, the pre-processing phase primarily focuses on preparing images according to the conditions of the network, for example, image compression, image scaling and image rotation [49]. However, applying contrast enhancement techniques is sometimes instructed to improve the detailed information given to the DL algorithm. In this context, one study involved contrast limited adaptive histogram equalization (CLAHE) to WL and NBI images to enhance the textural attributes in the input image [72]. CLAHE computes

histograms corresponding to the particular sections of the image and uses distribution parameters to define the shape of the histogram [80].

1.3.4 Conventional CAD systems for laryngeal endoscopic image

The development of conventional CAD for laryngeal endoscopic image analysis is continued after image pre-processing step by three more phases.

1.3.4.1 ROI segmentation: In conventional CAD architecture, ROI segmentation is mainly performed to focus on the characteristics of a specific structure in the larynx region and extract the features from it. The primary areas involved in this workflow are the vocal fold, laryngeal lesion, and vascular structures. On this subject, morphological region growing, morphological closing operation, morphological Black-Top-Hat operation, skeletonization operation, histogram of oriented gradient (HOG), Canny edge detector, and first-order derivatives of gaussian (FODG) are some of the proposed and validated approaches for ROI segmentation.

The proposed approach by Barbalata et al. studied two paths for segmenting lesions and vascular structures [59]. First, the lesion segmentation was implemented according to a proposed approach for polyp segmentation in endoscopic GI tract images [81]. Then, the Canny edge detector was used to find the boundaries of the lesion within the image based on discontinuities in brightness. The task was completed with a morphological closing operation to close the small gaps. The vessel segmentation process followed a path used for retinal blood vessel segmentation [82]. The study used the matched filter (MF) based on the FODG to extract the blood vessels and performed the final refinement on large and small vascular structures based on the Gabor filter and morphological Black-Top-Hat operation, respectively.

Another investigation on laryngeal endoscopic image analysis proposed a two-step process for automatic detection of the vocal fold along with an independent algorithm for vessel segmentation [60]. First, it applied the HOG algorithm [83] – a well-known traditional computer vision technique in object detection – to detect the vocal fold region based on the distribution of edge directions or intensity gradients. The process was followed by applying a region growing technique to segment the glottis in the vocal fold region, where threshold values were defined to avoid the distribution of the region's homogeneity. The vessel segmentation process started with a morphological region growing operation based on the seed points detected by the Canny edge detector. Then, the skeletonization of vessel was segmented by an iterative thinning operation and the final centerlines were generated by connecting the validated points [84].

1.3.4.2 Handcrafted feature extraction: The handcrafted feature extraction methods for laryngeal endoscopic analysis are categorized into three leading groups and are instructed as morphological, texture and statistical feature sets.

- The morphological features describe the geometrical characteristics of the segmented ROI, including the vocal fold and vessels. In the study by Barbalata et al., features related to the direction, width, and tortuosity of vascular structures were extracted from the segmented regions of the lesion. The width of vessels was established on the distance between the skeletons of one vessel using simple mathematical operations. The tortuosity of vessels was calculated based on the average value of angles among three different points on every skeleton [59]. In another study, Turkmen et al. considered the shape-related attributes of the vocal fold edge and the vocal fold's vascular characteristics. They defined the vocal fold edge curve based on the segmented glottic area and extracted four features, including size, location, splay portion, and symmetry of lesion. Moreover, they translated the visual elements of the segmented blood vessels and its orientation toward the lesion into the tortoise and longitudinal vascular vectors based on their angle to the baseline and finally computed four features from them [60].
- Texture features are placed in the spatial domain features group and are directly computed from the pixel values in the image. Moccia et al. implemented a texture-based feature extraction pipeline on endoscopic image patches, using Local Binary Patterns (LBP) and Gray Level Cooccurrence Matrix (GLCM) [61]. These strategies are invariant to some conditions under which larynx endoscopic images are captured, such as changes in endoscope pose and illumination conditions. LBP is a texture descriptor that labels the pixels of every patch by thresholding the neighborhood of each pixel, where it was calculated as a features vector in the form of a normalized histogram. GLCM evaluates the spatial relationship among pixels and estimates how frequently pair of pixels are present in a given direction and distance. Contrast, correlation, energy, and homogeneity are features derived from this descriptor and are also known as second-order statistical features [85].
- The last category directs to the statistical features explored by Moccia et al. for analyzing NBI patches [61]. Intensity mean, variance, and entropy were three first-order statistical features used in this study to represent the spatial distributions of the image patches.

1.3.4.3 ML classifiers: Supervised classifiers are the central part of the conventional CAD systems that involve the contribution of ML techniques. Every ML classifier learns the target population's characteristics according to the attributes and representation transformed into the input feature set. Although all

the ML classifiers follow the same objective, each has a unique set of hyperparameters that requires optimization to reach the highest performance. For this reason, different optimization techniques, such as manual search, grid search, randomized search, and Bayesian optimization [86], are commonly used for the optimization of both ML classifiers and DL networks to find the most optimum hyperparameters for creating the target predictive model. The optimization process is usually applied to the training set or part of it.

Among all the available ML classifiers, Support Vector Machine (SVM), k-Nearest Neighbors (kNN), Random Forests (RF), Naive Bayes (NB), Linear Discriminant Analysis (LDA), and Multilayer Perceptron (MP) are the applied approaches for laryngeal endoscopic image classification. Each ML classifier considers specific attributes of the feature set to perform the classification, leading studies to use more than one classifier in their development and select the one with the best performance.

- SVM performs classification by finding the optimal hyperplane in an N-dimensional space that maximizes the margin between the classes. The kernel function is the main hyperparameter of the SVM classifier. Linear, polynomial, and radial kernels are standard functions that map the data into some feature space. Then, the type of kernel function may introduce more hyperparameters for the SVM, such as regulation and kernel parameters [87].
- kNN is a nonparametric method and performs classification by finding the most similar data points in the training data and making an educated guess based on their categories. Therefore, the number of neighbors is the most critical hyperparameter of the kNN classifier. Moreover, it is recommended to optimize the distance metrics for selecting the combination of the neighborhood [88].
- In RF, the bootstrap sample, which comprises a sample of data drawn from a training set, is used to create an ensemble learning method that consists of a collection of decision trees. RF has several hyperparameters; however, the depth of trees and the number of estimators (trees) require the tuning process [89].
- NB is a probabilistic ML technique that is used as a classifier. It is based on probability models of Bayes theorem that incorporate strong independence assumptions [90].
- LDA is a technique for reducing dimensionality in which features are combined into linear combinations to represent or separate two or more groups [91].
- MLP is a feed-forward ANN supplement with at least an input layer, a hidden layer, and an output layer. Except for the input nodes, each node is

a neuron that uses a nonlinear activation function and utilizes backpropagation for training [92].

In the primary attempt in 2014 to develop a CAD system for laryngeal lesion detection and classification, Barbalata et al. applied the LDA classifier as the ML classifier. The confidence measure based on the malignancy probability that was computed from the feature values led to benign-malignant image classification [59].

In the subsequent investigation conducted in 2015, a group of five ML classifiers, including SVM, kNN, RF, NB, and MLP, were used to perform a binary image classification in the form of a hierarchical decision tree architecture. A specific feature set and defined classification scenario were portrayed in every tree node to arrive at a final decision representing laryngeal histopathology [60].

The last study in this context, from 2017, compares the classification performance of four ML classifiers during the laryngeal tissue classification in endoscopic NBI images. SVM with Gaussian Kernel, kNN, RF, and NB were individually trained and tested on the different mixtures of normalized feature sets. Moreover, one additional step, confidence estimation, was added to the classification workflow to calculate the procedure's reliability [61].

1.3.5 DL-based CAD systems for laryngeal endoscopic image

The rapid advancement of the DL in MIC was spread to the laryngeal endoscopic image analysis field in the last five years. DL takes advantage of learning multiple levels of features from the data by iterative adjustment of layers' weight. Therefore, a large number of data is needed to reach high-performance outputs, especially in complex tasks [54, 55].

1.3.5.1 DL model training: Training the CNN is the most crucial part of building a DL-based CAD. There are three paths to train the CNN model. In the first method, CNN is designed and trained from scratch. Although this path can result in high accuracy, it requires hundreds of thousands of labeled images and considerable computational resources. The second path relies on the transfer learning concept that allows reusing a pre-trained model as a starting point for a new classification task with comparatively little data. The pre-trained network is a network that has already been introduced to a specific data set and learned to extract valuable features from it. The data set used for the pre-training is not always the same as the actual data set for the second classification task, but the extracted features are similar in nature. This network can then be used as a starting point to learn a new classification task. Finally, the third method uses a pre-trained CNN to extract learned features for training the ML classifier. This path requires fewer data and computational resources than the last two approaches [93, 94].

1.3.5.2 DL model fine-tuning: Independent from the path that is followed for training the CNN, hyperparameter tuning is an inevitable step. The same optimization techniques used for ML classifiers are usually applied for CNN hyperparameter tuning. However, the hyperparameters are defined before the training and are divided into the variables that define the network structures and the parameters that determine how the network is trained.

Dropout and activation functions are the primary hyperparameters related to the CNN structure. The dropout helps to reduce the model's size and avoid overfitting, meaning the predictive model performs very well on the trained data but does not generalize enough to classify the new unseen data. The activation functions learn and predict continuous and complex connections between variables of the network [95, 96].

The principal hyperparameters to define the training process are the optimizer, loss function, learning rate, number of epochs, and batch size. Stochastic Gradient Descent (SGD) and Adam are two common examples of optimizers that aim to adjust the weights in the model to arrive at the possible highest accuracy or the lowest loss function. The learning rate describes how fast the network's parameters are updated to reach the minimum loss function. Usually, a lower learning rate increases the chance of meeting the minimum loss function, but it requires more time and computation resources. The batch size is the number of sub-samples of training data introduced to the network. Defining a bigger batch size affects slower learning but with lower variance in validation accuracy. The loss function is an evaluation method that measures the label and predicted output values to represent whether the model fits the data well. Finally, the number of epochs decides the number of iterations the learning algorithm will perform throughout the whole training set. Notably, a too-small and too-big number can end with underfitting – meaning that the model has not learned enough – and overfitting, respectively [95, 96].

1.3.5.3 DL architectures: Transfer learning integrated with the fine-tuning process is the most appropriate training path in forming the DL-based CAD for laryngeal endoscopic image analysis. All the networks involved in this field have been first trained on the ImageNet database [97], an image data set with more than 14 million data in 27 high-level categories organized hierarchically. The trained networks have learned to identify low-level and high-level features in the images of ImageNet. Therefore, they can use their weights as a starting point and get adopted via the fine-tuning operation for other image classification and object detection tasks. VGG Network (VGGNet), Inception Network, Residual Networks (ResNets), and Efficient Network (EfficientNet) are the four main pre-trained networks used for implementing DL-based CDA systems on laryngeal endoscopic images.

- VGGNet is an architecture focusing on depth as a vital element of CNNs [98]. It uses very small 3x3 convolution filters, one after the other, to increase the depth while lowering the parameters and improving the training time and performance. Moreover, the network incorporates 1x1 convolution filters as a linear transformation to make more non-linear predictions. There are variants of VGGNet, including VGG16 and VGG19, which differ only in the number of convolutional layers of the network (See Figure 1.7-b). The pre-trained VGG16 with around 18 million parameters was used in a study in 2022 for laryngeal image classification as binary benign-malignant and multi-class benign histopathology classification [71]. The analysis contained quite an extensive data set with 19353 labeled endoscopic image data; however, the authors do not give the characteristics of the adopted VGG16 network for this specific classification.
- Increasing the deep convolution layers in a model is not always beneficial, as the issue of overfitting and computational expenses may arise. So, other more optimum CNN architectural designs were started and presented, such as Inception Network [99]. This heavily engineered architecture introduced a new level of organization inside the network, called Inspection modules. The network follows the concept of placing multiple convolutional filters but with different sizes that operate on the same level. Therefore, the created architecture is wider rather than having deep layers. The continuous evolution of the Inception Network resulted in the creation of numerous versions of the network. InceptionV3, with 25 million parameters, is the most practical model of this category for laryngeal endoscopic image classification. One application of pre-trained InceptionV3 was reported in 2019. First, the model was modified for a multi-class classification task with four classes. Then, it was fine-tuned using GSD optimizer and cross-entropy loss function on a large data set with 13721 endoscopic larynx images [64]. The second study was conducted in 2021, and the pre-trained InceptionV3 model was used as the backbone of a DL-based CAD in which the model was adopted and fine-tuned on 4591 NBI images for laryngeal SCC diagnosis [69].
- One year after the introduction of Inception Network, ResNets stepped into the computer vision world [100]. The deepness level of ResNets is related to the network's capability to capture high or higher patterns. ResNets optimize toward zero, accelerating the convergence to the optimal point in the solution space instead of an actual number. Batch normalization is another exciting feature that is embedded in ResNets structure. It speeds up the convergence and, in doing so, reduces the training epochs required. It also has a regularization effect during the training phase. Several variants of ResNets architecture have a different number of layers (See Figure 1.7-a).

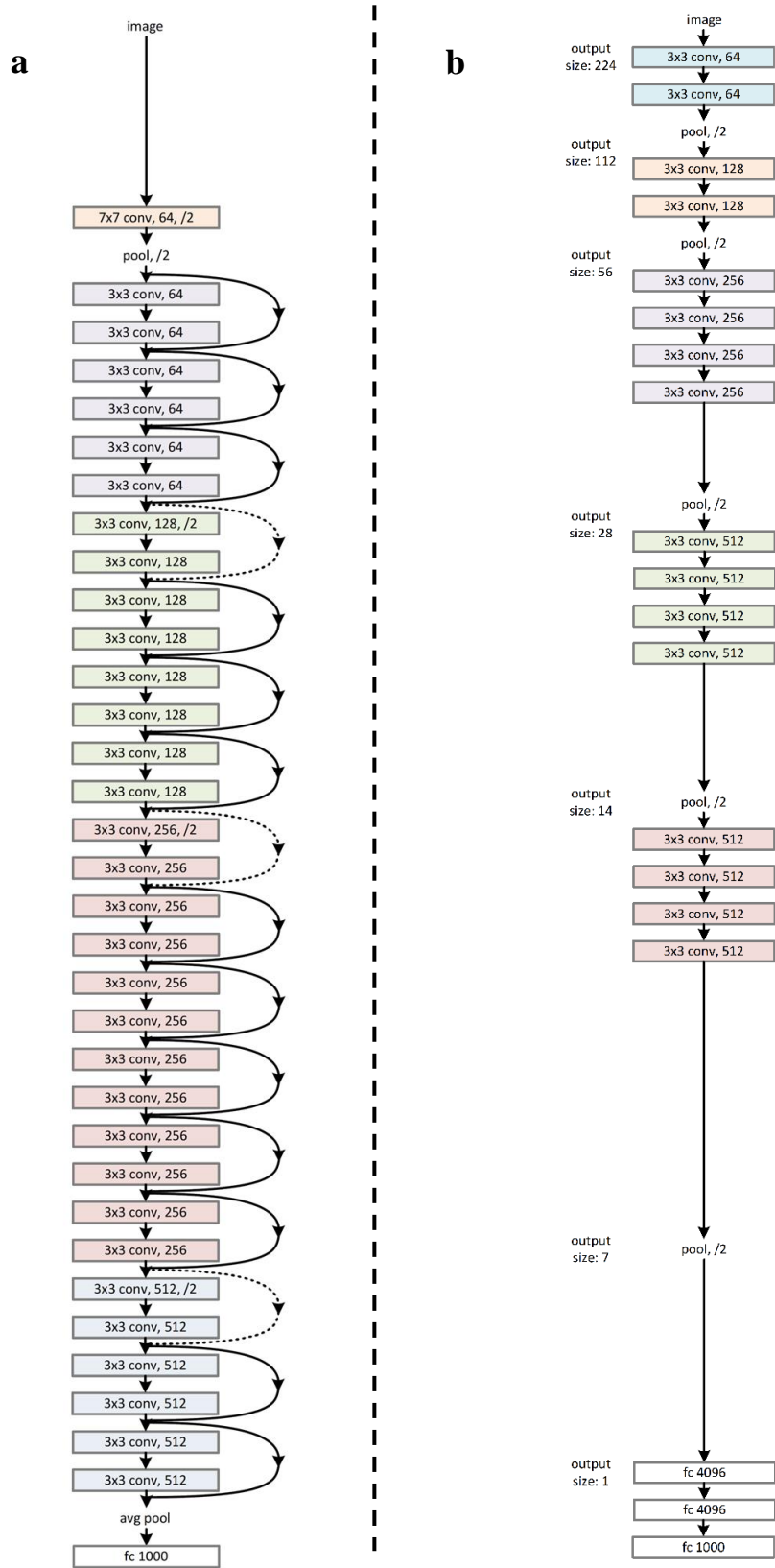


Figure 1.7. Examples of DL-based network architecture. (a): ResNets with 34 parameter layers, and (b): VGGNet with 19 convolutional layers [100].

ResNet34, RedNet50, ResNet101, and ResNet186 are the pre-trained models used in the context of DL-based CAD systems on larynx endoscopic images. The first study in 2020 informed the application of a pre-trained ResNet50 for extracting a learned feature map from the input image used for the laryngeal tumor detection task [65]. The model was part of a more extensive architecture called RetinaNet [101], which was adopted for the laryngeal tumor detection task. The second study's focus was on using a pre-trained ResNet101 model adopted for a multi-class classification of the endoscopic larynx images. GSD optimizer and cross-entropy loss function were selected for the training and fine-tuning process over 24667 endoscopic images [67]. Finally, the most recent study in this category from 2022 proposed a two-stream classification network in an upstream manner. The architecture adopted pre-trained ResNet18 and ResNet34 as the small and the large stream, respectively. The Adam optimizer was applied for these two models' training and fine-tuning steps on the open-access Laryngoscope8 data set [73].

- The traditional methods of scaling a CNN model are random. They are mainly used as a standalone technique focusing only on one aspect of the network such as depth or width and require manual tuning. As illustrated in Figure 1.8, the development of EfficientNet showed that a better scaling could be achieved by compound scaling, meaning a uniform scaling of all networking dimensions like depth, width, and resolution using a constant ratio [102]. Different versions of EfficientNet are generated by scaling up the baseline network. The pre-trained EfficientNetB0 was adopted and fine-tuned by Cho et al. using Adam optimizer. The data set of this study included 4106 endoscopic images designed for a multi-class classification task based on nine laryngeal histopathologies [68].

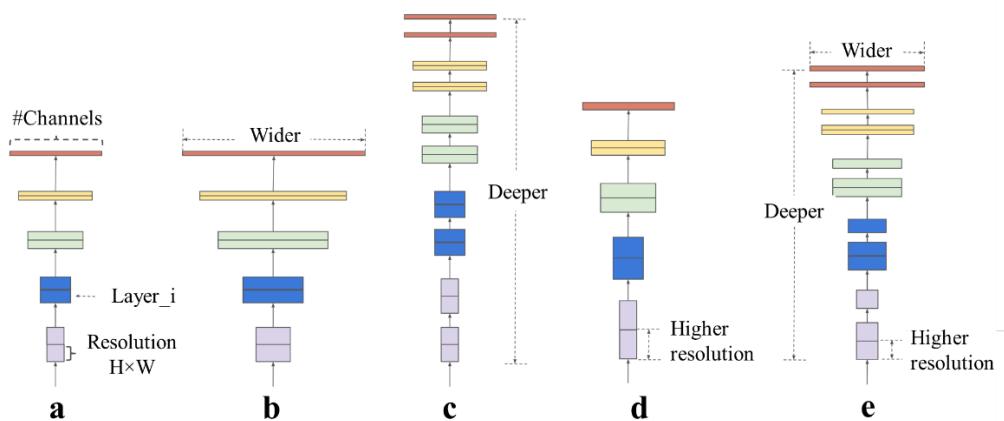


Figure 1.8. Model scaling in EfficientNet. (a): Baseline network, (b) to (d): Conventional scaling that only increases one dimension of network width, depth, or resolution, and (e): Compound scaling method [102].

1.3.6 Laryngeal endoscopic image and hybrid CAD systems

With the recent advances in MIC, multiple modified versions of the CAD system were introduced for MIA. The main characteristic of these architectures relies on combining the feature extraction and classification steps of conventional and DL-based CAD systems into one architecture. There are two examples of these systems in larynx endoscopic image analysis.

The first study was presented in 2019 and was developed based on the concept of extracting both handcrafted and learned features and using them for training a ML classifier. In this way, a more comprehensive range of attributes was fed into the model to learn the target population's aspects. LPB and first-order statistical features were included as handcrafted feature sets. Also, learned features were extracted from pre-trained variations of ResNets and Inception Network. The fine-tuning process was carried out on particular layers of each network, in which a stopping point was defined to extract the features. Finally, the SVM classifier with Radial Basis Function (RBF) was trained and validated for a multi-class classification task using an open-source data set of larynx NBI images for laryngeal tissue classification [62].

The second version of CAD systems was proposed based on the concept of ensemble modeling. It is reported that a single ML model can have bias and high variability. Therefore, generating two or more related but different models followed by aggregating the results into a single score can help to reduce the model's error and maintain the model's generalization. In this study, 8 different pre-trained CNN architectures including VGG16, VGG19, ResNet50 and ResNet101 were trained and fine-tuned on different data sets. Moreover, 12 handcrafted texture, color, and morphological features were extracted and used to train the SVM classifier separately. Finally, the prediction was made based on the ensemble model generated from all trained models using the weighted sum rule technique. This ensemble model was validated on an open-source data set of larynx NBI images for laryngeal tissue classification [66].

1.3.7 What is needed to improve the performance of CAD systems?

The application of MIC solutions for the endoscopic image analysis of the larynx faced considerable improvement in the last decade. The purpose of such explanations was first focused on providing a better understanding of presented information in the endoscopic images, such as structural characteristics, via traditional image processing techniques. These solutions were later extended into more advanced decision-based paths in the form of CAD systems built based on ML methods. All these techniques aimed to move toward providing objective assistance for laryngeal endoscopic image analysis. They also showed their potential to reduce the user-dependency level and consequently shorten the

learning curve for younger clinicians. With these strategies, CAD systems have claimed to facilitate and improve early diagnosis of laryngeal lesions to preserve the organ's functionality and reduce the number of unnecessary biopsies. However, the ML applicability is not well developed and integrated into actual clinical practice on laryngeal lesion assessment to truly implement what these systems contend [49, 51, 52].

ML development falls into the group of data-driven solutions. These approaches require extensive data sets to provide robust and reliable assistance. Furthermore, creating such a data set usually demands multi-center cooperation, a standardized image acquisition and evaluation strategy that must be included in the current ML-based laryngeal lesion assessment workflow. Besides, the endoscopic data for generating a data set is mainly post-extracted from video recordings. This process mandates frame adjustment and selection to include images with relevant anatomy and pathology in the data set. Therefore, a comprehensive technical and clinical effort is needed to create a proper data set. Moreover, ML methods are characterized by many parameters; therefore, their real-time application is computationally expensive and requires sufficient hardware and software infrastructures in clinical facilities.

But there are more challenging conditions related to using WLE and NBI for computer-based laryngeal image analysis that can affect the ML approaches or cannot be solved by the proposed methods [49, 51, 52].

It is known that the stable visualization of target anatomical structures in WLE and NBI images is challenging because it is essential to have the proper distance of the endoscope from the mucosa. This problem results in limited resolution in endoscopic images that can affect the implementation of the applied ML method. Nonetheless, this problem is addressed during the development of most ML-based techniques for laryngeal image analysis.

The second concern is directed to a more critical issue that can affect the performance of the ML approaches but cannot be solved with these solutions. Laryngeal lesions and their vascularization structures vary significantly in form, color, texture, and size. Therefore, stable, detailed, focused, and magnified visualizations of these particular structures are required for better understanding and assessing the pathological conditions. Unfortunately, the level of magnification and focus needed to achieve this goal is not covered by the commercially available endoscopic imaging systems in the form of normal WLE and NBI images for laryngeal lesion assessment. However, these concerns in the field of laryngeal lesion assessment are investigated through the integration of a new endoscopy-based imaging modality.

1.4 Magnifying endoscopy - a new advancement in laryngeal lesion assessment

The high-resolution and magnified endoscopic images can lead Otolaryngologists to better detection of laryngeal lesions during clinical examination and provide appropriate and adequate input data for developing CAD systems.

The application of High Definition (HD) endoscopy systems was first introduced to the endoscopic diagnosis of GI tract diseases. Then, with a gap of a few years, the first HD endoscopy systems were adopted for examination and diagnosis purposes in Otolaryngology. The systems could also be combined with enhancement imaging techniques for high-resolution visualization of the target lesion. However, this system integration in Otolaryngology still needed the component of detailed and focused visualization of the lesion that could be filled with magnifying imaging technology [103].

Magnifying Endoscopy (ME) is an optical and powerful imaging modality that can provide enlarged visualization of the microstructures of surface mucosa, including vessels and cellular nuclei. This diagnostic endoscopic technique offers a real-time representation of tissue characteristics with several hundred-fold magnifications. Therefore, minor structures missed by conventional endoscopic systems can be detected via this modality [104].

The ME can work based on two leading technologies. First is the electronic magnification that is usually embedded in the endoscopy system and expands the image to a certain level. However, image quality is highly deteriorated in electronic magnification because the image consists of fewer pixels at every step of magnification, resulting in low resolution. The second technique is optical magnification in the form of magnifying endoscopes. It provides high-magnified images using movable focus lenses that can move very close to the mucosal surface without losing the image resolution [105, 106].

The combination of ME with HD endoscopy systems and enhancement techniques has introduced a novel branch of Biologic Endoscopy known as Optical Biopsy that aims for early diagnosis of lesions without performing a surgical tissue excision [107, 108]. Integrating ME with NBI is the most well-known and applicable example of Optical Biopsy strategies. This approach was first established and operated in GI tract disease diagnosis [103]. The standalone application of high-resolution NBI can improve the contacts between the target lesion and background epithelium, visualized as a well-marked brownish area. After adding magnification, subtle changes in mucosa structures and vascular architecture can be identified. Therefore, the added value of the ME can provide a histological diagnosis for the early detection of pathological conditions. The ME with NBI has shown a powerful diagnostic performance in GI tract disease and

could be used as a promising screening technique for patients with medical conditions that have a high probability of evolving cancer [106, 109].

Considering the similarity of mucosal characteristics of the GI tract and larynx, the application of ME combined with high-resolution enhancement endoscopy was translated to the field of Otolaryngology in the last decade. This trend followed the main objectives of Optical Biopsy for laryngeal lesion assessment. Several studies reported combining magnifying GI endoscopy systems with NBI for laryngeal lesion assessment. The improved visualization of mucosal structures and vascularization networks in these investigations showed a more accurate endoscopic diagnosis of the pathological condition in larynx. Although GI endoscopic systems are powerful tools, they are not suitable for Otolaryngology procedures due to their large size, which makes them unsuitable for transnasal and transoral applications [107, 110, 111].

1.4.1 ME with contact endoscopy

Contact Endoscopy (CE) is an optical diagnostic imaging technique that allows in vivo and in situ examination of superficial layers of mucosal epithelium. A contact endoscope is a magnifying endoscope developed by KARL STORZ, Tuttlingen, Germany (Figure 1.9), for laryngeal applications using optical magnification technology with three 1x, 60x, and 150x magnification levels. The endoscope is commercially available with a straight-forward telescope at 0° or a forward-oblique telescope at 30° viewing angle in two different diameters and lengths. In clinical examination, the patients need to go through general anesthesia to introduce the contact endoscope via laryngoscope to the vocal fold region. For visualization, the endoscope should be in contact with mucosa and then can be moved gently over the target area [112, 113].

CE has a great history in laryngeal lesion assessment [114–116]. The early application of CE on the larynx focused on evaluating the cellular architecture of mucosal epithelium in the vocal fold. The mucosal surface was stained with 1% methylene blue, a nontoxic short-effect dye that transmits a dark blue color to the nucleus and a light blue color to the cytoplasm. Higher mitotic activity and greater concentration of nuclear material make mucosal epithelium with pathological conditions more stained. Therefore, Otolaryngologists or histopathologists could assess this real-time and non-invasive visualization of cellular architectures to perform the histological diagnosis of laryngeal lesions without surgical biopsy. Nevertheless, one of the main limitations of CE with methylene blue is related to the restricted penetration depth of the dye [117–119].

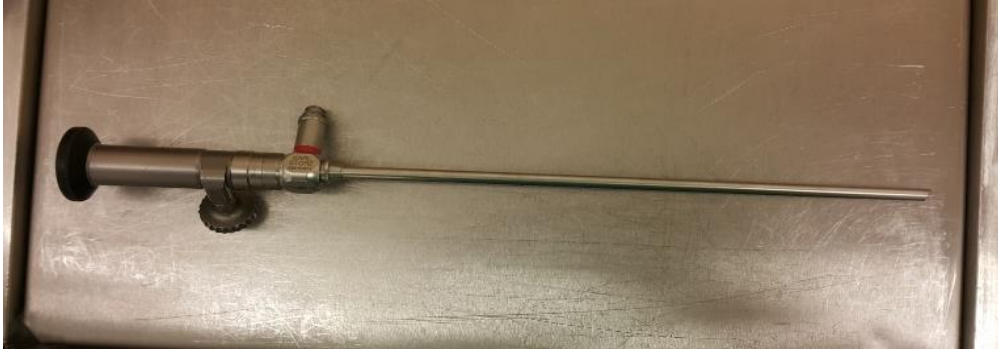


Figure 1.9. The contact endoscope with a forward-oblique telescope at 30° viewing angle.

The lesion's growth in every anatomical organ is followed by its vascularization, where micro blood vessels surrounding the tumor usually develop abnormally in higher quantity with more tortuosity. The CE technique without staining allows the detailed and magnified examination of this vasculature network of mucosal epithelium in the vocal fold. Moreover, the high magnification level of CE provides visualization of the distribution and dynamics of the microcirculation in the mucosal surface [112]. During the early application of the CE on clinical examination of larynx, evaluation of this microvasculature around the lesion was considered as the secondary source of information for the initial assessment of laryngeal lesions. However, recent advancements have shown that evaluation of the vascularization networks of superficial mucosal epithelium in vocal fold can provide Otolaryngologists with more critical information than the cellular architectures of the lesion [35]. Therefore, the CE technique could be the potential solution to the high-resolution and magnified endoscopic diagnosis of laryngeal lesions. This step-by-step progress of the CE technique and its integration into HD-enhanced endoscopy systems have introduced a new path to the domain of Optical Biopsy of laryngeal pathologies focusing on vascular architectures of superficial layers of mucosal epithelium.

1.4.2 Enhanced CE and vocal fold mucosal vascularization

As discussed in Section 1.2.2, the standalone enhanced endoscopy techniques have introduced the evaluation of vascular architecture as a new source of information to the clinical examination during the laryngeal lesion diagnosis procedure. However, assessing the most profound changes in abnormal microvascular structures required a more magnified view. The Enhanced CE (ECE) is the exact tool for this purpose by focusing on high-resolution and magnified visualization of the microvascular network of the superficial layer of mucosa. This endoscopic diagnosis technique in the larynx was first introduced in 2015 [120]. It combines the CE technique with well-known high-resolution enhancement endoscopy techniques, either NBI or SPIES. Figure 1.10 represents some examples of ECE imaging using NBI.

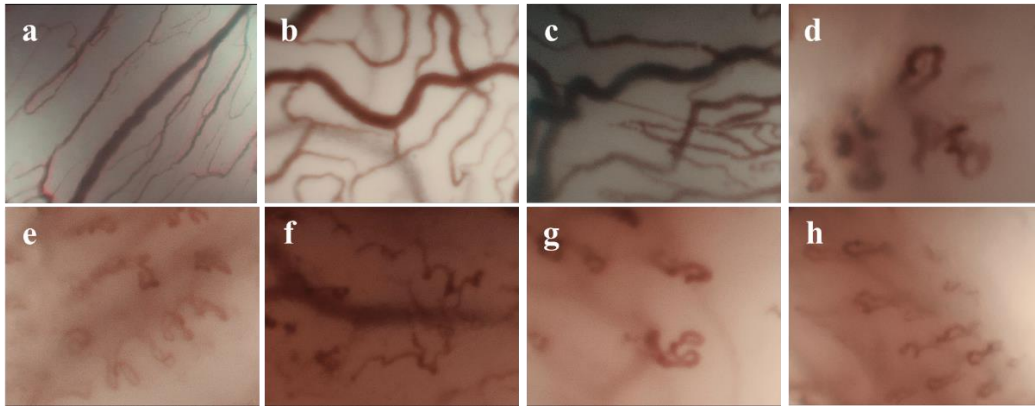


Figure 1.10. ECE imaging using NBI for different histopathologies of vocal fold. (a): Cyst, (b): Reinke's edema, (c): Hyperkeratosis, (d): Papillomatosis, (e): Low-Grade Dysplasia, (f): High-Grade Dysplasia, (g): Carcinoma in Situ, and (h): SCC [122].

The ECE can be added to the current clinical examination procedure of laryngeal lesions during clinical examination and after direct laryngoscopy using RE under general anesthesia (Figure 1.5). In this way, the Otolaryngologist first examines the overall anatomical structures and geometrical characteristics of the vocal fold, target lesion and surrounding regions and then moves to the application of ECE to get a more detailed examination of superficial mucosa of the lesion and its surrounding area, especially vascularization network. As presented in Figure 1.11, with this strategy, ECE allows the real-time study of multiple regions that can arrive at a non-invasive and faster histological diagnosis compared to surgical biopsy [121].

The diagnostic strategy of laryngeal lesions based on vascular architecture relies on the concept of neoangiogenesis. It means the development of new blood vessels from a pre-existing vasculature which plays a vital role in the growth of lesions. The abnormal tissue cells continuously release the factors that stimulate the unusual development of IPCLs to create enough suppliers for the tumor. Therefore, the level of development of a microvascular network of cancer is correlated to the degree of lesion progression. Furthermore, it means that the neoangiogenic stimulus increases from benign to pre-malignant and invasive cancer lesions resulting in the vascularization network with more levels of chaos and disorder [120, 123].

ECE can provide a detailed evaluation of these abnormal changes in vascular structures of the laryngeal mucosa. However, among the studies that correlate the level of disorder of vessels to the type of lesion and histopathology, there is only one guideline with a specific focus on vasculature architectures represented in ECE images. In this study, the vascular patterns are divided into five groups correlated to the diagnosis normal, inflammation, hyperplasia, mild to moderate dysplasia, and high-grade dysplasia/carcinoma in situ/ invasive carcinoma [120, 124].



Figure 1.11. Application of CE-NBI during clinical examination of laryngeal lesion.

Nevertheless, the ELS guideline is more studied and applied to standalone enhanced endoscopy and ECE images as a guidance tool for the clinical examination of laryngeal lesions. The main principle of this guideline is simple to implement and is expanded for both non-magnified and magnified visualization of the vascular structures in the superficial larynx mucosa. Moreover, it represents how the magnified view facilitates the correlation of LVC and PVC to the benign and malignant laryngeal lesions [121, 123, 125]. Therefore, the potential of the ECE imaging technique with a focus on the high-resolution presentation of vascular structures of mucosal epithelium boosted the expectancy of bringing the application of Optical Biopsy into the actual clinical settings for early assessment of pre-malignant and malignant lesions and avoiding the unnecessary invasive tissue biopsy.

1.4.3 Challenges to integrate ECE into the clinical setting

In the last two decades, significant improvements in endoscopic imaging facilitated the clinical examination of laryngeal lesions and provided the potential to perform Optical Biopsy. Potential advantages of this development are directed to the patients because high-resolution and magnifying visualization allow Otolaryngologists to better distinguish malignant lesions from non-cancerous ones without conducting surgical biopsy. In addition, this development can prompt diagnosis procedures in the early stages to reach optimum organ preservation.

Several studies have reported the high performance of ECE as an Optical Biopsy technique in diagnosing different types of mucosal lesions in the head and neck region [120, 121, 125]. CE, especially when coupled with the NBI enhancement technique, offers a wide range of advantages compared to the other Biologic

Endoscopy modalities. First, ECE offers magnified visualization of the examined region, such as the vascularization network of the lesion, which the normal WLE and NBI modalities cannot provide. Second, ECE resolves the issues created by the inconsistent distance between the endoscope tip and the examined region in standard endoscopic imaging, including limited resolution. Because the contact endoscope should have a fixed distance and direct contact with the tissue. Third, ECE is quick, simple, and produces real-time and repeatable results that can be reviewed as many times as needed. Moreover, it does not have the risks related to surgical biopsy [112]. However, it is recognized that the limiting factors of ECE prevented it from gaining an acceptable place in routine clinical practice for performing Optical Biopsy.

Increased levels of detailed information in ECE images raised significant problems in the visual assessment of the represented superficial mucosa. The ECE image interpretation drives toward a subjective process. Although several guidance tools were proposed to help Otolaryngologists with the understanding of vascularization networks, the similarities between these vascular structures in precursor and malignant laryngeal lesions raise a new challenge. For instance, as observed in laryngeal Papillomatosis (Figure 1.10-d), PVC with wide-angled turning points can be difficult to distinguish from PVC with narrow-angled turning points, as observed in pre-malignant and malignant histopathologies (Figure 1.10-h). Therefore, extracting the diagnostic information from ECE images highly depends on the experience of the Otolaryngologists. Several studies addressed the issue of subjective interpretation in ECE image analysis and reported moderate to high interobserver agreement in less experienced and experienced groups. Moreover, there is always a learning curve in this procedure that highlights the necessity of extensive training to minimize the risk of subjective evaluation [121, 125]. All these challenges can result in a complicated laryngeal lesions assessment that may cause an increased number of false positive diagnosis and potential over/ under treatments. The presence of these risks limited the application of ECE in research use and constituted primary impediment to the adoption of ECE as an Optical Biopsy technique in clinical settings.

Chapter 2 – Contributions

2.1 Motivation and contributions

As discussed in Chapter 1, finding a minimally invasive, fast, reliable, and accurate approach for early diagnosis remains the main challenge in laryngeal lesions assessment. Recent improvements in endoscopic imaging techniques and ML-based strategies using WLE and NBI images have aimed to address this challenge by moving towards performing Optical Biopsy. However, the challenge remained unsolved for the following reasons:

- There is a lack of magnified visualization of the vascularization network in normal WLE and NBI images that could provide valuable information about the status of the lesion. This issue could also adversely impact the performance of the developed ML-based methods that rely on this information.
- The proposed ML-based methods are intended for research purposes and contain intricate algorithms that may overlook critical aspects necessary for clinical application. Such elements include being easy to understand, easy to use, and adaptable to the standard clinical workflow.

The application of magnifying and enhanced endoscopic imaging techniques, mainly the CE-NBI modality, holds promise in providing a solution for performing Optical Biopsy. The main advantage of CE-NBI modality over standard endoscopic imaging techniques is related to the enlarged visualization of the vascularization network, providing principal complementary value for Otolaryngologists during laryngeal lesion assessment. However, interpreting complex information from CE-NBI images remains subjective due to the similarity of vascular patterns and imposes extensive learning on users. Therefore, CE-NBI has yet to be integrated into the standard clinical examination procedure for laryngeal lesions.

While there is currently no publicly available data set of CE-NBI images, several reports emphasized the clinical values of this modality in assessing laryngeal lesions. However, no studies have explored the technical aspects of CE-NBI modality by developing and validating computer-based techniques. This option could overcome the limitation of CE-NBI, making it a standard modality for performing Optical Biopsy and enabling early diagnosis of laryngeal lesions. To achieve this goal, this thesis aims to develop a pipeline for a CAD system on laryngeal lesion assessment using CE-NBI images, where the main focus is divided into two main categories:

1. The first main task of the thesis is related to CE-NBI data set generation to address the need for publicly available data in this area. The data was collected from the Department of Otorhinolaryngology, Head and Neck Surgery of Magdeburg University Hospital, along with an active collaboration with the clinical team that has rich experiences in endoscopic data collection as well as laryngeal lesion diagnosis and treatment. After passing the required ethical approval, data collection, preparation and annotation were conducted simultaneously during the four years of the project by the author of this thesis. Every video file of CE-NBI was first annotated manually to extract the time intervals with good image quality. Then, the frames were extracted from the defined intervals, and images with good resolution and a unique vascularization network were selected for every patient. Two experienced Otolaryngologists verified the chosen images and labeled them based on the type of vascular pattern. After this step, the CE-NBI image labeling was completed by annotating the data based on the diagnosed laryngeal histopathology, type of laryngeal lesion, and leukoplakia diagnosis. The final version of the generated data set – with 11144 CE-NBI images of 210 adult patients – is currently available in a public repository [122].

As a part of the first task and in collaboration with the clinical team, a clinical study was conducted on the part of the final data set to evaluate the subjective assessment of the CE-NBI images, analyze the intra-user variabilities during the visual assessment of the images and highlight the values of vascular structures on laryngeal lesion assessment [121]. The outcome of this evaluation was used as a guideline to plan the development strategy and design the pipelines of the CAD system according to the clinical requirements.

2. The second main task of the thesis is related to technical exploration of CE-NBI images in the form of developing the CAD system. This task addressed the need for technical exploration of this new imaging modality to overcome its limitation on laryngeal lesion assessment. For that, two pipelines for assessing laryngeal lesions based on image classification tasks with a focus on the vascular characteristics in CE-NBI images were developed:
 - a. Pipeline 1 is based on feature engineering techniques combined with ML classifiers for supervised classification of CE-NBI images. This strategy was selected because of two reasons. First, CE-NBI data were collected from patients with pathological conditions on the vocal fold. Some histopathologies, such as Reinke's Edema and Dysplasia are more common in this target patient group. Therefore, in the first two years of the project, the data set included a limited number of images with only specific histopathologies. Second, it

was crucial to develop techniques that are easy to understand in the clinical community to create the infrastructures and involve clinicians in the development phases as the primary users. Therefore, the development strategy in pipeline 1 was focused on the handcrafted features that could describe the characteristics visualized in CE-NBI images and could explain the physiological changes happening in superficial mucosal vessels with the appearance of pathological conditions. Pipeline 1 included two sets of self-explainable features related to the geometrical and textural characteristics of CE-NBI images.

- i. Method 1 focused on the geometrical attributes of vascularization networks in CE-NBI images based on visualized and understandable characteristics. The proposed approach includes a three-step image pre-processing strategy. Then, five geometrical indicators are defined to assess the level of disorder of vascularization networks, including two direction-based and three curvature-based indicators. In the next step, 24 Geometrical Features (GF) are defined after qualitatively analyzing the indicators' behavior on three main categories of vascular patterns. Finally, the supervised classification step takes place where four ML classifiers are trained and tested using GF for automatic CE-NBI image classification based on the vascularization networks as well as benign and malignant histopathologies. The outcome of this approach is presented in Contribution 1.

After discussions with the clinical team, the decision was made to stay with benign-malignant lesion classification for further development. The main reason for this decision was that the type of lesion is the first primary input for the Otolaryngologists during the clinical examination to proceed with the additional treatment plans.

In the next step of development in pipeline 1, the impact and value of Method 1 on the current workflow of clinical examination for laryngeal lesions are investigated. The manual and automatic image classification scenarios are planned to compare the performance of Method 1 to the assessment of less-experienced and experienced Otolaryngologists on CE-NBI image classification, focusing on vascular characteristics. The image data set for this evaluation contains 10 to 50 images per patient. However, three to five images per patient are selected for

the manual approach and later were used as the testing set in automatic classification. Four different supervised ML classifiers are trained on GF computed on training sets of CE-NBI images. First, the classification performance of the two approaches is compared separately. Then, an evaluation strategy is developed to compare the results of both methods based on the routine clinical examination procedure of laryngeal lesions. The result of this evaluation is presented in Contribution 2.

- ii. Method 2 explored textural characteristics of the CE-NBI images, focusing on simple and self-explainable features with minimum need for an image pre-processing step. This imaging modality visualizes a magnified representation of vascularization networks, which renders the application of traditional texture feature extraction techniques ineffective. In Method 2, every image is first divided into line profiles that are easy to process. As the pixel values of the vessels differed significantly from those of the image background, the line profiles deflected maintain-shaped behavior with various degrees of slopes on images with different vascularization networks. These behaviors resemble a cyclist stage profile where two Cyclist Effort Features (CyEff) are calculated based on the effort a cyclist should take to travel through each stage profile. The implementation of this approach was complex as there was a physics concept behind calculating the features, and there were many parameters to adjust. The supervised image classification is performed where four ML classifiers are trained and tested using CyEff to classify CE-NBI images into benign and malignant classes and then is compared with the performance of GF as well as standard texture-based features. The outcome of this study is presented in Contribution 3.
- b. Pipeline 2 is based on DL-based architectures for the supervised classification of CE-NBI images. The development strategy chosen for this pipeline addresses the challenge posed by an increased number of images in the data set. Although the data set started to have a wide variation of histopathologies, it raises concern about the complexity of the data due to the greater variety in patterns of vascularization networks. This issue affected the performance of the methods in pipeline 1, where the ML models could not generalize well on the new data. Therefore, there was necessary to change the

strategy to the DL-based methods for a more robust and objective assessment of CE-NBI images. But it was still in mind to choose a path that is simple, easy to understand, and optimal for real-time application in clinical examination.

- i. Method 3 focused on fully automatic CE-NBI image classification based on transfer learning and cut-off layer technique. By the time of development, the CE-NBI data set has reached approximately 8000 images, but this number was not sufficient to develop and train a DL-based model from scratch. Therefore, this option was discarded in favor of a transfer learning approach. After researching supervised classification methods of endoscopy images, pre-trained VGG19 and ResNet50 architectures are selected as two well-known and profound networks. However, the preliminary training and validation results lead to the decision to choose ResNet50 as the final architecture for Method 3. In three different experiments, the fine-tuning strategy and image augmentation techniques are applied to adopt the network's performance for CE-NBI image classification. Moreover, the ResNet50 is combined with the cut-off layer technique to optimize the network size to move toward real-time application. The result of this investigation is provided in Contribution 4.

Besides Method 3, the DL-based pipeline included the image classification strategy based on ensemble modeling implemented on the final CE-NBI data set. Since the results of this approach have yet to be published, Method 4 is not presented as a core contribution to this thesis. However, this strategy and its performance are discussed and demonstrated in Chapter 3 and Appendix B as Method 4.

Moreover, some parallel projects are ongoing on this topic, focusing on leukoplakia cases that are outside the domain of this thesis.

2.2 Contribution 1: Novel Automated Vessel Pattern Characterization of Larynx Contact Endoscopic Video Images

2.2.1 Summary

In this paper, we designed and implemented a set of handcrafted features – GF – to characterize the level of disorder of vessels in CE-NBI images. By using these features for different image classification tasks, we were aiming to see how these features can represent the geometrical characteristics of vascularization network and how these characteristics can be correlated to the type of laryngeal lesion and histopathology.

The main implementation block of this approach included the selection of proper image pre-processing techniques, designing indicators that characterize geometrical attributes of vascularization networks in CE-NBI images, extracting handcrafted features from the indicators, and performing four supervised image classification scenarios according to the type of vascularization networks, type of laryngeal lesion, as well as benign and malignant laryngeal histopathology.

The image pre-processing phase employed a three-step approach to enhance and segment vascularization networks in CE-NBI images. This step involved applying the Daubechies level 7 discrete wavelet transformation to the images, followed by the Frangi filter and skeletonization techniques. Subsequently, five indicators were computed based on the primary geometric characteristics of vascularization networks typically observed by Otolaryngologists in CE-NBI images. The direction-based indicators included the histogram of gradient direction (HGD) and rotational image averaging (RIA). Additionally, three indicators, namely angle (ANG), distance (DIS), and curvature (CUR), were introduced to capture the level of curvature exhibited by the vessels. Through qualitative analysis of the behavior of each indicator on three distinct types of vascularization networks, 24 handcrafted features were extracted using conventional mathematical and statistical operations. Finally, supervised classification scenarios were performed, where four ML classifiers were trained using GF and subsequently tested on various subsets of the data set.

2.2.2 Contribution

The author of this thesis initiated the main idea and conducted the design and implementation of vessels' pattern characterization indicators, handcrafted features, and image classification scenarios. Moreover, the author of this thesis reimplemented the image pre-processing techniques, performed the qualitative analysis of indicators' behavior, and evaluated features' performance on ML-based techniques. Finally, co-authors contributed to the data collection and preparation, results assessment, and manuscript revision.

2.2.3 Novel Automated Vessel Pattern Characterization of Larynx Contact Endoscopic Video Images

Nazila Esmaeili, Alfredo Illanes, Axel Boese, Nikolaos Davaris, Christoph Arens, and Michael Friebe



Novel automated vessel pattern characterization of larynx contact endoscopic video images

Nazila Esmaeili¹ · Alfredo Illanes¹ · Axel Boese¹ · Nikolaos Davaris² · Christoph Arens² · Michael Friebe¹

Received: 10 January 2019 / Accepted: 18 July 2019 / Published online: 27 July 2019
© The Author(s) 2019

Abstract

Purpose Contact endoscopy (CE) is a minimally invasive procedure providing real-time information about the cellular and vascular structure of the superficial layer of laryngeal mucosa. This method can be combined with optical enhancement methods such as narrow band imaging (NBI). However, these techniques have some problems like subjective interpretation of vascular patterns and difficulty in differentiation between benign and malignant lesions. We propose a novel automated approach for vessel pattern characterization of larynx CE + NBI images in order to solve these problems.

Methods In this approach, five indicators were computed to characterize the level of vessel's disorder based on evaluation of consistency of gradient and two-dimensional curvature analysis and then 24 features were extracted from these indicators. The method evaluated the ability of the extracted features to classify CE + NBI images based on the vascular pattern and based on the laryngeal lesions. Four datasets were generated from 32 patients involving 1485 images. The classification scenarios were implemented using four supervised classifiers.

Results For classification of CE + NBI images based on the vascular pattern, polykernel support vector machine (SVM), SVM with radial basis function (RBF), k-nearest neighbor (kNN), and random forest (RF) show an accuracy of 97%, 96%, 96%, and 96%, respectively. For the classification based on the histopathology, Polykernel SVM showed an accuracy of 84%, 86% and 84%, RBF SVM showed an accuracy of 81%, 87% and 83%, kNN showed an accuracy of 89%, 87%, 91%, RF showed an accuracy of 90%, 88% and 91% for classification between benign histopathologies, between malignant histopathologies and between benign and malignant lesions, respectively.

Conclusion These promising results show that the proposed method could solve the problem of subjectivity in interpretation of vascular patterns and also support the clinicians in the early detection of benign, pre-malignant and malignant lesions.

Keywords Contact endoscopy · Larynx · Vascular pattern · Feature extraction · Classification

Introduction

The larynx (voice box) is part of the head and neck region, and laryngeal cancer belongs to the most common cancer types with high incidence and mortality. Precancerous lesions such

as laryngeal dysplasia precede the development of laryngeal cancer. 85–95% of laryngeal cancers are squamous cell carcinomas (SCC) [1]. Early detection and diagnosis of suspicious mucosal lesions could provide an important opportunity to preserve the larynx and vocal fold function. Histopathological examination of suspicious laryngeal tissue using surgical biopsy is currently the gold standard for diagnosis, which is an invasive procedure and can cause serious problems for the patient [2].

The development of larynx endoscopy techniques provide a minimally invasive examination along with the possibility of early detection of vocal fold disorders. Barbalata and Mattos [3] proposed a method for laryngeal tumor detection and classification in narrow band imaging (NBI) endoscopic images. They reported an accuracy of 84.3% in recognizing malignant laryngeal tumors based on vascular characteriza-

This work was financially supported by the Federal Ministry of Education and Research (BMBF) in context of the 'INKA' project (Grand Number 03IPT7100X and by EFRE funding in context of the ego.-INKUBATOR program (ZS/2016/09//81061/IK 01/2015)).

✉ Nazila Esmaeili
nazila.esmaeili@ovgu.de

¹ INKA, Institute of Medical Technology, Otto-von-Guericke University Magdeburg, Magdeburg, Germany

² Department of Otorhinolaryngology, Head and Neck Surgery, Magdeburg University Hospital, Magdeburg, Germany

tion of the tumor. Turkmen et al. [4] proposed an approach to classify vocal fold disorders based on visible blood vessels and shape of vocal fold edges, with a sensitivity of 86%, 94%, 80%, 73%, and 76% for healthy, polyp, nodule, laryngitis, and sulcus vocalis classes, respectively. Moccia et al. [5] applied texture-based and first-order statistical features on 100×100 px patches in NBI endoscopic images to classify laryngeal tissue into four classes: tissue with intraepithelial papillary capillary loop (IPCL)-like vessels, leukoplakia, tissue with hypertrophic vessels and healthy tissue. They used support vector machine (SVM) classifier and reported achieved median classification recall of 93% with the best performing feature. In a recent study by Nanni et al. [6], an ensemble of convolutional neural networks (CNNs) and handcrafted features for bioimage classification was proposed. This ensemble obtained promising performance on the NBI endoscopic images dataset [5] with 97.33% accuracy to differentiate between four laryngeal tissue classes. Despite all the advantages, standalone application of normal white light video laryngoscopy or NBI cannot provide highly magnified visualization of color, contour, texture, and extent of mucosal lesions. For this reason, there is a need to have a technique that provides more precise evaluation of histopathology of laryngeal tissue for differential diagnosis of laryngeal cancerous lesions.

Contact endoscopy (CE) is an optical technique that allows detailed examination of the superficial layers of laryngeal mucosa providing a visualization of cells and vascular structures. This procedure is regularly performed using white light imaging, but it can also be combined with optical enhancement technologies like NBI. NBI is able to increase tissue contrast and to enhance the superficial vascular pattern at the site of examination [7]. The first application of CE in the larynx was reported in [8] and its efficiency was subsequently confirmed as a diagnostic tool in the evaluation of various pathologies in the larynx [9].

In the early application of larynx CE, the main focus was on finding histopathological information by evaluating the cellular architecture of the tissue. An example of that is the study [10], where a computer-assisted image analysis for diagnosis of precancerous and cancerous lesions based on the characterization of the cellular architecture in CE images was used. Recent studies showed that the evaluation of vascular patterns of the larynx superficial network can provide the surgeons more information than the cellular field. This is because the structure and the organization of blood vessels in the vocal fold is dynamic and undergoes significant changes in non-cancerous and cancerous stages [11,12]. Puxeddu et al. [13] visually classified vascular patterns in enhanced contact endoscopy (ECE) images into five categories for differential diagnosis between normal tissue and hyperplasia versus mild dysplasia and carcinoma. But,

there is no study on automatic classification of the CE vascular patterns.

It is recognized that the limiting factors of CE prevented it to gain acceptable place in routine clinical practice despite its potential advantages. Interpretation and evaluation of CE require extensive learning from the clinicians [14,15]. Studies showed that at the beginning of the training, there is a risk of subjective interpretation of vascular patterns [13,16]. This problem may cause an increased number of false positives which results in unnecessary biopsies [17]. Also, difficulty in differentiation between hyperplasia and mild-to-moderate dysplasia as well as an inability in differentiation of carcinoma in situ from carcinoma was reported [2,13,15].

The main objective of this work is to automatically characterize and assess vascular patterns in CE + NBI images to classify images based on the vascular pattern and laryngeal histopathology. For this, a new algorithm is proposed to evaluate the level of disorder in vessels based on the consistency of gradient direction and the vessels' curvature. Five indicators were computed after image preprocessing and vessel segmentation and then 24 features were extracted based on the qualitative properties of the indicators. The extracted features were fed into four different classifiers to classify images based on the vascular pattern and on larynx histopathology.

Material and methods

Data acquisition

Video scenes of 32 patients presenting different primary diagnosis were acquired during the examination of vocal folds with a frame rate equal to 30 frame per seconds (fps) in the department of Otorhinolaryngology at the University Hospital Magdeburg. A contact endoscope (KARL STORZ, Tuttlingen, Germany) in combination with an endoscopic imaging system (VISERA 4K UHD, Olympus, Japan) was used to capture the video scenes in Audio Video Interleave (AVI) format. In all procedures, the magnification of the contact endoscope was fixed at $60\times$ in order to have a fixed camera–tissue distance. For each patient, video segments where contact endoscope was used were manually extracted. Inside the video segments, we manually selected the intervals where the video quality was acceptable to see the vessels (good resolution without blur and artifacts). Then, one frame every three frames were extracted from the selected intervals in JPEG format images (1008×1280 px) and stored in the patient datasets to use them for the further processing.

Patients' data were pseudonymized, and only biopsy results were used as the final diagnosis for each patient, based on the classification of laryngeal histopathologies used by the medical doctors at the Magdeburg University Hospital (Fig. 1). In this classification, laryngeal histopathologies are

| Benign | | | | | Malignant | |
|--------|----------------------|----------------------|----------------|--------------------|------------------|-------------------|
| Cyst | Reinke's edema | Papilloma | | | Dysplasia severe | Carcinoma |
| Polyp | Chronic inflammation | Hyperkeratosis | Dysplasia mild | Dysplasia moderate | | Carcinoma in situ |
| | | Other benign lesions | | | | |

Fig. 1 A classification for laryngeal histopathology used at the University Hospital Magdeburg. The severity increases from left to right

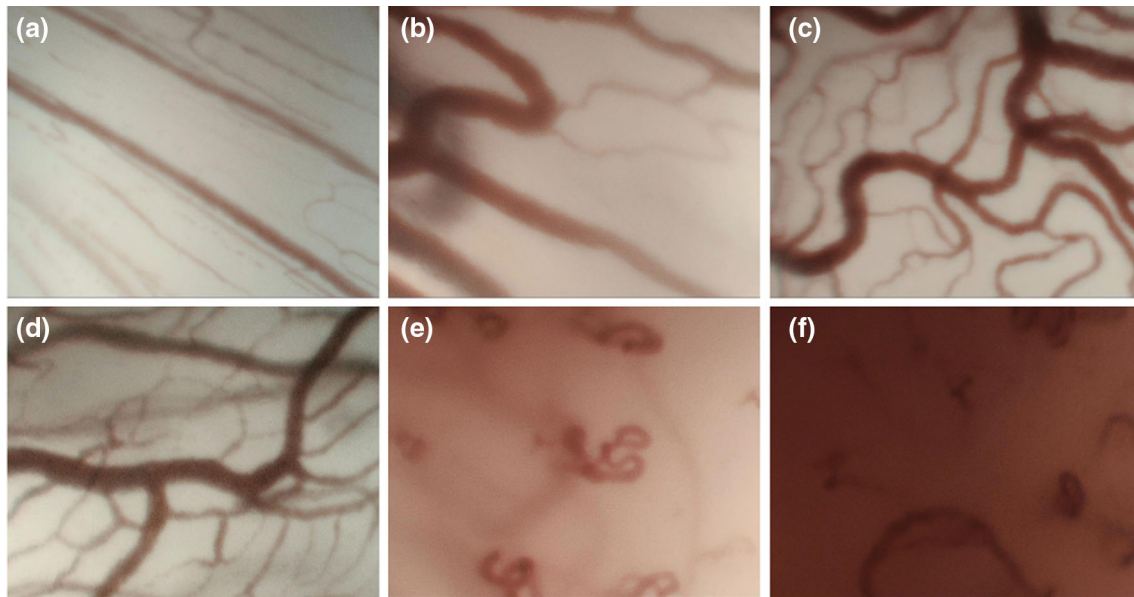


Fig. 2 Examples of CE + NBI images of six different cases: **a** healthy, **b** polyp, **c** reinke's edema, **d** dysplasia mild, **e** carcinoma in situ, **f** carcinoma

divided into two main groups: benign and malignant, which each of them subdivided in different histopathologies.

Image preprocessing and indicators extraction

Figure 2 shows examples of vocal fold images extracted from videos belonging to 6 different histopathologies. One of the main characteristics that clinicians observe in these images is the level of disorder in vessels. In conversation with our clinicians and based on recent publications, where vessel patterns were manually analyzed and classified [12–14], 5 indicators were proposed for characterizing vessel patterns. These indicators intended to take into account geometrical characteristics to assess the level of disorder of vascular patterns. The main idea was to extract features from the indicators and use them for classifying images according to the vessel's level of disorder and laryngeal histopathology. Figure 3 shows the main steps for the automatic feature extraction and classification procedures which are subsequently described in more details.

Image preprocessing and vessel segmentation

In order to remove the very low frequency trend in the image, a Daubechies level 7 discrete wavelet transformation was first applied to detrend each row and column of the image matrix [18]. Then a Frangi filter was used for vessel enhancement [19]. Frangi filter is a multiscale method using second order local structure of an image (Hessian) to find tubular structures as well as first-order transaction (gradient vector) to estimate the direction of these structures. In the image, vessels appear in different sizes. So it is important to have a measurement scale (Sigma) which varies within certain range in order to cover all different width and detect all vessels. The empirical tests performed in [19] showed that the range of Sigma between 1 to 8 can cover all the possible vascular structures. In this study, we have set the Sigma to the already tested values in order to extract the vessels in CE + NBI images. The resulting image was converted to a binary image followed by a skeletonization procedure using iterative thinning to reduce vessels to one-pixel-wide lines. This step resulted in

Fig. 3 Block diagram with the main steps of the proposed approach

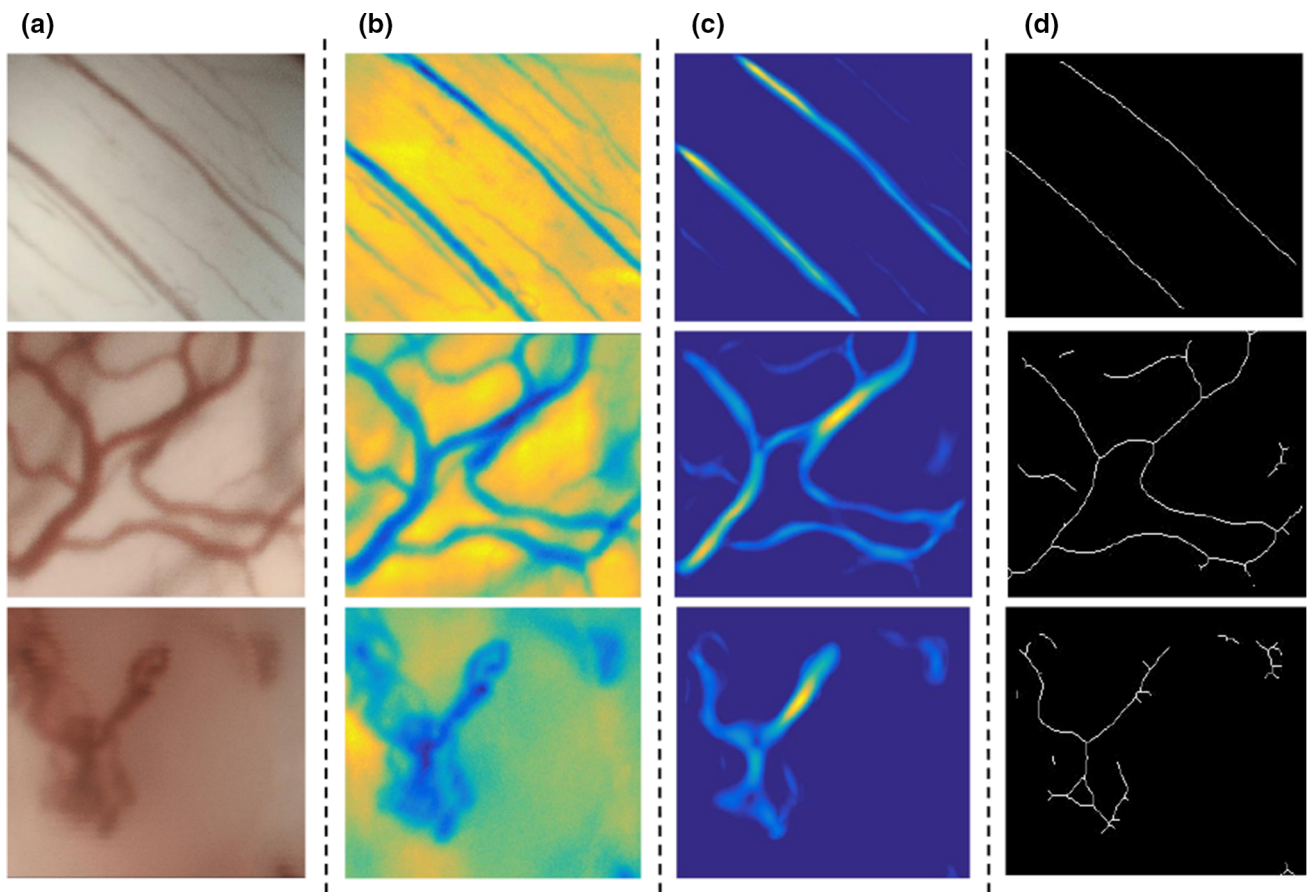
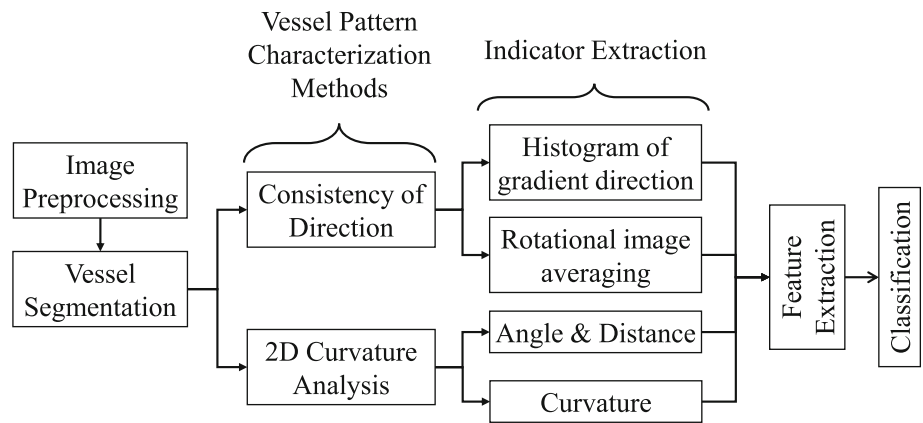


Fig. 4 Image preprocessing for three different vascular patterns in CE + NBI images: **a** original image, **b** homogenization, **c** Frangi filter, **d** skeletonization

three processed images: enhanced, filtered, and binary skeletonized referred as I_H , I_F and I_S , respectively (see examples in Fig. 4).

Image indicator extraction

As previously explained, five different indicators were computed to distinguish among the different vascular patterns

based on direction-based and curvature-based characteristics.

Direction-based indicators Two indicators were based on the consistency of the vessel direction, corresponding to histogram of gradient direction (HGD) and rotational image averaging (RIA). HGD was computed over the image I_H and RIA was computed over the image I_F .

The gradient of an image is a directional change in intensity. The image gradients are useful because the direction of gradients are more consistent (similar directions) in straight objects than in curved objects. For the HGD computation an algorithm was designed to compute magnitude and direction of the image gradients based on the method explained in [20]. The magnitude is computed to localize regions of significant gradient and the direction is used to compute an histogram of distribution of angles. The gradient direction was normalized based on the gradient magnitude values. In summary, the HGD indicator correspond to the normalized histogram of significant gradient directions.

RIA consists of computing the average over the rows of I_F for different rotation angles of the image. This average should be peaky when vessels are more or less parallel at a given angle and should show flatter behavior when the vessels are more curved. For that, the image was rotated from 0 to 360 degree in steps of 45 degrees, and at each rotation the average over the image was calculated as:

$$s_{\text{row}}^\theta(x) = \frac{1}{N} \times \sum_{y=1}^N I_F(x, y) \tag{1}$$

where s_{row}^θ is the resulting average row vector for the rotation angle θ , $I_F(x, y)$ represents the intensity value of the pixel at the location (x, y) and N is the number of rows of the image. The final RIA indicator correspond to the concatenation of each s_{row}^θ .

Curvature-based indicators For these indicators, vessel segments greater than 20 px in the image I_S were taken into account.

The first two indicators, angle (ANG) and distance (DIS), were computed from the distance and the angle between a defined reference point (A in Fig. 5) and each pixel belonging to the vessel’s skeleton ($C(x, y)$ in Fig. 5). The distance is simply calculated as the Euclidean distance between the reference and the skeleton point. For the angle computation a second reference point (B in Fig. 5) was defined and then the angle was computed between the vectors formed by the two reference and by the original reference with the skeleton point:

$$d(A, C) = \sqrt{(x_A - x_C)^2 + (y_A - y_C)^2} \tag{2}$$

$$\theta(A, C) = \arctan \left(\frac{\|\vec{AB} \times \vec{AC}\|}{\vec{AB} \cdot \vec{AC}} \right) \tag{3}$$

ANG and DIS correspond to vectors containing the resulting d and θ respectively for each pixel of a vessel segment. For each image an ANG and DIS vector per vessel segment is stored into a cell format.

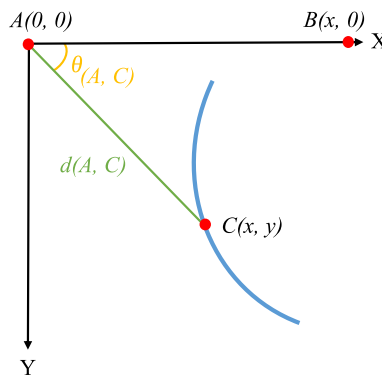


Fig. 5 Computation of indicators ANG and DIS

The third curvature-based indicator, curvature (CUR), was extracted directly from the level of curvature of the vessels. For each identified segment, the curvature at each pixel point is estimated using the method presented in [21], where a global approximation of tangents using a two linear digital straight segment is applied. The CUR indicator corresponded to the concatenation of the resulting curvatures of each identified vessel segment.

Results

Dataset generation

Four different datasets were generated in order to evaluate the performance of the proposed approach. The approach was first validated in terms of classifying CE + NBI images based on the vascular patterns. The reason of performing this test was to evaluate the ability of the algorithm to solve the problem of subjective interpretation of vascular patterns. Then the approach was validated in terms of its suitability to classify images based on the histopathologies of the larynx, with respect to level of disorder of vessels. These tests were performed to evaluate the ability of the algorithm to solve the problems related to difficulty in differentiation between benign and malignant lesions.

Dataset based on the degree of disorder of vascular patterns

Dataset I was generated to evaluate the performance of the proposed approach to differentiate between different degrees of disorder of vascular patterns. It included 1485 CE + NBI images from 32 patients and two medical experts came to a consensus to label them into three groups based on the vascular patterns: “order”, “disorder” and “very disorder”. “Order” vascular patterns relate to thin and parallel vessels. “Disorder” vascular patterns refer to longitudinal vascular changes.

Table 1 Histopathologies used for the generation of the three datasets

| Type of cancer | Histopathology | Patients | Images | Total |
|----------------|-------------------|----------|--------|------------------------|
| Benign | Cyst | 3 | 150 | 20 patients 890 images |
| | Polyp | 4 | 130 | |
| | Reinke's edema | 5 | 250 | |
| | Papilloma | 5 | 230 | |
| | Dysplasia mild | 3 | 130 | |
| Malignant | Dysplasia severe | 4 | 130 | 11 patients 465 images |
| | Carcinoma in situ | 4 | 155 | |
| | Carcinoma | 3 | 180 | |
| Total | | 31 | 1355 | – |

“Very disorder” vascular patterns involve perpendicular vascular pattern representing dilated IPCLs [12].

Dataset based on the histopathologies of the larynx

Three datasets were generated following the classification of laryngeal histopathologies (Fig. 1). Table 1 shows the different histopathologies including the number of patients and images per patient that were used to generate these datasets and to evaluate the performance of the proposed approach to differentiate between different laryngeal histopathologies:

- *Dataset II* CE + NBI images of the benign histopathologies. 20 patients with 890 images labeled into four groups: cyst, polyp & reinke's edema, papilloma, and dysplasia mild.
- *Dataset III* 465 CE + NBI images belonging to 11 patients diagnosed with malignant histopathologies labeled into three groups: dysplasia severe, carcinoma in situ and carcinoma.
- *Dataset IV* CE + NBI images belonging to 31 patients with benign and malignant histopathologies that included a total of 1355 images labeled into two groups: benign and malignant.

Qualitative analysis

Figure 6 shows the five indicators for three different vascular patterns. The indicators have qualitative characteristics that can be used to differentiate between different vascular patterns. A visual analysis of the indicators allows the following observation:

- The HGD indicator shows changes in the energy concentration with respect to the angle of the gradient vectors. Parallel vessel patterns show energy concentration of the gradient vector in two angles, while more chaotic vessel structures show a leakage in the energy distribution and even an equal distribution of energies (flat indicator)

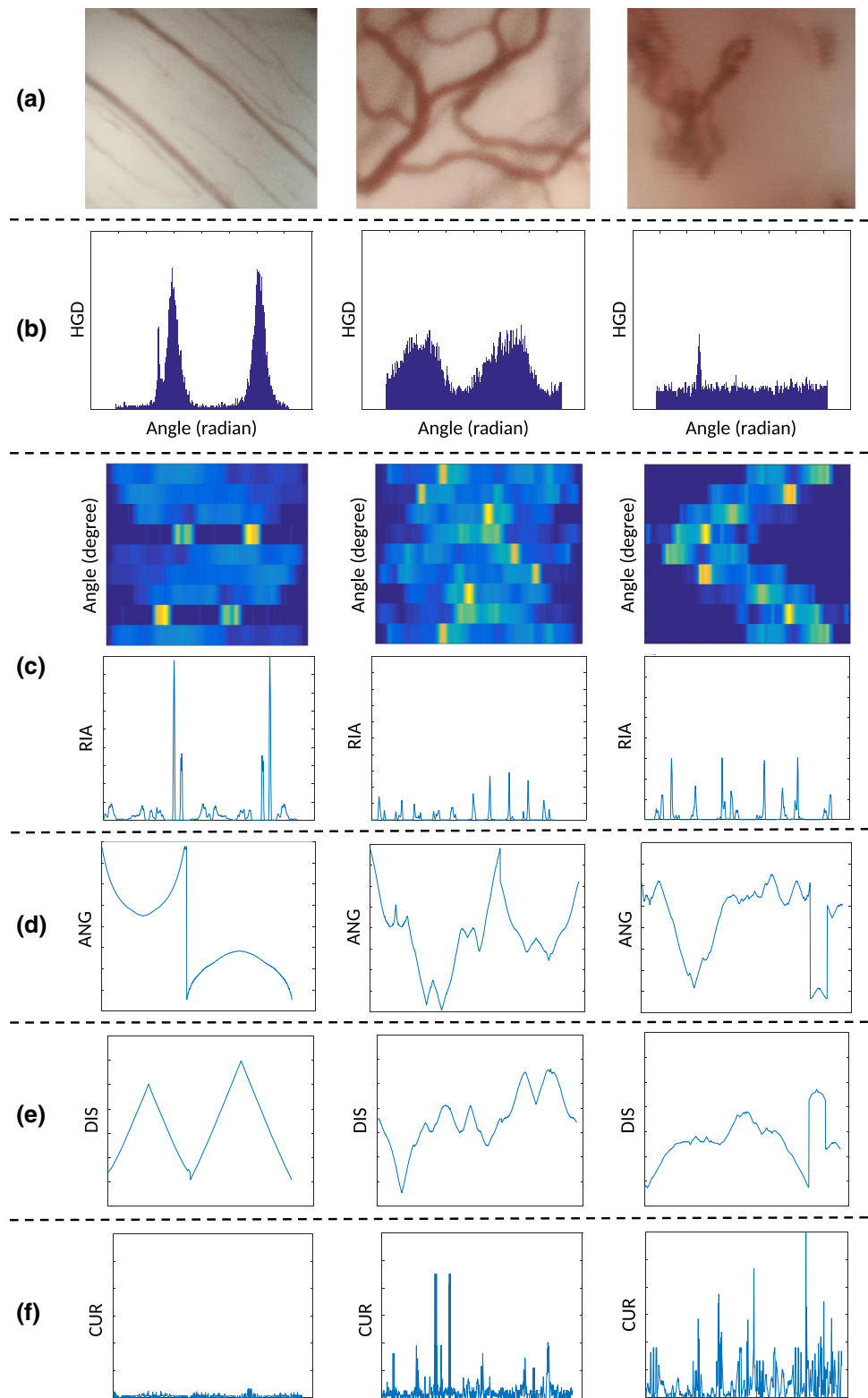
in the presence of spiral vessel patterns. It is possible to assume that the energy and energy-related characteristics of the HGD indicator can differentiate between different vascular patterns.

- The matrix of row averages for each rotation angle of the RIA indicator displays highly concentrated energies in two rotational angles when vessel patterns are parallel. This produces a final RIA containing a few number of main peaks of high amplitude. The more the vessel patterns become chaotic, the quantity of peaks and the energy leakage increase in the RIA. Energy-related features can therefore be used for characterizing vessel patterns.
- The displayed signal for both ANG and DIS indicators (Fig. 6) are a concatenated version for several vessel segments. This is why some signal discontinuities can be observed in the indicators. Disrespecting these discontinuities, we can observe that a vessel with significant curve patterns produces ANG and DIS indicators involving an increased number of changes per distance unit. This means that the quantity of changes of sign in their derivatives and the polynomial fitting errors will be higher for disorder patterns than for ordered ones, making it more suitable for distinguishing between patterns.
- CUR indicator variance increases when the vessel patterns become disorder. This is mainly because disorder patterns involve a higher number of loops and therefore more significant curvature's values. For this indicator the features are also based on energies and peaks in the signal and also statistical values as variance.

Following this analysis, 24 features extracted from the 5 indicators are proposed for assessing vessel patterns, explained in the following.

HGD features Four features are proposed from the HGD indicator. The first, second, and third features, F_1 , F_2 , and F_3 , are simply computed as the total energy and the minimal value of the HGD indicator, and as the difference between the maximum and minimum value of the indicator, respectively.

Fig. 6 Five indicators for three different vascular patterns in CE + NBI images: **a** original image, **b** HGD indicator, **c** RIA indicator, **d** ANG indicator, **e** DIS indicator, **f** CUR indicator



$$F_1 = \sum_{g=1}^{N_{\text{HGD}}} \text{HGD}^2(g) \quad (4)$$

$$F_2 = \min_g[\text{HGD}(g)] \quad (5)$$

$$F_3 = \max_g[\text{HGD}(g)] - \min_k[\text{HGD}(g)] \quad (6)$$

where g correspond to the sample index of the indicator and N_{HGD} correspond to the total number of samples of the HGD indicator. The fourth HGD indicator F_4 intends to assess localized energy concentration of the indicator's peaks. For that, significant peaks in HGD are first identified using a simple signal peak detector. Let on_i and off_i being the onset and offset of the HGD peak waveform i (HGDP_i) and n_p the number of significant peaks identified in the HGD indicator. Then F_4 is computed as the ratio between the sum of the energy of the peaks of HGD and its total energy.

$$F_4 = \frac{\sum_1^{n_p} \left[\sum_{\text{on}_i}^{\text{off}_i} \text{HGDP}_i^2 \right]}{F_1} \quad (7)$$

RIA features Four features are extracted from the RIA indicator. The first two features, F_5 and F_6 , are computed as the total energy and as the number of significant peaks in the RIA indicator, respectively.

$$F_5 = \sum_{g=1}^{N_{\text{RIA}}} \text{RIA}^2(g) \quad (8)$$

$$F_6 = \text{Peaks}[\text{RIA}] \quad (9)$$

where g correspond to the sample index of the indicator, N_{RIA} correspond to the total number of samples of the RIA indicator and Peaks denote a function for significant signal peaks detection using a standard peak detector. For the third RIA feature F_7 , a similar approach than for the computation of F_4 is proposed but using the average of the peak energies instead of the summation.

$$F_7 = \frac{\frac{1}{n_p} \sum_1^{n_p} \left[\sum_{\text{on}_i}^{\text{off}_i} \text{RIA}_i^2 \right]}{F_5} \quad (10)$$

The fourth RIA feature F_8 is computed as the average of the ratios between amplitude and width of each peak waveform of the indicator.

$$F_8 = \frac{1}{n_p} \sum_1^{n_p} \left[\frac{\text{Amplitude}(\text{RIAP}_i)}{\text{Width}(\text{RIAP}_i)} \right] \quad (11)$$

where RIAP_i correspond to the RIA peak waveform i , n_p to the number of identified RIA peaks and amplitude and

width correspond to functions that compute the amplitude and width of each peak waveform.

ANG features Six features are extracted from the ANG indicator. One of the main characteristic of the ANG and DIS indicators is the change of sign in the derivative. Therefore, the first four ANG features, F_9 , F_{10} , F_{11} and F_{12} , are computed by exploiting this characteristic. Let M be the number of vessel segments identified in an image. For each vessel segment m , the derivative of the ANG indicator is first computed using the derivative filter presented in [22]. Then, the number of changes of sign s_m is computed for each segment m and is used for computing the features. F_9 , F_{10} , F_{11} and F_{12} are computed as the mean of s_m , the total number of changes of sign in an image, the maximal and the median of s_m , respectively.

$$F_9 = \frac{1}{M} \sum_{m=1}^M s_m \quad (12)$$

$$F_{10} = \sum_{m=1}^M s_m \quad (13)$$

$$F_{11} = \max_m [s_m] \quad (14)$$

$$F_{12} = \text{median} [s_m] \quad (15)$$

Additionally, two features are computed based on the error e_m of a 3rd degree polynomial fitting for each vessel segment m .

$$F_{13} = \frac{1}{M} \sum_{m=1}^M e_m \quad (16)$$

$$F_{14} = \text{median} [e_m] \quad (17)$$

DIS features Six features were extracted from the DIS indicator (F_{15} to F_{20}) using the same equations which were explained for the ANG indicator.

CUR features Four features are proposed from the CUR indicator. The first three CUR features, F_{21} , F_{22} and F_{23} are simply computed as the total energy, the number of significant peaks and the variance of the CUR indicator.

$$F_{21} = \sum_{g=1}^{N_{\text{CUR}}} \text{CUR}^2(g) \quad (18)$$

$$F_{22} = \text{Variance} [\text{CUR}] \quad (19)$$

$$F_{23} = \text{Peaks} [\text{CUR}] \quad (20)$$

where g correspond to the sample index of the indicator and N_{CUR} correspond to the total number of samples of the CUR indicator. The fourth CUR feature F_{24} takes into account

the observation that more chaotic vessel patterns will result in a bigger number of curves whose curvature level also is bigger. This is why we proposed as feature the number of signal peaks times the amplitude of the peak.

$$F_{24} = n_p \times \sum_1^{n_p} \text{Amplitude (CUR)} \quad (21)$$

The approach was implemented in MATLAB R2016b and executed on a PC with a CPU operating at 2.30 GHz resulting in an execution time of 4.02 seconds per image for image preprocessing, indicator computation, and feature extraction.

Features classification performances

SVM, k-nearest neighbors (kNN), and random forests (RF) were used to classify CE + NBI images first based on three different vascular patterns (database I) and then based on the different histopathologies of the larynx (database II, III, and IV).

SVM performs classification by finding the hyperplane that maximizes the margin between the classes. The objective of the SVM algorithm is to find an optimal hyperplane in an N-dimensional space that distinctly classifies the data points. In this study, SVM with polykernel and radial basis function (RBF) kernel were used in order to classify linear and nonlinear separable data, respectively. The grid search method was used in order to optimize the SVM parameters using tenfold cross-validation and the classification was performed using a sequential minimal optimization (SMO) algorithm. In SVM Polykernel, there is one important parameter to optimize which is C , while in SVM with RBF kernel, there are two main parameters to optimize which are C and γ . C is the regulation parameter that controls the cost of misclassification on the training data and γ is the kernel parameter that defines how far the influence of a single training example reaches. In our study, we decided to make the range of C and γ from 0.01 to 1000 with 10 times increment. The SVM with Polykernel performed the best with $C = 1$ and SVM with RBF kernel showed the best performance with $C = 1$ and $\gamma = 0.01$. Furthermore, for solving the multi-class problem, a pairwise classifier trained the SVM to assign features into multi-class [23–25].

kNN is a nonparametric method and performs classification by finding the most similar data points in the training data and making an educated guess based on their classifications. The input consists of the k closest training examples in the feature space and the output is a class membership. In this study and in order to classify and assign a new sample to a new class, the distance of a sample was calculated using Euclidean distance algorithm. In kNN, k is the main parameter to optimize. For that, we used grid search method to find

Table 2 Classification results using Polykernel SVM classifier

| Database | Accuracy | Sensitivity | Specificity | AUC |
|-------------|----------|-------------|-------------|-------|
| Dataset I | 0.973 | 0.980 | 0.983 | 0.977 |
| Dataset II | 0.846 | 0.819 | 0.942 | 0.917 |
| Dataset III | 0.864 | 0.856 | 0.931 | 0.917 |
| Dataset IV | 0.847 | 0.806 | 0.868 | 0.837 |

the optimized value with a range of k from 1 to 10 with step size equal to 1 and used tenfold cross-validation to select the best value. The classifier showed the best performance with $k = 3$ [26].

RF is an ensemble learning method for classification that operates by constructing a multitude of decision trees at training time and outputting the class. In this study, RF was trained via the bagging method. Bagging consists of randomly sampling subsets of the training data, fitting a model to these smaller data sets, and aggregating the predictions. Hence, instead of searching greedily for the best predictors to create branches, it randomly samples elements of the predictor space, thus adding more diversity and reducing the variance of the trees at the cost of equal or higher bias. There are many parameters in RF that can be optimized. In this study, we optimized only two important parameters which were the depth of the trees and number of estimators. The depth of the trees specifies the maximum depth of each tree and the number of estimators specifies the number of trees in the forest of the model. We made the range for the depth of the trees from 1 to 10 with step size equal to 1 and for number of estimators from 10 to 100 with step size equal to 5. The optimum parameters that were obtained after using grid search method with tenfold cross-validation were the depth of 8 with 50 trees [27].

A 24-dimensional space was fed into each classifier. The selected classifiers were applied by employing WEKA 3.8.1 as a machine learning tool. For all classification scenarios, a tenfold cross-validation was used for testing as well as for hyperparameter tuning. In order to measure the performance of each classifier, a confusion matrix was computed for each classification scenario and the accuracy, sensitivity, specificity, and area under the curve (AUC) receiver operating characteristics (ROC) were obtained from it. Tables 2, 3, 4, and 5 illustrate the classification results for each classifier. As we used tenfold cross-validation for all the classification scenarios, the values presented in these tables are the average results.

Discussion

To our knowledge, this is the first study on automatic characterization of vascular patterns in CE + NBI images with

Table 3 Classification results using RBF SVM classifier

| Database | Accuracy | Sensitivity | Specificity | AUC |
|-------------|----------|-------------|-------------|-------|
| Dataset I | 0.968 | 0.976 | 0.978 | 0.973 |
| Dataset II | 0.816 | 0.757 | 0.926 | 0.901 |
| Dataset III | 0.873 | 0.864 | 0.931 | 0.921 |
| Dataset IV | 0.837 | 0.834 | 0.839 | 0.837 |

Table 4 Classification results using kNN classifier

| Database | Accuracy | Sensitivity | Specificity | AUC |
|-------------|----------|-------------|-------------|-------|
| Dataset I | 0.965 | 0.974 | 0.978 | 0.989 |
| Dataset II | 0.892 | 0.879 | 0.958 | 0.969 |
| Dataset III | 0.877 | 0.873 | 0.939 | 0.956 |
| Dataset IV | 0.912 | 0.871 | 0.933 | 0.953 |

Table 5 Classification results using RF classifier

| Database | Accuracy | Sensitivity | Specificity | AUC |
|-------------|----------|-------------|-------------|-------|
| Dataset I | 0.966 | 0.975 | 0.979 | 0.996 |
| Dataset II | 0.906 | 0.900 | 0.965 | 0.981 |
| Dataset III | 0.884 | 0.879 | 0.943 | 0.973 |
| Dataset IV | 0.911 | 0.939 | 0.858 | 0.979 |

classification of the images using a set of features describing the level of disorder of vascular patterns.

Regarding the evaluation of the vascular structure in the CE images, there is a study [13] which classified the vascular patterns in ECE images into five groups. This classification was matched to the final diagnosis, with accuracy in the differential diagnosis between normal tissue and hyperplasia versus mild dysplasia and carcinoma of 97.6%. The result of this classification was based on the experience of the clinicians with a risk of subjective interpretation of vascular patterns in CE images [13,16]. In contrast to that, we used an automated algorithm to characterize the level of disorder of vascular patterns. This method showed the ability to differentiate between three different vascular patterns with the accuracy, sensitivity, specificity, and AUC of over 96%. For the final diagnosis based on the vascular patterns in the ECE images of the vocal fold, [13,15] reported the difficulty in differentiation between hyperplasia and low to moderate dysplasia. In comparison with our study, a different classification was used for laryngeal histopathologies and the RF classifier showed the best results to differentiate between benign histopathologies with an accuracy of 90%, a sensitivity of 90%, a specificity of 96%, and AUC of 98%.

For the evaluation of the cellular architecture of the most superficial mucosa in the head and neck area, according to [28] CE has an accuracy of 72–92%, a sensitivity of 77–100%, and a specificity of 66–100% to diagnose benign

and malignant head and neck mucosal lesions. These results depend on the experience of clinicians and are based on the evaluation of cellular structures. Our proposed approach can perform an automatic differentiation between benign and malignant lesions based on vascular structure. All classifiers resulted in accuracy, sensitivity, specificity, and AUC of over 83%. The studies that focus on the cellular structure of the laryngeal tissue in CE reported the difficulty in diagnostic differentiation of carcinoma in situ from carcinoma [2], as well as dysplasia severe from carcinoma in situ and carcinoma [10,29]. Our approach has the potential to solve these problems by distinguishing between three malignant histopathologies. For that, RF classifier showed the best results with the accuracy, sensitivity, specificity, and AUC of 88%, 87%, 94% and 97%, respectively.

Conclusion

Based on the results, the presented approach could provide a confident way for clinicians to interpret vascular patterns in CE + NBI images with high accuracy. It also confirms the relevance of the vascular structures to the laryngeal histopathologies and to the stage of laryngeal cancer. Our approach has the potential to operate as an assisting system to help the clinicians make the final decision about the histopathology of the laryngeal tissue in the routine and surgical procedures.

As a first work in this field, our main objective was to propose an approach for characterization of the vascular patterns. Based on the discussion with clinicians and their requirements, we planned first to test the ability of the algorithm to differentiate between benign and malignant cases and then test the performance of the algorithm in each group (benign and malignant) to differentiate between different histopathologies. The next step of our work will be a multi-class classification with other features and considering all benign and malignant histopathologies. Further work is necessary to improve the results by computing more indicators, applying other feature extraction methods and implementing feature selection techniques to evaluate the influence of each class of features on the final results.

Also, the presented CE problems seem to be an ideal basis for machine learning approaches such as shown by [5,6] using texture-based features and CNNs. This is possible when a high amount of data is available. CE is not a routine procedure in the clinical settings which caused a limitation on the number of patients available for our study. This problem also led to other limitations in the variety of histopathologies, especially in the benign cases. Therefore, the classification scenario of the benign cases was conducted with only available histopathologies. Hence, increasing the number of images per each dataset, testing the algorithm

with other available classification of vascular patterns in CE images, and applying other classification methods are suggested for the future.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interest in this work.

Ethical approval The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

1. Markou K, Christoforidou A, Karasmanis I, Tsiropoulos G, Triaridis S, Constantinidis I, Vital V, Nikolaou A (2013) Laryngeal cancer: epidemiological data from Northern Greece and review of the literature. *Hippokratia* 17(4):313
2. Mishra A, Nilakantan A, Datta R, Sahai K, Singh SP, Sethi A (2012) Contact Endoscopy: a promising tool for evaluation of laryngeal mucosal lesions. *J Laryngol Voice* 2(2):53
3. Barbalata C, Mattos LS (2016) Laryngeal tumor detection and classification in endoscopic video. *IEEE J Biomed Health Inf* 20(1):322–332
4. Turkmen HI, Karsligil ME, Kocak I (2015) Classification of laryngeal disorders based on shape and vascular defects of vocal folds. *Comput Biol Med* 62:76–85
5. Moccia S, De Momi E, Guarnaschelli M, Savazzi M, Laborai A, Guastini L, Peretti G, Mattos LS (2017) Confident texture-based laryngeal tissue classification for early stage diagnosis support. *J Med Imaging* 4(3):034502
6. Nanni L, Ghidoni S, Brahmam S (2018) Ensemble of convolutional neural networks for bioimage classification. *Appl Comput Inform*. <https://doi.org/10.1016/j.aci.2018.06.002>
7. Yang SW, Lee YS, Chang LC, Hwang CC, Chen TA (2012) Diagnostic significance of narrow-band imaging for detecting high-grade dysplasia, carcinoma in situ, and carcinoma in oral leukoplakia. *Laryngoscope* 122:2754–2761
8. Andrea M, Dias O, Santos A (1995) Contact endoscopy of the vocal cord: normal and pathological patterns. *Acta Oto-Laryngol* 115(2):314–316
9. Arens C, Dreyer T, Glanz H, Malzahn K (2003) Compact endoscopy of the larynx. *Ann Otol Rhinol Laryngol* 112(2):113–119
10. Tamawski W, Frączek M, Jeleń M, Kręcicki T, Zalesska-Kręcicka M (2008) The role of computer-assisted analysis in the evaluation of nuclear characteristics for the diagnosis of precancerous and cancerous lesions by contact laryngoscopy. *Adv Med Sci* 53(2):221–227
11. Stefanescu DC, Ceachir OC, Zainea VI, Hainarosie M, Pietrosanu C, Ionita IG, Hainarosie R (2016) Methylene blue video contact endoscopy enhancing methods. *Rev Chim* 67:1558–1559
12. Arens C, Piazza C, Andrea M, Dikkers FG, Gi RETP, Voigt-Zimmermann S, Peretti G (2016) Proposal for a descriptive guideline of vascular changes in lesions of the vocal folds by the committee on endoscopic laryngeal imaging of the European Laryngological Society. *Eur Arch Oto-Rhino-Laryngol* 273(5):1207–1214
13. Puxeddu R, Sionis S, Gerosa C, Carta F (2015) Enhanced contact endoscopy for the detection of neoangiogenesis in tumors of the larynx and hypopharynx. *Laryngoscope* 125(7):1600–1606
14. Puxeddu R, Carta F, Ferreli C, Natalia C, Gerosa C (2018) Enhanced contact endoscopy (ECE) in head and neck surgery. *Endo-Press*
15. Carta F, Sionis S, Cocco D, Gerosa C, Ferreli C, Puxeddu R (2016) Enhanced contact endoscopy for the assessment of the neoangiogenetic changes in precancerous and cancerous lesions of the oral cavity and oropharynx. *Eur Arch Oto-Rhino-Laryngol* 273(7):1895–1903
16. Mannelli G, Cecconi L, Gallo O (2016) Laryngeal preneoplastic lesions and cancer: challenging diagnosis. *Qualitative literature review and meta-analysis*. *Crit Rev Oncol Hematol* 106:64–90
17. Piazza C, Cocco D, Del Bon F, Mangili S, Nicolai P, Peretti G (2011) Narrow band imaging and high definition television in the endoscopic evaluation of upper aero-digestive tract cancer. *Acta Otorhinolaryngol Ital* 31(2):70
18. Boese A, Illanes A, Balakrishnan S, Davaris N, Arens C, Friebe M (2018) Vascular pattern detection and recognition in endoscopic imaging of the vocal folds. *Curr Dir Biomed Eng* 4(1):75–78
19. Frangi AF, Niessen WJ, Vincken KL, Viergever MA (1998) Multiscale vessel enhancement filtering. In: *International conference on medical image computing and computer-assisted intervention*. Springer, Berlin, pp 130–137
20. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. In: *IEEE computer society conference on computer vision and pattern recognition, CVPR 2005, vol 1, pp 886–893*
21. Hermann S, Klette R (2003) Multigrid analysis of curvature estimators. *CITR, The University of Auckland, New Zealand*
22. Illanes A, Zhang Q, Medigue C, Papelier Y, Sorine M (2006) Multi-lead T wave end detection based on statistical hypothesis testing. In: *6th IFAC symposium on modelling and control in biomedical systems, MCBMS'06, pp 93–98*
23. Hsu CW, Chang CC, Lin CJ (2003) A practical guide to support vector classification. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
24. Ring M, Eskofier BM (2016) An approximation of the Gaussian RBF kernel for efficient classification with SVMs. *Pattern Recognit Lett* 84:107–113
25. Syarif I, Prugel-Bennett A, Wills G (2016) SVM parameter optimization using grid search and genetic algorithm to improve classification performance. *Telkonnika* 14(4):1502
26. Peterson LE (2009) K-nearest neighbor. *Scholarpedia* 4(2):1883
27. Breiman L (2001) Random forests. *Mach Learn* 45(1):5–32
28. Szeto C, Wehrli B, Whelan F, Franklin J, Nichols A, Yoo J, Fung K (2011) Contact endoscopy as a novel technique in the detection and diagnosis of mucosal lesions in the head and neck: a brief review. *J Oncol* 2011:196302
29. Mishra AK, Nilakantan A, Sahai K, Datta R, Malik A (2014) Contact endoscopy of mucosal lesions of oral cavity-preliminary experience. *Med J Armed Forces India* 70(3):257–263

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

2.3 Contribution 2: Laryngeal Lesion Classification Based on Vascular Patterns in Contact Endoscopy and Narrow Band Imaging: Manual versus Automatic Approach

2.3.1 Summary

In this paper, we designed and implemented a scenario for laryngeal lesion assessment based on manual and automatic CE-NBI image classification. We first aimed to compare the performance of both approaches on the classification of benign and malignant laryngeal lesions based on the vascular patterns in CE-NBI images. Then, we studied and evaluated the challenges of manual classification and showed how the GF combined with ML classifiers as a computer-based approach can assist Otolaryngologists in overcoming these problems.

The main implementation block of this approach focused on the automatic classification strategy that was designed based on the routine and standard clinical examination of the laryngeal lesions. It included updating the CE-NBI data set, improving the image pre-processing step, fine-tuning the ML classifiers, and comparison of manual and automatic approaches based on the level of agreement/disagreement between Otolaryngologists as well as misclassification levels of each approach based on the laryngeal histopathologies.

The CE-NBI subset 1 consisted of two to five randomly chosen CE-NBI images per patient, which were subjected to manual evaluation by Otolaryngologists. This subset was subsequently employed as the testing set for the automatic approach. In the manual approach, three specialists and three resident Otolaryngologists visually assessed the CE-NBI images and categorized the patients into benign and malignant groups based on the appearance of PVCs in the images. For the automatic approach, Method 1 was utilized for CE-NBI image classification. This method integrated the GF with four ML classifiers. Notably, the Frangi filter was substituted with the Jerman filter, and the tuning parameters of the ML classifiers were adjusted accordingly. Two procedures were developed and implemented to compare the results obtained from the manual and automatic approaches. The study indicated that the subjective evaluation during visual assessment could be reduced by Method 1, especially in case of disagreements among Otolaryngologists.

2.3.2 Contribution






The author of this thesis initiated the main idea and conducted the design and implementation of the automatic classification scenario. In addition, the adjustment of the image pre-processing techniques and hyperparameter tuning of four supervised ML classifiers have been done by the author of this thesis. Finally, co-authors helped conduct the manual classification scenario and contributed to the data collection and preparation, results assessment, and manuscript revision.

2.3.3 Laryngeal Lesion Classification Based on Vascular Patterns in Contact Endoscopy and Narrow Band Imaging: Manual versus Automatic Approach

Nazila Esmaeili, Alfredo Illanes, Axel Boese, Nikolaos Davaris, Christoph Arens, Nassir Navab, and Michael Friebe

Article

Laryngeal Lesion Classification Based on Vascular Patterns in Contact Endoscopy and Narrow Band Imaging: Manual Versus Automatic Approach

Nazila Esmaeili ^{1,*}, Alfredo Illanes ¹, Axel Boese ¹, Nikolaos Davaris ², Christoph Arens ², Nassir Navab ³ and Michael Friebe ^{1,4}

¹ INKA-Application Driven Research, Otto-von-Guericke University Magdeburg, 39120 Magdeburg, Germany; alfredo.illanes@med.ovgu.de (A.I.); axel.boese@ovgu.de (A.B.); michael.friebe@ovgu.de (M.F.)

² Department of Otorhinolaryngology, Head and Neck Surgery, Magdeburg University Hospital, 39120 Magdeburg, Germany; nikolaos.davaris@med.ovgu.de (N.D.); christoph.arenas@med.ovgu.de (C.A.)

³ Chair for Computer Aided Medical Procedures and Augmented Reality, Technical University Munich, 85748 Munich, Germany; navab@cs.tum.edu

⁴ IDTM GmbH, 45657 Recklinghausen, Germany

* Correspondence: nazila.esmaeili@med.ovgu.de

Received: 29 May 2020; Accepted: 18 July 2020; Published: 19 July 2020



Abstract: Longitudinal and perpendicular changes in the vocal fold's blood vessels are associated with the development of benign and malignant laryngeal lesions. The combination of Contact Endoscopy (CE) and Narrow Band Imaging (NBI) can provide intraoperative real-time visualization of the vascular changes in the laryngeal mucosa. However, the visual evaluation of vascular patterns in CE-NBI images is challenging and highly depends on the clinicians' experience. The current study aims to evaluate and compare the performance of a manual and an automatic approach for laryngeal lesion's classification based on vascular patterns in CE-NBI images. In the manual approach, six observers visually evaluated a series of CE+NBI images that belong to a patient and then classified the patient as benign or malignant. For the automatic classification, an algorithm based on characterizing the level of the vessel's disorder in combination with four supervised classifiers was used to classify CE-NBI images. The results showed that the manual approach's subjective evaluation could be reduced by using a computer-based approach. Moreover, the automatic approach showed the potential to work as an assistant system in case of disagreements among clinicians and to reduce the manual approach's misclassification issue.

Keywords: laryngeal cancer; contact endoscopy; narrow band imaging; automatic classification; feature extraction; machine learning

1. Introduction

Laryngeal cancer is the second most frequent malignant tumor of the head and neck region [1]. The vast majority of primary laryngeal cancers are Squamous Cell Carcinomas (SCC) arising from the epithelial lining of the larynx, mostly as a result of tobacco and alcohol consumption. A total of 40% of these cancers are diagnosed at an advanced stage, which is associated with a poorer prognosis and quality of life [2]. The early diagnosis of laryngeal cancer is crucial to reduce patient mortality and preserve vocal fold function.

Specific changes in the morphology and three-dimensional orientation of the vocal fold's sub-epithelial blood vessels have proved to be associated with the development of benign and malignant laryngeal lesions. Several approaches have been proposed to describe and classify these

vascular changes. Among the complex classification systems proposed by [3] and [4], the European Laryngological Society (ELS) introduced a simplified classification that divides vascular changes into longitudinal and perpendicular classes [5,6]. Longitudinal Vascular Changes (LVC) spread along the length and width of the vocal fold and can be observed in all kinds of benign or malignant lesions. On the contrary, Perpendicular Vascular Changes (PVC) develop perpendicularly towards the mucosa, as a result of neoangiogenesis in laryngeal Papillomatosis, pre-malignant and malignant histopathologies.

The endoscopic detection and evaluation of vascular changes can provide complementary diagnostic information for clinicians to detect and differentiate between benign and malignant laryngeal lesions [7]. As a minimally-invasive endoscopic technique, Contact Endoscopy (CE) can provide real-time visualization of cellular and vascular structures of the laryngeal mucosa [8,9]. For the purpose of detecting and evaluating superficial vascular changes, several enhanced endoscopic techniques such as Narrow Band Imaging (NBI) have been combined with CE to ease the detection of vascular changes [10]. The use of enhanced CE showed promising results in the assessment of vascular patterns followed by indicative of various laryngeal pathologies [4,11,12].

Clinicians can receive useful information about the type and suspected histopathology of laryngeal lesions by evaluating LVC and PVC in enhanced CE images; however, it is a challenging task for them. There are similarities between vascular patterns of benign and malignant laryngeal lesions. The PVC with wide-angled turning points, as observed in laryngeal Papillomatosis can be difficult to distinguish from PVC with narrow-angled turning points, as observed in pre-malignant and malignant histopathologies [5,12–14]. Hence, the interpretation of vascular patterns in enhanced CE images requires an extensive learning curve from the clinicians to reduce the risk of subjective evaluation that can cause potential problems in differentiation between benign and malignant laryngeal lesions [4,10,12,15,16].

In this study, we first aimed to present the results of manual versus automatic classification of benign and malignant laryngeal lesions based on the vascular patterns in CE-NBI images. We then evaluated the issues of manual classification and subsequently showed how a computer-based approach can assist the clinicians to overcome these problems. A manual and an automatic classification approach were defined to conduct this evaluation. In the manual approach, six experienced and less experienced otolaryngologists individually evaluated PVC and LVC in CE-NBI images of patients and classified them into benign and malignant groups. An updated version of the algorithm proposed in [17,18] with 24 features and four supervised classifiers has been used to classify CE-NBI images into benign and malignant groups. The results of the two approaches were compared in terms of classification sensitivity and specificity. The potential of an automatic approach to assist the clinicians is presented through two evaluation strategies.

2. Material and Methods

2.1. Data Acquisition

CE-NBI images were extracted from video scenes of adult patients who received a microlaryngoscopy for benign, pre-malignant or malignant lesions of the vocal folds. Video scenes were captured using an Evis Exera III Video System with integrated NBI-filter (Olympus Medical Systems, Hamburg, Germany) and a rigid 30-degree contact endoscope (Karl Storz, Tuttlingen, Germany) with a fixed magnification of 60 to have a fixed camera–tissue distance. For each video scene, we selected the time intervals where the video quality was good enough to visualize the vessels. Then, one in every ten frames was extracted from the selected intervals in JPEG format images (1008 × 1280 pixels) to have unique and non-redundant CE-NBI images.

2.2. Dataset Generation

The CE-NBI dataset included 1632 extracted images of 68 patients. The patients' data were pseudonymized. Based on the WHO classification [19], histological diagnoses were used to label images as belonging to a benign or a malignant class. Table 1 shows the histopathologies with the number of patients and images used for the generation of the dataset.

Two image subsets were created from the CE-NBI dataset. The *Subset I* included a series of two to five randomly selected CE-NBI images of each patient—total of 336 images, $\approx 20\%$ of the dataset. The *Subset II* included the rest of the CE-NBI images—a total of 1296 images, $\approx 80\%$ of the dataset, and was used as the training set of the automatic approach. The *Subset I* was evaluated by the otolaryngologists in the manual approach and then used as the testing set for the automatic approach. Figure 1 presents some examples of CE-NBI images with LVC and PVC belonging to the generated dataset.

Table 1. Histopathologies used for the generation of the dataset.

| Type of Lesion | Histopathology | Number of Patients | Number of Images |
|----------------|----------------------|--------------------|------------------|
| Benign | Cyst | 3 | 90 |
| | Polyp | 5 | 71 |
| | Reinke's edema | 12 | 329 |
| | Hyperkeratosis | 4 | 82 |
| | Squamous Hyperplasia | 3 | 75 |
| | Papillomatosis | 11 | 286 |
| | Amyloidosis | 2 | 32 |
| | Nodule | 1 | 26 |
| | Granuloma | 1 | 28 |
| | Fibroma | 1 | 2 |
| (Pre)Malignant | Mild Dysplasia | 3 | 77 |
| | Moderate Dysplasia | 2 | 49 |
| | Severe Dysplasia | 3 | 68 |
| | Carcinoma In Situ | 9 | 249 |
| | SCC | 8 | 168 |
| Total | | 68 | 1632 |

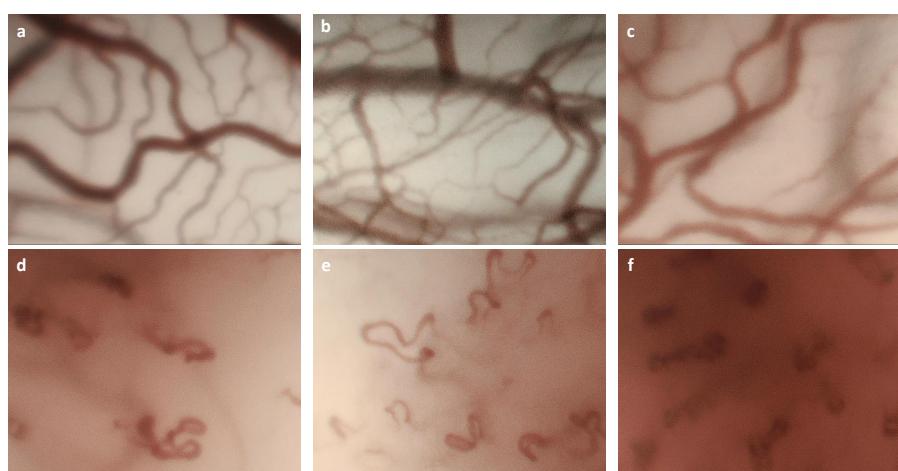


Figure 1. Examples of Longitudinal Vascular Changes (LVC) and Perpendicular Vascular Changes (PVC) in Contact Endoscopy (CE)-Narrow Band Imaging (NBI) images with different histopathologies: (a) Reinke's edema, LVC; (b) polyp, LVC; (c) amyloidosis, LVC; (d) severe dysplasia, PVC, (e) carcinoma in situ, PVC; (f) Squamous Cell Carcinomas (SCC), PVC.

2.3. Manual Approach

Three specialist and three resident otolaryngologists evaluated the images and classified the patients into benign and malignant groups. The residents had less than two years of experience in operating with CE-NBI images and the specialists worked for more than five years with such images. The otolaryngologists were blinded to the histologic diagnosis. They used the ELS guideline to independently visually evaluate the CE-NBI images of *Subset I* based on PVC appearance in the CE-NBI images, as explained in [12].

2.4. Automatic Approach

We used the algorithm presented in [17,18] to perform the automatic approach. The algorithm consists of a pre-processing step involving vessel enhancement and segmentation [20]. A feature extraction step was then applied to extract 24 geometrical features based on the consistency of gradient direction and the curvature level. Supervised classification step was conducted using the features and four classifiers to classify CE-NBI images into benign and malignant groups.

In this study, we made two main changes to the algorithm proposed in [17,18]. First, the Jerman filter [21] was used as pre-processing for the vessel enhancement step instead of the Frangi filter to overcome the problems related to the established enhancement function, not well adapted to natural variations of the vascular morphology. Second, the values of the tuning parameters of four classifiers including Support Vector Machine (SVM) with Polykernel and Radial Basis Function (RBF) [22], k-Nearest Neighbor (kNN) [23] and Random Forest Classifier (RFC) [24] were updated to have the optimum classification results with the current dataset.

In order to cover all the possible vascular structures, the vesselness parameter σ of the Jerman filter was set in the range of 0.5 mm to 2.5 mm with a step size of 0.5 mm. The parameter τ controlling the response uniformity was empirically set as 1.

The hyperparameter tuning process of all classifiers was updated using a grid search combined with 10-fold cross validation.

The performance of SVM is mainly affected by the regulation parameter (C) and kernel parameter (γ). The regulation parameter together with Polykernel and RBF controls the trade-off between achieving a low error in training data. γ determines how quickly class boundaries dissipate when they get far from the support vectors in SVM with RBF. The range of C and γ values were set within the range of 0.001 to 1000 with a ten-fold increment. The SVM with RBF completed the high overall performance with $C = 1$ and $\gamma = 0.01$ and SVM with Polykernel indicated the best results with $C = 1$.

Euclidean Distance was applied to calculate the distance of a sample in the case of kNN. To select the optimum k , a range from 1 to 20 with the step size equal to one were used. kNN confirmed the best performance at $k = 5$.

The optimization for RFC was done by adjusting the depth of trees and the number of estimators. The range of depth of the trees was set from 1 to 20 with step size equal to one. For the number of estimators, values from 10 to 100 with an increase of five was defined. The classifier gave the best performance at a depth of 8 with 55 trees.

In all classification scenarios, Subset I and Subset II were used as the testing and training sets, respectively. CE-NBI images were labeled as 0 for benign and as 1 for malignant groups. Each classifier was trained using the images' labels and feature vectors that were computed from the CE-NBI images of the training set. For the testing, the features vectors computed from the CE-NBI images of the testing set were fed into the predictive model of each classifier and then the expected labels were collected.

3. Evaluation Strategy

3.1. Classification Performances of Manual and Automatic Approaches

The global performances of the manual and automatic classification were evaluated using two classification measurements: sensitivity and specificity.

In the manual classification, the otolaryngologists assessed the set of CE-NBI images in the *Subset I* and classified each patient's image set as benign or malignant. Following [12], the PVC-positive patients with the malignant histological diagnosis were considered as true positive cases. With this assumption, a confusion matrix was created and the average value of sensitivity and specificity of all otolaryngologists, specialists and residents, was calculated using the following parameters:

- True Positive: PVC-positive patients with malignant lesions.
- True Negative: PVC-negative patients with benign lesions.
- False Negative: PVC-negative patients with malignant lesions.
- False Positive: PVC-positive patients with benign lesions.

In the automatic classification, the classifiers classified each CE-NBI image of *Subset I* as benign or malignant. A confusion matrix was calculated for each classifier using the predicted and actual labels of the images. Then, sensitivity and specificity were calculated using the following parameters:

- True Positive: actual image label is malignant, predicted image label is malignant.
- True Negative: actual image label is benign, predicted image label is benign.
- False Negative: actual image label is malignant, predicted image label is benign.
- False Positive: actual image label is benign, predicted image label is malignant.

Based on the descriptions above, the sensitivity and specificity values can show the performances of classifiers/otolaryngologists to correctly classify malignant and benign images/patients.

3.2. Comparison Procedure Between Manual and Automatic Classification

In a routine clinical procedure, the otolaryngologist evaluates a set of CE-NBI images of a patient and then identifies a patient's lesion as benign or malignant. For the manual classification in this work, the clinicians performed a similar routine, making a decision based on a set of images belonging to a patient. Since the automatic classification does not classify a patient but an image, in order to compare automatic to manual classification we made the following assumption: if a given classifier correctly classifies more than half of the images of a patient, then the patient is considered as a correct classification performed by this classifier. Following the assumption, two procedures for comparing between manual and automatic classification were proposed.

The first comparison procedure consists of comparing both approaches based on the level of agreement/disagreement between clinicians for classifying a patient as benign or malignant. In this aim, patients were divided into three categories:

- Category I includes 29 patients. All otolaryngologists correctly classified these patients.
- Category II includes 26 patients. One to five otolaryngologists correctly classified these patients.
- Category III includes 13 patients. All otolaryngologists misclassified these patients.

The second comparison procedure aims to compare manual and automatic classifications in terms of their misclassification levels depending on the histopathologies. This evaluation was performed to analyze the histopathologies in benign and malignant groups that caused significant difficulties for otolaryngologists and then to see how the automatic approach behaves with these cases.

We divided the patients into the 15 groups presented in Table 1. For each histopathology, a misclassification percentage was computed per patient for the automatic and manual classification as follows:

- Misclassification percentage of all otolaryngologists per patient in each histopathology group:

$$\left(\frac{\text{Number of doctor(s) who misclassified the patients}}{\text{Total number of doctors} \times \text{Total number of patients}} \right) \times 100 \quad (1)$$

where the total number of patients was the number of patients for the corresponding histopathology.

- Misclassification percentage of every classifier per patient in each histopathology group:

$$\left(\frac{\text{Number of misclassified patient(s)}}{\text{Total number of patients}} \right) \times 100 \quad (2)$$

- Misclassification percentage of all classifiers per patient in each histopathology group:

$$\left(\frac{\text{Number of misclassified patient(s) by all classifiers}}{\text{Total number of classifiers} \times \text{Total number of patients}} \right) \times 100 \quad (3)$$

4. Results and Discussion

Table 2 shows the global performances of the manual and automatic classification. In the manual approach, otolaryngology specialists showed a better performance than the otolaryngology residents. These results prove that the interpretation of CE-NBI images based on vascular patterns is subjective and highly depends on otolaryngologists' experience.

For the automatic approach, RFC with a sensitivity of 0.846 and SVM with RBF kernel with a specificity of 0.981 showed better results in comparison to the other classifiers.

The overall specificity values of otolaryngologists are low. This means that both groups had difficulties in distinguishing patients with benign histopathologies from malignant ones visually. This fact can be explained by the similarity between vascular patterns of benign and malignant histopathologies that can not be distinguished easily. For instance, Papillomatosis is a benign histopathology with similar vascular patterns than malignant histopathologies. This similarity leads to visually misclassify Papillomatosis as malignant. However, all four classifiers showed higher specificity than otolaryngologists proving the ability of automatic approach to overcome such a problem.

Table 2. General performance of manual and automatic approaches

| Classification Measurements | | Sensitivity | Specificity |
|---|----------------------------|-------------|-------------|
| Manual Classification (per patient) | Otolaryngology specialists | 0.955 | 0.727 |
| | Otolaryngology residents | 0.630 | 0.609 |
| | All otolaryngologists | 0.818 | 0.630 |
| Automatic Classification (per image) | SVM with polykernel | 0.830 | 0.882 |
| | SVM with RBF | 0.806 | 0.981 |
| | kNN | 0.814 | 0.863 |
| | RFC | 0.846 | 0.895 |

Figure 2 shows the detailed results of the first comparison procedure consisting of comparing both approaches based on the level of agreement/disagreement between clinicians for classifying a patient as benign or malignant. A first visual inspection shows that the classifiers individually misclassified 1 to 2 images in some patients at the Category I, where all otolaryngologists correctly classified these patients. Nevertheless, based on the assumption made in Section 3.2, the automatic approach did not misclassify any patient of this category.

For the patients belonging to Category II, both manual and automatic increased their misclassification levels compared to Category I. In the automatic approach, it is possible to observe that several images belonging to a patient can be misclassified. However, if we consider the automatic

classification per patient, only for one patient, two classifiers (SVM with polykernel and RFC) perform a misclassification. On the other hand, otolaryngologists showed a significant misclassification in some cases. For example, in the case of patients p26, p34 and p 72, five clinicians misclassified the patients, while the classifiers classified the patients correctly. These patients were diagnosed as Papillomatosis and Hyperkeratosis cases and belong to benign histopathologies. Figure 3a–c, displays the PVC vascular patterns in the CE-NBI images of these patients. As pointed out in the introduction, the difference between PVC in benign and in malignant histopathologies is not visually evident for the otolaryngologist. This causes a significant difficulty for the clinicians to distinguish benign from malignant cases based on the vascular patterns. Based on the results, the automatic approach showed the ability to identify this difference and then classify the patients correctly because it is capable of quantifying and differentiating these tiny differences. SVM with RBF did not show any misclassification per patient in this category.

For the Category III, where all otolaryngologists misclassified the patients, SVM with RBF misclassified fewer images compared to the other three classifiers. Concerning the classification per patient performed by the classifiers, it is possible to see that misclassifications were made for only two patients. Particularly, for patient p10 three classifiers failed in their classification. According to the histopathology, it corresponds to a patient presenting Hyperkeratosis. A set of CE-NBI images of this case is presented in Figure 3d. The type of vascular patterns of Hyperkeratosis can notably vary from one patient to another one. The CE-NBI dataset included 4 patients for this histopathology, presenting LVC and PVC vascular patterns. Due to this variation, the classifier's learning process using the proposed features [17,18] can be complicated. SVM with RBF showed no misclassification per patient in this Category.

These results show that the complexity of a manual analysis of a laryngeal lesion can be related to the type of histopathology and therefore we decided to perform a separated analysis based on the histopathology of the lesion. Table 3 presents the results of this second comparison procedure.

For the benign histopathologies, otolaryngologists showed high misclassification percentage of 83%, 77%, 46%, 33% and 27% for Fibroma, Papillomatosis, Hyperkeratosis, Squamous Hyperplasia and Polyp, respectively. Except for Fibroma, the misclassification level of each classifier is lower than the manual classification. Notably, in the case of Papillomatosis, the misclassification is significantly reduced in each classifier. If all classifiers are considered, the misclassification decreases from 77% to 7% in this histopathology. Papillomatosis causes classification difficulties to the otolaryngologists due to their vascular patterns that has similar characteristics to the malignant histopathologies. SVM with RBF and kNN seems to have the ability to solve this issue with 0% misclassification.

In the case of Fibroma, the misclassification percentage varied significantly among the four classifiers. This can be explained by the reduced number of images that the dataset contains for this type of histopathology (only one patient and two images).

In the malignant group, the otolaryngologists had the highest misclassification percentage of 61% for mild dysplasia. This histopathology can have PVC as well as LVC vascular patterns that usually appear in benign histopathologies. Hence, it is challenging for the otolaryngologists to classify patients with this condition as malignant visually. For this histopathology, the four classifiers performed well by classifying every patient correctly.

In general, SVM with RBF showed no patient misclassification for all histopathologies.

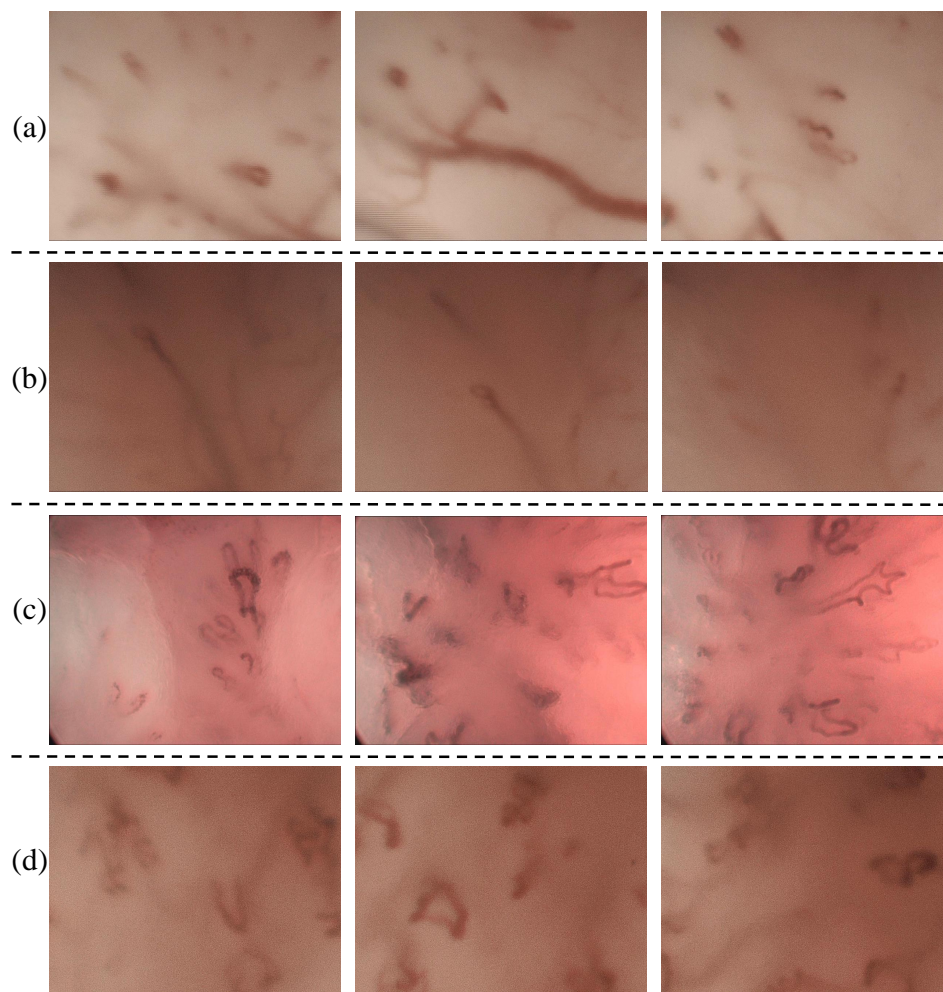


Figure 3. CE-NBI images of four patients from Category II and Category III: (a) p26, (b) p34, (c) p72 and (d) p10.

Table 3. Misclassification percentage of every histopathology category based on patient. C1 to C4 represent the four classifiers; C1: SVM with polykernel, C2: SVM with RBF, C3: kNN and C4: RFC.

| Type of Lesions | Histopathology | Man. and Auto. Classification (per Patient) | | | | | |
|-----------------|----------------------|---|-----|------|------|-----|-----------------|
| | | Doctors | C1 | C2 | C3 | C4 | All Classifiers |
| Benign | Cyst | 0% | 0% | 0% | 0% | 0% | 0% |
| | Polyp | 27% | 0% | 0% | 0% | 0% | 0% |
| | Reinke's edema | 7% | 0% | 0% | 0% | 0% | 0% |
| | Hyperkeratosis | 46% | 25% | 0% | 25% | 25% | 19% |
| | Squamous Hyperplasia | 33% | 0% | 0% | 0% | 0% | 0% |
| | Papillomatosis | 77% | 9% | 0% | 0% | 18% | 7% |
| | Nodule | 0% | 0% | 0% | 0% | 0% | 0% |
| | Granuloma | 0% | 0% | 0% | 0% | 0% | 0% |
| | Amyloidosis | 8% | 0% | 0% | 0% | 0% | 0% |
| Fibroma | 83% | 0% | 0% | 100% | 100% | 50% | |
| (Pre)Malignant | Mild Dysplasia | 61% | 0% | 0% | 0% | 0% | 0% |
| | Moderate Dysplasia | 17% | 0% | 0% | 0% | 0% | 0% |
| | Severe Dysplasia | 17% | 0% | 0% | 0% | 0% | 0% |
| | Carcinoma In Situ | 9% | 0% | 0% | 0% | 0% | 0% |
| | SCC | 25% | 0% | 0% | 13% | 0% | 3% |

5. Conclusions

Assessment of vascular patterns in CE-NBI images of vocal folds can provide valuable information for the clinicians to make the correct diagnostic decision before treatment. In this study, we showed how the evaluation of vascular patterns can be challenging for the otolaryngologists and how a computer-based approach can help clinicians ease this process.

In general, the otolaryngology specialists showed better classification performance than the residents in the manual approach. This proves that the interpretation of vascular patterns is subjective and depends on the clinicians' experience, as pointed out by several publications [4,10,12,15,16]. Both groups of otolaryngologists showed relatively low specificity on classifying a case as benign or malignant. This explains the difficulties in the visual classification of benign histopathologies. In the case of the benign group, otolaryngologists had the highest misclassification percentage for Papillomatosis and Hyperkeratosis. In the automatic approach, all four classifiers showed a higher specificity than both groups of otolaryngologists and showed significantly less misclassification percentage for Papillomatosis and Hyperkeratosis. The otolaryngology specialists showed significantly higher sensitivity than the residents. This means that specialists with more experience can easily detect PVC in CE-NBI images, while it is more challenging for the residents. In the malignant group, most of the misclassifications of otolaryngologists happened in the case of Mild Dysplasia and SCC. Although all classifiers showed lower sensitivity than otolaryngology specialists, they significantly reduced the misclassification percentage for Mild Dysplasia and SCC, compared to the otolaryngologists.

Two facts can explain the lower sensitivity and higher misclassification percentage that the classifiers show in the malignant group than the benign group. First, the CE-NBI dataset included more images in the benign group than in the malignant group (less training images were available for the malignant group). A significant part of CE-NBI images of the benign group belonged to the Papillomatosis with PVC patterns similar to those of malignant histopathologies. Second, the 24 features take only into account geometrical characteristics of the vascular patterns and no other characteristics that can also be important for the classification procedure. Due to these two points, it is possible that the algorithm shows some errors and classifies the CE-NBI images of malignant cases as benign. Hence, it is important to balance the number of CE-NBI images in the dataset for future works and develop new methods to improve the differentiation between wide and narrow angled points of PVCs.

The automatic approach showed its capacity to perform as an assistant system when there are disagreements among otolaryngologists or when they all misclassified the patients. SVM with RBF had the best performance and did not show any misclassification per patient in all the categories. This means that the combination of the proposed 24 features and SVM with RBF classifier, can provide valuable feedback for the clinicians to make decisions regarding the treatment planning. In general, the automatic approach has the potential to overcome the current issues in the field of enhanced CE and can operate as an assisting system to provide a more confident way for clinicians to learn as well as to make intraoperative decisions about the method and extent of surgical resection in patients with laryngeal cancer or benign vocal fold lesions in the routine surgical procedures.

Author Contributions: Conceptualization: N.E., A.I., A.B., N.D., C.A., N.N., M.F.; methodology: N.E., A.I.; software: N.E.; validation: N.E., A.I., N.D.; formal analysis: N.E., A.I.; investigation: N.E., N.D.; writing—original draft preparation: N.E.; writing—review and editing: A.I., A.B., N.D., C.A., N.N., M.F.; supervision, M.F., N.N.; project administration, N.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2020. *CA Cancer J. Clin.* **2020**, *70*, 7–30. [[CrossRef](#)]
2. Tamaki, A.; Miles, B.A.; Lango, M.; Kowalski, L.; Zender, C.A. AHNS Series: Do you know your guidelines? Review of current knowledge on laryngeal cancer. *Head Neck* **2018**, *40*, 170–181. [[CrossRef](#)] [[PubMed](#)]
3. Ni, X.; He, S.; Xu, Z.; Gao, L.; Lu, N.; Yuan, Z.; Lai, S.; Zhang, Y.; Yi, J.; Wang, X.; et al. Endoscopic diagnosis of laryngeal cancer and precancerous lesions by narrow band imaging. *J. Laryngol. Otol.* **2011**, *125*, 288–296. [[CrossRef](#)] [[PubMed](#)]
4. Puxeddu, R.; Sionis, S.; Gerosa, C.; Carta, F. Enhanced contact endoscopy for the detection of neoangiogenesis in tumors of the larynx and hypopharynx. *Laryngoscope* **2015**, *125*, 1600–1606. [[CrossRef](#)] [[PubMed](#)]
5. Arens, C.; Piazza, C.; Andrea, M.; Dikkers, F.G.; Gi, R.E.T.P.; Voigt-Zimmermann, S.; Peretti, G. Proposal for a descriptive guideline of vascular changes in lesions of the vocal folds by the committee on endoscopic laryngeal imaging of the European Laryngological Society. *Eur. Arch. Otorhinolaryngol.* **2016**, *273*, 1207–1214. [[CrossRef](#)]
6. Mehlum, C.S.; Døssing, H.; Davaris, N.; Giers, A.; Grøntved, Å.M.; Kjaergaard, T.; Möller, S.; Godballe, C.; Arens, C. Interrater variation of vascular classifications used in enhanced laryngeal contact endoscopy. *Eur. Arch. Otorhinolaryngol.* **2020**, 1–8.
7. Sun, C.; Han, X.; Li, X.; Zhang, Y.; Du, X. Diagnostic performance of narrow band imaging for laryngeal cancer: A systematic review and meta-analysis. *Otolaryngol. Head Neck Surg.* **2017**, *156*, 589–597. [[CrossRef](#)]
8. Andrea, M.; Dias, O.; Santos, A. Contact endoscopy during microlaryngeal surgery: A new technique for endoscopic examination of the larynx. *Ann. Otol. Rhinol. Laryngol.* **1995**, *104*, 333–339. [[CrossRef](#)]
9. Arens, C.; Dreyer, T.; Glanz, H.; Malzahn, K. Compact endoscopy of the larynx. *Ann. Otol. Rhinol. Laryngol.* **2003**, *112*, 113–119. [[CrossRef](#)]
10. Piazza, C.; Cocco, D.; Del Bon, F.; Mangili, S.; Nicolai, P.; Peretti, G. Narrow band imaging and high definition television in the endoscopic evaluation of upper aero-digestive tract cancer. *Acta Otorhinolaryngol. Ital.* **2011**, *31*, 70.
11. Arens, C.; Voigt-Zimmermann, S. Contact endoscopy of the vocal folds in combination with narrow band imaging (compact endoscopy). *Laryngo-Rhino-Otologie* **2015**, *94*, 150. [[PubMed](#)]
12. Davaris, N.; Lux, A.; Esmaili, N.; Illanes, A.; Boese, A.; Friebe, M.; Arens, C. Evaluation of vascular patterns using contact endoscopy and narrow-band imaging (CE-NBI) for the diagnosis of vocal fold malignancy. *Cancers* **2020**, *12*, 248. [[CrossRef](#)] [[PubMed](#)]
13. Šifrer, R.; Rijken, J.A.; Leemans, C.R.; Eerenstein, S.E.; van Weert, S.; Hendrickx, J.J.; Bloemena, E.; Heuveling, D.A.; Rinkel, R.N. Evaluation of vascular features of vocal cords proposed by the European Laryngological Society. *Eur. Arch. Otorhinolaryngol.* **2018**, *275*, 147–151. [[CrossRef](#)]
14. Šifrer, R.; Šereg-Bahar, M.; Gale, N.; Hočevnar-Boltežar, I. The diagnostic value of perpendicular vascular patterns of vocal cords defined by narrow-band imaging. *Eur. Arch. Otorhinolaryngol.* **2020**, *277*, 1–9. [[CrossRef](#)]
15. Mannelli, G.; Cecconi, L.; Gallo, O. Laryngeal preneoplastic lesions and cancer: Challenging diagnosis. Qualitative literature review and meta-analysis. *Crit. Rev. Oncol.* **2016**, *106*, 64–90. [[CrossRef](#)] [[PubMed](#)]
16. Puxeddu, R.; Carta, F.; Ferrel, C.; Chuchueva, N.; Gerosa, C. *Enhanced Contact Endoscopy (ECE) in Head and Neck Surgery*; Endo-Press: Tuttlingen, Germany, 2018.
17. Esmaili, N.; Illanes, A.; Boese, A.; Davaris, N.; Arens, C.; Friebe, M. A Preliminary Study on Automatic Characterization and Classification of Vascular Patterns of Contact Endoscopy Images. 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. 2703–2706.
18. Esmaili, N.; Illanes, A.; Boese, A.; Davaris, N.; Arens, C.; Friebe, M. Novel automated vessel pattern characterization of larynx contact endoscopic video images. *Int. J. Comput. Assist. Radiol. Surg.* **2019**, *14*, 1751–1761. [[CrossRef](#)]
19. Gale, N.; Hille, J.; Jordan, R.C.; Nadal, A.; Williams, M.D. Regarding Laryngeal precursor lesions: Interrater and intrarater reliability of histopathological assessment. *Laryngoscope* **2019**, *129*, E91–E92. [[CrossRef](#)]
20. Boese, A.; Illanes, A.; Balakrishnan, S.; Davaris, N.; Arens, C.; Friebe, M. Vascular pattern detection and recognition in endoscopic imaging of the vocal folds. *Curr. Dir. Biomed. Eng.* **2018**, *4*, 75–78. [[CrossRef](#)]

21. Jerman, T.; Pernuš, F.; Likar, B.; Špiclin, Ž. Enhancement of vascular structures in 3D and 2D angiographic images. *IEEE Trans. Med. Imaging* **2016**, *35*, 2107–2118. [[CrossRef](#)]
22. Hsu, C.W.; Chang, C.C.; Lin, C.J. *A Practical Guide to Support Vector Classification*; National Taiwan University: Taipei, Taiwan, 2003.
23. Peterson, L.E. K-nearest neighbor. *Scholarpedia* **2009**, *4*, 1883. [[CrossRef](#)]
24. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

2.4 Contribution 3: Cyclist Effort Features: A Novel Technique for Image Texture Characterization Applied to Larynx Cancer Classification in Contact Endoscopy—Narrow Band Imaging

2.4.1 Summary

In this paper, we designed and implemented a set of handcrafted features – CyEfF – to evaluate the textural characteristics of CE-NBI images. We aimed to see the significance of CyEfF in representing the textural characteristics of these forms of endoscopic images and how these two features and their combination with GF can be correlated to the type of laryngeal lesion. The main implementation block of this approach included selecting proper image pre-processing techniques, defining the CyEfF formulation, adjusting manual parameters for computing CyEfF, performing statistical ranking tests, and conducting supervised image classification scenarios according to the type of laryngeal lesion.

The main idea of CyEfF is to represent the CE-NBI image as a hilly surface, where different intensity profiles can be identified between defined starting and ending points. Each of these profiles can be considered a Tour de France stage trajectory where a cyclist needs to perform a specific effort to arrive at the finish line. When many cyclists travel through different trajectories, an average effort of all cyclists can be obtained to represent important textural characteristics of the image. Energy and power as two CyEfF were extracted based on this concept. The image pre-processed step using a Median filter was followed by features extraction, where some parameters were defined and set manually. The performance of this feature set was first studied using two statistical ranking tests. Then, two CE-NBI image classification scenarios were conducted where the standalone CyEfF and their combination with GF and conventional textural features were used to train four supervised ML classifiers and later were tested on CE-NBI image classification based on laryngeal lesions. The performance of CyEfF in this study showed that this feature set could describe the textural characterization of CE-NBI images with only two features, and their combination with GF could be part of a CAD system.

2.4.2 Contribution

The author of this thesis performed the design and implementation of CyEfF formulation and computation. Moreover, the reimplementation of the image pre-processing techniques, performing experiments for defining manual parameters, evaluation of features' performance based on features ranking techniques, as well as supervised ML-based image classification scenarios have been conducted by the author of this thesis. Furthermore, co-authors assisted in the conceptualization of the methodology and helped in data collection and preparation, results' assessment as well as revising the paper.

2.4.3 Cyclist Effort Features: A Novel Technique for Image Texture Characterization Applied to Larynx Cancer Classification in Contact Endoscopy—Narrow Band Imaging

Nazila Esmaeili, Axel Boese, Nikolaos Davaris, Christoph Arens, Nassir Navab, Michael Friebe, and Alfredo Illanes.

Article

Cyclist Effort Features: A Novel Technique for Image Texture Characterization Applied to Larynx Cancer Classification in Contact Endoscopy—Narrow Band Imaging

Nazila Esmaeili ^{1,2,*}, Axel Boese ¹, Nikolaos Davaris ³, Christoph Arens ³, Nassir Navab ², Michael Friebe ^{1,4} and Alfredo Illanes ¹

¹ INKA—Innovation Laboratory for Image Guided Therapy, Otto-von-Guericke University Magdeburg, 39120 Magdeburg, Germany; axel.boese@med.ovgu.de (A.B.); michael.friebe@ovgu.de (M.F.); alfredo.illanes@med.ovgu.de (A.I.)

² Chair for Computer Aided Medical Procedures and Augmented Reality, Technical University Munich, 85748 Munich, Germany; navab@cs.tum.edu

³ Department of Otorhinolaryngology, Head and Neck Surgery, Magdeburg University Hospital, 39120 Magdeburg, Germany; nikolaos.davaris@med.ovgu.de (N.D.); christoph.aren@med.ovgu.de (C.A.)

⁴ IDTM GmbH, 45657 Recklinghausen, Germany

* Correspondence: nazila.esmaeili@med.ovgu.de



Citation: Esmaeili, N.; Boese, A.; Davaris, N.; Arens, C.; Navab, N.; Friebe, M.; Illanes, A. Cyclist Effort Features: A Novel Technique for Image Texture Characterization Applied to Larynx Cancer Classification in Contact Endoscopy—Narrow Band Imaging. *Diagnostics* **2021**, *11*, 432. <https://doi.org/10.3390/diagnostics11030432>

Academic Editor: Maciej Misiolek

Received: 29 January 2021

Accepted: 26 February 2021

Published: 3 March 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: Background: Feature extraction is an essential part of a Computer-Aided Diagnosis (CAD) system. It is usually preceded by a pre-processing step and followed by image classification. Usually, a large number of features is needed to end up with the desired classification results. In this work, we propose a novel approach for texture feature extraction. This method was tested on larynx Contact Endoscopy (CE)—Narrow Band Imaging (NBI) image classification to provide more objective information for otolaryngologists regarding the stage of the laryngeal cancer. Methods: The main idea of the proposed methods is to represent an image as a hilly surface, where different paths can be identified between a starting and an ending point. Each of these paths can be thought of as a Tour de France stage profile where a cyclist needs to perform a specific effort to arrive at the finish line. Several paths can be generated in an image where different cyclists produce an average cyclist effort representing important textural characteristics of the image. Energy and power as two Cyclist Effort Features (CyEff) were extracted using this concept. The performance of the proposed features was evaluated for the classification of 2701 CE-NBI images into benign and malignant lesions using four supervised classifiers and subsequently compared with the performance of 24 Geometrical Features (GF) and 13 Entropy Features (EF). Results: The CyEff features showed maximum classification accuracy of 0.882 and improved the GF classification accuracy by 3 to 12 percent. Moreover, CyEff features were ranked as the top 10 features along with some features from GF set in two feature ranking methods. Conclusion: The results prove that CyEff with only two features can describe the textural characterization of CE-NBI images and can be part of the CAD system in combination with GF for laryngeal cancer diagnosis.

Keywords: texture feature extraction; classification; contact endoscopy; narrow band imaging; larynx

1. Introduction

Medical images contain crucial information that is analyzed by clinicians to find abnormalities and diagnose diseases. The level of tortuosity of anatomical structures such as blood vessels is one type of information that can be useful for clinicians. Vascular networks in tumors are irregular in size, shape, and branching pattern, lack the normal hierarchy, and do not display the recognizable features of arterioles, capillaries, or venules [1]. For example, in ophthalmology, retinal vascular tortuosity can be a potential indicator of diseases such as hypertension, diabetes, or atherosclerosis [2]. The changes in the organization and structure of the larynx vocal fold's blood vessels are directly related to the development of

benign and subsequent malignant laryngeal lesions. The manual assessment of vascular structures can, however, result in significant inter-observer variability and with that in subjective diagnosis [3,4].

Nowadays, Computer-Aided Diagnosis (CAD) systems use different feature extraction methods in combination with classification algorithms to assist clinicians in solving such problems. Features extraction is the process of generating features such as color, shape, and texture to describe the content of an image [5]. The significance of these features for describing image characteristics are of great importance and essential for the good performance of the CAD. There are several deep learning-based and hand-crafted feature extraction methods for medical image analysis. The deep learning-based approaches include the automatic features extraction and classification that mostly result in a high performance, but the majority of these approaches are computationally expensive to train, need lots of data and are known as the black art [6,7]. In the biomedical field, texture features are often used for characterizing an image using several hand-crafted feature extraction methods [8–14]. Although these methods have shown good performances for computing features, they have some drawbacks. Usually, a large number of features is needed for the classification, resulting in computationally expensive solutions. Moreover, most of the proposed features in the literature have limited or no meaning for the clinicians [5,15].

In this work, we propose a novel approach for image texture characterization. The main principle of the proposed approach is to consider an image as an irregular relief surface where different paths can be traced between a starting and an ending point. Each path can be thought of as a Tour de France course profile, where a cyclist needs to perform a specific effort to arrive at the finish line. The effort performed by a large number of cyclists following different paths in the hilly relief image can be representative of the image texture. Using this concept, we have extracted two features that we dubbed the Cyclist Effort Features (CyEff).

The usability of the proposed approach was tested to classify larynx Contact Endoscopy (CE)—Narrow Band Imaging (NBI) images into benign and malignant classes. CE-NBI is an enhanced endoscopic imaging technique that allows a detailed examination of laryngeal mucosa and provides more precise information about the structure of the superficial capillary network and sub-mucosal vessels in comparison to other endoscopic techniques. The visual evaluation of endoscopic images such as CE-NBI, is a subjective process causing difficulty for clinicians to recognize malignant lesions [3,16,17]. Several computer-based diagnosis approaches were applied to laryngeal endoscopic images to overcome this issue and present complementary information about the state of the larynx for clinicians [18]. Recent studies included a Deep Convolutional Neural Network (DCNN) using laryngoscopic images for larynx cancer detection [19], a set of texture-based features and Deep Learning-based descriptors extracted from endoscopic NBI images for laryngeal Squamous Cell Carcinoma (SCC) detection [20], a set of texture-based and first-order statistical features [21] plus an ensemble of Convolution Neural Networks (CNN) with texture and frequency domain based features [22] for larynx cancerous tissue classification using endoscopic NBI images, a set of features combined with supervised Machine Learning techniques for vascular patterns' assessment in CE-NBI images and laryngeal cancer diagnosis [23–25].

With the primary goal of this work to show the significance of the CyEff for classification purposes, we have compared the proposed features with two other sets, including 24 Geometrical Features (GF) [24] and 13 Entropy Features (EF) [21] that have been proposed in the literature for the larynx endoscopic image classification. The results showed that the classification performance of the two proposed CyEff is similar to the performance of other feature sets that includes a greater number of features and indicated the significance of the CyEff set on improving the classification performance of GF.

2. Method

2.1. Cyclist Effort Features Formulation

Figure 1 depicts the main idea behind the proposed feature extraction method. We show in Figure 1a two CE-NBI images of vessels. A 2-Dimensional (2D) grayscale image can be viewed as a 3-Dimensional (3D) surface by representing the intensity values of each pixel (being located in the x-y plane) along the z-axis. With that, we can consider each image as a hilly relief surface where a path can be traced between a starting and an ending point (see Figure 1b). We can imagine each of these paths as a Tour de France bicycle race stage. When a cyclist starts racing within an image, one trajectory of cyclist creates a sort of Tour de France stage profile (see Figure 1c). The cyclist needs to make an effort to accomplish each stage. This effort can be assessed by the energy that the cyclist spends and the associated cyclist's power. We can see in Figure 1c how these two trajectories can involve profiles that require a different degree of effort of a cyclist.

When a large number of cyclists, randomly distributed over the whole image, are performing different trajectories, an average effort of all cyclists can be obtained by computing average energy and power. This average effort can be representative of the image texture. As in the Tour de France, a stage can be classified as flat, mountainy or hilly. Our main idea using this new concept (cyclist energy and power features) is to classify texture in images since the average effort of cyclists in an image can vary according to its characteristic patterns.

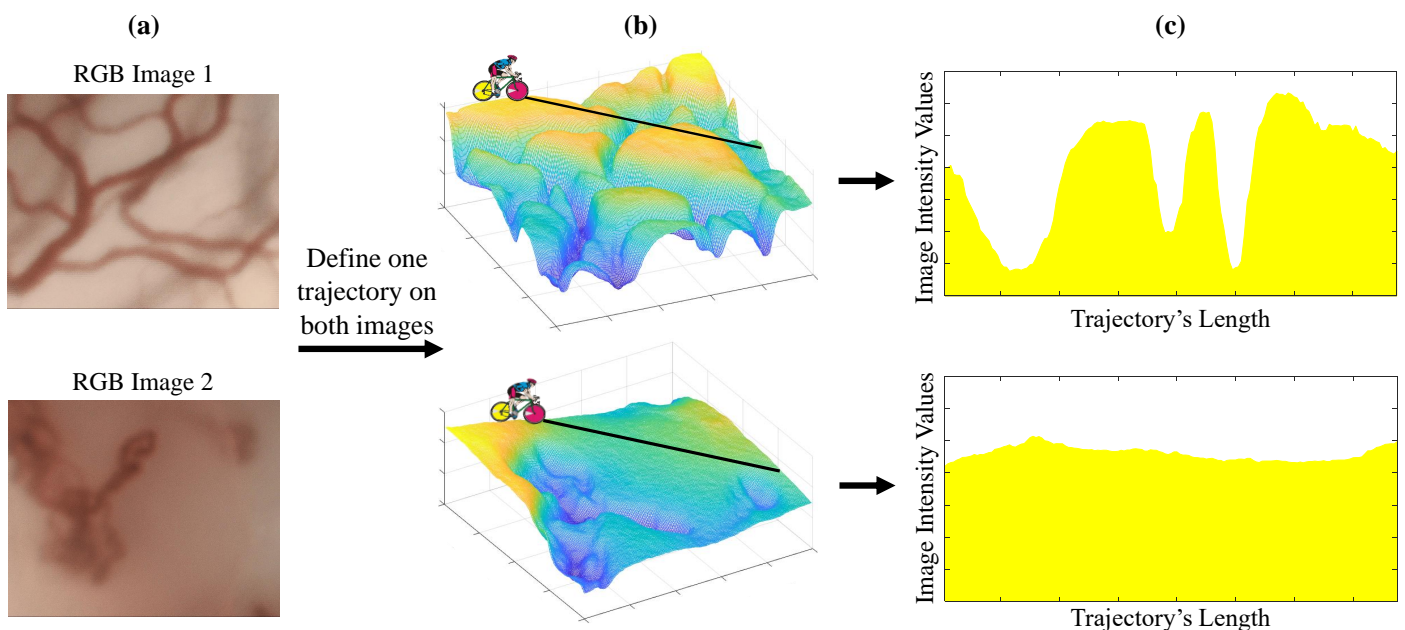


Figure 1. (a): RGB 2D image. (b): 3D representation of image. (c) Stage profile of similar trajectory on two images.

There are three primary forces that a cyclist must overcome in order to move forward [26]:

- Gravity Force (F_G): is one of the critical factors in cycling because a cyclist needs to fight against it cycling uphill. It can be calculated in metric units as $F_G = g \cdot \sin(\arctan(S)) \cdot m$, where S is the percentage grade to measure the steepness of a hill. g is the gravitational force constant and m is the combined weight of cyclist and bike.
- Rolling Resistance Force (F_R): is the friction between the tires and the road surface and is calculated in metric units as $F_R = g \cdot \cos(\arctan(S)) \cdot m \cdot C_r$, where C_r is a dimensionless parameter that captures the bumpiness of the road and the quality of tires.
- Air Resistance Force (F_A), which for the purposes of this work can be assumed to be a constant.

The total force resisting the cyclist is, therefore, $F_T = F_G + F_R + F_A$ and is the key parameter to calculate the cycling power and energy as:

$$P = F_T \cdot v \quad \text{and} \quad E = F_T \cdot v \cdot t \quad (1)$$

where v is the cycling velocity and t is the time duration of the cyclist's effort. These two parameters are used to compute the textural features proposed in this work.

2.2. Cyclist Effort Features Computation

The block diagram in Figure 2 shows the feature extraction process. Since the computation of cyclist power and energy requires the estimation of slopes in an image are known to be sensitive to high-frequency noise, the image is first pre-processed using a Median filter. For this filter, the kernel size was set empirically to 5×5 after visually evaluating the effect of three different kernel sizes on some randomly selected CE-NBI images. Then, different straight-line trajectories are generated inside the image between randomly selected starting and ending coordinate points. The trajectories need to include sufficient data from the image; hence each trajectory had at least 50-pixels length, equivalent to around 1% of the image's size. The pixel intensity values under each trajectory line are stored as vector arrays that correspond to race profiles.

Let $TP_k(i)$ be the pixel value of the trajectory profile vector k (with $k = 1, \dots, N_k$ and N_k corresponding to the total number of trajectories generated in an image) at the pixel index i ($i = 1, \dots, N_i$ with $N_i > 50$ being the length of the vector TP_k). For computing the cycling power and the cycling energy features of a full image, the power and energy of these individual TP_k trajectories should be first calculated. For that, each trajectory vector TP_k is first divided into N_s non-overlapped sections of length L . Then the power and energy of each one of these sections are computed using Equation (1). Figure 2 shows an example of the calculation process for the section $N_s = 15$. The section's slope percentage S_n and time interval t_n have to be estimated for each generated section n ($n = 1, \dots, N_s$). A trajectory can be seen as a curve in the 2-D plane, where the x-axis correspond to the pixel elements i and the y-axis correspond to the value of the vector $TP_k(i)$. Following this representation, let $A_n = (A_{nx}, A_{ny})$ and $B_n = (B_{nx}, B_{ny})$ being the starting and ending coordinate points, respectively, of the trajectory in section n . Then, the time interval can be computed as a simple ratio between a distance and the velocity as:

$$t_n = \frac{d(A_n, B_n)}{v} \quad (2)$$

where $d(A_n, B_n)$ corresponds to the Euclidean distance between A_n and B_n and v to the cyclist velocity. The section's slope percentage S_n can be calculated as the ratio between the y-axis jump between A_n and B_n and the length of the section n :

$$S_n = \frac{B_{ny} - A_{ny}}{L} \quad (3)$$

Using the estimated S_n and t_n it is possible to compute the power P_n and energy E_n of section n using the Equation (1). Then the power and energy of a trajectory TP_k are computed as:

$$P_k = \sum_{n=1}^{N_s} P_n \quad \text{and} \quad E_k = \sum_{n=1}^{N_s} E_n \quad (4)$$

P_P and E_E of the full image are computed as the average values of P_k and E_k , respectively.

The approach was implemented in MATLAB R2019a and executed on a PC with a CPU operating at 1.60 GHz resulting in an average execution time of 0.71 seconds per image.

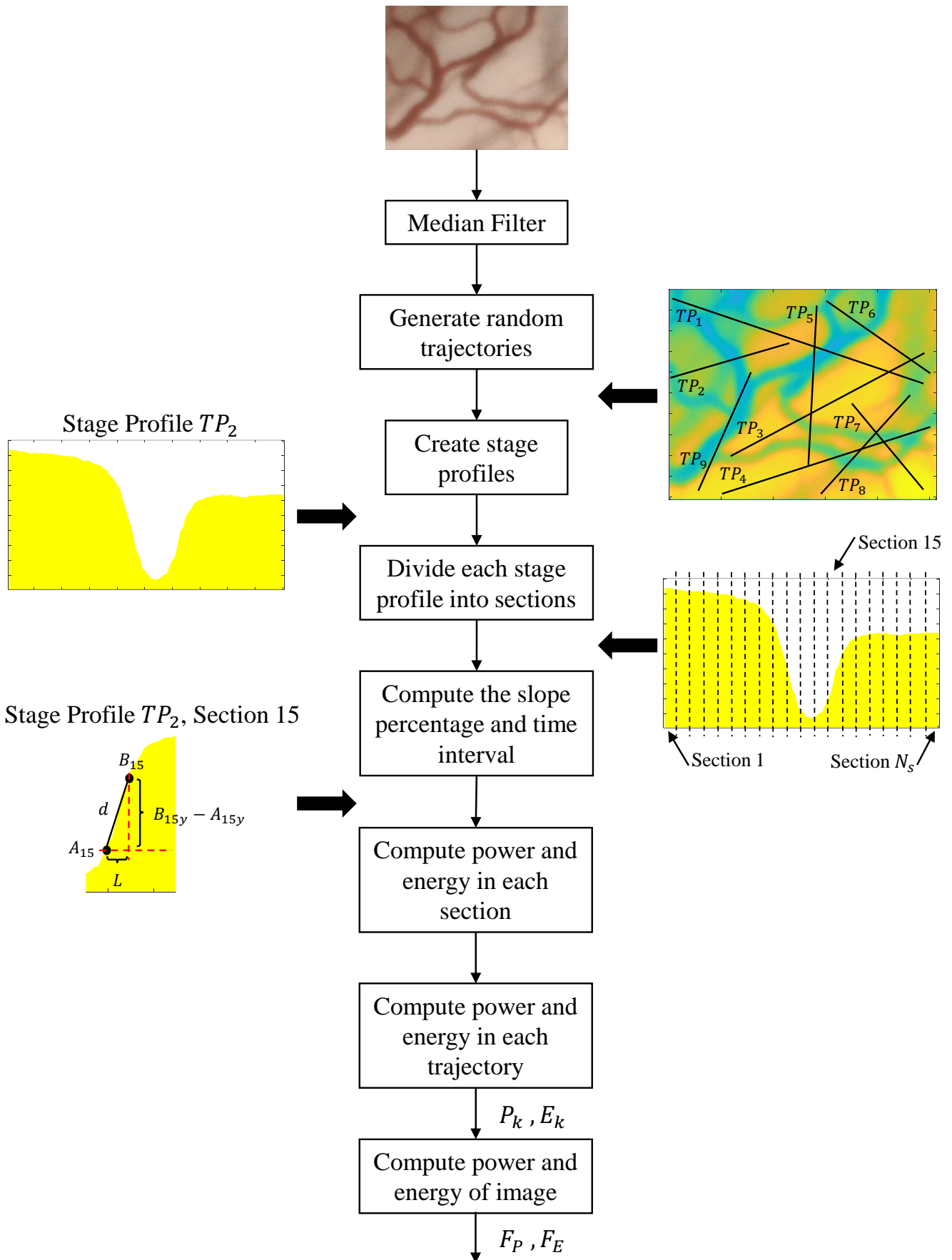


Figure 2. Feature extraction flowchart.

3. Experiments

3.1. Data Acquisition and Dataset Generation

The usability of the proposed method was evaluated in CE-NBI image classification. An updated version of the Dataset IV in [24] including 48 patients and 2701 CE-NBI images was used. Patients' data were anonymized and the biopsy results were used to label images into benign and malignant lesions according to the WHO classification [27]. The benign group involved images of patients with Cyst, Polyp, Reinke's edema, Papillomatosis and Mild Dysplasia. The malignant group included patients diagnosed with Severe Dysplasia, Carcinoma in situ and SCC.

For further parameter settings and feature evaluation procedure, 80% and 20% of CE-NBI images of the whole dataset were assigned to the training and testing sets, respectively. The images of patients were exclusively tied to separate sets in order to limit the chance of possible over-fitting. Training set was used for hyperparameter optimization as well as training process and testing set was used to evaluate the performance of the features.

3.2. Parameter Settings

Following the Equations (1) to (4) and their computation, six parameters needed to be defined. CE-NBI images in the training set were used to find the optimum number of trajectories N_k and the length of each section L . The values from 50 to 800 with step size of 50 were set to find the N_k . As the CyEff values of the selected images did not change significantly for $N_k > 500$, the number of trajectories N_k was set to 500. L was defined within the range of 1 to 10 pixels, with the step size was equal to one. The optimum CyEff values of the selected images were achieved at $L = 2$ pixels. To transform the pixel to the meter unit, we assumed that the longest path in the image is equal to the approximately longest path in Tour de France (200,000 m). The gravitational force g , the cyclist-bike weight m , and C_r related with tires and road characteristics are constant values ($g \approx 9.8 \text{ (m/s}^2\text{)}$, $m = 80 \text{ (Kg)}$ [26], $C_r = 0.005$ [26]). For this work, the cyclist velocity v can also be taken as a constant, and we have set this value to 11 (m/s), which corresponds to the average velocity in the Tour de France.

3.3. Feature Evaluation Procedure

The performance of the proposed features was compared with two other feature sets presented in the literature for classifying larynx endoscopic images: Geometrical Features (GF) and Entropy Features (EF).

- The GF set describes the level of disorder of vascular patterns in CE-NBI images [23,24]. This set of features intended to take into account geometrical characteristics of vessels including the consistency of gradient direction and the vessels' curvature and showed high performances on CE-NBI classification in different datasets [24,25].
- The EF set was used in combination with other types of features for classifying laryngeal tissue in NBI images. We converted each image into a grey-scale level and then divided it into seven different patch sizes of 50×50 , 100×100 , 150×150 , 200×200 , 250×250 , 300×300 pixels and the whole image. In each patch, the entropy was computed following [21] and stored in a matrix. The mean and variance were computed as features for each image.

GF includes 24 features ($F1$ to $F24$), EF 13 Features ($F25$ to $F37$) and CyEff two features ($F38$ and $F39$). In order to reduce the very-low frequency trends in the image that can affect the features computation, a homogenization filter was first applied to the image before the features' computation [24,28].

Two classification scenarios were conducted to evaluate the performance of the feature sets for the classification of CE-NBI images into benign and malignant classes. For that, four supervised classifiers including Support Vector Machine (SVM) with Polykernel and Radial Basis Function (RBF) [29], k-Nearest Neighbours (kNN) [30], and Random Forests (RF) [31] were used. First, each feature set was individually exposed to the classifiers to compare

their ability in classifying CE+NBI images. Second, the combinations of feature sets were created by adding EF and CyEfF to the GF. This scenario was performed to see how the proposed features (CyEfF) and the already used features for texture characterization in endoscopy images (EF) can improve the classification performance of the GF.

A 10-fold Cross-Validation with grid search method was used on training data and all feature sets for hyperparameter optimization. Then, the optimized parameters were applied to create the predictive model of classifiers for every feature sets. The features calculated from the CE-NBI images in the testing set with 10-fold Cross-Validation was used to evaluate these predictive models. A confusion matrix was computed in each testing scenario and the accuracy, sensitivity and specificity were obtained from it.

The optimization was conducted to find the value of the regulation parameter (C) and kernel parameter (γ) for the SVM classifier. The values within the range of 0.001 to 1000 with a ten-fold increment were assigned for both parameters. The SVM with Polykernel demonstrated the highest performance with $C = 1$ and the SVM with RBF indicated the best results with $C = 1$ and $\gamma = 0.01$.

For optimizing the kNN performance, the Euclidean distance was used as distance metric. Also, values within the range of 1 to 1000 with step size equal to one were used to select the optimum k . The optimum performance of the kNN classifier was obtained at $k = 10$.

The values of depth of trees and the number of estimators were adjusted to reach the optimized performance of RF. The number of estimators were defined within the range of 1 to 1000, with an increase of five. For the depth of the trees, values from 1 to 50 with step size equal to one were set. The classifier showed the highest overall performance at a depth of 7 with 60 trees.

Two feature ranking methods, including t -test [32] and Wilcoxon signed-rank test [33], were used to find the top-ranked features that have more influence on the classification results. The t -test investigates how significant the differences between groups are. It provides p -values as well. A p -value is the probability that the results from sample data occurred by chance. In most cases, p -value of 0.05 is accepted to mean the data is valid. Wilcoxon signed-rank test can be used to identify if samples from two independent yet related distributions are significantly different.

4. Results and Discussion

Figure 3 shows a qualitative example of one trajectory on four different CE-NBI images associated to benign and malignant lesions. Based on the E_k and P_k values in Figure 3c, the energy and power of the trajectory is significantly different between benign and malignant images. Furthermore, the F_E and F_P values as the two CyEfF show the variation between two groups of images.

Table 1 shows the classification results of the first scenario described in the previous section. The SVM classification results with Polykernel and RBF had the highest accuracy of 0.882 and 0.875 using CyEfF. With the kNN and RF, GF showed the highest performance with the accuracy of 0.885 and 0.920, respectively. According to the performed result, CyEfF, with only two features, achieved comparable results than GF and EF, which used 24 and 13 features, respectively.

In studies [24,25], a subset of current CE-NBI image dataset were used for classification of CE-NBI images into benign and malignant classes. In comparison to the results in [24], the CyEfF set with Polykernel and RBF SVM showed a better classification accuracy. Furthermore, CyEfF with Polykernel SVM and kNN showed higher sensitivity and specificity than the results presented in [25].

Table 2 presents the results of the second classification scenario in which the combination of GF and CyEfF showed better performance than the combination of GF and EF. The classification accuracy of four classifiers increased from 3 to 12 percent by adding the two proposed features to the 24 GF, in which the highest accuracy of 0.966 was achieved with the kNN classifier. The combination of GF and CyEfF with four classifiers showed

higher accuracy, sensitivity and specificity in comparison to the results in [24,25] to classify CE-NBI images into benign and malignant classes. Moreover, in comparison to the other texture-based feature extraction methods such as local binary patterns (LBP) and gray-level co-occurrence matrix (GLCM) that were applied to the laryngeal tissue classification in NBI laryngoscopy [21], the combination of CyEff and GE feature sets with Polykernel SVM showed the higher performance. These results prove the significant effect of the CyEff on improving the classification of CE-NBI images with the already used GF set.

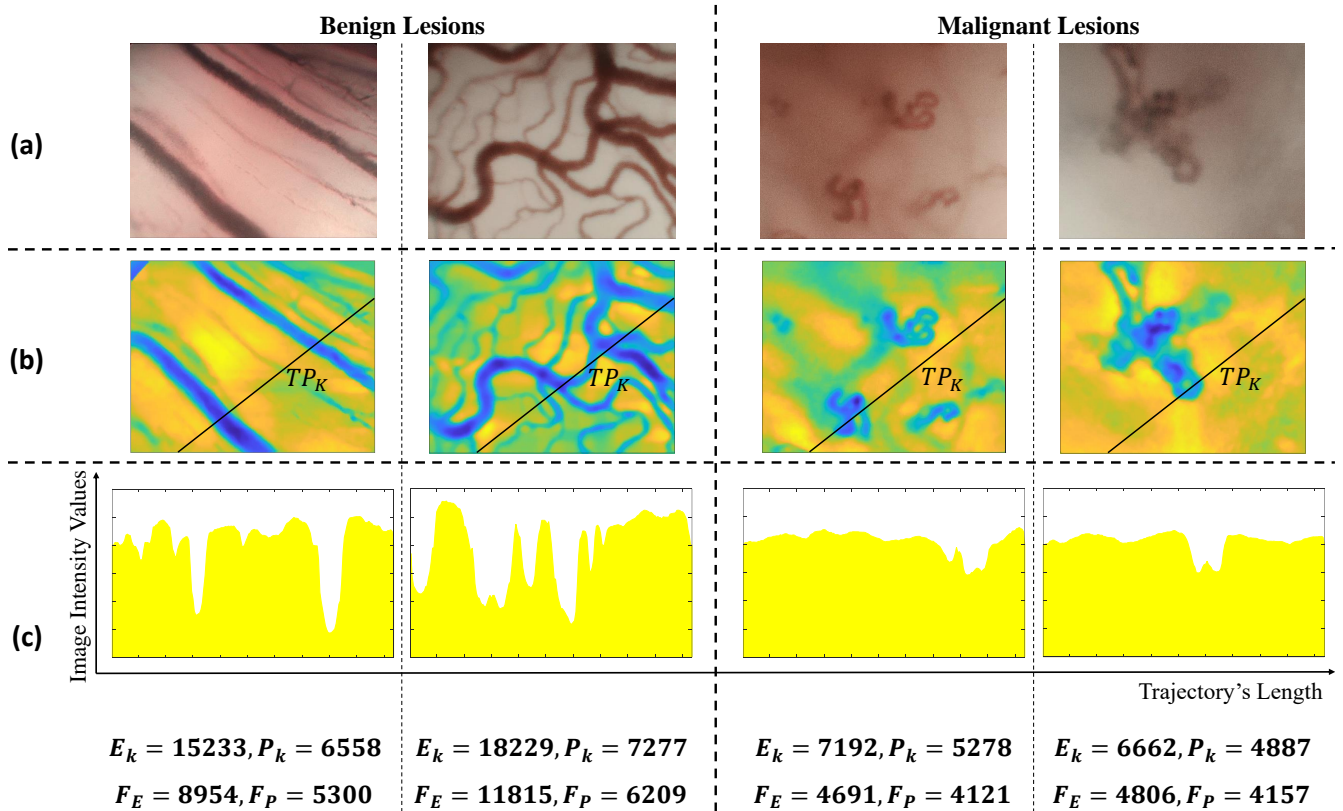


Figure 3. (a): Original CE-NBI image, (b): Pre-processed image with one random trajectory, (c): The stage profile of the random trajectory plus the cyclist’s energy and power values of the random trajectory and the 500 trajectories (whole image).

Table 1. Classification results of four classifiers using three features sets.

| Classifiers | Accuracy | | | Sensitivity | | | Specificity | | |
|---------------------|----------|-------|-------|-------------|-------|-------|-------------|-------|-------|
| | GF | EF | CyEff | GF | EF | CyEff | GF | EF | CyEff |
| SVM with Polykernel | 0.820 | 0.739 | 0.882 | 0.818 | 0.792 | 0.845 | 0.822 | 0.596 | 0.924 |
| SVM with RBF | 0.806 | 0.761 | 0.875 | 0.817 | 0.802 | 0.826 | 0.821 | 0.515 | 0.920 |
| kNN | 0.885 | 0.781 | 0.874 | 0.911 | 0.812 | 0.834 | 0.836 | 0.531 | 0.911 |
| RF | 0.920 | 0.788 | 0.859 | 0.935 | 0.801 | 0.831 | 0.892 | 0.538 | 0.886 |

Table 2. Classification results of four classifiers using combination of feature sets.

| Classifier | Accuracy | | Sensitivity | | Specificity | |
|---------------------|----------|----------|-------------|----------|-------------|----------|
| | GF+EF | GF+CyEff | GF+EF | GF+CyEff | GF+EF | GF+CyEff |
| SVM with Polykernel | 0.782 | 0.944 | 0.816 | 0.942 | 0.738 | 0.947 |
| SVM with RBF | 0.773 | 0.897 | 0.813 | 0.981 | 0.702 | 0.818 |
| kNN | 0.795 | 0.966 | 0.837 | 0.959 | 0.718 | 0.973 |
| RF | 0.808 | 0.956 | 0.831 | 0.952 | 0.724 | 0.961 |

In order to confirm the significance of the proposed CyEff, Table 3 shows 10 top-ranked features for each ranking method. Energy (F38) and power (F39) features are ranked as the top 10 features along with some features from GF set in both ranking methods. Figure 4a shows the box plot of energy and power features with the p -values equal to 4.5654×10^{-35} and 1.4419×10^{-32} , computed from the t -test, respectively. Based on these values, the proposed features showed a statistically significant difference between benign and malignant classes. Also, Figure 4a shows, that the range of energy and power features for benign and malignant classes are distinguishable. Figure 1b presents that the combination of energy and power features has a separation among benign and malignant classes. These results prove the influence and significance of the proposed CyEff on the classification results.

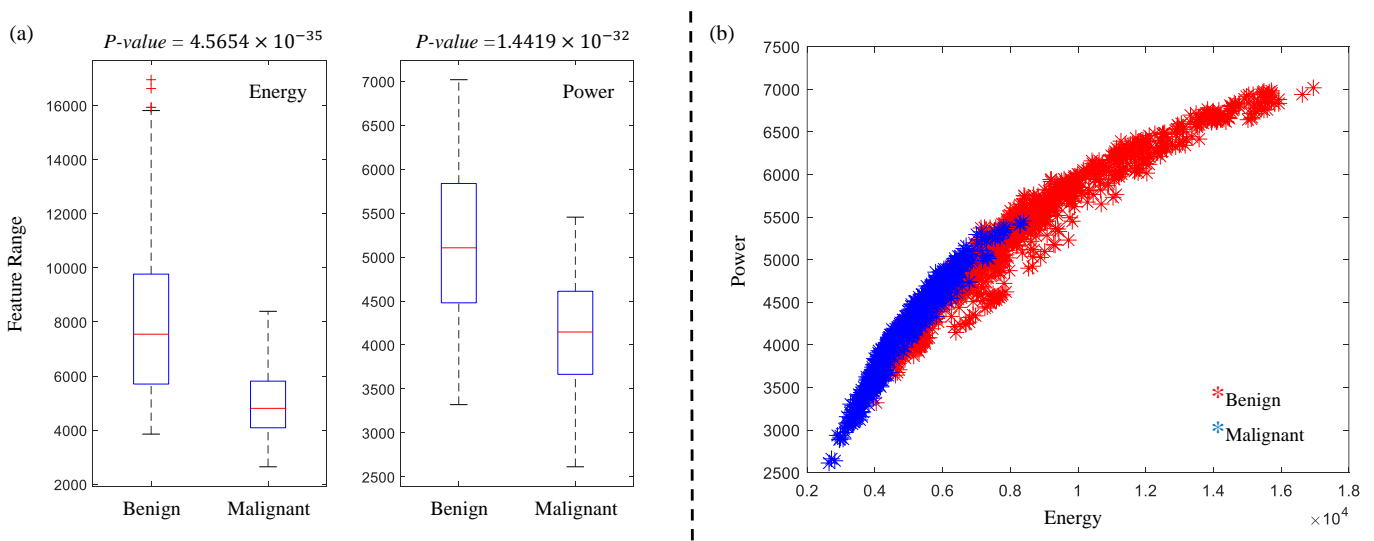


Figure 4. (a): Box plot of energy and power features. (b): Projected data points of benign and malignant classes using CyEff.

Table 3. Feature ranking results: F01-F24: GF, F25-F37: EF and F38, F39: CyEff.

| Method | Ranking | | | | | | | | | |
|----------------------|---------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | #1 | #2 | #3 | #4 | #5 | #6 | #7 | #8 | #9 | #10 |
| t -test | F38 | F21 | F39 | F14 | F24 | F22 | F09 | F20 | F17 | F15 |
| Wilcoxon signed-rank | F14 | F38 | F39 | F21 | F24 | F08 | F09 | F22 | F15 | F07 |

The very-low frequency trending characteristics of the image background are usually highly problematic for extracting features since significant pixel values involve progressive changes in the image plane that may affect the extraction of important texture information. In order to study how this type of noise can affect the proposed features, we have computed the method performances by removing the homogenization pre-processing stage. Results show that the classification performance does not significantly vary with or without the pre-processing stage. The accuracy of 0.868, 0.867, 0.850 and 0.872 using SVM with Polykernel and RBF, kNN and RFC were achieved without the pre-processing, respectively. In comparison to the results in Table 1, the accuracy varied only 1 to 2 percent.

5. Conclusions

CyEff approach is an understandable and intuitive method that showed promising results with less amount of data for training in comparison to other deep learning-based feature extraction methods. According to the presented results, CyEff can describe the textural characterization of CE-NBI images with only two features, which is one of the main advantages of this approach over other hand-crafted feature extraction methods. Moreover,

removing the pre-processing stage related to attenuation of very low-frequency trending characteristics of the image did not significantly affect the classification performance of proposed features. However, further evaluation should be conducted for this matter in future work.

As the focus of this paper is on the CE-NBI images, we compared only the performance of the proposed features with other research works in the field of CE-NBI endoscopic imaging modality [23–25]. For this reason, comparative experiments to already existing texture-based feature extraction methods on this dataset would be suggested for further development.

Based on the recent advances and improvements in the field of CNN-based approaches, there is a high probability that the application of these methods can result in better performance in CE-NBI classification and can overcome the critical limitations of the hand-crafted feature extraction methods. However, it will take a great amount of time to collect and label the data to develop such a method in the medical field for real clinical use. According to our knowledge, there is no CNN-based method for the classification of CE-NBI images in the literature. Hence, the comparison between deep learning-based and hand-crafted feature extraction methods for CE-NBI classification is necessary for future developments.

In spite of the technological advancements, differentiation between malignant and benign lesions in the larynx is difficult in reality, irrespective of the clinicians' level of experience. In addition, the subjectivity in laryngeal cancer diagnosis has been reported several times, resulting in invasive surgical biopsy and subsequent histological examination. CyEff in combination with GF as part of a CAD system can potentially solve these problems in CE-NBI image classification and help the clinicians to make final decisions about the stage of laryngeal cancer in the routine and surgical procedures.

With the primary objective of this work to present the significance of the CyEff for CE-NBI image classification, testing the proposed set of features in other imaging modalities is something that should be accomplished for future work. Based on the presented results, the proposed approach can be used as a new texture-feature extraction method in medical image analysis. For example, it can be applied to the fundus images, as the level of tortuosity of vessels in these images is also crucial for clinicians.

Author Contributions: Conceptualization, N.E., A.B., N.D., C.A., N.N., M.F., A.I.; Formal analysis, N.E., A.I.; Investigation, N.E., N.D.; Methodology, N.E., A.I.; Project administration, N.E.; Software, N.E.; Supervision, N.N., M.F.; Validation, N.E., A.B., A.I.; Writing—original draft, N.E.; Writing—review & editing, A.B., N.D., C.A., N.N., M.F., A.I. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

Ethical Approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee (number 49/18).

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|-------|-------------------------|
| CE | Contact Endoscopy |
| NBI | Narrow Band Imaging |
| CyEff | Cyclist Effort Features |
| GF | Geometrical Features |
| EF | Entropy Features |

| | |
|------|-----------------------------------|
| CAD | Computer-Aided Diagnosis |
| DCNN | Deep Convolutional Neural Network |
| CNN | Convolution Neural Networks |
| SCC | Squamous Cell Carcinoma |
| SVM | Support Vector Machine |
| RBF | Radial Basis Function |
| kNN | k-Nearest Neighbours |
| RF | Random Forests |

References

- Ribatti, D.; Nico, B.; Crivellato, E.; Vacca, A. The structure of the vascular network of tumors. *Cancer Lett.* **2007**, *248*, 18–23. [[CrossRef](#)] [[PubMed](#)]
- Ramos, L.; Novo, J.; Rouco, J.; Romeo, S.; Álvarez, M.D.; Ortega, M. Retinal vascular tortuosity assessment: inter-intra expert analysis and correlation with computational measurements. *BMC Med. Res. Methodol.* **2018**, *18*, 1–11. [[CrossRef](#)] [[PubMed](#)]
- Davaris, N.; Lux, A.; Esmaeili, N.; Illanes, A.; Boese, A.; Friebe, M.; Arens, C. Evaluation of vascular patterns using contact endoscopy and narrow-band imaging (CE-NBI) for the diagnosis of vocal fold malignancy. *Cancers* **2020**, *12*, 248. [[CrossRef](#)]
- Mehlum, C.S.; Døssing, H.; Davaris, N.; Giers, A.; Grøntved, Å.M.; Kjaergaard, T.; Möller, S.; Godballe, C.; Arens, C. Interrater variation of vascular classifications used in enhanced laryngeal contact endoscopy. *Eur. Arch. Oto-Rhino-Laryngol.* **2020**, 1–8. [[CrossRef](#)]
- Goel, R.; Kumar, V.; Srivastava, S.; Sinha, A. A review of feature extraction techniques for image analysis. *Int. J. Adv. Res. Comput. Commun. Eng.* **2017**, *6*, 153–155.
- Cai, L.; Gao, J.; Zhao, D. A review of the application of deep learning in medical image classification and segmentation. *Ann. Transl. Med.* **2020**, *8*, 713. [[CrossRef](#)] [[PubMed](#)]
- Maier, A.; Syben, C.; Lasser, T.; Riess, C. A gentle introduction to deep learning in medical image processing. *Z. Für Med. Phys.* **2019**, *29*, 86–101. [[CrossRef](#)] [[PubMed](#)]
- Khan, S.A.; Yong, S.P.; Janjua, U.I. A Comparative Evaluation of Features for Medical Image Modality Classification. *J. Teknol.* **2016**, *78*. [[CrossRef](#)]
- Gao, Y.; Fu, R.; Kuang, Y.; Lv, Q. Classification and Retrieval of Abdominal Medical Image Based on Gray Level Concurrence Matrix. *Chin. Med. Equip. J.* **2012**, *3*. [[CrossRef](#)]
- Jafarpour, S.; Sedghi, Z.; Amirani, M.C. A robust brain MRI classification with GLCM features. *Int. J. Comput. Appl.* **2012**, *37*, 1–5.
- Garra, B.S.; Krasner, B.H.; Horii, S.C.; Ascher, S.; Mun, S.K.; Zeman, R.K. Improving the distinction between benign and malignant breast lesions: the value of sonographic texture analysis. *Ultrasound Imaging* **1993**, *15*, 267–285. [[CrossRef](#)] [[PubMed](#)]
- Ko, B.C.; Kim, S.H.; Nam, J.Y. X-ray image classification using random forests with local wavelet-based CS-local binary patterns. *J. Digit. Imaging* **2011**, *24*, 1141–1151. [[CrossRef](#)] [[PubMed](#)]
- Nanni, L.; Lumini, A.; Brahmam, S. Local binary patterns variants as texture descriptors for medical image analysis. *Artif. Intell. Med.* **2010**, *49*, 117–125. [[CrossRef](#)] [[PubMed](#)]
- Acharya, U.R.; Chowriappa, P.; Fujita, H.; Bhat, S.; Dua, S.; Koh, J.E.; Eugene, L.; Kongmebhol, P.; Ng, K.H. Thyroid lesion classification in 242 patient population using Gabor transform features from high resolution ultrasound images. *Knowl.-Based Syst.* **2016**, *107*, 235–245. [[CrossRef](#)]
- Humeau-Heurtier, A. Texture feature extraction methods: A survey. *IEEE Access* **2019**, *7*, 8975–9000. [[CrossRef](#)]
- Puxeddu, R.; Sionis, S.; Gerosa, C.; Carta, F. Enhanced contact endoscopy for the detection of neoangiogenesis in tumors of the larynx and hypopharynx. *Laryngoscope* **2015**, *125*, 1600–1606. [[CrossRef](#)]
- Mannelli, G.; Cecconi, L.; Gallo, O. Laryngeal preneoplastic lesions and cancer: challenging diagnosis. Qualitative literature review and meta-analysis. *Crit. Rev. Oncol.* **2016**, *106*, 64–90. [[CrossRef](#)]
- Turkmen, H.I.; Karşlıgil, M.E. Advanced computing solutions for analysis of laryngeal disorders. *Med. Biol. Eng. Comput.* **2019**, 1–18. [[CrossRef](#)]
- Xiong, H.; Lin, P.; Yu, J.G.; Ye, J.; Xiao, L.; Tao, Y.; Jiang, Z.; Lin, W.; Liu, M.; Xu, J. Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images. *EBioMedicine* **2019**, *48*, 92–99. [[CrossRef](#)]
- Araújo, T.; Santos, C.P.; De Momi, E.; Moccia, S. Learned and handcrafted features for early-stage laryngeal SCC diagnosis. *Med. Biol. Eng. Comput.* **2019**, *57*, 2683–2692. [[CrossRef](#)]
- Moccia, S.; De Momi, E.; Guarnaschelli, M.; Savazzi, M.; Laborai, A.; Guastini, L.; Peretti, G.; Mattos, L.S. Confident texture-based laryngeal tissue classification for early stage diagnosis support. *J. Med. Imaging* **2017**, *4*, 034502. [[CrossRef](#)]
- Nannia, L.; Ghidoni, S.; Brahmam, S. Ensemble of convolutional neural networks for bioimage classification. *Appl. Comput. Inform.* **2020**. [[CrossRef](#)]
- Esmaeili, N.; Illanes, A.; Boese, A.; Davaris, N.; Arens, C.; Friebe, M. A Preliminary Study on Automatic Characterization and Classification of Vascular Patterns of Contact Endoscopy Images. 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. 2703–2706.
- Esmaeili, N.; Illanes, A.; Boese, A.; Davaris, N.; Arens, C.; Friebe, M. Novel automated vessel pattern characterization of larynx contact endoscopic video images. *Int. J. Comput. Assist. Radiol. Surg.* **2019**, *14*, 1751–1761. [[CrossRef](#)] [[PubMed](#)]

25. Esmaeili, N.; Illanes, A.; Boese, A.; Davaris, N.; Arens, C.; Navab, N.; Friebe, M. Laryngeal Lesion Classification Based on Vascular Patterns in Contact Endoscopy and Narrow Band Imaging: Manual Versus Automatic Approach. *Sensors* **2020**, *20*, 4018. [[CrossRef](#)]
26. Hannas, B.L.; Goff, J.E. Model of the 2003 Tour de France. *Am. J. Phys.* **2004**, *72*, 575–579. [[CrossRef](#)]
27. Gale, N.; Hille, J.; Jordan, R.C.; Nadal, A.; Williams, M.D. Regarding Laryngeal precursor lesions: Interrater and intrarater reliability of histopathological assessment. *Laryngoscope* **2019**, *129*, E91–E92. [[CrossRef](#)]
28. Boese, A.; Illanes, A.; Balakrishnan, S.; Davaris, N.; Arens, C.; Friebe, M. Vascular pattern detection and recognition in endoscopic imaging of the vocal folds. *Curr. Dir. Biomed. Eng.* **2018**, *4*, 75–78. [[CrossRef](#)]
29. Ben-Hur, A.; Weston, J. A user's guide to support vector machines. In *Data Mining Techniques for the Life Sciences*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 223–239.
30. Piegel, L.A.; Tiller, W. Algorithm for finding all k nearest neighbors. *Comput.-Aided Des.* **2002**, *34*, 167–172. [[CrossRef](#)]
31. Liaw, A.; Wiener, M. Classification and regression by randomForest. *R News* **2002**, *2*, 18–22.
32. Kim, T.K. T test as a parametric statistic. *Korean J. Anesthesiol.* **2015**, *68*, 540. [[CrossRef](#)] [[PubMed](#)]
33. Natarajan, S.; Lipsitz, S.R.; Fitzmaurice, G.M.; Sinha, D.; Ibrahim, J.G.; Haas, J.; Gellad, W. An extension of the Wilcoxon rank sum test for complex sample survey data. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **2012**, *61*, 653–664. [[CrossRef](#)] [[PubMed](#)]

2.5 Contribution 4: Deep Convolution Neural Network for Laryngeal Cancer Classification on Contact Endoscopy-Narrow Band Imaging

2.5.1 Summary

In this paper, we developed a fully automatic CNN approach based on transfer learning combined with a cut-off layer technique for CE-NBI image classification. The study's main objective was to evaluate the performance of a DL-based method on laryngeal lesion assessment as well as the capability of such a strategy to deal with the complexity of vascularization networks in CE-NBI images.

The main implementation block of this approach included selecting a pre-trained DL architecture, defining the fine-tuning strategy, implementing the cut-off layer technique, performing data augmentation, and conducting training, validation, and testing experiments to arrive at the optimum model.

The pre-trained ResNet50 model on the ImageNet database was adopted as the backbone of the DL-based approach. The fine-tuning technique wherein all the layers were fine-tuned was applied to this selected architecture. Moreover, the cut-off-layer technique to discard part of the network was integrated into the last layer in the feature extraction part of the network, where the classifier part begins. Three experiments were conducted to determine the ResNet50 model for CE-NBI classification based on laryngeal lesions. The main difference between these experiments was related to the strategy of data separation. Apart from this, a few experiments also considered different network hyperparameters and changes in the number of data using data augmentation techniques. In the training phase, binary cross entropy was used as a loss function along with SGD as the optimizer and early stoppage was set with a patience of 5 epochs to avoid possible overfitting. The fine-tuned model with a size equal to 1% of the complete ResNet50 architecture trained on the data set, including data augmentation, showed faster training with less prone to result in overfitting. Therefore, this model showed effective performance as part of the CAD system on laryngeal lesion assessment using CE-NBI images.

2.5.2 Contribution

The author of this thesis initiated the main idea and conducted the pre-trained DL architecture selection and fine-tuning of hyperparameters. Moreover, the author of this thesis designed the training and testing experiments, implemented the cut-off layer technique, and evaluated the performance of the trained models in different experiments on CE-NBI image classification. Finally, co-authors contributed to the data collection and preparation, conducting the training and testing experiments, results assessment, and manuscript revision.

2.5.3 Deep Convolution Neural Network for Laryngeal Cancer Classification on Contact Endoscopy-Narrow Band Imaging

Nazila Esmaeili, Esam Sharaf, Elmer Jeto Gomes Ataide, Alfredo Illanes, Axel Boese, Nikolaos Davaris, Christoph Arens, Nassir Navab, and Michael Friebe

Communication

Deep Convolution Neural Network for Laryngeal Cancer Classification on Contact Endoscopy-Narrow Band Imaging

Nazila Esmaeili ^{1,2,*}, Esam Sharaf ¹, Elmer Jeto Gomes Ataide ^{1,3}, Alfredo Illanes ¹, Axel Boese ¹, Nikolaos Davaris ⁴, Christoph Arens ⁵, Nassir Navab ² and Michael Friebe ^{1,6}

- ¹ INKA—Innovation Laboratory for Image Guided Therapy, Otto-von-Guericke University Magdeburg, 39120 Magdeburg, Germany; esam.sharaf@ovgu.de (E.S.); elmer.gomesataide@ovgu.de (E.J.G.A.); alfredo.illanes@med.ovgu.de (A.I.); axel.boese@med.ovgu.de (A.B.); michael.friebe@ovgu.de (M.F.)
 - ² Chair for Computer Aided Medical Procedures and Augmented Reality, Technical University of Munich, 85748 Munich, Germany; nassir.navab@tum.de
 - ³ Department of Nuclear Medicine, Medical Faculty, Otto-von-Guericke University Magdeburg, 39120 Magdeburg, Germany
 - ⁴ Department of Otorhinolaryngology, Head and Neck Surgery, Magdeburg University Hospital, 39120 Magdeburg, Germany; nikolaos.davaris@med.ovgu.de
 - ⁵ Department of Otorhinolaryngology, Head and Neck Surgery, Giessen University Hospital, 35392 Giessen, Germany; christoph.arens@hno.med.uni-giessen.de
 - ⁶ IDTM GmbH, 45657 Recklinghausen, Germany
- * Correspondence: nazila.esmaeili@med.ovgu.de



Citation: Esmaeili, N.; Sharaf, E.; Gomes Ataide, E.J.; Illanes, A.; Boese, A.; Davaris, N.; Arens, C.; Navab, N.; Friebe, M. Deep Convolution Neural Network for Laryngeal Cancer Classification on Contact Endoscopy-Narrow Band Imaging. *Sensors* **2021**, *21*, 8157. <https://doi.org/10.3390/s21238157>

Academic Editors: Sang Hyun Park, Manhua Liu and Dong Hye Ye

Received: 31 October 2021
Accepted: 3 December 2021
Published: 6 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: (1) Background: Contact Endoscopy (CE) and Narrow Band Imaging (NBI) are optical imaging modalities that can provide enhanced and magnified visualization of the superficial vascular networks in the laryngeal mucosa. The similarity of vascular structures between benign and malignant lesions causes a challenge in the visual assessment of CE-NBI images. The main objective of this study is to use Deep Convolutional Neural Networks (DCNN) for the automatic classification of CE-NBI images into benign and malignant groups with minimal human intervention. (2) Methods: A pretrained Res-Net50 model combined with the cut-off-layer technique was selected as the DCNN architecture. A dataset of 8181 CE-NBI images was used during the fine-tuning process in three experiments where several models were generated and validated. The accuracy, sensitivity, and specificity were calculated as the performance metrics in each validation and testing scenario. (3) Results: Out of a total of 72 trained and tested models in all experiments, Model 5 showed high performance. This model is considerably smaller than the full ResNet50 architecture and achieved the testing accuracy of 0.835 on the unseen data during the last experiment. (4) Conclusion: The proposed fine-tuned ResNet50 model showed a high performance to classify CE-NBI images into the benign and malignant groups and has the potential to be part of an assisted system for automatic laryngeal cancer detection.

Keywords: Deep Convolution Neural Network; contact endoscopy; narrow band imaging; classification; larynx; cancer

1. Introduction

Laryngeal cancer is one of the most common malignancies in the head and neck area, with a growing incidence rate every year [1]. The treatment options and prognosis depend on the cancer stage at the time of diagnosis. Precancer or early-stage laryngeal cancer is associated with high rates of laryngeal preservation, a local control rate of 87–89%, and a favorable prognosis [2]. On the other hand, advanced-stage cancer requires multi-modal treatment strategies resulting in significant toxicities and a poorer quality of life. Despite optimized treatment schemes, studies report high recurrence rates and a 5 year overall survival of 33–61% [3,4].

Nowadays, the endoscopic imaging modalities have become the standard procedure for screening and early diagnosis of laryngeal cancerous and precancerous lesions in clinical settings. These methods are widely applicable before performing a surgical biopsy for histological tissue examination in the context of the so-called optical biopsy [5,6]. As one of these techniques, the combination of Contact Endoscopy (CE) with Narrow Band Imaging (NBI) can represent an enhanced and magnified visualization of changes in the morphology and three-dimensional orientation of vocal fold's subepithelial blood vessels [7,8]. The visual evaluation of these vascular structures in CE-NBI images can provide complementary information for the diagnosis of laryngeal cancerous or precancerous lesions. However, the use of CE-NBI for diagnosis highly relies on the experience of the otolaryngologists and requires several years of training. This can result in a subjective decision process followed by an overtreatment or undertreatment planning [7,9,10].

The advanced development of feature engineering, Machine Learning (ML), and Deep Learning (DL) methods in the area of medical applications provides several paths to assist the clinicians and overcome such challenges in the clinical environments. In this regard, several computer-based approaches were used on the larynx endoscopic images. These methods can assist otolaryngologists by providing complementary information regarding the stage of the cancer and characteristics of the vascular trees and larynx epithelial tissue [11]. In the area of laryngoscopic and NBI image analysis, an ensemble of Convolutional Neural Networks (CNN) with texture and frequency-domain-based features [12] and a set of hand-crafted texture and first-order statistical features [13] were proposed for larynx cancerous tissue classification. A Deep Convolutional Neural Network (DCNN) achieved the overall accuracy of 86% to detect cancer, precancerous lesions, and normal tissues in larynx [14]. A image classification system based on CNN outperformed the manual assessment of trainees in discriminating cysts, granulomas, nodules, normal cases, palsies, papillomas, and polyps [15]. The combination of hand-crafted and DL-based features showed a median classification recall of 98% for the diagnosis of early stage Squamous Cell Carcinoma (SCC) in larynx [16]. Moreover, another CNN-based approach achieved an equivalent performance to otolaryngologists' predictions for the diagnosis of laryngeal SCC [17].

Given that there is a need for more magnified and enhanced endoscopic techniques such as CE-NBI images, two sets of hand-crafted features combined with ML techniques were proposed for the automatic assessment of these type of images. These methods have the potential to provide an evaluation of vascular characteristics [18,19], assist otolaryngologists when there are disagreements regarding the final diagnosis [20,21], and present a computer-based classification of benign and malignant laryngeal lesions [22]. However, these works exhibited certain drawback in terms of the multiple image preprocessing stages that resulted in the loss of information from the images as well as manual feature extraction processes. Additionally, these studies focused only on some specific characteristics of the CE-NBI images, such as vascular geometry and textural characteristics and not the structures as a whole.

The main objective of this study is to use a fully automatic CE-NBI endoscopic image-based DCNN approach for the classification of laryngeal lesions and provide an objective assessment for otolaryngologists during the treatment process. This is performed to circumvent the disadvantages posed by ML-based approaches and rather have an approach that is more streamlined and automatic with minimal human intervention in the classification of lesions. To our knowledge, this is the first study that applies a DCNN-based approach for larynx CE-NBI image classification. The proposed approach uses the transfer learning concept which includes a pretrained ResNet50 model instead of developing a network from the scratch. Moreover, the pretrained ResNet50 model was tuned and combined with cut-off-layer technique to achieve the optimum architecture for this classification task. The performance of the proposed approach was evaluated in three different experiments. Then, it was compared to the performance of the state-of-the-art methods in the area of CE-NBI image classification.

2. Materials and Methods

In this section, we highlight the aspects of data preparation, discuss the model architecture, and detail the steps carried out during the experiments and training of the DCNN.

2.1. Data Preparation

CE-NBI video scenes of 146 patients who went through a microlaryngoscopy procedure were captured using an Evis Exera III Video System with integrated NBI-filter (Olympus Medical Systems, Hamburg, Germany). This setup included a rigid 30-degree contact endoscope (Karl Storz, Tuttlingen, Germany) with a fixed magnification of $60\times$. Then, 8181 CE-NBI images were extracted from the videos as explained in Esmaili et al. [7,18]. We went through each video scene and manually selected the time intervals where the video quality was good enough to visualize the blood vessels. Then, one in every ten frames was automatically extracted from the selected intervals in JPEG format images (1008×1280 pixels) to have unique and nonredundant vascular pattern in CE-NBI images. All patients' data were pseudonymized, and only biopsy results were taken to label images into benign and malignant lesions according to the WHO classification [23]. The benign class has 5313 images of patients with histopathologies such as Cyst, Polyp, Reinke's edema, Papillomatosis, Hyperplasia, Hyperkeratosis, and Mild Dysplasia. The malignant group includes 2868 images of patients diagnosed with Moderate Dysplasia, Severe Dysplasia, and Carcinoma in situ and SCC. The data were preprocessed and prepped in terms of size before being used as an input for the DCNN.

2.2. Model Architecture

The DCNN architecture used in this study is discussed here. DCNNs have gained recognition due to their adaptability for image recognition problem statements. These networks also yield higher accuracies as compared to other ML methods, due to their ability to solve problems from end-to-end rather than breaking them down as in the case of ML.

Transfer learning concept has become an important part of the growth of DL-based approaches in the field of medical image classification. It provides the chance of reusing a pretrained model as a starting point for a new classification task with comparatively few data. The pretrained network is a network that has already been introduced to a specific dataset and learned to extract valuable features from it. The dataset used for the pretraining is not always the same as the actual dataset for the second classification task, but the extracted features are similar in nature. This network can then be used as a starting point to learn a new classification task. In this study, a pretrained ResNet50 on ImageNet [24] database was considered for CE-NBI image classification task. Residual Networks (ResNets) are considered as examples of very deep classic structures in the computer vision literature [25]. ResNet50 is 50 layers deep, and the deepness level is related to the network's capability to capture high (or higher) patterns. ResNets optimize toward zero, which in turn accelerates the convergence to the optimal point in the solution space, instead of a real number. Batch normalization is another interesting feature that is embedded in ResNet's structure. It speeds up the convergence and in doing so reduces the training epochs required. It also has a regularization effect during the training phase. Figure 1 shows the overall view of the proposed architecture.

The pretrained ResNet50 was combined with the fine-tuning strategy as well as cut-off-layer technique to obtain the optimum performance for CE-NBI image classification. Fine-tuning a pretrained DCNN is beneficial as it enables the user to speed up training and overcome smaller dataset sizes. The fine-tuning technique wherein all the layers were fine-tuned was adopted for this work. In order to account for the issue of overfitting of ResNet50, we proposed setting the cut-off-layer to discard part of the network. The cut-off layer is the last layer in feature extraction part of the network, where the classifier part begins. This layer tends to be where the activation occurs. While training the network, it was noted that overfitting occurred due to the large size of the original ResNet architecture.

Hence, the cut-off layer was set empirically. This resulted in several models with different layer counts and therefore feature counts depending on where the cut-off layer was set. The final cut-off layer was selected based on the overall performance of the network. Then, different variations of the model were implemented for having sufficient number of features in a trade off between the training stage success and generalization ability of the model on unseen images.

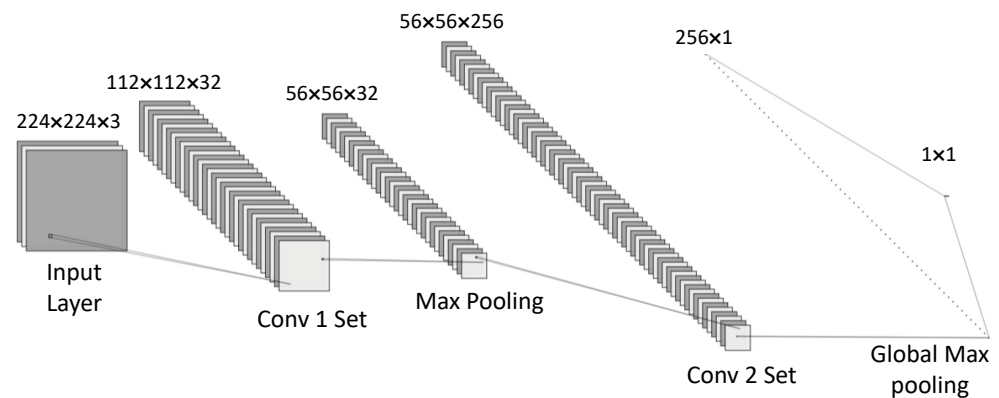


Figure 1. The overall architecture of the proposed approach.

2.3. Experiments

The experiments for this work were divided into three parts as shown in Table 1. A total of three experiments were conducted to determine the model most suitable for our problem statement. In the conducted experiments, a total of 72 models were trained and tested using the data collected. The main difference between these experiments was related to the strategy of data separation. Apart from this, a few experiments also took into consideration different network hyperparameters and changes in the volume of data. In Experiment 1, the separation into training and testing sets was performed randomly to form a 80–20 train-test split. Additionally, different cut-off-layer strategies and classifiers were tested in this experiment. In Experiment 2, we employed a manual method for splitting the training and testing data. This was performed so as to ensure that none of the test data were part of the training data as well as the images of patients exclusively tied to separate sets. Then, the best-performed model from Experiment 1 was tested in this experiment. In Experiment 3, data augmentation (vertical and horizontal flipping) was applied, and testing data selection criteria were kept the same as Experiment 2. The best-performed model from Experiment 1 was also tested under the specified condition of Experiment 3.

Table 1. The summary of three experiments classified according to the different conditions.

| Experiment | Data Augmentation | Cut-Off Layer | Classifier | Dataset Separation |
|--------------|-------------------|--|------------------------------|--------------------|
| Experiment 1 | No | conv2_block3_out (230 K parameters) | Global Max Pooling | Random |
| | | conv2_block3_out (230 K parameters) | Global Max Pooling + Dropout | |
| | | No cut-off (23.5 M parameters) | Global Max Pooling | |
| Experiment 2 | No | conv2_block3_out (230 K parameters) | Global Max Pooling | Manual |
| Experiment 3 | Yes | conv2_block3_out (230 K parameters) | Global Max Pooling | Manual |

2.4. Training Details

The ResNet50 model was adopted as the backbone for this work. Input images were resized to 224×224 pixels in the preprocessing stage. Data augmentation on the images was performed by employing the horizontal and vertical flipping methods. Binary cross entropy was used as a loss function along with Stochastic Gradient Descent (SGD) as the optimizer. The parameters were tuned as follows: `batch_size = 32`, `learning_rate = 0.001`, `decay = 1 \times 10^{-6}`, `momentum = 0.9`, `Nesterov momentum = True`. The cut-off layer was set at “conv2_block3_out” in an iterative process. Early stoppage was also set with a patience of 5 epochs. The network was trained for a total of 35 epochs and programmed using Python version 3.8.8. The study was carried out on a deep learning workstation with and Nvidia Quadro P6000 GPU. The 5-fold crossvalidation technique was used for validating the models.

2.5. Performance Metrics

The study used accuracy, sensitivity, and specificity as performance metrics. These are given below along with their formulas:

$$\text{Accuracy} = \frac{\text{TruePositives} + \text{TrueNegatives}}{\text{TotalNumberofImages}} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{TruePositives}}{\text{TruePositives} + \text{FalseNegatives}} \quad (2)$$

$$\text{Specificity} = \frac{\text{TrueNegatives}}{\text{TrueNegatives} + \text{FalsePositives}} \quad (3)$$

3. Results

The performance of the selected models from the three experiments are listed in Table 2. On average, 69.9 min was taken to execute the training and validation phase during different experiments, followed by a testing phase that took on average 52.3 s.

Table 2. Results of the selected models in each experiment. Metrics of the validation and testing phases are averages over five folds.

| Experiment | Model | Validation | | | Testing | |
|--------------|---------|------------|-------------|-------------|---------|----------|
| | | Accuracy | Sensitivity | Specificity | Loss | Accuracy |
| Experiment 1 | Model 5 | 0.979 | 0.967 | 0.986 | 0.06 | 0.991 |
| | Model 6 | 0.943 | 0.914 | 0.959 | 0.15 | 0.958 |
| | Model 7 | 0.967 | 0.960 | 0.974 | 0.11 | 0.984 |
| Experiment 2 | Model 5 | 0.976 | 0.958 | 0.985 | 0.07 | 0.929 |
| Experiment 3 | Model 5 | 0.925 | 0.888 | 0.960 | 0.20 | 0.835 |

Of the all models trained and tested, Models 5–7 showed the most promising results during Experiment 1. Model 5 achieved an accuracy, sensitivity, and specificity of 0.979, 0.967, and 0.986, respectively. When compared to the metrics produced by Model 6 and Model 7, these scores were higher in both the validation and testing phases. Figure 2 shows the comparison between the accuracy curves between Models 5 and 7 over 35 epochs for Experiment 1. It can be seen from the figure that the curves for Model 5 are more consistent as opposed to the curves seen in Model 7 in the this experiment. On the other hand, by visual evaluation of the graph, we can see that the accuracy achieved by Model 5 at epoch 5 is equal to 0.927, while Model 7 had a lower rate equal to 0.853% at the same epoch. Based on these evaluations, we decided to move forward with Model 5 and Global Max Pooling classifier for the following two experiments.

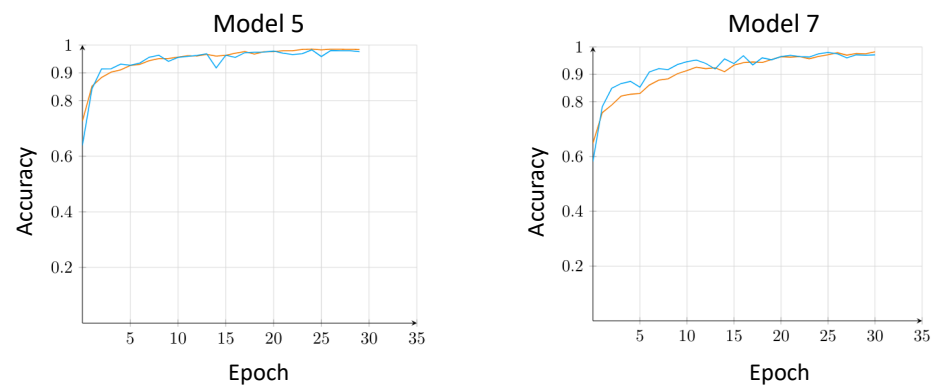


Figure 2. Comparison of the accuracy track between Model 5 and Model 7 in Experiment 1. Orange and blue lines represent the training and validation phase, respectively.

In Experiment 2, Model 5 exhibited marginally lower scores in terms of validation accuracy, sensitivity, and specificity where the testing data were manually selected so as to ensure they were not part of the training set. In this experiment, the deviation in accuracy value occurs between validation and testing scenarios because there is the possibility that the validation set is not representative to the testing dataset. This can lead to biased fine-tuned model to the validation set and possible overfitting in this scenario. Therefore, we moved on to Experiment 3 with Model 5 and Global Max Pooling classifier together with data augmentation techniques.

Model 5 in Experiment 3 exhibited an accuracy, sensitivity, and specificity of 0.925, 0.888, and 0.960, respectively, during the validation phase and an accuracy score of 0.835 in the testing scenario. Figure 3 depicts the examples of the classification given by Model 5. The top row of the Figure 3 corresponds to accurately classified images and the bottom row to inaccurately image classifications. The Perpendicular Vascular Changes (PVC) in laryngeal Papillomatosis can be difficult to visually distinguish from PVC in premalignant and malignant histopathologies [26]. Among the accurate classifications represented in Figure 3, it is significant to note that Model 5 was able to accurately differentiate such images where there were similar vascular structures but different histopathologies (malignant Carcinoma in situ vs. benign Papilloma). On the other hand, classification inaccuracies can arise due to the complexity of the vessel arrangements in the CE-NBI images. This issue was predicted in Experiment 3 as the testing data included a set of unseen and augmented images. Moreover, the dataset has a comprehensive selection of several histopathologies from different patients that can increase the chance of complexity during classification scenarios of the unseen and augmented data.

Figure 4 depicts the graphs of the accuracy and loss for Model 5 in Experiment 3. Both graphs follow a smooth ascend (accuracy) and descend (loss). From this, we can infer that the model followed a relatively stable training cycles through each of the epochs. The accuracy (training vs. validation) graph show a good fit overall for the model during the experiment. Although they meet in the end, the loss (training vs. validation) graph shows a much more erratic behavior during the epochs.

Figure 5 exhibits the confusion matrix of Model 5 in testing scenario of Experiment 3. The images in the benign and malignant groups were labeled as 0 and 1, respectively. With this explanation, it can be seen from this matrix that the number of misclassified images in the malignant group is more than the benign class.

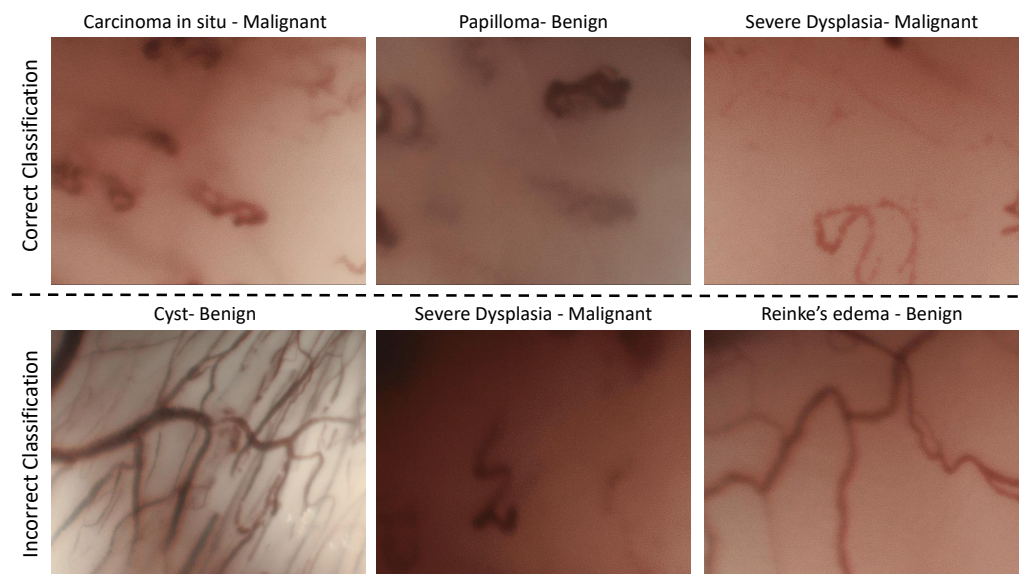


Figure 3. The example of correct and incorrect classification of CE-NBI images in Experiment 3.

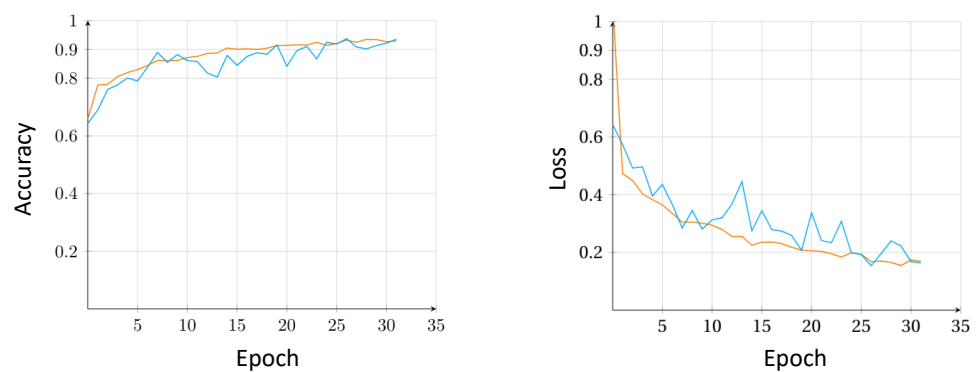


Figure 4. The accuracy and loss graphs of model Model 5 in Experiment 3. Orange and blue lines represent the training and validation phase, respectively.

| | | Predicted Class | | Total |
|--------------|---|-----------------|------|-------|
| | | 0 | 1 | |
| Actual Class | 0 | 2012 | 152 | 2164 |
| | 1 | 484 | 964 | 1448 |
| Total | | 2496 | 1116 | |

Figure 5. Confusion matrix testing scenario of Experiment 3.

4. Discussion

In this study, a fully automatic DCNN-based approach using a pretrained and fine-tuned ResNet50 architecture was adopted and evaluated on CE-NBI images for the benign and malignant laryngeal lesion classification. To the best of our knowledge, no previous study has applied DCNN-based models on larynx CE-NBI images for any classification or segmentation purposes. Considering the presented results, the DCCN-based approach has

the potential to differentiate malignant lesions from several benign ones in CE-NBI images with high performance and can provide a more consistent interpretation and an objective decision-making process for clinicians.

The application of DCNN-based methods has brought effective solutions in the area of image analysis for a better understanding of image content. Together with the development of these techniques, the concept of transfer learning has introduced a new perception to deal with the problem of a limited number of images for training these models. It allows reusing the pretrained models for a similar task, such as image classification. Among DCNNs that achieved significant outcomes, AlexNet [27], VGGNets [28], InceptionNets [29], and ResNets [25] are some well-known pretrained models for medical image classification. These architectures were developed for certain purposes and have shown their own strengths and limitations. Depending on the area of application as well as the type of imaging modality, each of these networks has shown the ability to provide a better understanding of the patients' status for the clinicians [30–32]. Among them, the ResNet convolutional networks are the most popular as they can offer very deep architectures with shortcut connections to solve the vanishing gradient problem. Moreover, the batch normalization features in these networks can speed up the convergence and reduce the required training epochs [25]. In the area of medical image analysis, ResNet34 was evaluated to determine the class of laryngeal Stimulated Raman Scattering (SRS) images based on normal or neoplastic classes. This architecture showed the rapid and automated recognition on the validation set with an accuracy of 0.959 [33]. In another study, a fine-tuned ResNet50 network was used for classifying multimodal images of breast tissues into normal, fat, and cancerous. Using leave-one-patient-out crossvalidation, the model achieved the mean sensitivity of 0.862 on the validation images [34]. In addition, fine-tuned ResNet50, InceptionV2, and SqueezeNet models were selected to multiclassify laryngoscopy frames into four classes and were achieved the macroaverage AUC (Area Under the Curve) of 0.998, 0.989, and 0.999, respectively [35]. In a recent evaluation, ResNet50 and ResNet101 architectures were part of an ensemble model that was applied for cancer tissue classification in larynx NBI images. The combination of this ensemble model with a series of hand-crafted features achieved the classification accuracy of 0.954 [12]. Considering the proven performance of ResNet convolutional networks in medical image classification tasks as well as the advantages of these architectures over other networks, the pretrained ResNet50 was used for our evaluation. This network utilized images in the pretraining step that displayed a pattern similar to that of the blood vessels as used in this study.

After the evaluation, the outcomes of three different experiments, the fine-tuned ResNet50 model from the Experiment 3 was proposed as the final architecture from 72 total models. This model achieved the mean accuracy, sensitivity, and specificity of 0.925, 0.888, and 0.960 in the validation phase and the mean accuracy of 0.835 from the testing scenario. Although this model showed lower performance than the tested models in Experiments 1 and 2, it was evaluated in a more realistic scenario. One of the major benefits of this model over the latest DCNN-based methods is the size of the fine-tuned ResNet50 model. The application of the cut-of-layer technique resulted in a smaller model that only has the size equal to $\approx 1\%$ of the full ResNet50 architecture (1.96 Megabytes versus 180.65 Megabytes). In addition, the smaller architecture showed faster training with less prone to result in overfitting. Earlier, it was mentioned that the chance of overfitting increases while using the ResNet50 architecture. Hence, apart from cut-off-layer technique, other strategies such as including a larger number of images, performing data augmentation, and early stopping were also employed to avoid the overfitting of ResNet50 in this study.

In comparison to the other works in the area of laryngeal cancer detection and classification, we used the CE-NBI images as the imaging modality. NBI imaging enables a highly contrasted visualization of vascular structures. The essential advantage of CE-NBI over the normal white light laryngoscopy is the highly magnified visualization of vascular patterns that results in a more precise evaluation of laryngeal lesions [7].

In this study, there is a slight data imbalance between the number of benign and malignant images in the CE-NBI image dataset ($\approx 60\%$ benign vs. $\approx 40\%$ malignant). This issue could be solved by using a two-fold data augmentation approach where the data augmentation is first performed to balance the data and then the second augmentation is applied to the entire dataset as a whole. However, this can increase the risk of redundancies especially in the case of CE-NBI images as vascular patterns are already very similar. For this reason, we chose not to tamper with the imbalance issue because it is not significantly greater than it would affect the performance of the network. Moreover, the data, as they are, are representative of the true clinical scenario where there is often an imbalance in the data collected. This dataset includes around 8000 CE-NBI images from a wide range of various histopathologies in both benign and malignant groups, which is a comparable number of data in comparison to other studies where the endoscopy-based imaging techniques were used for similar classification tasks in the larynx. The number of images on these evaluations ranges from a minimum of 330 to a maximum of 14,000 [12–16,35]. This maximum number exists because multiple clinical centers were in the data collection process simultaneously [14]. On the other hand, the subsets of this CE-NBI image dataset were used to develop and test multiple hand-crafted feature extraction and ML methods for laryngeal cancer classification [18,20,22]. In this respect, the recent work reported the classification accuracy of 0.966 using two feature sets combined with k-Nearest Neighbors (kNN) classifier [22]. Even though this method outperformed the proposed model, it included three different image preprocessing stages, needed the manual parameter selections, and was tested on a smaller dataset.

As was mentioned before, the benign lesions show similar vascular patterns to the malignant ones in CE-NBI image analysis. The visual evaluation of this cases can cause one of the serious problems in the clinical environment which is the differentiation between benign and malignant lesions [20]. In the present study, the achieved specificity was higher than the sensitivity values in all experiments. This outcome can emphasize the ability of the proposed model to overcome this issue and assist otolaryngologists to also evaluate benign cases more confidently.

5. Conclusions

In summary, a CE-NBI endoscopic image-based DCNN model was developed and tested through a fine-tuned ResNet50 architecture. The proposed model had a high performance for the automatic classification of laryngeal cancerous lesions and showed comparable performance to the studies in the area of larynx CE-NBI image classification, as was explained in the previous section. The proposed structure is significantly smaller than the full ResNet50 architecture as a result of the cut-off-layer technique. Moreover, no over- and under-fitting were observed in the final architecture. The proposed model has the potential to be a solution for the subjective assessment of the benign and malignant laryngeal lesions in clinical settings and reduce the chance of performing an invasive surgical biopsy. This effective solution can be part of the Compute-Aided-Diagnosis (CAD) system that assists otolaryngologists during the decision-making process and improves the optical diagnosis rate of larynx cancer.

To improve the performance of the proposed model, more investigations are planned for multidomain feature extraction methods (DCNN combined with hand-crafted features) as well as the development of ensemble DCNN models for the future work. Moreover, it is essential to continue further development on a multiclassification scenarios to differentiate between different laryngeal histopathologies and improve the application of optical biopsy in the clinical settings.

Author Contributions: Conceptualization, N.E., E.S., E.J.G.A., A.I., A.B., N.D., C.A., N.N. and M.F.; methodology, N.E., E.S. and E.J.G.A.; software, E.S. and E.J.G.A.; validation, N.E., E.S., E.J.G.A., A.I., A.B. and N.D.; formal analysis, N.E., E.S., E.J.G.A., A.I., A.B. and N.D.; investigation, N.E., E.S., E.J.G.A., A.I., A.B.; data curation, N.E., N.D. and C.A.; writing—original draft preparation, N.E., E.S. and E.J.G.A.; writing—review and editing, N.E., E.S., E.J.G.A., A.I., A.B., N.D., C.A., N.N. and M.F.; visualization, N.E., E.S. and E.J.G.A.; supervision, M.F. and N.N.; project administration, N.E. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The research related to human use complies with all the relevant national regulations and institutional policies and was performed in accordance with the tenets of the Helsinki Declaration and has been approved by the authors' institutional review board or equivalent committee (number 49/18).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data that support the findings of this study are part of the research project and are not publicly available.

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

References

1. Wang, J.Y.; Zhang, Q.W.; Wen, K.; Wang, C.; Ji, X.; Zhang, L. Temporal trends in incidence and mortality rates of laryngeal cancer at the global, regional and national levels, 1990–2017. *BMJ Open* **2021**, *11*, e050387. [[CrossRef](#)]
2. Guimarães, A.V.; Dedivitis, R.A.; Matos, L.L.; Aires, F.T.; Cernea, C.R. Comparison between transoral laser surgery and radiotherapy in the treatment of early glottic cancer: A systematic review and meta-analysis. *Sci. Rep.* **2018**, *8*, 11900. [[CrossRef](#)] [[PubMed](#)]
3. García-León, F.J.; García-Esteba, R.; Romero-Tabares, A.; Borrachina, J.G.M. Treatment of advanced laryngeal cancer and quality of life. Systematic review. *Acta Otorrinolaringol.* **2017**, *68*, 212–219. [[CrossRef](#)] [[PubMed](#)]
4. Elicin, O.; Giger, R. Comparison of current surgical and non-surgical treatment strategies for early and locally advanced stage glottic laryngeal cancer and their outcome. *Cancers* **2020**, *12*, 732. [[CrossRef](#)] [[PubMed](#)]
5. Missale, F.; Taboni, S.; Carobbio, A.L.C.; Mazzola, F.; Berretti, G.; Iandelli, A.; Fragale, M.; Mora, F.; Paderno, A.; Del Bon, F.; et al. Validation of the European Laryngological Society classification of glottic vascular changes as seen by narrow band imaging in the optical biopsy setting. *Eur. Arch. Oto-Rhino-Laryngol.* **2021**, *278*, 2397–2409. [[CrossRef](#)]
6. Lauwerends, L.J.; Galema, H.A.; Hardillo, J.A.; Sewnaik, A.; Monserez, D.; van Driel, P.B.; 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](#)]
7. Davaris, N.; Lux, A.; Esmaeili, N.; Illanes, A.; Boese, A.; Friebe, M.; Arens, C. Evaluation of Vascular Patterns using Contact Endoscopy and Barrow-Band Imaging (CE-NBI) for the Diagnosis of Vocal Fold Malignancy. *Cancers* **2020**, *12*, 248. [[CrossRef](#)]
8. Puxeddu, R.; Sionis, S.; Gerosa, C.; Carta, F. Enhanced contact endoscopy for the detection of neoangiogenesis in tumors of the larynx and hypopharynx. *Laryngoscope* **2015**, *125*, 1600–1606. [[CrossRef](#)]
9. Mannelli, G.; Ceconi, L.; Gallo, O. Laryngeal preneoplastic lesions and cancer: Challenging diagnosis. Qualitative literature review and meta-analysis. *Crit. Rev. Oncol./Hematol.* **2016**, *106*, 64–90. [[CrossRef](#)]
10. Mehlum, C.S.; Døssing, H.; Davaris, N.; Giers, A.; Grøntved, Å.M.; Kjaergaard, T.; Möller, S.; Godballe, C.; Arens, C. Interrater variation of vascular classifications used in enhanced laryngeal contact endoscopy. *Eur. Arch. Oto-Rhino-Laryngol.* **2020**, *277*, 2485–2492. [[CrossRef](#)]
11. Singh, V.P.; Maurya, A.K. Role of Machine Learning and Texture Features for the Diagnosis of Laryngeal Cancer. *Mach. Learn. Healthc. Appl.* **2021**, 353–367. [[CrossRef](#)]
12. Nannia, L.; Ghidoni, S.; Brahnam, S. Ensemble of convolutional neural networks for bioimage classification. *Appl. Comput. Inform.* **2020**, *17*, 19–35.
13. Moccia, S.; De Momi, E.; Guarnaschelli, M.; Savazzi, M.; Laborai, A.; Guastini, L.; Peretti, G.; Mattos, L.S. Confident texture-based laryngeal tissue classification for early stage diagnosis support. *J. Med. Imaging* **2017**, *4*, 034502. [[CrossRef](#)]
14. 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](#)] [[PubMed](#)]
15. 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.
16. Araújo, T.; Santos, C.P.; De Momi, E.; Moccia, S. Learned and handcrafted features for early-stage laryngeal SCC diagnosis. *Med. Biol. Eng. Comput.* **2019**, *57*, 2683–2692. [[CrossRef](#)]

17. Hu, R.; Zhong, Q.; Xu, Z.; Huang, L.; Cheng, Y.; Wang, Y.; He, Y. Application of deep convolutional neural networks in the diagnosis of laryngeal squamous cell carcinoma based on narrow band imaging endoscopy. *Chin. J. Otorhinolaryngol. Head Neck Surg.* **2021**, *56*, 454–458.
18. Esmaeili, N.; Illanes, A.; Boese, A.; Davaris, N.; Arens, C.; Friebe, M. Novel Automated Vessel Pattern Characterization of Larynx Contact Endoscopic Video Images. *Int. J. Comput. Assist. Radiol. Surg.* **2019**, *14*, 1751–1761. [[CrossRef](#)] [[PubMed](#)]
19. Esmaeili, N.; Illanes, A.; Boese, A.; Davaris, N.; Arens, C.; Friebe, M. A Preliminary Study on Automatic Characterization and Classification of Vascular Patterns of Contact Endoscopy Images. 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. 2703–2706.
20. Esmaeili, N.; Illanes, A.; Boese, A.; Davaris, N.; Arens, C.; Navab, N.; Friebe, M. Laryngeal Lesion Classification based on Vascular Patterns in Contact Endoscopy and Narrow Band Imaging: Manual versus Automatic Approach. *Sensors* **2020**, *20*, 4018. [[CrossRef](#)]
21. Esmaeili, N.; Illanes, A.; Boese, A.; Davaris, N.; Arens, C.; Navab, N.; Friebe, M. Manual versus Automatic Classification of Laryngeal Lesions based on Vascular Patterns in CE+NBI Images. *Curr. Dir. Biomed. Eng.* **2020**, *6*, 70–73. [[CrossRef](#)]
22. Esmaeili, N.; Boese, A.; Davaris, N.; Arens, C.; Navab, N.; Friebe, M.; Illanes, A. Cyclist Effort Features: A Novel Technique for Image Texture Characterization Applied to Larynx Cancer Classification in Contact Endoscopy—Narrow Band Imaging. *Diagnostics* **2021**, *11*, 432. [[CrossRef](#)] [[PubMed](#)]
23. Gale, N.; Hille, J.; Jordan, R.C.; Nadal, A.; Williams, M.D. Regarding Laryngeal precursor lesions: Interrater and intrarater reliability of histopathological assessment. *Laryngoscope* **2019**, *129*, E91–E92. [[CrossRef](#)]
24. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Li, K.; Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
25. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
26. Arens, C.; Piazza, C.; Andrea, M.; Dikkers, F.G.; Gi, R.E.T.P.; Voigt-Zimmermann, S.; Peretti, G. Proposal for a descriptive guideline of vascular changes in lesions of the vocal folds by the committee on endoscopic laryngeal imaging of the European Laryngological Society. *Eur. Arch. Oto-Rhino-Laryngol.* **2016**, *273*, 1207–1214. [[CrossRef](#)]
27. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1097–1105. [[CrossRef](#)]
28. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
29. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
30. Sarvamangala, D.; Kulkarni, R.V. Convolutional neural networks in medical image understanding: A survey. *Evol. Intell.* **2021**, *1*–22. [[CrossRef](#)]
31. Upreti, M.; Pandey, C.; Bist, A.S.; Rawat, B.; Hardini, M. Convolutional Neural Networks in Medical Image Understanding. *Aptisi Trans. Technopreneurship (ATT)* **2021**, *3*, 6–12. [[CrossRef](#)]
32. Yadav, S.S.; Jadhav, S.M. Deep convolutional neural network based medical image classification for disease diagnosis. *J. Big Data* **2019**, *6*, 1–18. [[CrossRef](#)]
33. Zhang, L.; Wu, Y.; Zheng, B.; Su, L.; Chen, Y.; Ma, S.; Hu, Q.; Zou, X.; Yao, L.; Yang, Y.; et al. Rapid histology of laryngeal squamous cell carcinoma with deep-learning based stimulated Raman scattering microscopy. *Theranostics* **2019**, *9*, 2541–2554. [[CrossRef](#)]
34. Ali, N.; Quansah, E.; Köhler, K.; Meyer, T.; Schmitt, M.; Popp, J.; Niendorf, A.; Bocklitz, T. Automatic label-free detection of breast cancer using nonlinear multimodal imaging and the convolutional neural network ResNet50. *Transl. Biophotonics* **2019**, *1*, e201900003. [[CrossRef](#)]
35. Galdran, A.; Costa, P.; Campilho, A. Real-Time Informative Laryngoscopic Frame Classification with Pre-Trained Convolutional Neural Networks. In Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), Venice, Italy, 8–11 April 2019; pp. 87–90.

Chapter 3 – Conclusion and future work

In this thesis, two novel pipelines of a CAD system for laryngeal lesion assessment based on CE-NBI images have been developed. Chapter 1 started with an introduction to the general clinical procedure for laryngeal lesion assessment and larynx cancer diagnosis. Then, the endoscopic imaging techniques – as the main component in this procedure – were described, and their roles and importance were demonstrated. In between, the main pains and obstacles in the current procedure of laryngeal lesion assessment were highlighted. Next, the general concept and structure of the CAD system and its application on laryngeal cancer diagnosis using standard WL and NBI endoscopic images were described. Conventional and DL-based CAD systems – as the two prominent architectures – were investigated in detail, and their primary deficiency in dealing with the current clinical pains in laryngeal lesion assessment was presented. Lastly, the concept of CE combined with enhanced endoscopy imaging was introduced. Moreover, the value of this modality with a focus on the vascularization network was presented, and its main limitations to improve the procedure of laryngeal lesion assessment were highlighted.

Many malignant laryngeal cases originate from precursor lesions, ranking LC in a group of cancers qualified for early detection. Early diagnosis of laryngeal lesions can result in favorable treatment outcomes and improve patients' quality of life. However, it requires a precise clinical and histopathological examination of the lesion and surrounding tissues. Nowadays, WL and NBI endoscopic images are the standard imaging tools for pre-/ intra-/ post-operative clinical examination of laryngeal lesions. They can visualize the examined region as an image or video, providing valuable information about mucosal and vascular changes of laryngeal tumors. On the other hand, surgical biopsy is the standard approach for histopathological examination of the laryngeal lesions that define the final diagnosis decision. Optical Biopsy was introduced to meet the clinical needs in the current workflow of laryngeal lesion assessment to move toward a less invasive diagnosis approach. This concept aims to provide a reliable, fast, and real-time endoscopy-based assessment of pathological conditions in the larynx to provide early diagnosis and reduce the number of unnecessary surgical biopsies. However, Optical Biopsy is not integrated into the standard examination workflow of laryngeal lesions. The limitation of WL and NBI endoscopic imaging modalities could explain the main reason behind this matter. These techniques cannot provide detailed and magnified visualization of mucosal and vascular transitions of the epithelial layer. Several guidelines and instruction tools were proposed to assist Otolaryngologists in interpreting the information related to the mucosal and vascular changes in WL and NBI endoscopic images. Nevertheless, they raised the

issues of a long learning process, subjective assessment, and increased false positive diagnoses. Multiple research studies were conducted on developing CAD systems based on classification or segmentation tasks on WL and NBI endoscopic images for a more objective assessment of laryngeal lesions and LC diagnosis. However, they could not provide the platform for integrating Optical Biopsy into the standard examination workflow because they were developed for research intentions. In addition, they dismissed the factors such as usability engineering practice to meet the users' needs for implementing a computer-based solution as a medical product in standard clinical procedures.

In recent years, CE-NBI modality has gained prominence in dealing with issues related to the WL and NBI endoscopic imaging modalities along with focusing on examining sub-epithelial vascularization networks in Optical Biopsy. CE-NBI provides magnified and enhanced visualization of these vascular networks, enabling a more detailed examination of lesions. However, visual interpretation of CE-NBI images poses a challenge due to the similarity and complexity between the vascular patterns of benign and malignant pathologies. This issue leads to a subjective diagnosis that requires significant expertise from Otolaryngologists. Therefore, CE-NBI remains primarily a research tool for laryngeal lesion assessment. To overcome the limitations of this imaging modality and harness its clinical and technical potentials, this thesis endeavors to introduce two pipelines of a CAD system for automatic assessment of laryngeal lesions in image classification tasks focusing on the vascular characteristics represented in CE-NBI images.

The first task of the thesis was dedicated to collecting and preparing the data required for any technical exploration of CE-NBI images. The second task is then devoted to designing and developing two pipelines for the CAD system:

- Pipeline 1 is based on feature engineering techniques combined with ML classifiers.
- Pipeline 2 is based on pre-trained DL-based architectures.

In pipeline 1, we introduced two novel sets of handcrafted features that described the geometrical characteristics and textural attributes of vascularization networks in CE-NBI images. The GF and CyEfF were combined with four distinct supervised ML classifiers to classify benign-malignant laryngeal lesions. Despite the limited available data, both standalone and combined use of these feature sets demonstrated high performance in CE-NBI image classification. Furthermore, the performance confirmed the correlation between the changes in the morphology of sub-epithelial blood vessels of the vocal fold and the type of laryngeal lesion. The main image misclassifications in these methods occurred when benign and malignant cases exhibited similar PVC in vascularization networks.

In pipeline 2, the strategy was changed to DL-based approaches to address the increasing similarity of vascular patterns in the CE-NBI data set. The first approach in this pipeline included a pre-trained and fine-tuned ResNet50 architecture combined with a cut-off layer technique to reduce the network size while optimizing the performance that adopts sufficiently for real-time clinical applications. The second approach, Method 4, was developed based on the concept of ensemble modeling to combine the power of different networks for CE-NBI image classification. The transfer learning concept was again applied to this development phase, where seven pre-trained DL-based architectures were selected and subjected to preliminary training and validation. Based on the results of this evaluation, EfficientNetB0V2 and DenseNet121 were chosen as the first two networks, and ResNet50V2 was added as the third model. Finally, the implementation step, including the training, validation, and fine-tuning of every network along with building the ensemble model, was conducted on the final CE-NBI data set. One of the challenges of this implementation was related to the imbalanced data in benign and malignant classes. Therefore, data augmentation techniques were combined in the development and were applied to the training set. In general, the main image misclassification in pipeline 2 was raised for these methods due to the complexity and variety of the vascular networks' arrangements in the CE-NBI image data set.

The main difference between the approaches used in pipeline 1 and 2 development can be described as follows.

- According to the development strategy:
 - In pipeline 1, we focused on the specific sources of information – as feature engineering techniques – in CE-NBI image classification. The first method was designed based on geometrical attributes of vascularization networks, and the second method was focused on the textural characteristics in CE-NBI images. On the other hand, the DL-based architectures developed and validated in pipeline 2 considered the entire image as the primary source of information and showed well generalization to new data.
 - For a reason explained in the previous point, implementing the methods in pipeline 1 required the application of two- to three-stage image pre-processing to enhance the sources of information that we were aiming to use for the features extraction step, such as blood vessels. Although Method 2 in pipeline 1 (CyEffF) showed high performance with only one-stage wavelet-based image pre-processing, the strategies implemented and validated in pipeline 2 did not require any image pre-processing techniques.
 - The methods in pipeline 1 were focused on feature engineering techniques. Although such methods provide complete control for data transformation and feature computation based on the visible

characteristics in CE-NBI images, they usually require hard coding for manual parameter selections. On the other hand, the techniques in pipeline 2 included an automatic and fast-forwarded feature extraction and classification process. Considering this option, the DL-based methods had better capability to explore the data where the underlying patterns or characteristics of the image were complex or difficult to capture.

- According to the CE-NBI data:
 - In pipeline 1, a limited number of images was used to develop and validate two sets of handcrafted features. The training, validation and testing process of ML classifiers using these sets of features included a maximum of 3000 CE-NBI images. On the other hand, developing the first DL-based technique started with around 8000 images and was expanded to about 11000 data for the development of DL-based ensemble model.
 - Due to fewer malignant histopathologies than benign ones, the number of CE-NBI images in the malignant group was always less than those in the benign class. The development of methods in pipeline 1 was affected by this issue in some specific classification scenarios. Nevertheless, this problem became more evident during the implementation of DL-based strategies in pipeline 2. Therefore, image data augmentation techniques were implemented on training data to tackle this issue.

All the methods implemented in pipelines 1 and 2 showed accuracy, sensitivity, and specificity higher than 80% in all CE-NBI image classification scenarios. Focusing on the benign-malignant laryngeal lesion assessment, each of the proposed methods or their fusion could play the role of a CAD system to assist and guide Otolaryngologists in evaluating CE-NBI images. With the valuable source of information provided in magnified and enhanced endoscopic images as CE-NBI, the application of a CAD system has the potential to deal with the issue of subjective assessment of laryngeal lesions, which could result in a less invasive, fast, and accurate Optical Biopsy diagnosis leading to earlier detection of laryngeal cancer.

In pipeline 1, the combination of GF and CyEff feature sets with the kNN ML classifier resulted in an accuracy of 96%. On the other hand, the pre-trained and fine-tuned ResNet50 in pipeline 2 showed an accuracy of 83%. Moreover, the ensemble model generated out of ResNet50V2, EfficientNetB0V2, and DenseNet121 architectures in this pipeline demonstrated an accuracy of 92%.

Each method has advantages and drawbacks in CE-NBI image classification based on benign-malignant laryngeal lesions. In pipeline 1, we arrived at the best CE-

NBI image classification performance with the combination of 24 GF and 2 CyEff fed into a classical and standard kNN classifier. The implementation was executed on around 3000 CE-NBI images on a PC with a CPU operating at 2.30 GHz resulting in an average feature computation time of 2.36 seconds per image and 12 minutes of training and validation. In pipeline 2, we first fine-tuned every pre-trained architecture on the entire CE-NBI data set with about 11000 images. Then, the implementation was carried out on a DL workstation with Nvidia Quadro P6000 GPU, where on average, 30 to 69 minutes were taken to execute the training and validation phase for different architectures.

Apart from the sources of information considered for CE-NBI image classification in each pipeline, increasing the complexity and similarity among vascularization networks affected the performance of feature engineering techniques combined with ML classifier as the final trained model did not generalize well to the new data. The data used to develop DL-based models included images from more patients with several cases per histopathology. Although the data set had a different variety of histopathology and vascularization networks, the DL-based architectures were generalized well during the training step. Moreover, we managed the problem of overfitting during the development of DL-based methods by performing standard approaches such as data augmentation, defining early stopping and avoiding overlap between training and testing data sets.

This project highlighted the value of the CE-NBI images as a minimally invasive, magnified, and enhanced endoscopy imaging technique for providing diagnostic information on performing Optical Biopsy. The value of this kind of medical data collected over several years was shared with the scientific community as a publicly available CE-NBI data set. Moreover, our collaborative effort with the clinical team illustrated the subjective evaluation of CE-NBI images during the current clinical examination of laryngeal lesions. This investigation highlighted the existing challenges and needs in clinical practice and guided us to focus on easy-to-understand development strategies with real-time functionality. For that, we focused on handcrafted features in pipeline 1 that capture the discernible characteristics of the images as perceived by the Otolaryngologist. Additionally, several strategies, such as cut-off layer technique, were applied in pipeline 2 to optimize the training and testing phases of DL architectures. Finally, the project presented the effectiveness of using multiple ML-based solutions to address clinical issues by providing more objective and reliable assistance in interpreting the data presented in CE-NBI images.

The long-term goal of this project is to peruse a clinical trial study to:

- Determine the effect of the CAD system in facilitating the clinical examination of laryngeal lesions via Optical Biopsy.

- Evaluate the impact of the CAD system on improving the comfortability and confidence of Otolaryngologists – especially the less-experienced groups – in the decision-making procedure.
- Study the role of the CE-NBI images combined with the CAD system in promptly diagnosing early-stage laryngeal cancer.

For this purpose, we need to finalize the architecture of the CAD system. According to the performance of each method, two main options are proposed for the final architecture of the CAD system, including the fusion of techniques introduced in pipeline 1 and 2 or selecting the final ensemble DL-based methods of pipeline 2. Therefore, further improvements are needed to be made in future studies.

One of this project's main learning points was understanding that the standalone application of CE-NBI imaging modality could not visualize the information provided by normal WLE and NBI imaging techniques. CE-NBI imaging could only represent magnified and enhanced visualization of a small region in the target lesion. At the same time, standard WLE and NBI modalities can provide an overall representation of the target tissue that highlights different critical characteristics of the lesion, such as shape, color, and size. Therefore, the value of the normal WLE and NBI in the clinical examination of laryngeal lesions could not be ignored, and the application of magnified and enhanced endoscopy techniques such as CE-NBI as well as the developed computer-based solutions in this project are added values to such procedures.

Another learning point of this project was achieved via close collaboration with our clinical team. Initially, we focused on developing computer-based methods to address a critical clinical need. Moreover, we learned to emphasize ways that are easy to understand by the clinical community and, more importantly, are feasible to be used and integrated into the current workflow of the clinical examination of laryngeal lesions.

And the last learning point refers to the concept of "sharing." Presenting and discussing our development results with the clinical and technical communities helped us to find the optimum development path during this project. It also encouraged us to share the CE-NBI data set we generated during the last years in a public repository.

The outcome of this thesis contributes to the research area of CAD systems that aim to improve the minimally invasive larynx lesion assessment procedure. This thesis proposes a series of steps to facilitate the integration of CE-NBI imaging into the routine clinical examination of laryngeal lesions. In the future vision, the application of magnified and enhanced endoscopic imaging to perform Optical Biopsy should be achieved not only with CE-NBI but also with other imaging

modalities, such as microscopic imaging that can provide more stable and high-resolution visualization of the target examined area in the larynx. Moreover, applying magnified and enhanced endoscopic images in a CAD system for Optical Biopsy should not be limited to clinical settings with a fully equipped operating room. Instead, the solutions should be feasible to implement in every clinical examination setup to support clinicians in making more informed decisions regarding patient care and improving clinical outcomes.

Chapter 4 – References

- [1] W. Flynn and P. Vickerton, “Anatomy, Head and Neck, Larynx Cartilage,” 2020.
- [2] N. Gale, M. Poljak, and N. Zidar, “Update from the 4th edition of the World Health Organization classification of head and neck tumours: what is new in the 2017 WHO blue book for tumours of the hypopharynx, larynx, trachea and parapharyngeal space,” *Head and neck pathology*, vol. 11, pp. 23–32, 2017.
- [3] D. E. Johnson, B. Burtness, C. R. Leemans, V. W. Y. Lui, J. E. Bauman, and J. R. Grandis, “Head and neck squamous cell carcinoma,” *Nature reviews Disease primers*, vol. 6, no. 1, pp. 1–22, 2020.
- [4] F. Bootz, “S3-Leitlinie Diagnostik, Therapie und Nachsorge des Larynxkarzinoms,” *Der Radiologe*, vol. 60, no. 11, pp. 1052–1057, 2020.
- [5] N. Gale, J. Hille, R. C. Jordan, A. Nadal, and M. D. Williams, “Regarding Laryngeal precursor lesions: Interrater and intrarater reliability of histopathological assessment,” *The Laryngoscope*, vol. 129, no. 3, E91-E92, 2019.
- [6] P. K. Doloi and S. Khanna, “A study of management of benign lesions of the larynx,” *International Journal of Phonosurgery & Laryngology*, vol. 1, no. 2, pp. 61–64, 2011.
- [7] R. Nocini, G. Molteni, C. Mattiuzzi, and G. Lippi, “Updates on larynx cancer epidemiology,” *Chinese Journal of Cancer Research*, vol. 32, no. 1, p. 18, 2020.
- [8] A. Aupérin, “Epidemiology of head and neck cancers: an update,” *Current opinion in oncology*, vol. 32, no. 3, pp. 178–186, 2020.
- [9] C. S. Mehlum, S. R. Larsen, K. Kiss, A. M. Groentved, T. Kjaergaard, S. Möller, and C. Godballe., “Laryngeal precursor lesions: Interrater and intrarater reliability of histopathological assessment,” *The Laryngoscope*, vol. 128, no. 10, pp. 2375–2379, 2018.
- [10] P. L. Sun and H. W. Gao, “Pathological diagnosis and progression of laryngeal precursor lesions and squamous cell carcinoma,” *Zhonghua bing li xue za zhi= Chinese journal of pathology*, vol. 50, no. 11, pp. 1311–1315, 2021.
- [11] C. E. Steuer, M. El-Deiry, J. R. Parks, K. A. Higgins, and N. F. Saba, “An update on larynx cancer,” *CA: a cancer journal for clinicians*, vol. 67, no. 1, pp. 31–50, 2017.
- [12] N. Daneshi, M. Fararouei, M. Mohammadianpanah, M. Zare-Bandamiri, S. Parvin, and M. Dianatinasab, “Effects of different treatment strategies and tumor stage on survival of patients with advanced laryngeal carcinoma: a 15-year cohort study,” *Journal of Cancer Epidemiology*, vol. 2018, 2018.

- [13] I. Hermanns, R. Ziadat, P. Schlattmann, and O. Guntinas-Lichius, "Trends in Treatment of Head and Neck Cancer in Germany: A Diagnosis-Related-Groups-Based Nationwide Analysis, 2005-2018," *Cancers*, vol. 13, no. 23, p. 6060, 2021.
- [14] A. T. Utami, P. Aditya, N. A. Aroeman, and Y. A. Dewi, "Laryngeal cancer treatment: An update review," *High Technology Letters*, vol. 27, no. 11, pp. 330–339, 2021.
- [15] C. Hrelec, "Management of Laryngeal Dysplasia and Early Invasive Cancer," *Current treatment options in oncology*, vol. 22, no. 10, pp. 1–11, 2021.
- [16] M. Tusaliu, I. Tita, D. Tuas, R. Ranete, and C. Goanta, "Total laryngectomy with voice prosthesis-,Gold standard "treatment for patients with advanced laryngeal cancer," *Archives of the Balkan Medical Union*, vol. 55, no. 3, pp. 527–531, 2020.
- [17] X. Mimica, M. Hanson, S. G. Patel, M. McGill, S. McBride, N. Lee, L. A. Dunn, J. R. Cracchiolo, J. P. Shah, R. J. Wong, and I. Ganly., "Salvage surgery for recurrent larynx cancer," *Head & neck*, vol. 41, no. 11, pp. 3906–3915, 2019.
- [18] F. J. Wippold and H. S. Glazer, "Diagnostic imaging of the larynx," *Cummings C. Otorhinolaryngology Head and Neck Surgery 4th ed. Philadelphia: Elsevier Mosby*, 2005.
- [19] D. D. Deliyiski and R. E. Hillman, "State of the art laryngeal imaging: research and clinical implications," *Current opinion in otolaryngology & head and neck surgery*, vol. 18, no. 3, p. 147, 2010.
- [20] A. Bozzato, L. Pillong, B. Schick, and M. M. Lell, "Aktuelle Bildgebung bei Diagnostik und Therapieplanung des Larynxkarzinoms," *Der Radiologe*, vol. 60, no. 11, pp. 1026–1037, 2020.
- [21] A. Boese *et al.*, "Endoscopic Imaging Technology Today," *Diagnostics*, vol. 12, no. 5, p. 1262, 2022.
- [22] Z. He, P. Wang, Y. Liang, Z. Fu, and X. Ye, "Clinically available optical imaging technologies in endoscopic lesion detection: current status and future perspective," *Journal of healthcare engineering*, vol. 2021, 2021.
- [23] K. G. Tulaci, E. Arslan, T. Tulaci, and H. Yazici, "Which one is favorable in the elderly? Transoral rigid laryngoscopy or transnasal flexible fiberoptic laryngoscopy," *American Journal of Otolaryngology*, vol. 41, no. 6, p. 102660, 2020.
- [24] N. H. Hassan, R. Usman, M. Yousuf, A. N. Ahmad, and I. Hirani, "Transoral flexible laryngoscope biopsy: Safety and accuracy," *World journal of otorhinolaryngology-head and neck surgery*, vol. 5, no. 1, pp. 30–33, 2019.
- [25] H. W. Schutte *et al.*, "Digital video laryngoscopy and flexible endoscopic biopsies as an alternative diagnostic workup in laryngopharyngeal Cancer: a prospective clinical study," *Annals of Otolaryngology, Rhinology & Laryngology*, vol. 127, no. 11, pp. 770–776, 2018.

- [26] N. Davaris, S. Voigt-Zimmermann, S. Kropf, and C. Arens, “Flexible transnasal endoscopy with white light or narrow band imaging for the diagnosis of laryngeal malignancy: diagnostic value, observer variability and influence of previous laryngeal surgery,” *European Archives of Oto-Rhino-Laryngology*, vol. 276, no. 2, pp. 459–466, 2019.
- [27] C. Sun, X. Han, X. Li, Y. Zhang, and X. Du, “Diagnostic performance of narrow band imaging for laryngeal cancer: a systematic review and meta-analysis,” *Otolaryngology-Head and Neck Surgery*, vol. 156, no. 4, pp. 589–597, 2017.
- [28] B. Popek, K. Bojanowska-Poźniak, B. Tomasik, W. Fendler, J. Jeruzal-Świątecka, and W. Pietruszewska, “Clinical experience of narrow band imaging (NBI) usage in diagnosis of laryngeal lesions,” *Otolaryngologia Polska*, vol. 73, pp. 18–23, 2019.
- [29] C. Piazza, D. Cocco, L. de Benedetto, F. Del Bon, P. Nicolai, and G. Peretti, “Narrow band imaging and high definition television in the assessment of laryngeal cancer: a prospective study on 279 patients,” *European Archives of Oto-Rhino-Laryngology*, vol. 267, no. 3, pp. 409–414, 2010.
- [30] M. Dobre, M. Poenaru, N. C. Balica, and C. I. Doros, “Detection of early laryngeal cancer and its precursor lesions by a real-time autofluorescence imaging system,” *Rom J Morphol Embryol*, vol. 55, no. 4, pp. 1377–1381, 2014.
- [31] A. O. H. Gerstner *et al.*, “Hyperspectral imaging of mucosal surfaces in patients,” *Journal of biophotonics*, vol. 5, no. 3, pp. 255–262, 2012.
- [32] L. Staníková, R. Walderová, D. Jančatová, M. Formánek, K. Zeleník, and P. Komínek., “Comparison of narrow band imaging and the Storz Professional Image Enhancement System for detection of laryngeal and hypopharyngeal pathologies,” *European Archives of Oto-Rhino-Laryngology*, vol. 275, no. 7, pp. 1819–1825, 2018.
- [33] Y. C. Lee, Y.-G. Eun, and I.-S. Park, “The Value of I-Scan Image-Enhanced Endoscopy in the Diagnosis of Vocal Cord Leukoplakia,” *Journal of The Korean Society of Laryngology, Phoniatics and Logopedics*, vol. 29, no. 2, pp. 98–102, 2018.
- [34] B. Regeling *et al.*, “Hyperspectral imaging using flexible endoscopy for laryngeal cancer detection,” *Sensors*, vol. 16, no. 8, p. 1288, 2016.
- [35] C. Arens *et al.*, “Proposal for a descriptive guideline of vascular changes in lesions of the vocal folds by the committee on endoscopic laryngeal imaging of the European Laryngological Society,” *European Archives of Oto-Rhino-Laryngology*, vol. 273, no. 5, pp. 1207–1214, 2016.
- [36] X. G. Ni *et al.*, “Endoscopic diagnosis of laryngeal cancer and precancerous lesions by narrow band imaging,” *The Journal of Laryngology & Otology*, vol. 125, no. 3, pp. 288–296, 2011.

- [37] C. Lin, S. Zhang, L. Lu, M. Wang, and X. Qian, “Diagnostic value and pathological correlation of narrow band imaging classification in laryngeal lesions,” *Ear, Nose & Throat Journal*, vol. 100, no. 10, pp. 737–741, 2021.
- [38] V. M. Joshi, V. Wadhwa, and S. K. Mukherji, “Imaging in laryngeal cancers,” *Indian journal of radiology and imaging*, vol. 22, no. 03, pp. 209–226, 2012.
- [39] B. Jaipuria, D. Dosemane, P. M. Kamath, S. S. Sreedharan, and V. S. Shenoy, “Staging of laryngeal and hypopharyngeal cancer: Computed tomography versus histopathology,” *Iranian Journal of Otorhinolaryngology*, vol. 30, no. 99, p. 189, 2018.
- [40] S. L. van Egmond, I. Stegeman, F. A. Pameijer, J. J. Bluemink, C. H. Terhaard, and L. M. Janssen, “Systematic review of the diagnostic value of magnetic resonance imaging for early glottic carcinoma,” *Laryngoscope Investigative Otolaryngology*, vol. 3, no. 1, pp. 49–55, 2018.
- [41] H. Irjala, N. Matar, M. Remacle, and L. Georges, “Pharyngo-laryngeal examination with the narrow band imaging technology: early experience,” *European Archives of Oto-Rhino-Laryngology*, vol. 268, no. 6, pp. 801–806, 2011.
- [42] M. Żurek, A. Rzepakowska, E. Osuch-Wójcikiewicz, and K. Niemczyk, “Learning curve for endoscopic evaluation of vocal folds lesions with narrow band imaging,” *Brazilian Journal of Otorhinolaryngology*, vol. 85, pp. 753–759, 2019.
- [43] F. Campo, V. D’Aguanno, A. Greco, M. Ralli, and M. de Vincentiis, “The Prognostic Value of Adding Narrow-Band Imaging in Transoral Laser Microsurgery for Early Glottic Cancer: A Review,” *Lasers in Surgery and Medicine*, vol. 52, no. 4, pp. 301–306, 2020.
- [44] D. H. Kim, Y. Kim, S. W. Kim, and S. H. Hwang, “Use of narrowband imaging for the diagnosis and screening of laryngeal cancer: A systematic review and meta-analysis,” *Head & neck*, vol. 42, no. 9, pp. 2635–2643, 2020.
- [45] A. de Vito, G. Meccariello, and C. Vicini, “Narrow band imaging as screening test for early detection of laryngeal cancer: a prospective study,” *Clinical otolaryngology*, vol. 42, no. 2, pp. 347–353, 2017.
- [46] H. I. Turkmen, M. E. Karşligil, and I. Kocak, “Classification of laryngeal disorders based on shape and vascular defects of vocal folds,” *Computers in biology and medicine*, vol. 62, pp. 76–85, 2015.
- [47] P. Liang, Y. Cong, and M. Guan, “A computer-aided lesion diagnose method based on gastroscopimage,” in *2012 IEEE International Conference on Information and Automation*, 2012, pp. 871–875.
- [48] J. Yanase and E. Triantaphyllou, “A systematic survey of computer-aided diagnosis in medicine: Past and present developments,” *Expert Systems with Applications*, vol. 138, p. 112821, 2019.
- [49] H. Ali, M. Sharif, M. Yasmin, M. H. Rehmani, and F. Riaz, “A survey of feature extraction and fusion of deep learning for detection of abnormalities

- in video endoscopy of gastrointestinal-tract,” *Artificial Intelligence Review*, vol. 53, no. 4, pp. 2635–2707, 2020.
- [50] N. Obukhova, A. Motyko, A. Pozdeev, and B. Timofeev, “Review of noise reduction methods and estimation of their effectiveness for medical endoscopic images processing,” in *2018 22nd Conference of Open Innovations Association (FRUCT)*, 2018, pp. 204–210.
- [51] H. Fujita, “AI-based computer-aided diagnosis (AI-CAD): the latest review to read first,” *Radiological physics and technology*, vol. 13, no. 1, pp. 6–19, 2020.
- [52] C. S. Bang, J. J. Lee, and G. H. Baik, “Computer-aided diagnosis of esophageal cancer and neoplasms in endoscopic images: a systematic review and meta-analysis of diagnostic test accuracy,” *Gastrointestinal Endoscopy*, vol. 93, no. 5, pp. 1006–1015, 2021.
- [53] F. Renna *et al.*, “Artificial Intelligence for Upper Gastrointestinal Endoscopy: A Roadmap from Technology Development to Clinical Practice,” *Diagnostics*, vol. 12, no. 5, p. 1278, 2022.
- [54] A. Paderno, F. C. Holsinger, and C. Piazza, “Videomics: bringing deep learning to diagnostic endoscopy,” *Current opinion in otolaryngology & head and neck surgery*, vol. 29, no. 2, pp. 143–148, 2021.
- [55] A. Paderno *et al.*, “Artificial intelligence in clinical endoscopy: Insights in the field of videomics,” *Frontiers in Surgery*, vol. 9, 2022.
- [56] H. Turkmen and M. E. Karşligil, “Advanced computing solutions for analysis of laryngeal disorders,” *Medical & biological engineering & computing*, vol. 57, no. 11, pp. 2535–2552, 2019.
- [57] Y. Bensoussan, E. B. Vanstrum, M. M. Johns III, and A. Rameau, “Artificial Intelligence and Laryngeal Cancer: From Screening to Prognosis: A State of the Art Review,” *Otolaryngology-Head and Neck Surgery*, 01945998221110839, 2022.
- [58] M. Żurek, K. Jasak, K. Niemczyk, and A. Rzepakowska, “Artificial Intelligence in Laryngeal Endoscopy: Systematic Review and Meta-Analysis,” *Journal of Clinical Medicine*, vol. 11, no. 10, p. 2752, 2022.
- [59] C. Barbalata and L. S. Mattos, “Laryngeal tumor detection and classification in endoscopic video,” *IEEE journal of biomedical and health informatics*, vol. 20, no. 1, pp. 322–332, 2014.
- [60] H. I. Turkmen, M. E. Karşligil, and I. Kocak, “Classification of laryngeal disorders based on shape and vascular defects of vocal folds,” *Computers in biology and medicine*, vol. 62, pp. 76–85, 2015.
- [61] S. Moccia *et al.*, “Confident texture-based laryngeal tissue classification for early stage diagnosis support,” *Journal of Medical Imaging*, vol. 4, no. 3, p. 34502, 2017.

- [62] T. Araújo, C. P. Santos, E. de Momi, and S. Moccia, “Learned and handcrafted features for early-stage laryngeal SCC diagnosis,” *Medical & biological engineering & computing*, vol. 57, no. 12, pp. 2683–2692, 2019.
- [63] M.-H. Laves, J. Bicker, L. A. Kahrs, and T. Ortmaier, “A dataset of laryngeal endoscopic images with comparative study on convolution neural network-based semantic segmentation,” *International journal of computer assisted radiology and surgery*, vol. 14, no. 3, pp. 483–492, 2019.
- [64] H. Xiong *et al.*, “Computer-aided diagnosis of laryngeal cancer via deep learning based on laryngoscopic images,” *EBioMedicine*, vol. 48, pp. 92–99, 2019.
- [65] A. Inaba *et al.*, “Artificial intelligence system for detecting superficial laryngopharyngeal cancer with high efficiency of deep learning,” *Head & neck*, vol. 42, no. 9, pp. 2581–2592, 2020.
- [66] L. Nanni, S. Ghidoni, and S. Brahnham, “Ensemble of convolutional neural networks for bioimage classification,” *Applied Computing and Informatics*, 2020.
- [67] J. Ren *et al.*, “Automatic recognition of laryngoscopic images using a deep-learning technique,” *The Laryngoscope*, vol. 130, no. 11, E686-E693, 2020.
- [68] W. K. Cho *et al.*, “Diagnostic Accuracies of Laryngeal Diseases Using a Convolutional Neural Network-Based Image Classification System,” *The Laryngoscope*, vol. 131, no. 11, pp. 2558–2566, 2021.
- [69] Y. He *et al.*, “A deep convolutional neural network-based method for laryngeal squamous cell carcinoma diagnosis,” *Annals of Translational Medicine*, vol. 9, no. 24, 2021.
- [70] L. Yin, Y. Liu, M. Pei, J. Li, M. Wu, and Y. Jia, “Laryngoscope8: Laryngeal image dataset and classification of laryngeal disease based on attention mechanism,” *Pattern Recognition Letters*, vol. 150, pp. 207–213, 2021.
- [71] M. E. Dunham, K. A. Kong, A. J. McWhorter, and L. K. Adkins, “Optical biopsy: automated classification of airway endoscopic findings using a convolutional neural network,” *The Laryngoscope*, vol. 132, S1-S8, 2022.
- [72] M. A. Azam *et al.*, “Deep Learning Applied to White Light and Narrow Band Imaging Videolaryngoscopy: Toward Real-Time Laryngeal Cancer Detection,” *The Laryngoscope*, vol. 132, no. 9, pp. 1798–1806, 2022.
- [73] S. Wang, Y. Chen, S. Chen, Q. Zhong, and K. Zhang, “Hierarchical dynamic convolutional neural network for laryngeal disease classification,” *Scientific Reports*, vol. 12, no. 1, pp. 1–7, 2022.
- [74] D. A. van Dyk and X.-L. Meng, “The art of data augmentation,” *Journal of Computational and Graphical Statistics*, vol. 10, no. 1, pp. 1–50, 2001.
- [75] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
- [76] P. Refaeilzadeh, L. Tang, and H. Liu, “Cross-validation,” *Encyclopedia of database systems*, vol. 5, pp. 532–538, 2009.

- [77] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*, 1998, pp. 839–846.
- [78] A. M. Mendrik, E.-J. Vonken, A. Rutten, M. A. Viergever, and B. van Ginneken, "Noise reduction in computed tomography scans using 3-D anisotropic hybrid diffusion with continuous switch," *IEEE transactions on medical imaging*, vol. 28, no. 10, pp. 1585–1594, 2009.
- [79] T. M. Lehmann and C. Palm, "Color line search for illuminant estimation in real-world scenes," *JOSA A*, vol. 18, no. 11, pp. 2679–2691, 2001.
- [80] G. Yadav, S. Maheshwari, and A. Agarwal, "Contrast limited adaptive histogram equalization based enhancement for real time video system," in *2014 international conference on advances in computing, communications and informatics (ICACCI)*, 2014, pp. 2392–2397.
- [81] S. Gross *et al.*, "Polyp segmentation in NBI colonoscopy," in *Bildverarbeitung für die Medizin 2009*, Springer, 2009, pp. 252–256.
- [82] B. Zhang, L. Zhang, L. Zhang, and F. Karray, "Retinal vessel extraction by matched filter with first-order derivative of Gaussian," *Computers in biology and medicine*, vol. 40, no. 4, pp. 438–445, 2010.
- [83] S. Stefanou and A. A. Argyros, "Efficient scale and rotation invariant object detection based on hogs and evolutionary optimization techniques," in *International Symposium on Visual Computing*, 2012, pp. 220–229.
- [84] H. I. Turkmen, M. E. Karsligil, and I. Kocak, "Classification of vocal fold nodules and cysts based on vascular defects of vocal folds," in *2013 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, 2013, pp. 1–6.
- [85] A. Humeau-Heurtier, "Texture feature extraction methods: A survey," *Ieee Access*, vol. 7, pp. 8975–9000, 2019.
- [86] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.
- [87] C.-W. Hsu, C.-C. Chang, C.-J. Lin, and others, *A practical guide to support vector classification*. Taipei, Taiwan.
- [88] L. E. Peterson, "K-nearest neighbor," *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.
- [89] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [90] D. D. Lewis, "Naive (Bayes) at forty: The independence assumption in information retrieval," in *European conference on machine learning*, 1998, pp. 4–15.
- [91] S. Balakrishnama and A. Ganapathiraju, "Linear discriminant analysis-a brief tutorial," *Institute for Signal and information Processing*, vol. 18, no. 1998, pp. 1–8, 1998.

- [92] F. Murtagh, “Multilayer perceptrons for classification and regression,” *Neurocomputing*, vol. 2, 5-6, pp. 183–197, 1991.
- [93] A. Mathew, P. Amudha, and S. Sivakumari, “Deep learning techniques: an overview,” in *International conference on advanced machine learning technologies and applications*, 2021, pp. 599–608.
- [94] D. Shen, G. Wu, and H.-I. Suk, “Deep learning in medical image analysis,” *Annual review of biomedical engineering*, vol. 19, p. 221, 2017.
- [95] R. Andonie, “Hyperparameter optimization in learning systems,” *Journal of Membrane Computing*, vol. 1, no. 4, pp. 279–291, 2019.
- [96] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A survey of convolutional neural networks: analysis, applications, and prospects,” *IEEE transactions on neural networks and learning systems*, 2021.
- [97] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, 2009, pp. 248–255.
- [98] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [99] C. Szegedy *et al.*, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [100] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [101] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.
- [102] M. Tan and Q. Le, “Efficientnet: Rethinking model scaling for convolutional neural networks,” in *International conference on machine learning*, 2019, pp. 6105–6114.
- [103] S.-H. Ho *et al.*, “Development of image-enhanced endoscopy of the gastrointestinal tract: a review of history and current evidences,” *Journal of clinical gastroenterology*, vol. 52, no. 4, pp. 295–306, 2018.
- [104] M. Akarsu and C. Akarsu, “Evaluation of new technologies in gastrointestinal endoscopy,” *JSLS: Journal of the Society of Laparoendoscopic Surgeons*, vol. 22, no. 1, 2018.
- [105] M. J. Bruno, “Magnification endoscopy, high resolution endoscopy, and chromoscopy; towards a better optical diagnosis,” *Gut*, vol. 52, suppl 4, iv7-iv11, 2003.
- [106] M. Muto, T. Horimatsu, Y. Ezoe, S. Morita, and S. Miyamoto, “Improving visualization techniques by narrow band imaging and magnification endoscopy,” *Journal of gastroenterology and hepatology*, vol. 24, no. 8, pp. 1333–1346, 2009.

- [107] P. Lukes *et al.*, “The role of NBI HDTV magnifying endoscopy in the prehistologic diagnosis of laryngeal papillomatosis and spinocellular cancer,” *BioMed research international*, vol. 2014, 2014.
- [108] C. Piazza, F. D. Bon, G. Peretti, and P. Nicolai, “‘Biologic endoscopy’: optimization of upper aerodigestive tract cancer evaluation,” *Current opinion in otolaryngology & head and neck surgery*, vol. 19, no. 2, pp. 67–76, 2011.
- [109] A. Boeriu *et al.*, “Narrow-band imaging with magnifying endoscopy for the evaluation of gastrointestinal lesions,” *World journal of gastrointestinal endoscopy*, vol. 7, no. 2, p. 110, 2015.
- [110] Hosono, Hiroshi and Katada, Chikatoshi and Okamoto, Tabito and Ichinoe, Masaaki and Sakamoto, Yasutoshi and Matsuba, Hiroki and Kano, Koichi and Ishido, Kenji and Tanabe, Satoshi and Koizumi, Wasaburo and others, “Usefulness of narrow band imaging with magnifying endoscopy for the differential diagnosis of cancerous and noncancerous laryngeal lesions,” *Head & neck*, vol. 41, no. 8, pp. 2555–2560, 2019.
- [111] C. Piazza, F. Del Bon, G. Peretti, and P. Nicolai, “Narrow band imaging in endoscopic evaluation of the larynx,” *Current opinion in otolaryngology & head and neck surgery*, vol. 20, no. 6, pp. 472–476, 2012.
- [112] A. Mishra *et al.*, “Contact Endoscopy-A promising tool for evaluation of laryngeal mucosal lesions,” *Journal of Laryngology and Voice*, vol. 2, no. 2, p. 53, 2012.
- [113] C. Szeto, B. Wehrli, F. Whelan, J. A. Franklin, J. Y. Nichols, and K. Fung., “Contact endoscopy as a novel technique in the detection and diagnosis of mucosal lesions in the head and neck: a brief review,” *Journal of oncology*, vol. 2011, 2011.
- [114] M. Andrea, O. Dias, and A. Santos, “Contact endoscopy during microlaryngeal surgery: a new technique for endoscopic examination of the larynx,” *Annals of Otolaryngology, Rhinology & Laryngology*, vol. 104, no. 5, pp. 333–339, 1995.
- [115] E. Carriero, J. Galli, G. Fadda, S. Di Girolamo, F. Ottaviani, and G. Paludetti, “Preliminary experiences with contact endoscopy of the larynx,” *European Archives of Oto-Rhino-Laryngology*, vol. 257, no. 2, pp. 68–71, 2000.
- [116] C. Arens, T. Dreyer, H. Glanz, and K. Malzahn, “Compact endoscopy of the larynx,” *Annals of Otolaryngology, Rhinology & Laryngology*, vol. 112, no. 2, pp. 113–119, 2003.
- [117] D. Cikojević, I. Glunčić, and V. Pešutić-Pisac, “Comparison of contact endoscopy and frozen section histopathology in the intra-operative diagnosis of laryngeal pathology,” *The Journal of Laryngology & Otolaryngology*, vol. 122, no. 8, pp. 836–839, 2008.

- [118] A. Warnecke *et al.*, “Contact endoscopy for the evaluation of the pharyngeal and laryngeal mucosa,” *The Laryngoscope*, vol. 120, no. 2, pp. 253–258, 2010.
- [119] W. Tarnawski, M. Frączek, M. Jeleń, T. Kręcicki, and M. Zalesska-Kręcicka, “The role of computer-assisted analysis in the evaluation of nuclear characteristics for the diagnosis of precancerous and cancerous lesions by contact laryngoscopy,” *Advances in Medical Sciences (De Gruyter Open)*, vol. 53, no. 2, 2008.
- [120] R. Puxeddu, S. Sionis, C. Gerosa, and F. Carta, “Enhanced contact endoscopy for the detection of neoangiogenesis in tumors of the larynx and hypopharynx,” *The Laryngoscope*, vol. 125, no. 7, pp. 1600–1606, 2015.
- [121] N. Davaris, A. Lux, N. Esmaili, A. Illanes, A. Boese, M. Friebe, and C. Arens., “Evaluation of vascular patterns using contact endoscopy and narrow-band imaging (CE-NBI) for the diagnosis of vocal fold malignancy,” *Cancers*, vol. 12, no. 1, p. 248, 2020.
- [122] N. Esmaili, N. Davaris, A. Boese, A. Illanes, M. Friebe, and C. Arens, *Contact Endoscopy – Narrow Band Imaging (CE-NBI) Data Set for Laryngeal Lesion Assessment*. Zenodo, doi: 10.5281/zenodo.6674034.
- [123] P. Kántor and P. Stanek, “Narrative Review of Classification Systems Describing Laryngeal Vascularity Using Advanced Endoscopic Imaging,” *Journal of Clinical Medicine*, vol. 12, no. 1, p. 10, 2023.
- [124] R. Puxeddu, F. Carta, C. Ferreli, N. Chuchueva, and C. Gerosa, *Enhanced contact endoscopy (ECE) in head and neck surgery*. Endo-Press Tuttlingen, Germany, 2018.
- [125] C. S. Mehlum, H. Døssing, N. Davaris, A. Giers, Å. M. Grøntved, T. Kjaergaard, S. Möller, C. Godballe, and C. Arens., “Interrater variation of vascular classifications used in enhanced laryngeal contact endoscopy,” *European Archives of Oto-Rhino-Laryngology*, vol. 277, no. 9, pp. 2485–2492, 2020.

Appendix A – Abstract of publications not discussed in the dissertation

Training of a Novel Artificial Intelligence Algorithm on the First Online Database of Laryngeal Vessels of the Vocal Folds using Contact Endoscopy and Narrow Band Imaging

Nikolaos Davaris, Nazila Esmaili, Alfredo Illanes, Axel Boese, Michael Friebe, Christoph Arens

Abstract

Introduction: Intraoperative images of vocal fold vessels using contact endoscopy and narrow band imaging (CE-NBI) have already been successfully used for endoscopic differentiation between benign and malignant vocal fold lesions and for training artificial intelligence (AI) algorithms. The first online database of such CE-NBI images was published in 2022 to promote cooperation between laryngological centers and the further development of AI-based approaches.

Material and methods: The online database contains 11,144 CE-NBI images from 210 patients with histologically proven benign and (pre)malignant vocal fold lesions. In the present study, 80% of these images were used for training and 20% for testing a novel AI-based (Convolutional Neural Network-CNN) approach to differentiate between benign and malignant laryngeal lesions. Finally, the sensitivity, specificity and accuracy of the method in the automated classification of the test images were calculated.

Results: The developed algorithm was trained with the CNN-based AI approach using 8,915 CE-NBI images from the online database. Applied to the 2,229 test images, a sensitivity of 82.2%, a specificity of 90.2% and an accuracy of 87.8% could be reached.

Conclusion: The results of the presented AI-based approach regarding the diagnostic quality of the method are comparable to previously published studies on the manual or automated evaluation of CE-NBI images. The online database is a valuable tool for the further development of AI algorithms in the diagnosis of vocal fold lesions.

Use of Artificial Intelligence (AI) for the Intraoperative Evaluation of Vocal Fold Leukoplakia

Nikolaos Davaris, Nazila Esmaili, Alfredo Illanes, Axel Boese, Michael Friebe, Christoph Arens

Abstract

Introduction: Assessing vocal fold leukoplakia can be challenging despite modern endoscopic methods. The characterization of the morphology of adjacent vocal fold vessels is of great importance but depends heavily on the clinical experience of the observer. Intraoperative contact endoscopy with Narrow Band Imaging (NBI-CE) enables optimized visualization of vascular changes while the data generated can well be used for an automated evaluation using Artificial Intelligence (AI) methods.

Methods: In the present study, the adjacent vessels of 40 vocal cord leukoplakias were recorded intraoperatively using NBI-CE. The generated data was evaluated using machine learning methods with the classification scenarios Support Vector Machine with Polynomial Kernel (SVM) and k-Nearest Neighbor (kNN). After the histology was obtained, the sensitivity, specificity and accuracy of both classifiers were calculated in the classification between benign and malignant findings.

Results: In total, 1998 contact endoscopy images were evaluated in 16 benign and 24 malignant leukoplakias. The vascular changes could be mathematically characterized by the algorithms as an increase in the disorder of the gradient vector and the level of curvature. The sensitivity, specificity and accuracy of the automated classification were 100%, 77.2% and 90.6% for the SVM and 100%, 79.8% and 91.7% for the kNN.

Conclusion: The use of methods of AI and machine learning allows an automated evaluation of the vascular changes in vocal cord leukoplakia. The algorithms used can support doctors in the clinical characterization of leukoplakia as potentially benign or malignant.

Testing of a Novel Approach for an Automated Classification of Compact Endoscopic Vascular Patterns in Laryngeal Lesions

Nikolaos Davaris, Nazila Esmaili, Alfredo Illanes, Axel Boese, Michael Friebe, Christoph Arens

Abstract

Introduction: The combination of contact endoscopy and narrow band imaging (compact endoscopy) is suitable for the examination of laryngeal vascular patterns and provides information on the dignity of the lesions. However, the evaluation of the shape of the vessel is partly subjective and dependent on experience. On the other hand, vascular changes can be characterized mathematically as an increase in the disorder of the gradient vector and the curvature of the blood vessels.

Methods: We have tested a novel approach to the automated classification of compact endoscopic vessel patterns using image and signal processing techniques. Videoendoscopic data from 22 patients were evaluated. First, the automated classification of the studied samples was tested in three groups: ordered patterns, disordered patterns, and patterns with a high degree of disorder. Furthermore, the allocation into four classification scenarios was tested on the basis of histological diagnoses.

Results: A total of 907 compact endoscopic images were evaluated. Of these, 40% were for training and 60% for testing. Sensitivity was 94%, specificity 97% and accuracy 94% for the automated classification of vascular patterns in one of the three groups. The classification according to histological diagnoses was achieved with a sensitivity of 85%, a specificity of 94% and an accuracy of 84%.

Conclusions: It was shown that the automated classification of compact endoscopic vascular patterns is feasible. The algorithm can assist physicians in clinical decisions and can be used in diagnostics and tumor follow-up of laryngeal dysplasia and carcinoma.

Vibro-Acoustic Sensing of Instrument Interactions as a Potential Source of Texture-related Information in Robotic Palpation

Thomas Sühn, Nazila Esmaeili, Sandeep Y Mattepu, Moritz Spiller, Axel Boese, Robin Urrutia, Victor Poblete, Christian Hansen, Christoph H Lohmann, Alfredo Illanes, Michael Friebe

Abstract

The direct tactile assessment of surface textures during palpation is an essential component of open surgery that is impeded in minimally invasive and robot-assisted surgery. When indirectly palpating with a surgical instrument, the structural vibrations from this interaction contain tactile information that can be extracted and analyzed. This study investigates the influence of the parameters contact angle α and velocity \vec{v} on the vibro-acoustic signals from this indirect palpation. A 7-DOF robotic arm, a standard surgical instrument, and a vibration measurement system were used to palpate three different materials with varying α and \vec{v} . The signals were processed based on continuous wavelet transformation. They showed material-specific signatures in the time–frequency domain that retained their general characteristic for varying α and \vec{v} . Energy-related and statistical features were extracted, and supervised classification was performed, where the testing data comprised only signals acquired with different palpation parameters than for training data. The classifiers support vector machine and k-Nearest Neighbors provided 99.67% and 96.00% accuracy for the differentiation of the materials. The results indicate the robustness of the features against variations in the palpation parameters. This is a prerequisite for an application in minimally invasive surgery but needs to be confirmed in realistic experiments with biological tissues.

Surgeons' Requirements for a Surgical Support System to Improve Laparoscopic Access

Moritz Spiller, Marcus Bruennel, Victoria Grosse, Thomas Sühn, Nazila Esmaeili, Jessica Stockheim, Salmal Tural, Roland Croner, Axel Boese, Michael Friebe, Alfredo Illanes

Abstract

Creating surgical access is a critical step in laparoscopic surgery. Surgeons have to insert a sharp instrument such as the Veress needle or a trocar into the patient's abdomen until the peritoneal cavity is reached. They solely rely on their experience and distorted tactile feedback in that process, leading to a complication rate as high as 14% of all cases. Recent studies have shown the feasibility of surgical support systems that provide intraoperative feedback regarding the insertion process to improve laparoscopic access outcomes. However, to date, the surgeons' requirements for such support systems remain unclear. This research article presents the results of an explorative study that aimed to acquire data about the information that helps surgeons improve laparoscopic access outcomes. The results indicate that feedback regarding the reaching of the peritoneal cavity is of significant importance and should be presented visually or acoustically. Finally, a solution should be straightforward and intuitive to use, should support or even improve the clinical workflow, but also cheap enough to facilitate its usage rate. While this study was tailored to laparoscopic access, its results also apply to other minimally invasive procedures.

Towards an Intraoperative Feedback System for Laparoscopic Access with the Veress Needle

Moritz Spiller, Nazila Esmaeili, Thomas Sühn, Axel Boese, Salmal Tural, Michael Friebe, Alfredo Illanes

Abstract

About 50 % of complications during laparoscopy occur when surgical access is created. The Veress needle and proposed technical alternatives do not provide reliable information to support the surgeons in guiding the needle, or the feedback is not clearly perceivable. Based on acoustic emissions, Surgical Audio Guidance (SURAG) proposes a non-invasive and efficient way to enhance the perception of guidance information through acoustic and visual feedback displayed in real-time. This article demonstrates that the developed feedback matches the information about tissue layer crossings provided by force measurements. This indicates that SURAG can provide an effective means to make laparoscopic access more precise and safe, especially in pediatric surgery, where space for placing the needle is minimal.

Surgical Audio Guidance SurAG: Extracting Non-invasively Meaningful Guidance Information during Minimally Invasive Procedures

Alfredo Illanes, Thomas Sühn, Nazila Esmaili, Iván Maldonado, Anna Schaufler, Chien-Hsi Chen, Axel Boese, Michael Friebe

Abstract

In this work we summarize applications of a novel approach for providing complementary information for guiding medical interventional devices (MID) and that have been recently published by our research team. This approach consist of using an audio sensor located in the proximal end of the MID in order to extract meaningful information concerning the interaction between the tip of the instrument and the tissue. The approach was successfully evaluated with different setups and MIDs.

Thyroid Ultrasound Texture Classification using Autoregressive Features in Conjunction with Machine Learning Approaches

Prabal Poudel, Alfredo Illanes, Elmer JG Ataide, Nazila Esmaeili, Sathish Balakrishnan, Michael Friebe

Abstract

The thyroid is one of the largest endocrine glands in the human body, which is involved in several body mechanisms like controlling protein synthesis, use of energy sources, and controlling the body's sensitivity to other hormones. Thyroid segmentation and volume reconstruction are hence essential to diagnose thyroid related diseases as most of these diseases involve a change in the shape and size of the thyroid over time. Classification of thyroid texture is the first step toward the segmentation of the thyroid. The classification of texture in thyroid Ultrasound (US) images is not an easy task as it suffers from low image contrast, presence of speckle noise, and non-homogeneous texture distribution inside the thyroid region. Hence, a robust algorithmic approach is required to accurately classify thyroid texture. In this paper, we propose three machine learning based approaches: Support Vector Machine; Artificial Neural Network; and Random Forest Classifier to classify thyroid texture. The computation of features for training these classifiers is based on a novel approach recently proposed by our team, where autoregressive modeling was applied on a signal version of the 2D thyroid US images to compute 30 spectral energy-based features for classifying the thyroid and non-thyroid textures. Our approach differs from the methods proposed in the literature as they use image-based features to characterize thyroid tissues. We obtained an accuracy of around 90% with all the three methods.

Parametrical Modelling for Texture Characterization—A Novel Approach Applied to Ultrasound Thyroid Segmentation

Alfredo Illanes, Nazila Esmaeili, Prabal Poudel, Sathish Balakrishnan, Michael Friebe

Abstract

Texture analysis is an important topic in Ultrasound (US) image analysis for structure segmentation and tissue classification. In this work a novel approach for US image texture feature extraction is presented. It is mainly based on parametrical modelling of a signal version of the US image in order to process it as data resulting from a dynamical process. Because of the predictive characteristics of such a model representation, good estimations of texture features can be obtained with less data than generally used methods require, allowing higher robustness to low Signal-to-Noise ratio and a more localized US image analysis. The usability of the proposed approach was demonstrated by extracting texture features for segmenting the thyroid in US images. The obtained results showed that features corresponding to energy ratios between different modelled texture frequency bands allowed to clearly distinguish between thyroid and non-thyroid texture. A simple k-means clustering algorithm has been used for separating US image patches as belonging to thyroid or not. Segmentation of thyroid was performed in two different datasets obtaining Dice coefficients over 85%.

Electrochemotherapy: A Review of Current Status, Alternative IGP Approaches, and Future Perspectives

Nazila Esmaeili, Michael Friebe

Abstract

The efficiency of electroporation (EP) has made it a widely used therapeutic procedure to transfer cell killing substances effectively to the target site. A lot of researches are being done on EP-based cancer treatment techniques. Electrochemotherapy (ECT) is the first EP-based application in the field of drug administration. ECT is a local and nonthermal treatment of cancer that combines the use of a medical device with pharmaceutical agents to obtain local tumor control in solid cancers. It involves the application of eight, 100 μ s, pulses at 1 or 5000 Hz frequency and specified electric field (V/cm) with a median duration of 25 minutes. The efficacy of chemotherapeutic drugs increases by applying short and intense electrical pulses. Several clinical studies proposed ECT as a safe and complementary curative or palliative treatment option (curative intent of 50% to 63% in the treatment of Basal Cell Carcinoma (BCC)) to treat a number of solid tumors and skin malignancies, which are not suitable for conventional treatments. It is used currently for treatment of cutaneous and subcutaneous lesions, without consideration of their histology. On the contrary, it is also becoming a practical method for treatment of internal, deep-seated tumors and tissues. A review of this method, needed instruments, alternative image-guided procedures (IGP) approaches, and future perspectives and recommendations are discussed in this paper.

Appendix B – Comparison of related works

Table B.1. Comparison of related works in the context of laryngeal endoscopic image classification.

| Study | Type of data | Number of data | Approach | Classification Tasks | Accuracy | Sensitivity | Specificity | Recall | AUC |
|-------|--------------------------------------|-------------------------------|---------------------------------------|--|----------|-------------|-------------|--------|-----|
| [59] | NBI (Extracted from video) | 60 patients 120 images | Handcrafted features + ML classifier | Lesion detection (Healthy vs. malignant cases) | 84.30% | --- | --- | --- | --- |
| [60] | WL (Extracted from video) | 70 patients 124 images | Handcrafted features + ML classifiers | Disorder detection (Healthy vs. non-healthy cases) | --- | 0.81 | --- | --- | --- |
| [61] | NBI (Extracted from video) | 33 patients 330 images | Handcrafted features + ML classifiers | Cancer diagnosis (SCC cases with four tissue classes) | --- | --- | --- | 93% | --- |
| [71] | WL (Still & extracted from video) | 19353 images | DL model, Pre-trained VGG16 | Benign cases with five classes | 80.80% | --- | --- | --- | --- |
| [71] | WL (Still & extracted from video) | 19353 images | DL model, Pre-trained VGG16 | Benign vs. malignant cases | 93% | --- | --- | --- | --- |
| [64] | WL (Still image) | 1816 patients 13721 images | DL model, Pre-trained Inception V3 | Benign & normal vs. precancerous & malignant cases | 0.867 | 0.731 | 0.922 | --- | --- |
| [64] | WL (Still image) | 1816 patients 13721 images | DL model, Pre-trained Inception V3 | Benign, normal, precancerous, and malignant cases | 0.745 | --- | --- | --- | --- |

| | | | | | | | | | |
|------------------------|----------------------------------|-------------------------------|---|---|--------|--------|--------|-------|-------|
| [69] | NBI (Still image) | 4591 patients 4591 images | DL model, Pre-trained InceptionV3 | SCC diagnosis (Benign vs. SCC cases) | --- | --- | --- | --- | 0.873 |
| [65] | NBI (Still image) | 374 patients 2400 images | DL model, Pre-trained ResNet50 | SCC diagnosis (Normal vs. SCC cases) | 97.25% | 95.50% | 98.38% | --- | --- |
| [67] | WL (Still image) | 9231 patients 24667 images | DL model, Pre-trained ResNet101 | Laryngeal neoplasms (Benign, precancerous lesions, and cancer cases) | 96.24% | --- | --- | --- | --- |
| [73] | WL (Still images) | 1950 patients 3057 images | DL model, Pre-trained ResNet18 and ResNet34 | Laryngeal disease (Seven pathological plus healthy cases) | 75.27% | --- | --- | --- | 0.91 |
| [68] | WL (Extracted from video) | 4106 patients 4106 images | DL model, Pre-trained EfficientNetB0 | Laryngeal disease (Eight pathological plus healthy cases) | 0.88 | --- | --- | --- | --- |
| [62] | NBI (Extracted from video) | 33 patients 330 images | Handcrafted & learned features + ML classifier | Cancer diagnosis (SCC cases with four tissue classes) | 98% | --- | --- | --- | --- |
| [66] | NBI (Extracted from video) | 33 patients 330 images | Handcrafted & learned features + ML classifier | Cancer diagnosis (SCC cases with four tissue classes) | 95.45% | --- | --- | --- | --- |
| Pipeline 1 Method 1 | CE-NBI | 32 patients 1485 images | GF + ML classifier | Lesion assessment (Vascular patterns) | 0.973 | 0.980 | 0.983 | 0.977 | --- |
| Pipeline 1 Method 1 | CE-NBI | 20 patients 890 images | GF + ML classifier | Lesion assessment (Benign cases) | 0.906 | 0.900 | 0.965 | 0.981 | --- |

| | | | | | | | | | |
|--------------------------------------|--------|------------------------------|--|--|-------|-------|-------|-------|-----|
| Pipeline 1 Method 1 | CE-NBI | 11 patients 465 images | GF + ML classifier | Lesion assessment (Malignant cases) | 0.884 | 0.879 | 0.943 | 0.973 | --- |
| Pipeline 1 Method 1 | CE-NBI | 31 patients 1355 images | GF + ML classifier | Lesion assessment (Bening vs. malignant cases) | 0.912 | 0.871 | 0.933 | 0.953 | --- |
| Pipeline 1 Method 1 | CE-NBI | 68 patients 1632 images | GF + ML classifier | Lesion assessment (Bening vs. malignant cases) | --- | 0.846 | 0.895 | --- | --- |
| Pipeline 1 Method 2 | CE-NBI | 48 patients 2701 images | CyEff + ML Classifier | Lesion assessment (Bening vs. malignant cases) | 0.875 | 0.826 | 0.920 | --- | --- |
| Pipeline 1 Method 1 + Method 2 | CE-NBI | 48 patients 2701 images | CyEff & GF + ML Classifier | Lesion assessment (Bening vs. malignant cases) | 0.966 | 0.959 | 0.973 | --- | --- |
| Pipeline 2 Method 3 | CE-NBI | 146 patients 8181 images | DL model, Pre-trained ResNet50 | Lesion assessment (Bening vs. malignant cases) | 0.835 | --- | --- | --- | --- |
| Pipeline 2 Method 4 | CE-NBI | 210 patients 11144 images | DL model, Pre-trained DenseNet121 | Lesion assessment (Bening vs. malignant cases) | 0.876 | 0.820 | 0.901 | --- | --- |
| Pipeline 2 Method 4 | CE-NBI | 210 patients 11144 images | DL model, Pre-trained EfficientNetB0V2 | Lesion assessment (Bening vs. malignant cases) | 0.878 | 0.822 | 0.902 | --- | --- |
| Pipeline 2 Method 4 | CE-NBI | 210 patients 11144 images | DL model, Pre-trained ResNet50V2 | Lesion assessment (Bening vs. malignant cases) | 0.907 | 0.846 | 0.934 | --- | --- |

| | | | | | | | | | |
|------------------------|--------|------------------------------|---|--|-------|-------|-------|-----|-----|
| Pipeline 2 Method 4 | CE-NBI | 210 patients 11144 images | Ensemble model, Pre-trained DenseNet121, EfficientNetB0V2, and ResNet50V2 | Lesion assessment (Bening vs. malignant cases) | 0.928 | 0.877 | 0.951 | --- | --- |
|------------------------|--------|------------------------------|---|--|-------|-------|-------|-----|-----|

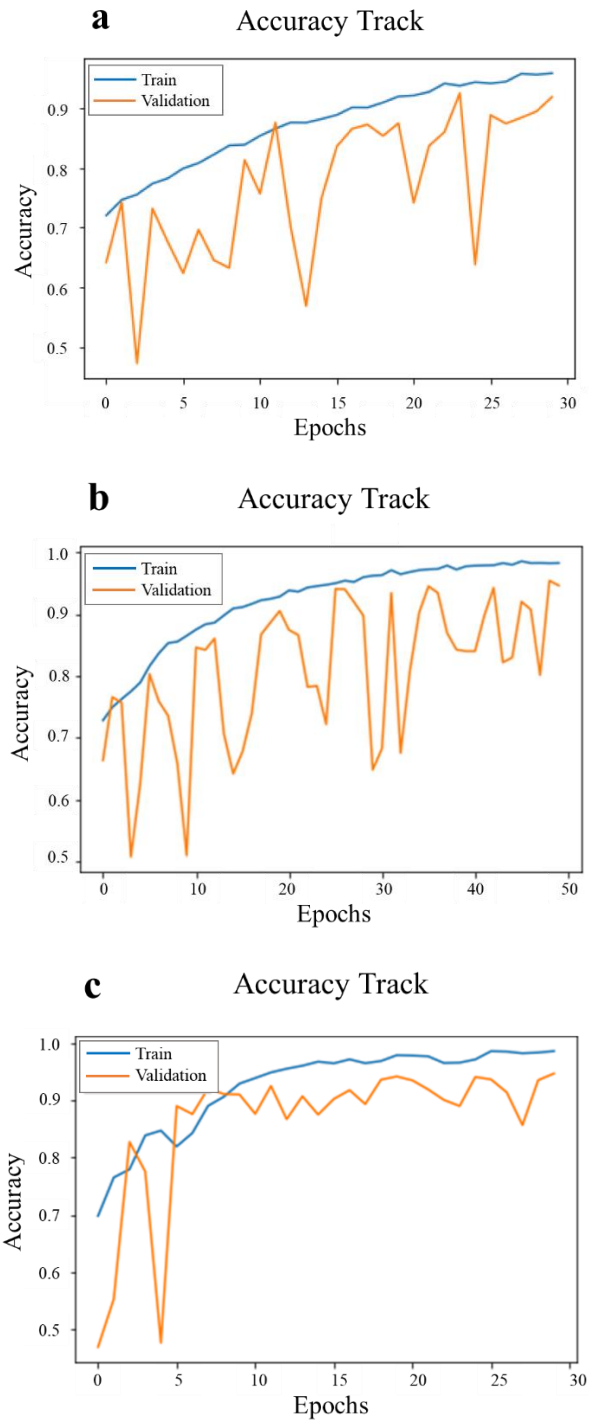


Figure B.1. Training and validation results of Method 4 on final CE-NBI image data set with 11144 images - accuracy graph of final models. (a): DenseNet121, (b): ResNet50V2, and (c): EfficientNetB0V2.

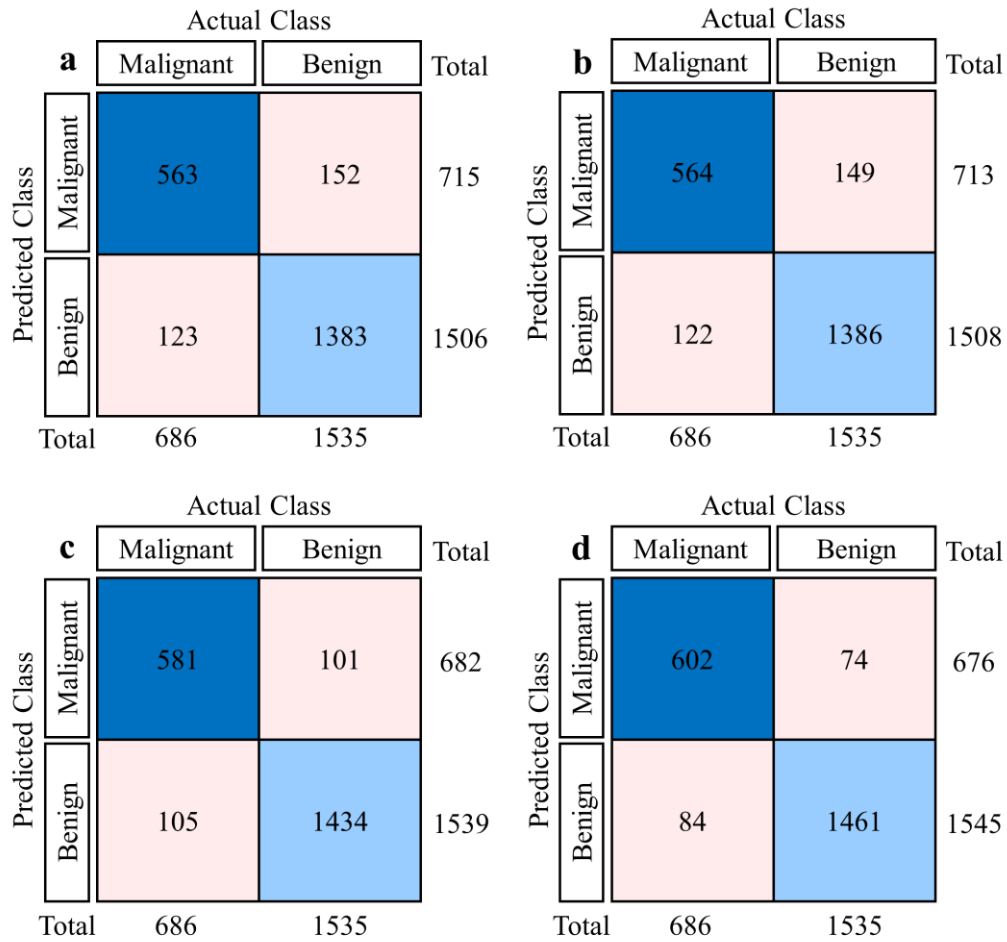


Figure B.2. Testing results of Method 4 on final CE-NBI image data set with 11144 images - confusion matrix. (a): DenseNet121, (b): EfficientNetB0V2, (c): ResNet50V2, and (d): Ensemble model.

