Design and implementation of a deep learning based CAD to empower cytology

Mauro Giuseppe Camporeale

Design and implementation of a Deep Learning based CAD to empower cytology

Thesis submitted for the degree of Philosophiae Doctor

Interuniversity Ph.D. Program in Industry 4.0
Università degli Studi di Bari “Aldo Moro”

Tutors
Prof. *Giovanni Dimauro*
Dr. *Rosalia Maglietta*

2025

Dissertation submitted for the degree of *Philosophiae Doctor*
Interuniversity Ph.D. Program in Industry 4.0

Cycle:
38th (2022, November 1 – 2025, October 31)

Administrative Headquarters: Politecnico di Bari – Department of Electrical and Information Engineering

Hosting University:
University of Bari – Department of Computer Science

Title: Design and implementation of a Deep Learning based CAD to empower cytology

Ph.D Candidate:
Mauro Giuseppe Camporeale

Academic Discipline:
Information Processing Systems (Italian code: ING-INF/05)

Tutors:
Prof. Giovanni Dimauro, University of Bari (Bari, Italy)
Dr. Rosalia Maglietta, CNR-STIIMA (Bari, Italy)

Coordinator:
Prof. Engr. Salvatore Digiesi, Politecnico di Bari (Bari, Italy)

External Reviewers:
Prof. Matteo Gelardi, Unit of Otolaryngology, Department of Clinical and Experimental Medicine, University of Foggia, Foggia, Italy

Prof. Dr. Ir. U. Kaymak, JADS, Eindhoven University of Technology, 'S Hertogenbosch, Netherlands

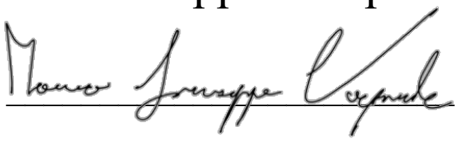
Last version:
12 23, 2025

Department of Electrical and Information Engineering
Industry 4.0 Ph.D. Program
SSD: ING-INF/05–INFORMATION PROCESSING SYSTEMS

Final Dissertation

Design and implementation of a Deep Learning based CAD to empower cytology

by
Mauro Giuseppe Camporeale:



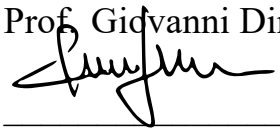
Referees:

Prof. Matteo Gelardi

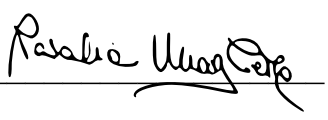
Prof. Uzay Kaymak

Supervisors:

Prof. Giovanni Dimauro



Prof. Rosalia Maglietta



Coordinator of Ph.D Program:

Prof. Salvatore Digiesi

Course n°38, 01/11/2022-31/10/2025

Abstract

This thesis presents a comprehensive investigation into the application of artificial intelligence and computer vision techniques to nasal cytology, with the goal of developing interpretable and clinically usable tools for diagnostic support. Nasal cytology is a valuable but underutilized method for assessing nasal mucosa conditions and inflammatory patterns. Despite its diagnostic potential, its manual nature, high inter-operator variability, and time-consuming analysis have limited its adoption in routine clinical practice. The work carried out in this thesis addresses these challenges by integrating medical expertise, machine learning methodologies, and software engineering principles into a unified, automated, and interpretable computational pipeline.

The research is structured around three main contributions. The first concerns the creation of a dedicated, annotated dataset for nasal cytology, one of the first publicly available resources in this field. The dataset provides a foundation for reproducible experiments and benchmarking of modern object detection models, establishing a common reference point for future studies. Building upon this resource, the second contribution introduces a novel interpretable deep learning pipeline for automatic cell detection and classification. Leveraging prototype-based reasoning, the system not only achieves strong predictive accuracy but also offers visual interpretability of its decisions, enabling clinicians to understand the model's reasoning process and validate its outputs in a transparent manner.

The third contribution extends the use of artificial intelligence beyond static cytological images to the functional analysis of ciliated cells through the estimation of ciliary beat frequency (CBF). The proposed method, implemented in the DeepCilia system, combines object detection and signal processing to provide fast and accurate CBF measurements from microscopy videos. This innovation enables reproducible and objective assessment of mucociliary clearance function, with potential implications in the diagnosis of disorders such as primary ciliary dyskinesia and chronic rhinosinusitis.

In addition to the methodological advances, the thesis demonstrates how these research components can be effectively integrated into a web-based clinical platform. The developed application provides physicians with an intuitive interface to manage patients, analyze cytological samples, visualize results, and store diagnostic data securely. The design emphasizes usability and data protection, employing server-side JWT-based authentication, HTTPS encryption, and a relational database architecture optimized for scalability and reliability.

Overall, the thesis contributes to the digital transformation of nasal cytology, providing a full pipeline that spans from dataset design to deployable software. The outcomes highlight how interpretable and reliable AI systems can bridge the gap between computational research and clinical practice, paving the way for the adoption of automated cytological analysis in real-world diagnostic workflows. Future developments include multi-center dataset expansion, clinical validation, and the extension of the platform to other cytological domains.

This work demonstrates that the convergence of deep learning, interpretability, and human-

centered design can produce practical, trustworthy, and scalable solutions, offering a model for how artificial intelligence can be responsibly integrated into medical diagnostics.

Declaration by the author

This thesis is composed of my original work, and contains no material previously published or written by another person except where due reference has been made in the text. I have clearly stated the contribution by others to jointly-authored works that I have included in my thesis.

I have clearly stated the contribution of others to my thesis as a whole, including statistical assistance, survey design, data analysis, significant technical procedures, professional editorial advice, financial support and any other original research work used or reported in my thesis. The content of my thesis is the result of work I have carried out since the commencement of my higher degree by research candidature.

Publications included in this thesis

- [1] **M. G. Camporeale**, U. Kaymak, G. Dimauro, Prototype Trees to Develop an Interpretable Automated Pipeline for Nasal Cytology, *2025 IEEE International Conference on Fuzzy Systems (FUZZ)*, 1-7, 2025

- [2] G. Dimauro, N. Barbaro, **M. G. Camporeale**, V. Fiore, M. Gelardi, M. Scalera, DeepCilia: Automated, deep-learning based engine for precise ciliary beat frequency estimation, *Biomedical Signal Processing and Control*, vol. 90, 105808, 2024

Submitted manuscripts included in this thesis

- [3] **M. G. Camporeale**, G. Dimauro, M. Gelardi, G. Iacobellis, M. S. Ladisa, S. Latrofa, N. Lomonte, A Nasal Cytology Dataset for Object Detection and Deep Learning, *arXiv preprint arXiv:2404.13745*, 2024

Other publications during candidature

- [4] G. Dimauro, D. Caivano, P. Di Pilato, A. Dipalma, **M. G. Camporeale**, A systematic mapping study on research in anemia assessment with non-invasive devices, *Applied Sciences*, vol. 10, n°14, 4804, 2020

- [5] G. Dimauro, M. E. Griseta, **M. G. Camporeale**, F. Clemente, A. Guarini, R. Maglietta, An intelligent non-invasive system for automated diagnosis of anemia exploiting a novel dataset, *Artificial Intelligence in Medicine*, vol. 136, 102477, 2023

- [6] G. Dimauro, **M. G. Camporeale**, A. Dipalma, A. Guarini, R. Maglietta, Anaemia detection based on sclera and blood vessel colour estimation, *Biomedical Signal Processing and Control*, vol. 81, 104489, 2023

- [7] **M. G. Camporeale**, L. Colizzi, N. Lomonte, A. Ragone, Development of a Desktop Application to Enable Doctors to Remotely Monitor Patients' Hematological Parameters, *International Conference on Product-Focused Software Process Improvement*, 48-59, 2023

- [8], L. Simone, **M. G. Camporeale**, N. Lomonte, G. Dimauro, V. Gervasi, An efficient deep learning approach for arrhythmia classification using 3D temporal SVCG, *2023 IEEE International Conference on Digital Health (ICDH)*, 234-239, 2023

- [9] V. S. Barletta, **M. G. Camporeale**, N. Lomonte, M. Scalera, E. Gentile, DeskEmo, a multi-platform desktop software for remote control and monitoring of hematological parameters, *Italian Forum of Ambient Assisted Living*, 224-232, 2023

- [10] **M. G. Camporeale**, F. Clemente, G. Dimauro, N. Lomonte, R. Maglietta, C. Pasciolla, D. Sacco, G. M. Zaccaria, Highly reliable personalized noninvasive hemoglobin estimation by using Vision Transformers and dual fine-tuning, *Computers in Biology and Medicine*, vol. 197, 111026, 2025
- [11] L. Simone, **M. G. Camporeale**, V. M. Rubino, V. Gervasi, G. Dimauro, Interpretable Early Detection of Parkinson's Disease Through Speech Analysis, *International Conference on Artificial Intelligence in Medicine*, 373-378, 2025

Contents

List of Figures

List of Tables

List of Abbreviations and Symbols

Technical Abbreviations	
AI	Artificial Intelligence
XAI	Explainable Artificial Intelligence
ML	Machine Learning
DL	Deep Learning
OD	Object Detection
ROI	Region of Interest
ANN	Artificial Neural Networks
CNN	Convolutional Neural Network
R-CNN	Region-based Convolutional Neural Network
DETR	DEtection TRansformer
YOLO	You Only Look Once
CT	Computed Tomography
CAD	Computer-Aided Diagnosis
k-NN	k-Nearest Neighbors
SVM	Support Vector Machines
ReLu	Rectified Linear Unit
SGD	Stochastic Gradient Descent
CAM	Class Activation Mapping
SMOTE	Synthetic Minority Oversampling Technique
NMCD	Nasal Mucosa Cell Dataset
BB	Bounding Box
AP	Average Precision
mAP	mean Average Precision
IoU	Intersection over Union
AUC	Area Under Curve
MBE	Mean Bias Error
RMSE	Root Mean Square Error
GT	Ground Truth
STFT	Short-Time Fourier Transform

FFT	Fast Fourier Transform
PITT	Pixel Intensity Trough-Time
FES	Forward Exponential Smoothing
BES	Backward Exponential Smoothing
SAVA	Sisson-Ammons Video Analysis
WFA	Whole-Field Analysis
OF	Optical Flow
FPS	Frame per Second
API	Application Programming Interface
SSL	Secure Sockets Layer
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
DB	Data Base
CLI	Command Line Interface
JWT	JSON Web Token
XSS	Cross-Site Scripting

Medical Abbreviations	
DNA	Deoxyribonucleic Acid
RNA	Ribonucleic Acid
ATP	Adenosine triphosphate
MGG	May-Grünwald-Giemsa
IgA	Immunoglobulin A
GM-CSF	Granulocyte-Macrophage Colony-Stimulating Factor
LTC	Leukotriene
PGD	Prostaglandin
CBF	Ciliary Beat Frequency
PCD	Primary Ciliary Dyskinesia

Generic Abbreviations	
FDA	Food and Drug Administration
EMA	European Medicine Agency
GDPR	General Data Privacy Regulation
HIPAA	Health Insurance Portability and Accountability Act

Chapter 1

Introduction

In recent years, artificial intelligence (AI) and, in particular, deep learning have profoundly transformed many domains of science, technology, and everyday life [12, 13]. Among the areas most impacted is medicine, where AI methods are increasingly being adopted to assist in diagnosis, treatment planning, and the development of personalized healthcare solutions [14]. Advances in computer vision and pattern recognition have enabled machines to achieve, and in some cases surpass, human-level performance in image classification tasks, opening the door to new applications in medical imaging. From radiology to pathology, the ability of deep learning models to analyze complex visual data has shown the potential to improve accuracy, reduce variability, and support clinicians in making more informed decisions [15].

Despite these advances, the integration of AI into clinical practice often faces two key challenges: the availability of well-curated datasets and the translation of research prototypes into tools that are usable in real-world contexts. Without sufficiently large and representative datasets, models risk being biased, underperforming, or failing to generalize across clinical scenarios. Furthermore, even when models demonstrate strong performance in experimental settings, they frequently remain confined to the research domain due to the lack of accessible interfaces and deployment pipelines. Bridging this gap between algorithm development and practical clinical utility is therefore an important step toward the effective adoption of AI in medicine [16].

Nasal cytology, a diagnostic discipline focused on the cellular analysis of the nasal mucosa, represents a domain where such AI-based support systems could be of great value. It is an established yet underutilized diagnostic method, providing insights into allergic, infectious, and inflammatory diseases of the upper respiratory tract [17]. However, cytological evaluation remains highly dependent on specialized expertise and manual interpretation, which can be time-consuming and subject to inter-observer variability. This makes it a promising candidate for digital innovation through artificial intelligence.

In this context, exploring how AI can enhance the diagnostic workflow in nasal cytology is both timely and necessary. By investigating how modern computational methods can contribute

to greater efficiency, consistency, and accessibility, research in this field has the potential to strengthen the role of cytology in clinical practice and extend its use to a broader range of medical settings. The work presented in this thesis is situated within this broader landscape of innovation, aiming to contribute to the ongoing dialogue between technological progress and clinical applicability.

1.1 Main Contribution

The adoption of artificial intelligence in medicine requires not only the design of accurate models but also careful attention to the availability of data, the interpretability of predictions, and the practicality of deployment. While deep learning has demonstrated remarkable results in many medical imaging domains, its application to nasal cytology remains limited, primarily due to the absence of dedicated datasets and the lack of end-to-end pipelines that extend from data collection to clinical usability. This thesis addresses these gaps by building a comprehensive workflow that moves beyond algorithm development to encompass the broader requirements of real-world adoption.

The main contributions of this work can be summarized as follows:

- **Creation of a dedicated dataset for nasal cytology.** A new image dataset was collected and annotated, covering diverse cellular types and pathological conditions, providing one of the first structured resources for applying deep learning to nasal cytology.
- **Development of deep learning model.** A Deep Learning (DL) architecture was fine-tuned on the nasal-cytology dataset, selecting the model between many state of the art detection models.
- **Interpretable model proposal.** A prototype-based approach was designed and implemented to classify nasal cytology images, combining strong predictive performance with visual explanations that make the decision process transparent to clinicians.
- **Deployment through a web application.** The trained model was integrated into a user-friendly web platform, enabling clinicians to upload and analyze cytological images, thus demonstrating how research can be translated into practical diagnostic support tools.

1.2 Thesis Outline

This thesis is organized into five chapters:

- **Chapter 1, Introduction:** introduces the motivation, objectives, and structure of the research, outlining the role of artificial intelligence in improving nasal cytology.

- **Chapter 2, Background:** provides the medical and technical background, describing the nasal mucosa, the rhinocytologic process, and the computational foundations of image analysis, machine learning, and deep learning.
- **Chapter 3, Main Contribution:** presents the main contributions through three studies: the development of interpretable AI models for cytology, the creation of a dedicated dataset and benchmarks, and the estimation of ciliary beat frequency using deep learning.
- **Chapter 4, WebApp Developing and Documentation**: describes the design and implementation of the web-based application that integrates all developed tools into a user-friendly clinical platform.
- **Chapter 5, Conclusion and Future Works**: finally concludes the thesis, summarizing the main results and discussing future research directions toward clinical validation and broader applicability.

Chapter 2

Background

The development of automated tools for cytological analysis requires a solid understanding of both the medical context in which these tools operate and the computational techniques that enable their implementation. For this reason, the background of this thesis is presented in two complementary parts. The first part is devoted to the medical background, providing an overview of the nasal mucosa, the different cellular populations that can be identified through rhinocytology, and the diagnostic relevance of cytological examination. This section establishes the biological foundation and clarifies why cytology is a valuable but also labor-intensive and error-prone method, which may benefit from computational support.

The second part presents the technical background, tracing the evolution of computer vision methods applied to medical imaging. It begins with the era of classical image processing, proceeds through the emergence of machine learning, and culminates in the rise of deep learning as the current state of the art. Particular attention is given to issues of interpretability, data availability, and the translational challenge of moving from research prototypes to clinically usable applications.

Taken together, these two perspectives provide the conceptual framework for the thesis. The medical background highlights the diagnostic importance and practical limitations of rhinocytology, while the technical background introduces the computational methods that can address these challenges. This dual foundation motivates the main contribution of the work, namely the design of interpretable, automated pipelines for cytological analysis supported by modern machine learning techniques.

2.1 Medical Background

Cytology, the study of cells, is a cornerstone of modern medicine, providing critical insights into both normal physiology and pathological processes. By analyzing cell morphology, structure, and composition, cytology allows clinicians to detect infections, inflammation, neoplastic changes, and other disease processes at an early stage. Traditionally applied in a variety of fields, from

oncology to hematology, cytology offers a minimally invasive yet highly informative window into tissue health and function.

Within the upper respiratory tract, the nasal mucosa represents a dynamic and complex interface between the external environment and the host immune system. The nasal epithelium is constantly exposed to allergens, pathogens, pollutants, and irritants, which can induce cellular and inflammatory responses. Nasal cytology, the microscopic examination of cells obtained from the nasal cavity, has emerged as an essential diagnostic tool for assessing these responses. By evaluating the presence, proportion, and morphology of epithelial cells, neutrophils, eosinophils, mast cells, and other inflammatory components, clinicians can gain a detailed understanding of the underlying mechanisms driving nasal diseases.

The clinical relevance of nasal cytology is particularly evident in conditions such as allergic rhinitis, chronic rhinosinusitis, vasomotor rhinitis, and other inflammatory or infectious disorders of the upper airway. In these contexts, cytological analysis not only supports differential diagnosis but also informs treatment decisions, disease monitoring, and prognosis. Beyond diagnostic applications, nasal cytology can provide valuable insights into pathophysiological processes, such as the recruitment and activation of inflammatory cells, tissue remodeling, and epithelial barrier dysfunction.

Despite its utility, nasal cytology presents certain challenges. Sample collection requires standardized techniques to ensure cellular integrity, and accurate interpretation demands both technical expertise and experience. Moreover, the process is often labor-intensive and subject to inter-observer variability, which can limit its widespread adoption and reproducibility in clinical practice. Nevertheless, advances in cytological methodology, staining techniques, and microscopic analysis continue to enhance the precision and applicability of nasal cytology as a diagnostic and research tool.

This section follows the exposition of the "Atlante di Citologia Nasale" (M. Gelardi, 2012) [18].

2.1.1 The Nasal Mucosa and its Cells

As discussed, nasal cytology provides a powerful window into the cellular composition and health of the nasal mucosa, offering insights into both normal physiology and pathological conditions of the upper respiratory tract. To fully appreciate the relevance of cytological analysis, it is essential to understand the anatomical and functional context in which these cells reside.

The nasal mucosa is not merely a passive lining of the nasal cavity; it is a highly specialized and dynamic tissue, constantly interacting with the external environment. Its structure and cellular composition are finely tuned to fulfill critical roles in air filtration, humidification, immune defense, and sensory perception. By examining the cells of the nasal mucosa, clinicians can detect subtle changes indicative of inflammation, infection, or allergic reactions, which in turn informs diagnosis and treatment.

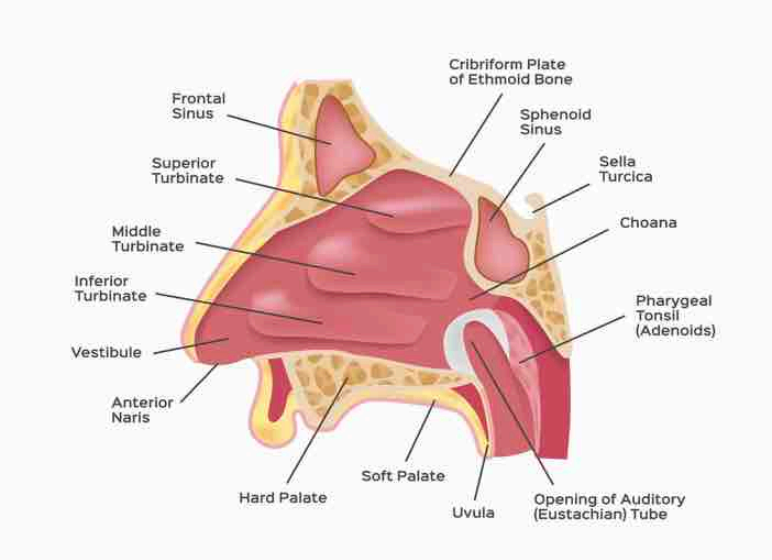


Figure 2.1: A cross section of the nasal cavity.

Nasal Mucosa

The nasal mucosa is a specialized lining of the nasal cavity that serves as the first point of contact between inspired air and the respiratory system. Far from being a simple protective barrier, it is a complex and dynamic tissue that integrates mechanical, immunological, and sensory functions. Its organization reflects the dual need to process large volumes of air efficiently while defending against environmental insults such as dust, microorganisms, and allergens.

Structurally, the nasal mucosa can be divided into two main layers (Figure 2.2):

- *Epithelium* - The epithelial layer covers the surface of the nasal cavity and serves as both a barrier and a functional interface with the environment. Its main roles are to protect underlying tissues, support the clearance of particles, and facilitate the initial interactions between inhaled substances and the local immune system. The epithelium rests on a basement membrane that provides structural support and separates it from the underlying connective tissue.

- *Lamina Propria* - Beneath the epithelium lies the lamina propria, a connective tissue layer that contains a rich network of blood vessels, nerves, and extracellular matrix components. This layer provides mechanical support, nourishes the epithelium, and contributes to the regulation of temperature and humidity of inspired air. It also houses glands and secretory structures that produce mucus and other protective secretions.

Physiological Functions of the Nasal Mucosa

The structural organization of these layers allows the nasal mucosa to perform its essential functions efficiently. The combination of surface epithelium and underlying connective tissue ensures that inhaled air is filtered, humidified, and warmed before reaching the lower respiratory

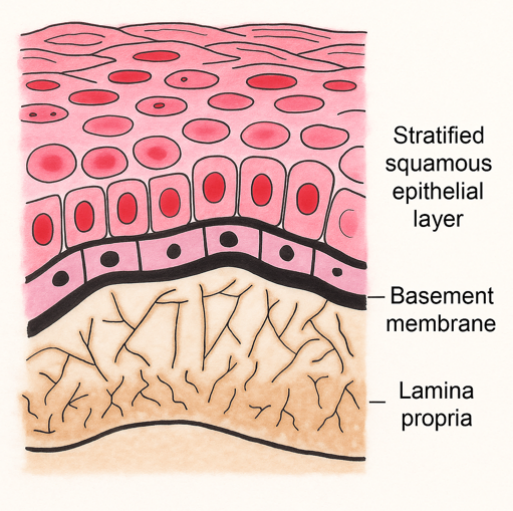


Figure 2.2: The Various layers of the nasal mucosa.

tract. Additionally, the vascularized lamina propria supports rapid immune responses and helps maintain tissue homeostasis despite continuous exposure to environmental challenges.

The nasal mucosa is a highly specialized tissue that plays a central role in maintaining the integrity of the respiratory system and ensuring the quality of inspired air. It performs multiple functions that are tightly integrated with its structural organization. One of its primary roles is the filtration and conditioning of air before it reaches the lower respiratory tract. Particulate matter, dust, and microorganisms are trapped at the mucosal surface, preventing their entry into the lungs. Simultaneously, the rich vascular network and tissue architecture help warm and humidify inspired air, optimizing it for effective gas exchange in the lower airways.

Beyond its mechanical functions, the nasal mucosa serves as a critical barrier and supports local immune defense. Its structure provides protection for deeper tissues against environmental insults, while also facilitating the detection and neutralization of potentially harmful agents. This combination of physical barrier and immunological activity ensures that the respiratory system can maintain health and respond to pathogens efficiently.

In addition to protective roles, the nasal mucosa contributes to sensory perception. It can detect chemical signals and physical stimuli from the environment, helping trigger reflexes such as sneezing or increased mucus secretion that protect the airways. In specialized regions, such as the superior nasal cavity, the mucosa supports olfaction, allowing the detection of odorants and contributing to the overall sensory experience of the environment.

The efficiency of the nasal mucosa arises from the close integration of its structural components with its physiological roles. The epithelial layer, basement membrane, and underlying connective tissue work together to filter, condition, and defend against inhaled particles, while vascular and secretory elements maintain tissue homeostasis and enable dynamic responses to environmental changes and injury. Understanding these structural and functional characteristics provides a necessary foundation for examining the cellular composition of the nasal mucosa in detail, which will be addressed in the following sections.

The Cell

Before addressing the specific cellular populations that can be identified within the nasal mucosa, it is essential to recall the fundamental organization of the cell itself. Every cell, regardless of its specialization, shares a common structural framework that includes a nucleus, a cytoplasm enriched with organelles, and a surrounding plasma membrane. This general architecture underlies the diverse functions performed by the cells of the respiratory epithelium and provides the basis for understanding the morphological and functional differences observed in cytological preparations. By outlining the universal features of cellular structure, we can more clearly appreciate the variations that distinguish the different cytotypes present in nasal cytology.

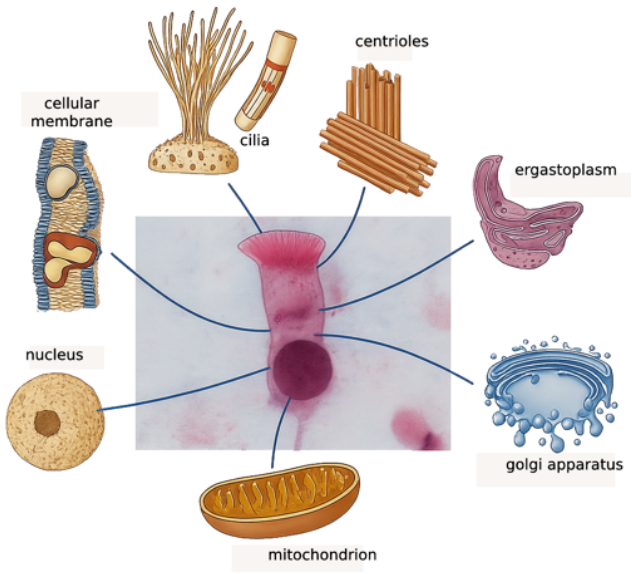


Figure 2.3: Cell organelles.

The nucleus

The nucleus contains the functional program of the cell. It governs the specificity of cellular synthesis, regulates metabolism, and directs the transmission of hereditary information. This is made possible by the presence of deoxyribonucleic acid (DNA), which constitutes the major component of the nucleus and determines cellular specificity.

A constant ratio is generally maintained between nuclear volume and cytoplasm. In the interphase resting stage, the nucleus is structurally composed of four fundamental elements: chromatin, nucleoplasm (or “nuclear sap”), the nuclear membrane, and the nucleoli.

Chromatin

The appearance of chromatin varies according to cell lineage and maturation stage. In younger cells, the nucleus is relatively large, with the nuclear-to-cytoplasmic ratio shifted in favor of the nucleus. In contrast, mature cells typically display condensed chromatin aggregated into clumps known as chromocenters. These aggregates are often located along the inner surface of

the nuclear envelope (membrane-associated chromatin). Conversely, other regions remain less condensed, reflecting areas of active metabolic transcription.

From a functional perspective, chromatin can be distinguished into two categories: euchromatin, the genetically active and metabolically dynamic fraction, and heterochromatin, the metabolically inert fraction. Euchromatin appears as light, dispersed areas within the nucleoplasm, conferring a more transparent and less optically dense aspect to the nucleus. Sex chromatin is detectable in females and usually manifests as a small chromatin mass with triangular or lenticular morphology. In segmented neutrophils, it is often observed as a drumstick-shaped nuclear appendage.

Nucleoplasm

The nucleoplasm, also referred to as the nuclear matrix, may be compared to the cytoplasmic matrix (cytoplasm). It appears optically clear, often containing fine granules or thin filamentous structures. It constitutes the medium in which chromatin and nucleoli are suspended. The nucleoplasm contains structural proteins, enzymes, metabolites, and ions, although it lacks independent metabolic activity. Its volume tends to decrease with cellular aging, leading to a reduction in the so-called "nuclear sap." Under pathological conditions, it may either accumulate excessively (nuclear edema) or, conversely, disappear almost completely (nuclear pyknosis).

Nuclear membrane

The nucleus is surrounded by a double-layered structure known as the nuclear membrane, which undergoes substantial variations according to the functional state of the cell. It is continuous with the lumen of the endoplasmic reticulum, thereby establishing a structural and functional continuity between these compartments.

The nuclear envelope contains circular openings of approximately 50 nm in diameter, known as nuclear pores. These structures mediate the bidirectional exchange of materials between the nucleus and cytoplasm. Under physiological conditions, pores are typically closed by a diaphragm-like structure. In certain states, the nucleoplasm may pass through these pores, whereas chromocenters are generally retained within the nucleus. In some cells, nuclear pores may occupy up to 10% of the nuclear membrane surface, while in others they are exceedingly sparse. Overall, their number tends to decrease progressively as the cell matures.

Nucleolus

The nucleolus is an intranuclear region where ribonucleic acid (RNA) is synthesized; for this reason, it is more prominent in metabolically active or reactive cells.A single nucleus may contain one or several nucleoli, and both their position and volume are highly variable. The characteristic intense blue staining (basophilia) of the nucleolus depends on the amount of RNA contained within it. In some cases, nucleoli may appear colorless; under such conditions, they are recognizable by the presence of a thicker rim of surrounding chromatin. Conversely, they

may sometimes be entirely obscured by nucleolus-associated chromocenters.

It is important to emphasize that the presence of a nucleolus—or even multiple nucleoli—does not, by itself, indicate malignancy. Instead, malignancy is suggested by the presence of enlarged nucleoli, irregularly shaped nucleoli, or multiple prominent nucleoli.

The Cytoplasm

The cytoplasm represents the internal environment of the cell. It is composed primarily of water, organic and inorganic molecules, and ions, which provide the essential substrates for the chemical reactions underlying cellular metabolism. Under the light microscope, the cytoplasm appears as a homogeneous area, while the cytoplasmic matrix is not morphologically distinguishable. In contrast, electron microscopy reveals the presence of organelles and the extremely fine structures that constitute it.

Within the cytoplasm, regions of clear cytoplasm devoid of granulations may be observed. Among these, the most relevant is the centrosome, a space containing organelles such as centrioles and the Golgi apparatus.

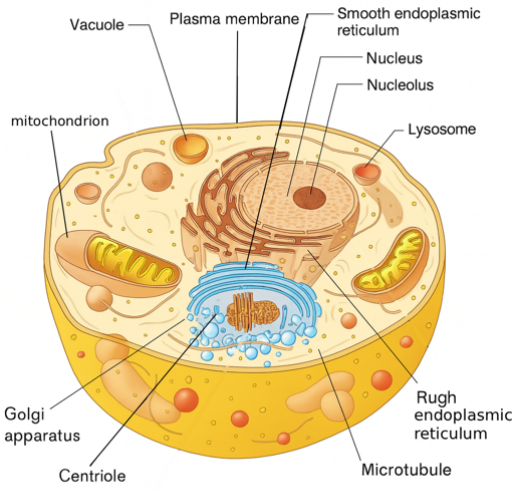


Figure 2.4: Schema of a cell.

Golgi apparatus

The Golgi apparatus is composed of several flattened cisternae, closely juxtaposed to one another and organized into a curved or semicircular structure. These cisternae form a hollow region from which numerous transport and secretory vesicles bud. Each Golgi stack, or dictyosome, typically contains between 2 and 8 cisternae. Functionally, the Golgi apparatus is central to the processing and packaging of proteins synthesized in the rough endoplasmic reticulum. At its periphery, numerous vesicles detach from the cisternae, carrying elaborated products such as immunoglobulins, specific secretory granules, and lysosomal enzymes. The Golgi apparatus

also contains numerous enzymes involved in carbohydrate and mucopolysaccharide metabolism, reflecting its critical role in glycosylation and secretion.

Centrioles

In mitotic cells, centrioles are always present. They appear as small cylinders about 150 µm long, with a variable length between 300 and 500 µm, arranged perpendicularly to each other. Their wall is made up of nine groups of triplet microtubules disposed in a circular fashion. Centrioles take part in the mechanisms that characterize mitosis and cell motility. At the end of cell division, they multiply by self-reproduction.

Microtubules

Microtubules are tubular structures with a diameter of about 200 nm and a wall thickness of around 4 nm. They are made up of 13 protofilaments, each with the same diameter, forming a hollow cylinder. Together they constitute the cytoskeleton, which gives the cell its shape.

During mitosis, microtubules form the spindle fibers. In interphase, they are concentrated around the nucleus, then spread throughout the cytoplasm. For a period after mitosis, once the cells have separated, they remain connected by a thin bundle of microtubules before the final separation.

Ribosomes

Ribosomes are particles about 140 Å in diameter, present in all cells (except red blood cells), and composed mainly of RNA and proteins. They consist of two subunits of different molecular weight: 40S and 60S. They are located in the nucleus, nucleolus, mitochondria, and especially in the cytoplasm, where their number determines the degree of basophilia of the cell. Ribosomes receive messenger RNA and transfer RNA and play an essential role in protein synthesis.

In the cytoplasm, ribosomes may be isolated (monoribosomes), in groups of three to ten elements, or in larger aggregates (polyribosomes), in which several ribosomes are connected by a single messenger RNA filament. Ribosomes may also be attached to the outer surface of the rough endoplasmic reticulum.

Endoplasmic reticulum

The endoplasmic reticulum is a system of interconnected tubules and flattened sacs, with or without ribosomes. It can therefore appear smooth or rough, that is, covered with ribosomes. In the rough endoplasmic reticulum, proteins synthesized by ribosomes accumulate. These are then transferred to other cellular sites, particularly the Golgi apparatus. The smooth endoplasmic reticulum, on the other hand, is involved in lipid metabolism. In the past, cytologists referred to the rough endoplasmic reticulum as ergastoplasm, describing it as a layered, basophilic structure.

Microfilaments

Microfilaments generally appear as fibrils or groups of fibrils about 5 µm in diameter, often located near the nucleus or along the nuclear envelope. They belong to the actin–myosin group and play an important role in cellular and intracellular movements. They are not visible with the light microscope.

Mitochondria

Mitochondria are round or oval structures, sometimes rod-shaped, about 5–10 µm long and 0.5–1 µm wide. All cell types contain a relatively constant number of mitochondria, which usually corresponds to the energy demands of the cell: the greater the functional activity, the higher the number of mitochondria.

They are delimited by a double membrane; the inner membrane forms numerous folds, or cristae, where the enzymes responsible for energy production (ATP) are located. Inside, mitochondria contain proteins, phospholipids, and ribosomes, as well as their own DNA (mitochondrial chromosome). Under the light microscope, mitochondria are not visible. In cells with strongly stained cytoplasm, however, they can sometimes be identified thanks to their negative image.

Lysosomes

Lysosomes are granules containing proteolytic enzymes. They are activated after the rupture of the lysosomal membrane, which usually occurs within a phagocytic vacuole. The main lysosomal enzymes include ribonuclease, deoxyribonuclease, lipase, and various proteolytic enzymes. Under both light and electron microscopy, lysosomes appear as granules of varying size and density; many are very small and not visible.

Secondary lysosomes form by the fusion of vacuoles with phagosomes, giving rise to digestive vacuoles. Depending on their enzyme content, lysosomes can appear differently stained: some are neutrophilic, others eosinophilic or basophilic.

Vacuoles

Vacuoles may be distinguished as: Contractile vacuoles, roughly corresponding to the contractile vacuoles of protozoa; and Lipid vacuoles, which are rare under physiological conditions but frequent in many pathological states and during cellular aging in vivo.

They are multivesicular bodies, whose function is not well understood, though they probably derive from pinocytosis. Such vacuolar structures, when large enough to be seen under the light microscope, appear as empty spaces on smears.

Intracytoplasmic inclusions

Inclusions are storage secretions temporarily unused by the cell. Their content can be highly variable: proteins, hemosiderin, mucopolysaccharides, and others. In pathological conditions, inclusions may occupy the entire cytoplasm.

Under the light microscope, they can usually be visualized only with special fixations and stains. With May-Grünwald-Giemsa (MGG) staining, most intracytoplasmic inclusions become clearly detectable.

Ground substance

The term ground substance refers to the space between cellular organelles, which under current observation methods cannot be visualized as a distinct structure. It is composed of water and soluble molecules, which are lost with conventional fixation and staining techniques.

Nevertheless, ground substance plays a significant role in both physiology and pathology of the cell, as it constitutes the microenvironment that surrounds the organelles and coordinates their function.

The Cellular Membrane

Electron microscopy has clearly defined the structure known as the cell membrane, which delimits the cell.

Like all biological membranes, it consists of two layers of phospholipids into which proteins are inserted, with a total thickness of about 8 µm. Its function, however, cannot be considered in isolation, since it depends on interactions with the immediately underlying layer. This layer, facing the cytoplasm, contains enzymes and contractile proteins; the outer layer, facing the extracellular space, contains receptors and molecules that determine cellular specificity.

With this organization, the cell membrane carries out essential functions: regulating exchanges with the external environment, capturing signals that regulate cellular activity and transmitting specific traits that allow the cell to be distinguished from others. With the light microscope, only the most prominent modifications of the cell membrane can be appreciated, while microvilli and microplicae are visible in almost all cells.

Cilia

Cilia are specializations of the cell surface, structurally similar to centrioles. Their role is primarily motor, although in some cases they may also serve a sensory function. Their number varies: in the tracheal epithelium, a single cell may have 200–250 cilia; in the nasal mucosa, about 100. Each cilium measures approximately 5–10 µm in length and 0.2 µm in diameter. In general, a cilium appears as a protrusion of the cell membrane containing a small amount of hyaloplasm.

Cilia display a typical microtubule arrangement known as the 9+2 pattern: nine peripheral doublets surrounding a central pair. The subunits are composed of bundles of 11–12 protofilaments, each formed by globular subunits arranged longitudinally. Each peripheral doublet consists of a complete subfibre A and an incomplete subfibre B. Subfibre A bears two lateral projections (dynein arms) oriented toward subfibre B of the adjacent doublet. The two central microtubules remain separated and are surrounded by a sheath.

At the base of the cilium, the peripheral microtubules are anchored to a basal body (centriole-like structure), while the central pair terminates freely. From the basal body, the cilium continues into the cytoplasm as rootlet fibrils (ciliary rootlets) with a periodic cross-striation of about 50–70 nm. In addition to tubulin, cilia contain a special protein, dynein, rich in ATPase, which is essential for motility.

The beating of cilia originates from the basal region. Cilia are capable of rapid, synchronous movements due to the sliding of their microtubules. The beating cycle consists of: a propulsive stroke (active or rapid phase), which generates forward movement and a return stroke (slow or recovery phase), which resets the cilium's position. During beating, cilia often bend slightly in the same direction, producing a coordinated wave-like motion called the metachronal rhythm. Cilia may also present small club-shaped projections at their tips, about 25–60 nm thick.

Microvilli

Microvilli are protrusions of the cell membrane of varying length, generally cylindrical in shape. They are present in variable numbers in many cell types and serve to increase the free surface area of the cell, thereby intensifying exchange processes between the external environment and the cell interior.

Microvilli are not precursors of cilia. Their role is instead to help maintain the moisture essential for ciliary function and to prevent surface dehydration.

Concluding this subsection, it is important to state that understanding the general structure and function of cells provides a foundation for exploring tissue-specific adaptations. In the context of the nasal mucosa, variations in cellular architecture and specialization enable the tissue to perform complex roles in barrier defense, secretion, and immune surveillance, which will be discussed in the following section on typical cells of the nasal mucosa.

2.1.2 Typical Cells of the Nasal Mucosa

Having outlined the general architecture of the cell, attention can now be turned to the typical cellular populations of the nasal mucosa. This epithelium, which forms the first line of defense against inhaled particles, allergens, and pathogens, is composed of a variety of specialized cell types. Each of these cells contributes in a distinct way to the maintenance of mucosal integrity and to the physiological processes of mucociliary clearance, secretion, and immune surveillance. By examining their morphology and cytological features, it becomes possible to understand not only their normal function but also the alterations that occur in disease states.

Ciliated Cell

The ciliated cell represents the most differentiated cytotype and is the most abundant within the nasal mucosa (accounting for about 80%).

They usually are intercalated among non-ciliated cells and submucosal gland ducts. They have a polygonal shape, measuring approximately 15–20 µm in height, with a nucleus located higher than the basal membrane. This nuclear position contributes to the overall architecture of the epithelial mucosa The apical surface carries approximately 100–250 cilia, each measuring 10–15 µm in length and about 0.2 µm in diameter, as well as 300 microvilli.

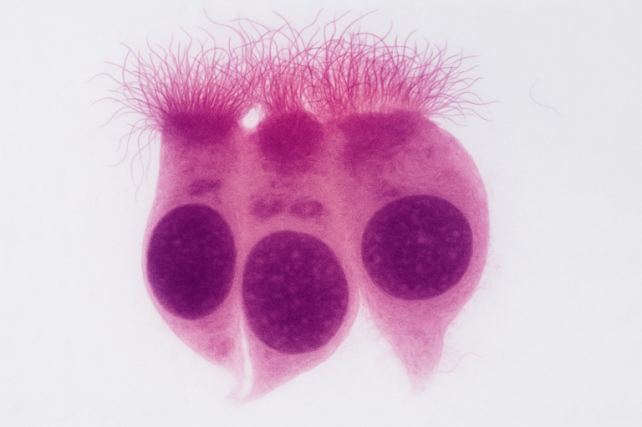


Figure 2.5: Ciliated cells.

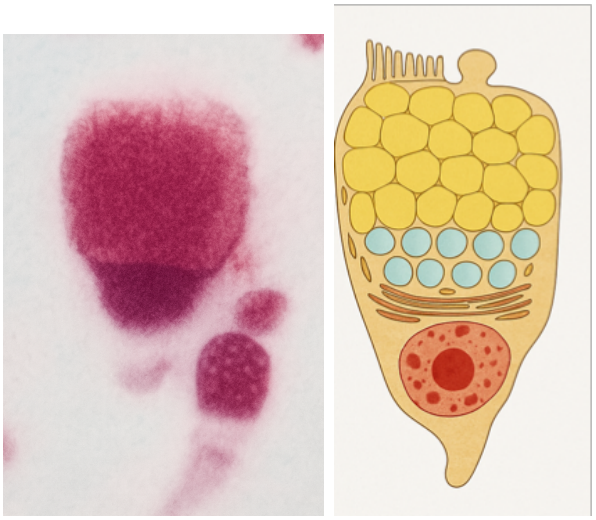
At low magnification (400×), only the apical part of the cell and the ciliary apparatus can be observed. At higher magnification, the cytoplasm and nucleus are visible. Finally, the basal portion—often referred to as the 'foot' of the ciliated cell—can be identified as the narrowest part in direct contact with the basal membrane. Electron microscopy studies have demonstrated that ciliated cells lack any specific anchoring structures to the basal membrane; instead, their adherence to the underlying layer is ensured by desmosomal junctions with basal cells. This applies not only to ciliated cells but also to muciparous (goblet) cells and striated cells. In nasal cytology images, Ciliated epithelial cells are generally identifiable by their elongated shape and the presence of a distinct ciliary border at the apical surface, which appears as a dense, finely textured band

Muciparous Cell

The muciparous cell (goblet cell) is a unicellular gland interspersed among the cells of the pseudostratified respiratory epithelium. It is a mucus-secreting gland that produces a viscous liquid rich in mucin, a protein-polysaccharide that, upon contact with water, forms mucus. Its surface is covered with numerous microvilli, whose size may vary according to the cell's secretory activity. At the apical portion, there is always a small depression called the stoma, through which mucin granules are released to the surface by exocytosis. Mucin granules mainly contain mucoproteins, glycoproteins, and sulfomucins. Unlike serous gland secretions, which are less viscous, these secretions are dense and rich in glycoproteins such as IgA.

The nucleus is generally located basally, while mucinogen granules—the biochemical precursors of mucin—are situated above it. In mature cells, these granules give the characteristic goblet-like shape (hence the term goblet cell). Like ciliated cells, muciparous cells rest on the

basal membrane only through their narrow basal cytoplasmic portion, which is elongated and fusiform. When mucus accumulation is considerable, the nucleus is displaced even further toward the base of the cell. In this condition, a characteristic chromatin reinforcement of the nuclear margin is observed, particularly evident in cytological preparations. Under light microscopy, the mucin granules often appear vacuolated, as they are degraded by fixation procedures. Muciparous cells tend to stand out in cytology images due to their swollen, vacuolated cytoplasm, giving them a rounded or pear-like contour, often with a basally displaced nucleus. A chromatin reinforcement line (dark violet line) is often found on the border of these cells cytoplasm.



(a) (b)

Figure 2.6: Muciparous Cell (a), with its composition schema (b).

Muciparous cells play a crucial physiological role: their secretory activity ensures proper hydration of the mucosal surface, which is essential for maintaining effective mucociliary clearance. Although their density in the respiratory tract decreases with age, especially after adolescence, their number can markedly increase when the airways are chronically exposed to irritants.

In regions where nasal glands are scarce, muciparous cells constitute the primary source of mucus, which is essential for maintaining the mucous coat of the respiratory surface.

Striated Cells

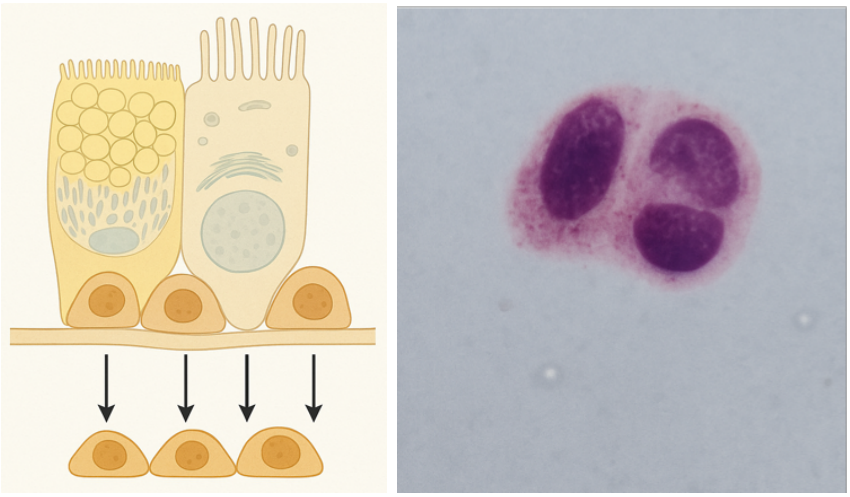
The striated cell is a columnar cell which, like the previously described ciliated and muciparous cells, has its nucleus displaced toward the basal pole. At the ultrastructural level, its apical plasmalemma appears raised into a regular series of microvilli, each containing microfilaments. Other bundles of microfilaments extend deeper into the cytoplasm, reaching cisternae of the rough endoplasmic reticulum, which are abundant together with glycogen granules.

The functional significance of these features is still not entirely clear. Some authors suggest that striated cells may be absorptive cells, possibly with a chemosensory role. Others hypothesize that they could represent progenitors of ciliated cells, though no conclusive studies have yet confirmed this hypothesis. As mentioned in subsection 2.1.1, the microvillar structure increases

the cell's surface area and prevents dehydration of the mucosal surface, maintaining the humidity essential for ciliary function.

Basal Cells

So named because it adheres to the basal membrane without reaching the superficial portion of the airway epithelium, the basal cell is relatively small. Its cytoplasm contains bundles of tonofilaments, mitochondria, rough endoplasmic reticulum, and a nucleus with dense chromatin, disproportionately large in relation to the cytoplasmic volume).



(a) (b)

Figure 2.7: Basal Cell: positioning schema (a), microscopic observation (b).

Basal cells have long been considered the progenitors of muciparous cells and ciliated epithelial cells of the airways. However, data supporting this role remain controversial. It appears that basal cells primarily serve to anchor columnar cells to the basal membrane. This is supported by the fact that columnar cells do not possess hemidesmosomes and instead attach to the basal lamina through adhesion molecules, with basal cells interposed between them. Moreover, the number of basal cells correlates positively with the height of columnar epithelial cells: the taller the columnar cells, the smaller their contact area with the basal membrane, and thus the greater the number of basal cells required to support adhesion. In nasal cytology images, basal cells appear smaller and darker, with high nuclear-to-cytoplasmic ratios and a compact architecture near the epithelial base

Neutrophil Granulocyte

The neutrophil granulocyte is a round-shaped cell, with dimensions ranging from 12 to 14 µm. Its maturation process, from myeloblast to mature neutrophil, takes about 10 days. At the end of this period, the cell forms part of a reserve pool, which can be mobilized rapidly when needed. In peripheral blood, there are approximately 700 million neutrophils per kilogram of body weight; each neutrophil remains in circulation for about 10 hours. After this period, the

cell performs its functions of phagocytosis and killing, destroying infectious agents (such as microorganisms) or toxic substances (foreign material or necrotic products).

Neutrophils are distinguished by their multilobed nucleus, a characteristic that has earned them the traditional designation of polymorphonuclear cells. The nucleus is segmented into lobes connected by thin chromatin strands. These lobes often take on shapes resembling “S,” “Y,” “Z,” or “E,” which become especially visible on smears, depending on the state of the cell. When observed in living cells, the lobes are arranged in a circular fashion, with the centrosome consistently located at the geometric center of the cell. The chromatin, composed of large masses separated by small clear spaces, is very dense. In pathological conditions (particularly in states of abnormal activity), nuclear appendages may appear (drumstick appendages).

The number of nuclear lobes correlates with the cell’s age: In young neutrophils, the nucleus appears elongated or slightly indented but not segmented. In mature neutrophils, the nucleus develops into 3–4 lobes. In aging neutrophils, nuclear segmentation may increase, producing up to 5 or more lobes, occasionally connected by very thin filaments.

Neutrophils and eosinophils in cytology images display round morphology but can be distinguished by their nuclear structure and granularity: neutrophils show multilobulated nuclei, while eosinophils exhibit bilobed nuclei and coarse, refractile granules.

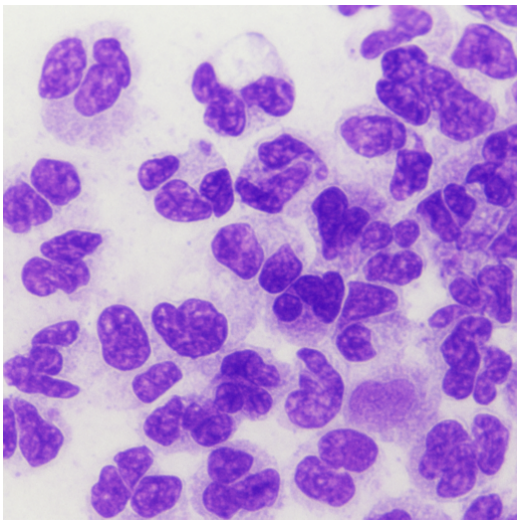


Figure 2.8: Neutrophil granulocytes in a classical cluster configuration.

The cytoplasm contains neutrophilic granules intermingled with some azurophilic granules. The number of granules is variable: under normal conditions they may be few, but sometimes abundant granulation is seen, and in rare cases, the cytoplasm may show an excessive number of apparently normal granules. Staining is also variable, but this is difficult to assess precisely, since the small size of the granules is at the limit of light microscopy resolution. In addition, some cells classified as “neutrophils” may occasionally be found to contain granules of slightly different staining properties.

One of the principal functions of neutrophils is the ingestion and destruction of microbes . The term phagocytosis, coined in the early 20th century, describes this process. Neutrophils can

phagocytose both cells and inert elements, such as dust particles or toxic substances. Thus, their functions are diverse, but their main activity is grouped under the broader term phagocytosis. Once phagocytosis occurs, cytoplasmic granules are discharged into the phagocytic vacuole, releasing lytic enzymes with strong bactericidal properties. This destructive activity, known as killing, includes mechanisms such as the formation of hydrogen peroxide, free radicals, and other highly reactive substances. In addition, neutrophils release a series of enzymes, including lysozyme, collagenase, elastase, phosphatase, and lactoferrin, which contribute to the digestion of ingested material. This process culminates with the expulsion of the digested material by exocytosis, while toxic products are neutralized by vesicular enzymes.

Occasional neutrophils in rhinocytograms are considered normal, but their increase is clinically significant. Indeed, in recent years, the frequency of high neutrophil counts in subjects from industrialized cities has risen, reflecting increased exposure to environmental irritants, toxic substances, and particulate matter. In such cases, neutrophils are recruited even in the absence of bacteria or fungal spores, with no other cytopathic signs. Their role is then mainly to phagocytose toxic or particulate material. Because the permanence of these agents on the mucosal surface must be minimized, neutrophils are appropriately equipped to release enzymes and reactive substances that facilitate clearance.

An increase in neutrophils is also observed in allergic rhinitis. In fact, in both seasonal and perennial forms—and especially in perennial allergic rhinitis—neutrophils frequently represent the predominant cell population.

Electron microscopy studies have demonstrated the presence of electron-dense and electron-lucent material within phagolysosomes of these neutrophils. This material corresponds to immune complexes (antigen–antibody), which are abundant on the mucosal surface in such pathologies.

Eosinophil Granulocyte

In 1879, Paul Ehrlich, while studying the staining properties of aniline dyes on peripheral blood smears, observed that a particular type of leukocyte, containing cytoplasmic granules, bound strongly to acidic dyes. Because the granules displayed a marked affinity for eosin—a red dye derived from the Greek eos (dawn)—he introduced the term eosinophil.

Eosinophils are post-mitotic leukocytes of the bone marrow. They develop under the influence of interleukins and GM-CSF (Granulocyte-Macrophage Colony-Stimulating Factor). After maturing in the bone marrow for about 3–4 days, eosinophils are released into circulation, where they have a lifespan of approximately one day, after which they migrate by diapedesis into tissues. In tissues, eosinophils can survive for an additional 8–12 days. Their survival is largely dependent on cytokines and growth factors. Eosinophils are particularly capable of tissue infiltration, which contributes to the pathogenesis of various clinical syndromes.

In healthy individuals, eosinophils account for 1–6% of total leukocytes. However, in allergic diseases, parasitic infestations, inflammatory and neoplastic conditions, pulmonary diseases, and

syndromes such as non-allergic rhinitis with eosinophilia, values may rise to 15–30% or higher.

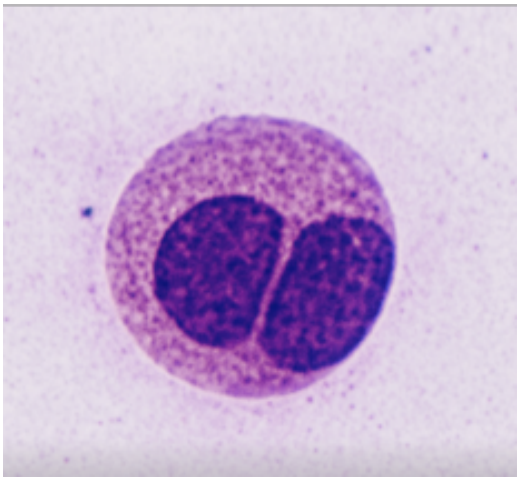


Figure 2.9: Eosiniphil granulocyte.

Structurally, eosinophils resemble other granulocytes. In fresh preparations, they are slightly larger than neutrophils (9 µm in diameter), but on smears they measure 12–15 µm due to cell flattening. The nucleus typically has two lobes, sometimes three. Chromatin is usually dense and clumped (heterochromatic). The cytoplasm is filled with large acidophilic granules, more voluminous than those of neutrophils. These granules are generally spherical or rhomboid, ranging from 0.5 to 1.5 µm in diameter.

Eosinophil functions are mediated by multiple components: Membrane constituents, such as Charcot–Leyden protein, that is frequently observed in the sputum of patients with bronchial asthma. Granular constituents, including major basic protein and eosinophil cationic protein. Non-granular constituents, such as enzymes which play a regulatory role in allergic inflammation. Arachidonic acid metabolites, acting as potent mediators, including bronchoconstrictors. Additionally, eosinophils release numerous cytokines which regulate eosinophil activation and chemotaxis.

Recent studies have highlighted the special role of major basic protein, which is capable of damaging epithelial junctions of the respiratory mucosa, inducing increased permeability and facilitating allergen penetration. This effect is especially evident in asthma-associated conditions.

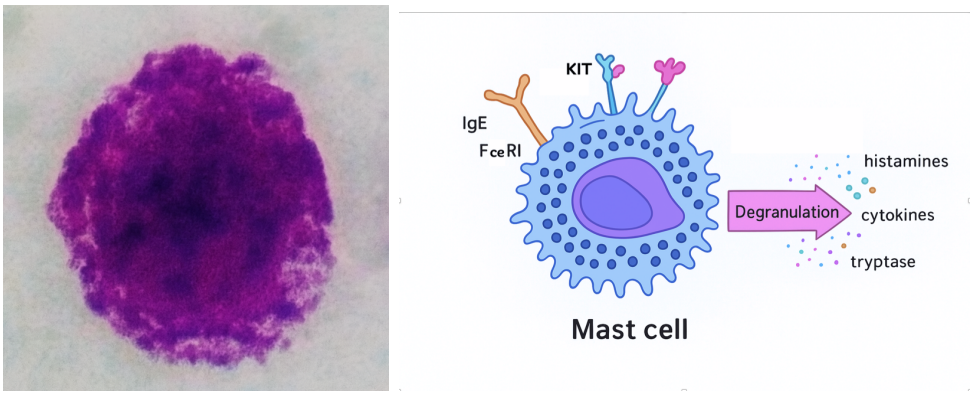
The lifespan of eosinophils in tissues is not fully established; it is estimated to be between 3 and 5 days, depending on the tissue. The mechanisms of their disappearance remain unclear: some are phagocytosed, as shown by the presence of eosinophil granules inside macrophages, others undergo cytolysis and extracellular release of their granule content.

From a diagnostic-morphological perspective, it is crucial to recognize that not all eosinophils display their typical appearance. Occasionally, they may show a single-lobed nucleus, rather than the usual bilobed form. In allergic conditions—especially pollen-related allergies—eosinophils may appear in full integrity, or in partial/complete degranulation stages. Another relevant feature is the presence of Creola bodies (clusters of epithelial cells accompanied by eosinophils),

frequently observed in patients with allergic rhinitis and asthma. Finally, the identification of massive extracellular eosinophilic granulation, even in the absence of recognizable intact cells, provides a strong diagnostic marker of acute allergic episodes with cellular degeneration.

Mast Cells

Mast cells, also known as mastocytes, are distributed throughout all mammalian tissues. They are found on mucosal and serosal surfaces, within lymphoid and connective tissues, and are associated with the nervous system, blood vessels, and neoplastic tissues. Their tissue localization distinguishes them from basophils, which are circulating leukocytes that infiltrate tissues only during inflammation. In the bronchial mucosa and nasal mucosa, mast cells are located at the interface between the external environment and the internal milieu, where they interact with inhaled allergens. Their role in allergic reactions has been demonstrated by the recovery of mast cells in lavage fluids after allergen provocation testing. Mast cells originate from bone marrow precursors and migrate into the bloodstream before localizing in tissues, where their maturation is influenced by environmental factors.



(a) (b)

Figure 2.10: A mast cell (a), and its schema (b).

The concept of mast cell histology dates back to the pioneering work of Paul Ehrlich (1870s), who demonstrated their property of metachromatic staining with certain dyes. This property remains relevant: toluidine blue and similar stains are still widely used today to highlight mast cells and basophils. In metachromatic staining, the basic dye interacts with the acidic proteoglycans of mast cell granules. While useful, this technique presents limitations: it does not always allow for clear morphological differentiation, the degree of staining can vary considerably, results may be difficult to interpret and could even lead to miscounting of mast cells.

Both mast cells and basophils express high-affinity receptors for IgE antibodies. This feature explains their central role in allergic responses and in the initiation of immediate hypersensitivity reactions. Immunological stimulation of human mast cells induces exocytosis of preformed mediators stored in granules, together with the synthesis of new eicosanoids (derived from arachidonic acid metabolism). Mast cells are thus considered key effector cells in acute allergic

inflammation. Their mediators—including histamine, tryptase, PGD, and LTC—are responsible for the early events of the inflammatory cascade.

Mast cell mediators trigger the symptoms of immediate hypersensitivity reactions, which occur shortly after allergen exposure. In the nasal mucosa, these mediators induce: Rhinorrhea, Nasal obstruction, Sneezing, primarily through the action of histamine on mucosal microvessels and sensory nerve endings. The release of histamine and other mediators marks the immediate phase of the allergic response (within minutes).

The concomitant production of LTC, PGD, chemotactic factors, and cytokines initiates a second wave of immune activation, characterized by: Leukocyte adhesion and migration, Priming of inflammatory cells, including lymphocytes and eosinophils. This gives rise to the late phase of the reaction, which develops hours after exposure and is associated with sustained inflammation and symptom persistence. Once migrating leukocytes have reached the site of allergen stimulation, their interaction with the allergen or with high local concentrations of cytokines induces the release of mast cell mediators.

Ultimately, mast cells can be regarded as one of the primary effector cells responsible for triggering both: Immediate allergic reactions, and Chronic allergic inflammation. The terminal phase of mast cell activity is characterized by massive degranulation and the release of both primary and secondary chemical mediators.

Mast cells and lymphocytes generally appear as isolated round elements, but mast cells show cytoplasmic granulation whereas lymphocytes have a dense nucleus with scant cytoplasm.

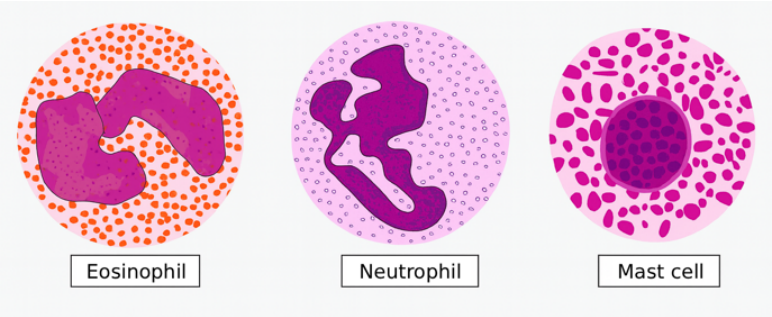


Figure 2.11: Lysosomal granules can be distinguished thanks to specific stainings (MGG).

Lymphocytes

Lymphocytes are circulating elements primarily responsible for cell-mediated and humoral immune responses. In early childhood, lymphocytes account for 30–70% of all white blood cells in peripheral blood. In adults, they represent 10–30% of circulating leukocytes.

From a morphological perspective, lymphocytes are categorized as: Small lymphocytes, Large lymphocytes and Activated lymphocytes.

Small Lymphocytes are the most common form, encompassing both T and B cells. They display a round nucleus with a thin rim of basophilic cytoplasm, often scanty. Nuclear chromatin is dense and deeply stained, sometimes with irregular margins or slight protrusions. Cytoplasmic

coloration ranges from pale to deep blue. The typical diameter is 7–10 µm. Functionally, small lymphocytes are in a resting phase, but are primed to become activated upon contact with specific antigens.

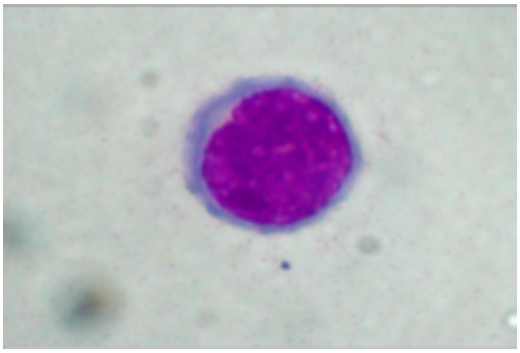


Figure 2.12: Lymphocyte.

Large lymphocytes account for 5–10% of circulating lymphocytes. Their diameter is up to two to three times greater than that of small lymphocytes. Nuclear morphology, chromatin distribution, and cytoplasmic staining are similar to those of small lymphocytes. Nucleoli are often visible. The cytoplasm is more abundant, with a bluish tint. Many of these belong to the T-cell lineage, particularly cytotoxic T cells and natural killer cells, which play a central role in immune defense mechanisms.

When lymphocytes become activated, the cell size increases significantly. The nucleus becomes round or oval, sometimes lobulated, with dispersed chromatin and occasionally visible nucleoli. The cytoplasm is abundant, basophilic, and often contains vacuoles and pseudopodia.

Activated lymphocytes respond to specific antigenic stimuli, and because of this variability they display a wide range of morphologies. This variability can make their correct identification challenging. For this reason, some authors prefer the term “lymphoid variants” rather than “atypical lymphocytes,” since their atypical appearance is simply due to morphological changes following antigenic stimulation.

Depending on the tissue where they reside, lymphocytes act as effectors of different immune activities: Some function as mediators, others as antibody producers. In general, they are classified into two main groups: *T lymphocytes* (cell-mediated immunity), *B lymphocytes* (humoral immunity)

T lymphocytes are thymus-dependent cells and play a central role in cell-mediated immunity. They directly attack non-self targets, controlling their growth and multiplication. They cooperate with other immune cells to coordinate defense against both viral and bacterial infections. Among T cells, the most important are the helper T cells (Th cells). Th1 cells release IL-2, interferon-γ, and TNF, mediating cellular immune responses. Th2 cells release IL-4, IL-5, and IL-13, which promote humoral immunity and allergic responses. Th0 cells can secrete both sets of cytokines, bridging Th1 and Th2 responses. Thus, T lymphocytes are essential in maintaining the balance between protective immunity and pathological reactions, such as allergies and autoimmune diseases.

B lymphocytes are the principal mediators of humoral immunity. Some stem cell precursors migrate from the blood into specialized lymphoid organs, where they interact with epithelial cells and complete their differentiation. Once differentiated, B lymphocytes are capable of antigen recognition. Upon stimulation, they become blast cells with basophilic cytoplasm and active chromatin, rapidly proliferating into plasmablasts and plasma cells that secrete antibodies. B lymphocytes play a particularly important role in the pathogenesis of allergic rhinitis and other allergic disorders.

2.1.3 Rhinocytologic Technique

The description of the main cellular components of the nasal mucosa provides the foundation for understanding their appearance in cytological preparations. To translate this anatomical and functional knowledge into clinical practice, it is necessary to examine how these cells can be collected, processed, and evaluated. The rhinocytologic technique represents the practical application of this knowledge, allowing the microscopic identification of the same epithelial and inflammatory elements directly from nasal secretions.

The entire process is divided into several sequential steps – from sampling to microscopic reading – each of which must be performed correctly to ensure reliable and reproducible results.

Sampling

The first phase consists of collecting cellular material from the nasal mucosa. Several techniques are available, each with specific advantages and limitations:

- Nasal blowing: the patient blows nasal secretions into a sterile container. This technique is harmless and very quick, but the sample may be poorly representative and sometimes contaminated by saprophytic bacteria.

- Nasal washing: a sterile saline solution (NaCl 0.9%) is instilled into the nasal cavities and then aspirated. This method provides a larger quantity of sample and is particularly useful for cytometric studies.

- Nasal swab: a cotton swab, usually moistened beforehand, is gently rubbed on the mucosa, especially at the level of the middle turbinate. It is a simple and widely used technique, useful to detect the presence or absence of specific cell populations.

- Nasal brushing: a small nylon brush is rotated along the septum and inferior turbinate. This method collects superficial epithelial cells with minimal discomfort.

- Nasal scraping: special instruments are used to gently scrape the superficial layer of the mucosa. It is safe, minimally invasive, and provides a well-represented sample of the epithelial surface.

- Biopsy: although it does not strictly belong to cytological techniques, biopsy may be performed in selected cases when a histological examination is required.

The choice of sampling method depends on the age of the patient, the degree of cooperation, and the diagnostic needs. The most common sampling sites are the septum and the inferior turbinate, where secretions are more easily collected.

Processing

Once collected, the sample is transferred onto a glass slide. The mucus must be spread evenly to avoid excessive thickness, which could interfere with observation. Each slide, often bearing a frosted band, is labeled with patient information and the date of sampling.

Fixation

Fixation preserves the morphological integrity of the cells. It can be achieved by simple air-drying, especially recommended for May-Grünwald-Giemsa (MGG) staining, or by immersion in fixative liquids such as acetone, methanol, ethanol 95%, or Mota's basic acetate. Correct fixation is essential to avoid morphological artifacts and ensure accurate staining.

Staining

Staining represents the most crucial step of the preparation, as it highlights the structural characteristics of the cells.

The most commonly used technique is the May-Grünwald-Giemsa (MGG) stain, which allows an excellent definition of the cellular populations present in nasal secretions. Other methods include Hansel, Wright, Papanicolaou, Toluidine Blue, Leishman, Alcian Blue, and Hematoxylin-Eosin.

MGG provides a polychromatic staining: nuclei appear red-violet, basophilic cytoplasm shades from light to dark blue, and eosinophilic cytoplasm takes on a typical orange hue. This wide range of colors allows an easy and reliable identification of neutrophils, eosinophils, mast cells, lymphocytes, as well as bacteria, spores, and fungi.

Slide mounting

After staining, slides must be mounted with synthetic resins that seal the cover glass onto the slide, ensuring long-term preservation. The application of a patient-identifying label completes this phase and guarantees traceability of the sample.

Microscopic observation

Cytological examination is the most delicate and demanding step of rhinocytology. After an initial low-power screening to evaluate staining quality and sample distribution, the analysis

progresses to higher magnifications, ending with oil immersion (1000x) for precise assessment of nuclear and cytoplasmic features.

This phase is long and repetitive, as many microscopic fields must be scanned to obtain a representative overview. Each field requires careful classification of all visible elements, a process that, repeated dozens of times per slide, makes the work highly time-consuming. Fatigue inevitably develops, and the repetitive nature of the task renders the human observer prone to errors such as misclassification or missed rare cells. Even trained professionals can show variability in their results, both between observers and within the same observer at different times.

For these reasons, microscopic observation is not only the most critical but also the most error-sensitive step in the rhinocytology process, underscoring the importance of rigorous training and, increasingly, the support of digital or automated systems to improve consistency.

Quantitative and Semi-Quantitive Reading

Once the observation fields have been selected, the quantitative phase begins. This consists of carefully counting and recording the number of cells belonging to each population at high magnification, usually across 50 microscopic fields, in order to produce percentages that reflect the relative cellular composition. This approach provides a rigorous framework, but it is demanding and can be affected by inter-observer variability, since even slight differences in what is counted as a valid cell may influence the final percentage values.

Nasal Cytology: semi-quantitative reading						
		microscopic fields (1.000x immersion)	0 ..... | ..... 10 ..... | ..... 20 ..... | ..... 30 ..... | ..... 40 ..... | 50			
	0 NR	+ (rare)	++ (some)	+++ (numerous)	++++ (very numerous)	
Ciliated Cells	0	1-100	101-200	210-300	>300	SIS+ % CCF % PNL %
Muciparous Cells	0	1-100	101-200	210-300	>300	
Metaplastic Cells	0	1-100	101-200	210-300	>300	
Neutrophils	0	1-20	21-40	41-100	>100	
Eosinophils	0	1-5	6-10	11-30	>30	Degr. 0
Mast Cells	0	1-5	6-10	11-30	>30	Degr. 0
Lynphocytes	0	1-5	6-10	11-30	>30	Att. %
Bacteria	0	| | |	|| || ||	||| ||| |||	|||||||||||| ||||||||||||	Intra 0 Extra 0
Spores	0	| | |	|| || ||	||| ||| |||	|||||||||||| ||||||||||||	Ife 0 MI 0

Figure 2.13: Semi-quantitative reading table.

Alongside this rigorous approach, rhinocytology often employs semi-quantitative "grading" systems, which classify the increase of specific cell populations into ranges (for example, from grade 1+ to 4+). These systems have the advantage of being more flexible and less dependent on absolute precision in counting. Because the cytotypes differ in granularity, staining affinity, and distribution, small errors in the exact number of cells do not necessarily compromise diagnostic accuracy. Instead, the key lies in recognizing significant shifts in proportion—such as a clear predominance of eosinophils in allergic rhinitis or neutrophils in infectious processes.

This makes semi-quantitative reading particularly useful in routine practice: while not mathematically exact, it remains highly informative, clinically reliable, and better suited to the time constraints and variability of everyday cytological analysis.

2.1.4 Conclusion

The medical background outlined in this chapter has provided the necessary foundation for understanding the biological and diagnostic context in which this work is situated. Beginning with the anatomy and physiology of the nasal mucosa, we have examined its structural organization, the fundamental architecture of the cell, and the specialized epithelial and immune populations that characterize this tissue. Each cytotype—whether ciliated, secretory, basal, or inflammatory—plays a distinct role in mucosal defense and homeostasis, and their identification is central to interpreting nasal pathology.

The rhinocytologic technique translates this knowledge into clinical practice through a standardized sequence of sampling, processing, staining, and microscopic evaluation. While highly informative, this method is also time-consuming and sensitive to human variability, especially in the phases of microscopic observation and quantitative or semi-quantitative interpretation. Nonetheless, rhinocytology remains a minimally invasive, reproducible, and valuable diagnostic tool for the assessment of allergic, infectious, and inflammatory diseases of the upper respiratory tract.

This medical overview serves as the clinical framework for the second part of the background, which addresses the technical aspects. In the following section, attention will shift to artificial intelligence and machine learning, exploring how computational methods can complement and enhance cytological analysis by reducing workload, improving consistency, and ultimately supporting the integration of rhinocytology into broader clinical practice.

2.2 Technical Background

2.2.1 Artificial Intelligence in Medicine

Artificial Intelligence (AI) has become one of the most influential technological drivers of innovation in the twenty-first century, with applications spanning from industry and finance to education and healthcare [12, 19]. In medicine, AI refers to the use of computational techniques designed to mimic, support, or augment human cognitive functions such as perception, reasoning, and decision-making [20]. The motivation for incorporating AI into medicine is clear: modern healthcare generates vast amounts of complex data, ranging from clinical notes and laboratory results to high-resolution imaging and genomic sequences. Human experts, while capable of nuanced interpretation, face limitations in speed, consistency, and scalability. AI systems, by contrast, can process large volumes of data rapidly and can be optimized for specific tasks, making them valuable allies in diagnostic and therapeutic decision-making [21].

The idea of computer-assisted medicine is not new. Early attempts at AI in healthcare can be traced back to the 1960s and 1970s, when rule-based expert systems were developed [15]. MYCIN [22], for example, was one of the earliest expert systems designed to assist in the diagnosis of bacterial infections and to recommend antibiotic treatments. These systems relied

on predefined sets of rules and logical inference engines, essentially encoding expert knowledge into structured decision trees. While innovative at the time, such approaches faced critical limitations: they required extensive manual encoding of knowledge, they struggled to handle uncertainty and variability in clinical data, and they were difficult to scale as medical knowledge evolved. This contributed to the so-called "AI winters" of the late twentieth century, where enthusiasm waned due to unmet expectations [23].

The resurgence of AI in medicine began in the 1990s and early 2000s, coinciding with advances in statistical machine learning [24]. Techniques such as support vector machines [25], random forests [26], and Bayesian networks [27] offered more flexible ways of modeling data compared to rule-based systems. In medical imaging, handcrafted features such as texture descriptors, edge maps, and wavelet transforms were combined with these algorithms to classify radiological scans or histopathological samples [28, 29]. These approaches achieved some success, but they were highly dependent on expert-driven feature engineering, which limited their generalizability. For example, a texture descriptor optimized for mammography might not transfer effectively to cytology images, as the relevant features differ substantially.

The deep learning revolution of the 2010s fundamentally transformed this landscape [30]. With the introduction of deep neural networks capable of learning hierarchical representations directly from raw data, AI began to show unprecedented performance in image classification, speech recognition, and natural language processing. In medicine, the impact was immediate and profound. Convolutional Neural Networks (CNNs) were rapidly applied to tasks such as detecting diabetic retinopathy in retinal fundus images, identifying skin cancer from dermoscopic photographs, and classifying lung nodules in Computed Tomography (CT) scans [31–33]. These models often achieved accuracy comparable to, and in some cases surpassing, that of experienced specialists. Importantly, deep learning eliminated the need for manual feature engineering: rather than specifying what features to look for, models could discover discriminative representations automatically during training.

The adoption of AI in healthcare has been motivated by several key benefits. First, AI systems can improve diagnostic accuracy by detecting subtle patterns that might be overlooked by human observers [34, 35]. Second, they can reduce inter-observer variability, providing standardized assessments across different institutions and practitioners [36]. Third, they can enhance efficiency by accelerating the interpretation of data, thereby reducing the workload of clinicians and enabling more timely decision-making [37]. These advantages are particularly valuable in resource-limited settings where specialist expertise may be scarce.

Despite these advances, the integration of AI into clinical practice remains a work in progress. Several challenges have tempered initial enthusiasm. One of the most significant is the issue of generalizability: models trained on data from one institution often perform poorly when applied to data from another, due to differences in patient populations, imaging protocols, or equipment [38]. Another challenge is interpretability: deep learning models, while accurate, are often perceived as "black boxes," raising concerns among clinicians who require transparent

reasoning to trust and adopt AI-based recommendations [39]. Additionally, ethical and legal questions remain unresolved, including responsibility for errors, data privacy, and the regulatory frameworks governing the use of AI in medical devices.

Nevertheless, the trajectory of AI in medicine continues to be one of growth and refinement. Ongoing research aims to address these challenges through the development of interpretable models, robust training techniques such as transfer learning and federated learning, and standardized evaluation protocols that ensure fairness and reproducibility. Moreover, the increasing availability of large, annotated medical datasets and advances in computational infrastructure suggest that the role of AI in healthcare will only expand in the coming years.

2.2.2 Machine Learning and Deep Learning Basics

Classical Computer Vision

Before the advent of Machine Learning and Deep Learning, computer vision in medicine was primarily based on classical image processing techniques. These approaches relied on mathematical transformations and handcrafted algorithms, rather than data-driven models, to extract meaningful information from medical images. Although limited compared to modern methods, they established the conceptual and methodological foundations upon which current artificial intelligence tools have been built.

The earliest applications concentrated on improving the quality of digital medical images. Modalities such as radiography, computed tomography, ultrasound, and magnetic resonance imaging often produced data affected by noise, low contrast, or acquisition artifacts. To mitigate these problems, digital image enhancement techniques were developed, including histogram equalization to adjust contrast [40], Gaussian and median filters to suppress noise [41, 42], and sharpening filters to emphasize fine details [43]. These procedures did not interpret the image per se, but they improved its readability for the clinician and facilitated subsequent algorithmic analysis.

Alongside enhancement, researchers focused on detecting structural boundaries and regions of interest. Edge detection quickly became a cornerstone of early computer vision in medicine, with operators such as Sobel [44], Prewitt [45], and the widely adopted Canny algorithm [46] used to delineate anatomical contours and pathological structures. Accurate boundary detection was essential, for instance, to outline lung nodules in chest X-rays, vascular structures in angiography, or tumor margins in mammography. Segmentation methods complemented edge detection by separating distinct regions within an image. Simple thresholding approaches, such as Otsu's method [47], were employed to distinguish tissue types or lesions from background. More advanced techniques, including region growing [48], active contours [49], and watershed algorithms [50], were gradually adopted to handle more complex or irregular structures. In cytological applications, segmentation was particularly relevant to isolate nuclei, separate overlapping cells, and quantify features associated with atypia.

Once regions of interest were identified, they could be described quantitatively through morphological and statistical analysis. Classical computer vision emphasized the extraction of geometric descriptors such as area, perimeter, circularity, and eccentricity, as well as textural features derived from gray-level co-occurrence matrices, such as Haralick descriptors [51]. Mathematical morphology operations like erosion, dilation, opening, and closing were used to refine cell boundaries or highlight particular structural elements within histological and cytological samples. These descriptors provided numerical summaries that could be compared across patients and conditions, thereby enabling the first steps toward automated classification. Importantly, however, the analysis remained deterministic: features were chosen manually, and thresholds for classification were fixed by expert knowledge rather than learned from data.

By integrating enhancement, segmentation, and feature extraction, the first computer-aided diagnosis (CAD) systems were developed during the 1980s and 1990s [52]. These early CAD tools were particularly influential in radiology and cytology. In mammography, for example, algorithms combining edge detection and segmentation were used to highlight microcalcifications [53], while in Pap smear cytology, morphological analysis of nuclear size and shape supported the identification of suspicious cells [54]. Such systems were not designed to replace clinicians but to act as decision-support tools, flagging areas that warranted closer inspection or providing quantitative measures that could assist in clinical judgment.

Despite their promise, classical computer vision systems faced important limitations. Their performance was highly sensitive to the quality of image acquisition and often required extensive parameter tuning. Differences in patient anatomy, imaging protocols, and pathological presentation could lead to inconsistent results. Moreover, the handcrafted nature of the features made these systems rigid and poorly adaptable to the variability inherent in medical data. As a result, while they offered valuable assistance, they lacked the robustness and scalability needed for widespread clinical adoption.

Nevertheless, the legacy of these methods is significant. Classical image processing and early computer vision introduced key concepts—enhancement, segmentation, feature extraction—that remain integral to modern pipelines. Many of these techniques are still embedded within preprocessing stages of contemporary Machine Learning workflows. More importantly, they demonstrated that computational tools could augment human expertise in medicine, reducing workload and increasing diagnostic consistency. This realization paved the way for the shift toward data-driven methods, where algorithms could learn features directly from medical images rather than relying solely on handcrafted descriptors. The classical era of computer vision thus represents the essential first chapter in the history of medical artificial intelligence, one that set the stage for the transformative impact of Machine Learning and Deep Learning in the decades that followed.

Machine Learning

The limitations of classical computer vision approaches, which depended on handcrafted descriptors and threshold-based decision rules, motivated the development of machine learning (ML) as a more adaptive paradigm. Rather than prescribing a fixed set of rules, ML algorithms learn patterns directly from data, thereby capturing the variability of medical images. In cytology, where staining conditions, cell morphology, and sample preparation vary widely, this adaptability represented a decisive advantage over deterministic methods.

In its simplest form, supervised machine learning can be formalized as a mapping function:

$$f : R^n \to Y \tag{2.1}$$

where an input vector $x \in R^n$ represents the extracted features, in our case feature related to a cell (e.g., nuclear size, eccentricity, texture descriptors), and the output $y \in Y$ corresponds to its class (e.g., normal, neutrophil, eosinophil, abnormal). The function f is learned from labeled examples (x_i, y_i) provided in a training dataset.

Early applications of ML in medical imaging relied heavily on algorithms such as k-nearest neighbors (k-NN) [55], decision trees [56], random forests [26], and particularly support vector machines (SVMs) [25] . These models were trained using manually engineered features, which condensed morphological and textural properties of cytological images into numerical vectors. For instance, Pap smear cytology often used nuclear texture descriptors or chromatin granularity as features, while nasal cytology might consider cell size or cytoplasmic coloration intensity [57].

SVMs gained prominence because of their ability to separate complex classes in high-dimensional feature spaces. The classifier determines an optimal separating hyperplane defined by:

$$f(x) = sign(w \cdot x + b) \tag{2.2}$$

where w is the weight vector, b the bias, and x the feature vector. The learning process minimizes a hinge loss function subject to a margin constraint, ensuring robust separation between classes. In cytology, this approach was applied to tasks such as distinguishing between normal and atypical epithelial cells, with SVMs outperforming simple rule-based systems [25].

Random forests represented another powerful family of algorithms. By aggregating the predictions of multiple decision trees trained on random subsets of features and samples, they achieved robustness against noise and variability [26]. In cytological applications, random forests were used to classify inflammatory cells versus epithelial cells, or to predict malignancy risk, leveraging their ensemble nature to reduce the effect of sample-specific artifacts [58].

The emergence of artificial neural networks (ANNs) in the 1990s expanded the expressive capacity of ML. These models compute successive transformations of the input vector through layers of interconnected units, with each unit applying a nonlinear activation function. For a hidden layer, the operation can be written as:

$$h^{(l)} = \sigma(W^{(l)} h^{(l-1)} + b^{(l)}) \tag{2.3}$$

where $W^{(l)}$ and $b^{(l)}$ are the learnable parameters of layer l, $h^{(l-1)}$ is the input from the previous layer, and σ is a nonlinear activation functions such as Rectified Linear Unit (ReLU) or sigmoid. Neural networks allowed for the approximation of highly nonlinear relationships, but early models were shallow and limited by computational resources and the small size of annotated medical datasets.

Supervised ML models are trained by minimizing a cost function that quantifies the discrepancy between predicted outputs and true labels. For classification problems common in cytology, the cross-entropy loss is widely used:

$$L = -\sum_{i=1}^{N} y_i log(\hat{y_i}) \tag{2.4}$$

where y_i is the true label (often one-hot encoded) and $\hat{y_i}$ the predicted probability distribution over classes. Minimization of this loss adjusts the model's parameters to improve prediction accuracy.

Despite their promise, machine learning systems in cytology faced several challenges. Traditional algorithms such as SVMs and random forests remained dependent on handcrafted features, which required domain expertise and often failed to capture the full biological variability of cellular morphology. Neural networks alleviated this dependency but introduced new requirements: large annotated datasets, substantial computational power, and careful strategies to prevent overfitting. In practice, cytological data are expensive to annotate, since labeling requires skilled pathologists. This annotation bottleneck limited the scale of training sets and thus the performance of ML models [59].

Another critical limitation was generalizability. Models trained on cytology slides from one laboratory often failed when applied to data from another, due to differences in staining protocols, imaging equipment, or sample preparation. This phenomenon, known as domain shift, highlighted the fragility of ML approaches in medical contexts [60]. Efforts such as data augmentation, feature normalization, and transfer learning were developed to mitigate this problem, but full standardization across laboratories remains a challenge.

Nonetheless, the adoption of ML represented a clear milestone in medical computer vision. By moving beyond rigid rules and enabling algorithms to learn from examples, it became possible to model subtle morphological variations and classify cells with greater accuracy. In cytology, this meant distinguishing inflammatory subtypes, recognizing atypical cells, and quantifying relative cell populations with a degree of reliability that classical methods could not match. Although still constrained by data availability and variability, machine learning established the framework that deep learning would later expand, opening the way toward fully automated and scalable analysis of cytological preparations.

Deep Learning

The evolution from machine learning to deep learning was driven by the recognition that many of the most relevant features in medical imaging cannot be adequately captured by

handcrafted descriptors [61]. Classical machine learning algorithms, while powerful, still relied on human-engineered features such as shape or texture statistics, which limited their capacity to model the full variability of cytological morphology. Deep learning overcame this limitation by enabling algorithms to learn hierarchical feature representations directly from raw data, thus eliminating the need for manual feature design. This development marked a transformative shift in computer vision for medicine, including cytology, where cells exhibit high variability in staining, morphology, and structural complexity.

At its core, deep learning refers to the use of artificial neural networks with multiple layers of processing, capable of extracting increasingly abstract representations of the input. The basic unit of a neural network remains the neuron [62], which computes a weighted sum of its inputs followed by a nonlinear activation:

$$h_j = \sigma(\sum_i w_{ij} x_i + b_j) \tag{2.5}$$

where x_i are the inputs, w_{ij} the learnable weights, b_j the bias and σ a nonlinear activation such as the rectified linear unit (ReLU). By stacking many such layers, the network learns complex hierarchical mappings from raw images to clinical categories.

A breakthrough for medical image analysis came with convolutional neural networks (CNNs), which exploit the spatial structure of images. Unlike fully connected networks, CNNs apply convolutional filters to local image regions, enabling the detection of edges, textures, and patterns regardless of their position. Formally, a convolutional layer computes:

$$h_{i,j}^{(k)} = \sigma((W^{(k)} * x)_{i,j} + b^{(k)}) \tag{2.6}$$

where $W^{(k)}$ is the kernel of the k-th filter, $*$ denotes the convolution operation and (i, j) indexes spatial location. This mechanism allows the network to progressively build complex representations: early layers detect simple edges, intermediate layers capture cellular structures such as nuclei, and deeper layers identify complex morphologies or pathological alterations.

Pooling operations, such as max-pooling, further introduce translation invariance and reduce dimensionality, while fully connected layers at the end of the network integrate the learned features for classification [63]. This represented a qualitative leap over traditional ML pipelines, where feature design often constrained model performance.

The loss function used for training CNNs in cytology tasks is typically categorical cross-entropy, which is similar to the cross-defined in the previous section, but adds an external summation that cycles for each of the classes C

$$L = -\sum_{i=1}^{N} \sum_{c=1}^{C} y_{i,c} log(\hat{y_{i,c}}) \tag{2.7}$$

Optimization is generally performed with stochastic gradient descent (SGD) or one of its variants (Adam, RMSProp), updating weights via backpropagation.

Deep learning architectures such as U-Net further expanded the scope of cytological analysis by addressing segmentation tasks. U-Net, introduced in 2015 [64], employs an encoder-decoder structure with skip connections, making it possible to precisely delineate cellular boundaries even in cluttered or overlapping fields. This architecture became particularly relevant in digital cytology, where accurate segmentation is a prerequisite for reliable quantification of cell populations.In cytology, U-Net-like models have been applied to automatically segment cell nuclei [65].

The advantages of deep learning in cytology are significant. By learning features directly from data, CNNs and related architectures adapt to staining variability, morphological diversity, and preparation artifacts that would confound traditional methods [66]. They also scale well to high-resolution images [67], capturing subtle chromatin patterns and cytoplasmic granularity that may elude human observers. These strengths explain the rapid rise of deep learning in digital pathology and cytology over the past decade.

Nevertheless, deep learning is not without limitations. The most critical is the need for large annotated datasets, which are costly and time-consuming to obtain in medicine [68]. Cytological annotation requires expert pathologists, whose availability is limited. This data bottleneck is compounded by issues of class imbalance, since pathological cells are often much rarer than normal ones. Another limitation is computational demand: training deep CNNs requires powerful GPUs and large memory resources, which may not be accessible in all research or clinical environments.

Interpretability also remains a challenge [39]. While CNNs achieve high accuracy, their decision-making process is often opaque, raising concerns about trust and accountability in clinical practice. Methods such as saliency maps and class activation mappings (CAMs) have been proposed to visualize which regions of a cytology image contribute most to a prediction, but these remain approximations rather than transparent explanations.

Finally, deep learning models are sensitive to domain shift [69]. Differences in staining protocols, slide preparation, or imaging hardware can significantly degrade performance when a model trained in one laboratory is applied in another. This issue has spurred research into domain adaptation and federated learning, which aim to increase generalizability across institutions.

Despite these challenges, deep learning has fundamentally reshaped cytological analysis. It has moved the field from handcrafted descriptors and shallow classifiers to end-to-end systems capable of learning directly from raw data. In the context of nasal cytology, deep learning holds particular promise for automating cell identification, quantification, and classification, thereby reducing the burden on human observers and minimizing errors associated with fatigue and subjectivity. By integrating biological knowledge with computational advances, deep learning represents the current frontier of medical computer vision and forms the technical foundation for the approaches explored in this thesis.

2.2.3 Interpretability and Explainability in Medical AI

As reported at the end of the previous section, despite the impressive performance of deep learning systems in image classification, segmentation, and diagnosis, a major barrier to DNN clinical adoption is the so-called "black box" problem [70]. While neural networks can achieve accuracies that rival or even surpass those of human experts, the internal decision-making process of these models is often opaque. For clinicians, who must justify their diagnostic conclusions and therapeutic choices, a system that produces accurate but uninterpretable outputs is of limited practical value. In medicine, accuracy alone is insufficient; transparency, accountability, and trust are equally essential.

The term interpretability refers to the degree to which a human can understand the reasons behind a model's output [71]. Explainability, closely related, encompasses the techniques and methods used to make the internal mechanisms of complex models more transparent [72]. In the clinical setting, these concepts are tied not only to usability but also to ethics, legal frameworks, and regulatory approval. A physician must be able to explain why an AI tool suggested a particular diagnosis, both to patients and to other healthcare professionals. Similarly, regulatory agencies such as the Food and Drug Administration (FDA) for the United States or the European Medicine Agency (EMA) for Europe increasingly require some degree of model interpretability before approving AI-based diagnostic systems [73].

Several approaches have been developed to address this challenge. One of the earliest and most widely used methods is the generation of saliency maps or heatmaps, which highlight the regions of an image that most influenced the network's decision [74]. In cytology, such visualizations can be used to indicate which cell clusters or morphological structures the algorithm deemed most relevant, helping clinicians validate whether the model's reasoning aligns with established medical knowledge. Techniques like Gradient-weighted Class Activation Mapping (Grad-CAM) [75] have proven particularly effective in producing intuitive, human-readable explanations of CNN outputs.

Beyond saliency maps, more structured approaches to interpretability have emerged. Prototype learning is one such method, where models learn to associate new cases with a set of prototypical examples stored within the network [76]. For instance, a system trained on nasal cytology could justify its classification of a sample as "eosinophilic rhinitis" by directly comparing it with prototypical images of eosinophilic smears in its internal memory. This approach provides not only a classification but also a tangible explanation in the form of comparable reference cases, which is often more convincing to clinicians than abstract numerical scores.

Another important development has been the use of attention mechanisms [77]. Originally developed in natural language processing, attention has been adapted to computer vision to allow models to dynamically focus on the most relevant parts of an image during inference. In medical AI, attention maps can highlight diagnostically meaningful regions, effectively combining prediction with a degree of justification. Such models bridge the gap between performance and

interpretability, offering explanations that are inherently integrated into the model architecture.

One additional direction gaining momentum is the development of linguistically grounded explainable models. Fuzzy logic systems, for example, encode knowledge using human-readable rules and linguistic variables (e.g., "high cell density," "moderate eosinophil presence"), enabling a direct link between input features and clinical reasoning [78]. By structuring decision processes around interpretable rules rather than latent representations, fuzzy models can offer explanations that mirror the kind of qualitative reasoning familiar to clinicians. Linguistic summarization techniques similarly aim to translate quantitative model outputs into concise, natural-language statements, helping to contextualize numerical predictions within medically meaningful descriptions [79, 80]. More recently, large language models (LLMs) have begun to play a role in improving model transparency: they can generate narrative justifications, synthesize evidence, and articulate model reasoning in clinically relevant terms. Although these approaches are still developing, especially in regulated medical environments, their potential to produce explanations that align closely with human communication styles makes them a promising complement to visual interpretability methods [81, 82].

Interpretability is not only a technical challenge but also a matter of clinical trust. Physicians are unlikely to integrate AI systems into routine practice if they cannot reconcile the model's predictions with their own reasoning. Misaligned or misleading explanations can erode confidence, while transparent and clinically meaningful justifications can foster adoption. Moreover, the interpretability of medical AI has direct implications for regulatory compliance, ethical practice, and patient communication. A clinician must be able to explain not just what the AI predicted, but why the prediction was made, especially in contexts where treatment decisions have significant consequences.

2.2.4 Data Challenges in Medical Imaging

One of the defining characteristics of artificial intelligence in medicine is its reliance on large volumes of high-quality data. The performance of machine learning and deep learning models depends critically on the availability of representative training datasets. Unlike conventional software systems, which are designed explicitly by programmers, learning-based systems acquire their functionality through exposure to examples. This makes the dataset not only a resource but the very foundation of the system's capabilities. In medical imaging, however, the construction of suitable datasets is fraught with challenges that are both technical and organizational.

A first and fundamental issue is the quality and consistency of annotations [83]. Deep learning models are typically trained in a supervised manner, which requires each image or region of interest to be labeled according to diagnostic categories. In the context of nasal cytology, for instance, smears must be annotated with the presence and proportion of various cell types, as well as pathological indicators such as eosinophilia or neutrophilia. Such annotations demand the expertise of trained cytologists or pathologists, whose time is limited and costly. Moreover, inter-

observer variability can introduce inconsistencies, as two experts may not agree on borderline cases. This variability complicates the creation of gold-standard datasets and can propagate uncertainty into the model's training process.

Another major challenge is the limited size of medical datasets [84]. Unlike natural image repositories such as ImageNet, which contain millions of labeled examples, cytology datasets are often constrained to hundreds or thousands of samples. This limitation arises from the difficulty of collecting and annotating medical data, as well as from privacy and ethical considerations that restrict open sharing. Small datasets increase the risk of overfitting, where models memorize specific training examples rather than learning generalizable patterns. Overfitting reduces the reliability of AI systems when applied to new, unseen cases, a particularly dangerous outcome in clinical practice.

In addition to dataset size, there is the problem of class imbalance [85]. In many medical applications, pathological cases are far less frequent than normal or benign ones, reflecting real-world prevalence rates. For nasal cytology, certain inflammatory patterns or rare cellular abnormalities may be underrepresented compared to common findings. If left unaddressed, this imbalance can bias the model toward majority classes, leading to poor sensitivity in detecting rare but clinically significant conditions. Several strategies have been proposed to mitigate this problem, including data resampling, the generation of synthetic examples through methods such as Synthetic Minority Oversampling Technique (SMOTE), and the use of class-weighted loss functions that penalize misclassification of minority classes more heavily.

Variability in imaging conditions further complicates the situation [86]. Medical images can differ significantly depending on the equipment used, the staining protocol, the resolution of the microscope, and even the technician's handling of samples. In cytology, small differences in stain intensity or slide preparation can alter the appearance of cells in ways that challenge both human observers and machine learning systems. A model trained on data from a single laboratory may perform poorly when applied to data from another, a problem known as domain shift. Techniques such as data augmentation—introducing controlled variations into the training data—and domain adaptation—adjusting models to new data distributions—have emerged as crucial strategies to address this issue [87].

One promising approach to overcoming dataset limitations is transfer learning. This technique involves pretraining a model on a large dataset from a related domain (for example, ImageNet [88] or histopathology slides) and then fine-tuning it on the smaller target dataset. By leveraging general features such as edges, textures, and basic shapes learned in the source domain, the model requires fewer examples in the medical domain to achieve robust performance. Transfer learning has become standard practice in medical AI, including cytology, where annotated datasets are often scarce.

Finally, the challenges of data in medical imaging are not limited to technical considerations; they are also deeply tied to ethical and legal issues. Patient privacy is a central concern, and strict regulations such as the General Data Privacy Regulation (GDPR) in Europe or Health

Insurance Portability and Accountability Act (HIPAA) in the United States govern the handling of medical data. Anonymization procedures, while necessary, can complicate data sharing and limit collaborative research. Furthermore, biases in datasets—reflecting demographic, institutional, or regional imbalances—can propagate into models, potentially exacerbating disparities in healthcare delivery [73]. Ensuring fairness, inclusivity, and representativeness in medical AI datasets is therefore both a technical and an ethical imperative.

2.2.5 From Models to Applications

While the development of increasingly sophisticated models has transformed the landscape of medical artificial intelligence, it is essential to recognize that research prototypes alone are not sufficient to impact clinical practice. The ultimate value of medical AI lies not only in achieving high accuracy in controlled experiments but in translating these results into usable applications that can be integrated into healthcare workflows, tested in real-world conditions, and validated against clinical outcomes. Without this translational step, even the most promising algorithms risk remaining confined to academic publications rather than improving patient care.

There are several reasons why this transition from models to applications is so critical. First, medicine is an applied science where the practical utility of a method outweighs its theoretical elegance. Second, the medical domain is characterized by variability—across patients, laboratories, imaging devices, and staining protocols—that cannot be fully captured in isolated datasets. Applications provide a platform for prospective testing, where models are challenged under heterogeneous real-world conditions and their robustness can be truly assessed. Finally, clinical adoption requires usability: physicians and laboratory technicians need interfaces that are intuitive, reliable, and seamlessly integrated into existing systems, rather than raw models that require specialized technical knowledge.

However, the transition from research to application involves significant challenges. A primary difficulty is regulatory and ethical compliance. Unlike research settings, clinical tools must meet strict standards for safety, reliability, and transparency. Medical AI applications must undergo rigorous validation, often through multi-center clinical trials, before they can be approved for use. Closely linked to this is the issue of explainability: clinicians need to trust the system's outputs, which means that black-box predictions from complex models such as deep neural networks are not sufficient on their own. Mechanisms for interpretability, audit trails, and error reporting are required to bridge the gap between algorithmic predictions and clinical decision-making.

Another challenge lies in the technical integration of models into healthcare infrastructure. Laboratory information systems, electronic health records, and imaging workflows are highly standardized and often resistant to rapid change. Embedding AI applications into these systems requires not only technical compatibility but also institutional support and training for end-users. Moreover, computational constraints must be considered: models that demand high-end

GPUs or large memory resources may be impractical in many clinical settings, particularly in low-resource contexts.

The background section has provided a dual perspective, medical and technical, necessary to contextualize the contributions of this thesis. On the medical side, the anatomy of the nasal mucosa, its cellular composition, and the principles of rhinocytology have been reviewed to clarify the clinical relevance of automated cytological analysis. On the technical side, the historical trajectory of computer vision in medicine has been traced, emphasizing the evolution from classical image processing to machine learning and deep learning, while addressing issues of interpretability, data scarcity, and application-oriented development. Together, these perspectives highlight both the need and the opportunity for automated pipelines capable of assisting clinicians in the time-consuming and error-prone process of cytological examination. Against this background, the next section will present the main contribution of this thesis: the design, implementation, and validation of novel AI-based approaches for nasal cytology, accompanied by the construction of a dedicated dataset and the development of a usable application framework.

Chapter 3

Main Contribution

3.1 Nasal Cytology Dataset and DL Baseline

The first core contribution of this thesis is represented by the article *A Nasal Cytology Dataset for Object Detection and Deep Learning* [3]. This work addresses one of the major bottlenecks in medical AI applied to cytology: the scarcity of open datasets and standardized benchmarks for training and evaluating computational models. While deep learning has shown strong results in other domains of cytology, progress in nasal cytology has been hampered by the lack of publicly available, well-annotated data and by the absence of reproducible baselines.

To overcome these limitations, this study introduces the Nasal Mucosa Cell Dataset (NMCD), a novel resource containing more than 10,000 annotated cells across 500 images. The dataset was designed to capture the diversity of nasal mucosa cytotypes, including epithelial, muciparous, inflammatory, and immune cells, as well as relevant artifacts, and was carefully curated in collaboration with clinical experts. Alongside the dataset, the paper provides an initial benchmarking analysis based on state-of-the-art object detection architectures, including YOLOv8, DETR, and Faster R-CNN. These experiments offer a first quantitative reference for automated cell detection and recognition in nasal cytology.

The significance of this contribution lies not only in the creation of a dataset but also in the establishment of a common experimental ground for the research community. By proposing a shared benchmark, the study opens the way for more consistent comparisons across methods and fosters reproducibility in a field that has so far lacked standardized resources. Within the scope of this thesis, this paper represents the methodological starting point, laying the foundation for subsequent developments in interpretable pipelines and application-ready tools.

3.1.1 Introduction

In recent years, artificial intelligence (AI) algorithms have pervasively spread among many aspects of everyday life, including health and medicine [89–91]. In particular, deep learning (DL) models have been applied with great success to several clinical tasks concerning image

diagnosis [92, 93]. One of these clinical tasks is to use AI to address cytology, which is generally defined as the practice of observing and analyzing cells within specimens sampled from an organ or tissue, and to detect anomalies and possible ongoing diseases [94]. With deep architectures, several cytological pipelines have been successfully constructed and automated by AI algorithms [95–99], providing important support instruments to doctors. Several examples involving different organs/tissues and pathologies are: cervix liquid cytology to diagnose and prevent brain cancer [100–102], bronchoalveolar lavage fluid cytology [103] to detect various respiratory illnesses, blood cytology [104] to detect hematologic disorders and urine cytology [105] for the identification of atypical and malignant cells that provide risk profiles for oncological diseases. Similar works of veterinary science include [102] and [106], which address automated cytological practices for dog skin pathologies and endometriosis in dairy cows, respectively.

Although good results have been obtained with these approaches, an important bottleneck is still represented by the lack of openly available data related to experimental papers, mainly due to legal restrictions, patient privacy issues or clinical/academic policies [107, 108]; in particular, for cytology some public datasets exist, mainly concerning cervical tissue [109]. An important update has recently been given by the Live-Cell dataset [110] and EVICAN dataset [111], both consisting of (unlabeled) microscope images where cells are annotated within a rectangle referred to as a "bounding box", which represents a general purpose benchmark for cell segmentation algorithms, such as [112]. These data have great potential, for use in pretraining DL models, i.e. learning a transferable embedding space [109] to subsequently fine-tune new tissue-specific tasks for which only a limited number of samples are available.

This work contributes to several open challenges by presenting a novel machine learning-based approach to aid in the automated detection and classification of nasal mucosa cells. The DETR [113], YOLO [114] and Faster R-CNN [115] models have shown good performance in cell detection and their accurate classification, revealing great potential for accelerating the work of rhinology experts.

A valuable contribution of this article is the publication of the Nasal Mucosa Cell Dataset (NMCD) [116], a novel dataset, composed of more than 10.000 instances of cells of the nasal mucosa, distributed throughout 500 images. This new open dataset is provided as a benchmark for training and deploying computer vision or AI models focused on cytological practices involving the nasal mucosa. This dataset allows to explore both the cell detection task and cell classification task, which will be from now on referred to as the cell recognition task, hopefully contributing to the development of AI-aided medicine research [117] and to the still largely unexplored potentialities of AI applied to rhinology [118, 119].

In general, a comparison of the results obtained with different datasets and experimental sets that are not homogeneous has little significance. Instead of offering comparisons with other studies, it is useful to define and propose a common experimental basis. We believe that this study can be considered a new starting point for the community working on this topic and can promote further studies to obtain the best possible results and package a very useful system for

doctors.

This document is organized as follows: in the next subsection, a background about nasal cytology and cells of nasal mucosa is provided. In Section 2, we describe the dataset construction pipeline, while in Section 3 the chosen models and metrics are briefly described. In Section 3.1.4 the experiments are described: they were carried out with the DETR, YOLO and Faster R-CNN models to be used as an initial benchmark for cell detection and the cell recognition tasks. Finally, Section 3.1.5 and 3.1.6 discuss the results and conclusions of this work, respectively.

Nasal Cytology

Nasal cytology, or rhinology, is a subfield of otolaryngology focused on microscopic observation of nasal mucosa samples; this subfield is at recognizing different types of cells to identify and diagnose ongoing pathologies [94, 120].

This methodology has good accuracy in diagnosing rhinitis and infections and is very inexpensive and accessible without any instrument more complex than a microscope, even optical ones. Mucosa samples are obtained noninvasively, just using a simple swab, to be then smeared onto a glass (fixation) and colored with staining, in our case, the May-Grunwald-Giemsa (MGG), before being observed under a microscope.

Cells of the nasal mucosa are distinguishable by specific characteristics of each cytotype. In the present study, the following cells are observed:

Epithelial cells: These cells are the main components of the nasal mucosa and constitute 80% of the observed cytotype (Fig. 3.1a) in the exams of helthy patients. Their presence is not associated with ongoing pathologies.

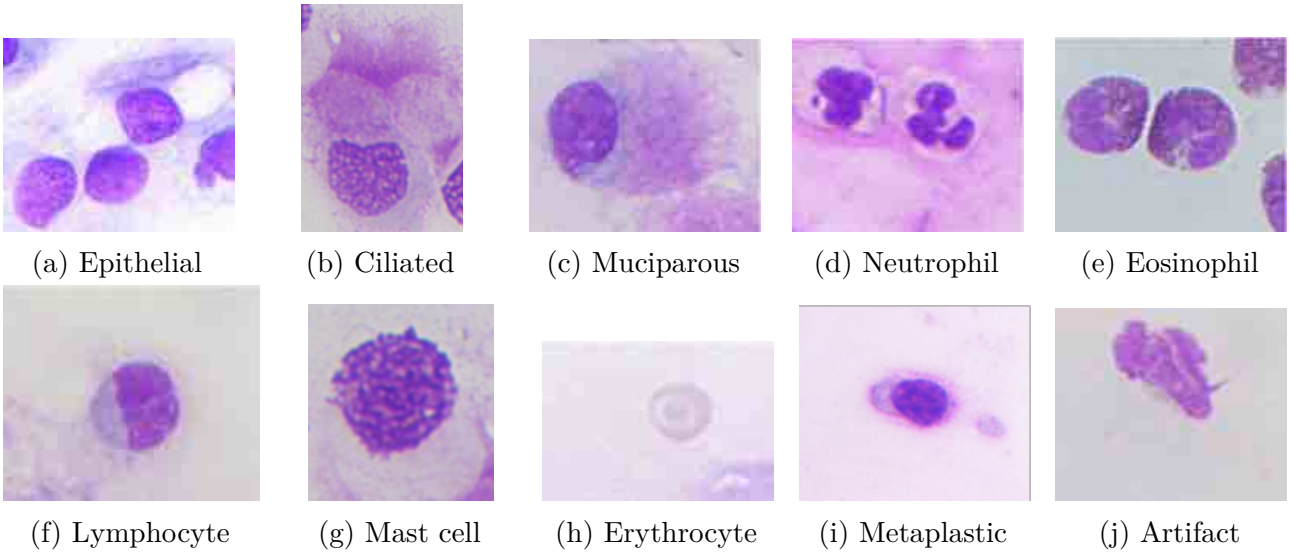
Ciliated cells: In the epithelium cell family, this subtypes is also observed, characterized by their "tailed-like" shape (Fig. 3.1b).

Metaplastic cells: These cells are another sub-type of epithelial cells, characterized by their round shape. Their presence is usually associated with ongoing inflammatory reactions.

Muciparous cells: These calciform mucous-secreting cells (Fig. 3.1c) are characterized by a bilobed shape with a chromatin-reinforced membrane. An increase in muciparous cells results in increased mucus production, a symptom of nasal pathologies with chronic trends, such as dust mite allergies.

Neutrophils: These granulocytes have several nucleoli and a round shape (Fig. 3.1d). Their main function is to phagocytize germs. An increase in their number should always be kept under control, as it is an immune response indicator.

Eosinophils: These polynuclear granulocytes are slightly larger than neutrophils (Fig. 3.1e). MGG staining tends to highlight eosinophil grains with an orange color. Allergic diseases are associated with an increase in their population (15-30%).



(a) Epithelial (b) Ciliated (c) Muciparous (d) Neutrophil (e) Eosinophil

(f) Lymphocyte (g) Mast cell (h) Erythrocyte (i) Metaplastic (j) Artifact

Figure 3.1: Examples of observable items with MGG staining.

Lymphocytes: These white blood cells (Fig. 3.1f) are responsible for the immune response. Their large nucleus is surrounded by a thin cytoplasmic "light blue" rim.

Mast cells: The nuclei of these large oval cells are covered with intense basophil granules (Fig. 3.1g). Their presence in the nasal mucosa is caused by ongoing allergies.

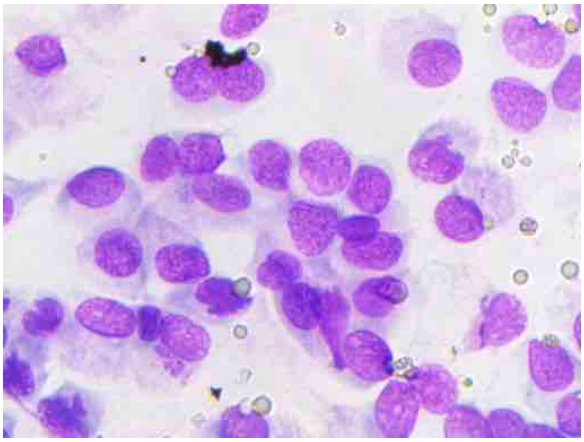
Erythrocytes: The presence of these red blood cells in rhinological specimens may be attributed to pathologies, previous internal nose wounds, or even small blood loss during the smear process (Fig. 3.1h).

Artifacts: We designated this name, for an object with a morphology similar to that of a cell but it is not one; an obect with this characteristics is considered an artifact. Examples of artifacts may be pollen pieces (Fig. 3.1j) or dirt spots on the slide.

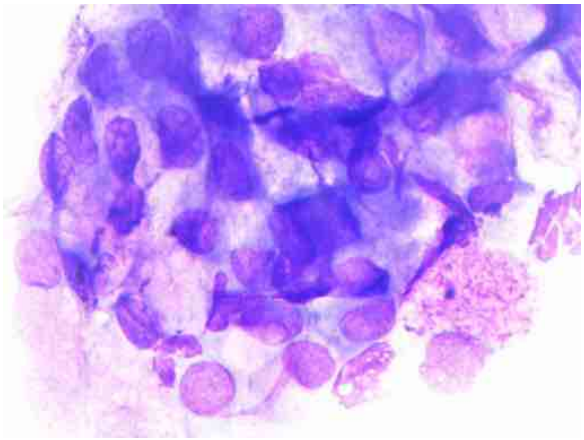
3.1.2 Dataset Construction

The construction of the NMCD dataset is the result of intense work and collaboration between otolaryngologists and computer scientists who, convinced of the great contribution that artificial intelligence can make to this branch of medicine, decided to make material available to the scientific community to allow them to challenge and confront each other in this new application field.

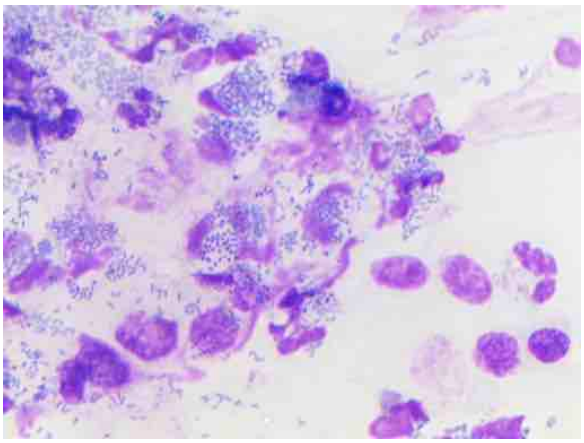
The data were collected from 14 rhinological slides collected at the Rhinology Clinic of the Otolaryngology Department of the University of Bari. The direct smear technique was used for collection and MGG staining was performed. An optical microscope ProWay XSZPW208T (Figure 3.2a) with 1000x zoom, equipped with a 3MP DCE-PW300 camera (Figure 3.2b) was selected to acquire 50 images (microscope fields) from each slide: this specific quantity was chosen since it is defined in the Rhino cytology protocol. Thus 700 images with a size of 1024×768 were obtained.



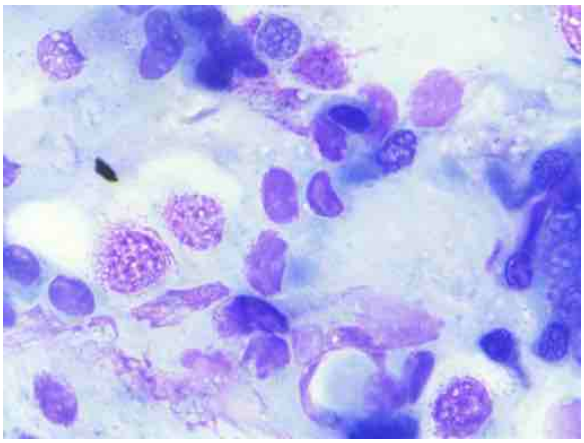
(a) A field with presence of Óartifacts



(b) Dense cluster of cells



(c) A bacterial colony



(d) Cells covered with biofilm

Figure 3.2: Examples of difficult OD cases

The annotations were drawn manually by trained people, one of whom is among the most experienced Italian rhinocytolgists, using the Roboflow platform [121]. Multiple annotators took part in the annotation process, and in case of conflicting annotations, after a brief discussion, the annotation was decided by majority voting.

During this phase, a dropping policy was followed, and images were discarded whenever we detected the following:

1) Sampling noise (i.e. dirt on the slide or blurred photos)

2) Duplications of large cytoplasmic areas already present in other images

3) Too dense and indistinct clusters of cells

A total of 200 cytological fields were pruned,resulting in 500 images. A bounding box (BB) was manually drawn on each cell in the images, to which a label was attached to specify the class to which the cell belonged. As the cells were generally round, the smallest rectangular area that enclosed them was marked as the bounding box.

Therefore, it is possible to identify overlaps between BBs in images because of the proximity between the cells and the rectangular structure of the box. Labeling procedures resulted in more than 10,000 BBs corresponding to cells.With respect to Roboflow, annotations were made

Number of occurrences				
Class	TR	VAL	TS	Total
artifacts	820	98	100	1018
erythrocytes	37	6	5	48
eosinophils	420	54	54	528
epithelial cells	4062	503	495	5060
epithelial ciliated cells	93	11	11	115
lymphocytes	94	11	12	117
mast-cells	15	2	2	19
metaplastic cells	184	23	22	229
muciparous cells	403	51	50	504
neutrophils	2583	323	308	3214

Table 3.1: Distribution of the 10 considered classes in the dataset.

available in any standard annotation format required for computer vision algorithms, like Pascal Voc, Coco and JSON.

The 500 microscopic field images were divided into training, validation, and test sets (80%- 10%- 10%) using the stratified holdout strategy to maintain the same class distribution within the three sets. More details on the class distribution are provided in Table 3.1.

3.1.3 Methods

To provide a baseline that is as accurate as possible for cell detection and cell recognition tasks, we tested and compared three different but equally high-performance architectures: one of the latest version of YOLO (YOLOv8) [114], DETR [113], a transformer model specifically designed for object detection (OD); and also the latest object detection model from the family of Region Based Convolutional Neural Networks (R-CNN), Faster R-CNN [115].

YOLO v8

One of the latest iteration of YOLO developed by Ultralytics [114] introduces new features and improvements to enhance the overall performance, flexibility and efficiency. This model supports a wide range of computer vision tasks, including object detection. The version of YOLOv8 used in this paper is YOLOv8x, which has the largest number of parameters and thus more chances to learn to distinguish between the classes.

The DETR model

Explicit post-processing operations that are quite common in object detection, such as duplicate suppression or anchoring, are avoided, and an end-to-end model (direct set prediction) is implemented. These operations were developed in [113] with the base idea of simultaneously processing the whole image and predicting the entire set of bounding boxes and classification labels (unique loss and computable in parallel for each image).

The architecture consists of a Convolutional Neural Network (CNN) backbone (in this study ResNet-50 [122]), to embed visual features, an encoder and a decoder, both of which are transformers based on multihead self-attention, which allows to always keep the whole image as context. The produced embeddings are fed to three ReLU-activated feed forward neural layers which predict the center coordinates, height and width of the bounding boxes, with their classes through one last softmax layer. The number of predicted bounding boxes N is generally larger than the actual number of objects of interest; hence as in many other OD models, one additional "background" class is used to discard meaningless boxes. Training occurs by optimizing a set-oriented cost function, which is called the Hungarian loss.

Faster RCNN

Region-based Convolutional Neural Networks (R-CNN) [123] are a group of machine learning models designed for computer vision tasks, particularly object detection and localization. Their primary function is to process an input image and generate a set of bounding boxes, each enclosing an object and specifying its category. R-CNN models typically use selective search techniques on feature maps produced by a CNN.

When given an input image, R-CNN utilizes the "selective search" algorithm to identify regions of interest (ROIs), which are rectangular areas that potentially outline objects within the image. Each ROI is then passed through a neural network to extract feature representations. These features are subsequently analyzed by an ensemble of support vector machine (SVM) classifiers, which determine the type of object (if any) present in each ROI.

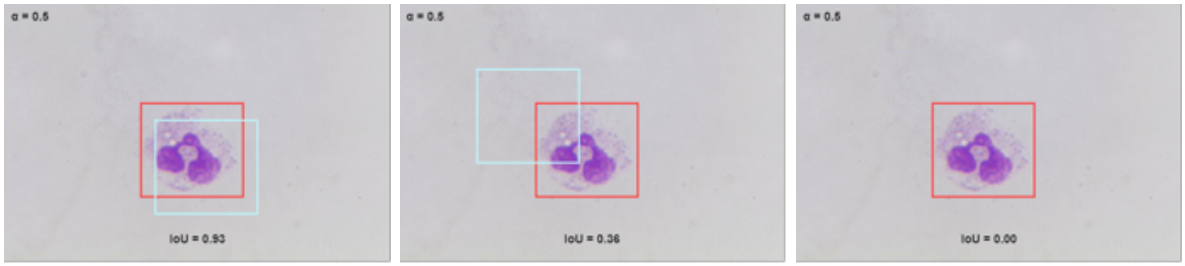
This was optimized in Fast R-CNN that runs the neural network once on the whole image instead of once for each ROI.At the end of the network is a ROIPooling module, which slices out each ROI from the network's output tensor, reshapes it, and classifies it. Then Faster R-CNN integrated the ROI generation into the neural network itself [115].

Metrics

To evaluate the performance of the three tested models, we rely on their specific evaluators, who report the results in terms of average precision (AP) when referring to a single class and mean average precision (mAP) when referring to multiple or all cytotypes. To better grasp some concepts about AP, we introduce a few notes about intersection over union (IoU).

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{3.1}$$

The IoU (Eq. 3.1) quantifies the overlap between the prediction and the annotation by dividing the intersection area between the two bounding boxes (A and B in Eq. 3.1) by their union. An IoU threshold is defined to distinguish between true/false positives/negatives (Figure 3.3).



(a) True Positive (b) False Positive (c) False Negative

Figure 3.3: Example of sorting for IoU threshold = 0.5

For for both models, the performances of the evaluators are compared in terms of average precision (Eq. 3.2).

$$AP = \int_0^1 prec(rec)d(rec) \tag{3.2}$$

With a fixed IoU threshold, the precision and recall can be calculated and then a precision-recall curve is extracted. For both evaluators the AP was then calculated as the area under curve (AUC) for the precision-recall curve. We consider the AP50 which is the AUC of the precision-recall curve with an IoU threshold of 0.5 and the AP50-95 which is the average between APs with IoU thresholds varying from 0.5 to 0.95 with a step of 0.05.

$$mAP = \frac{1}{n}\sum_{k=1}^{k=n} AP_k \tag{3.3}$$

The mean average precision (Eq. 3.3) is defined as the arithmetic average between the APs of each of the classes (in Eq. 3.3 $n = 10$, which is the number of classes).

3.1.4 Experiments and Results

Two different experiments were carried out, both involving the three aforementioned models.

The first experiment, which is referred to as "cell recognition", is aimed at evaluating the performance of the model in identifying and classifying cells into their respective cytotypes. As previously described, this task is very difficult even for specialists, given the strong interclass similarity between some cytotypes and the high intraclass variance in the appearance of cells that belong to the same cytotype.

The second experiment focused on establishing how well the three models can detect cells in the cytologic field without assigning them to a specific cytotype. This task will be referred to as "the cell detection task".

This additional task involved the generalized use of our dataset, as the authors did in [111] and [110] where they developed OD models capable of detecting cells independently of their staining, morphology and tissue.

Finely learning pretrained models to detect cells within nasal mucosa would hence make our dataset eligible for training the general purpose "large cell detection model".

Among the most diffucult tasks that must be accomplished is the ability to address the dense cell clusters like the one shown in Figure 3.2b, which is very common in the nasal mucosa and is difficult to analyze even for expert rhinologists. From an OD point of view, such phenomena involves many bounding boxes that are very densely distributed and usually overlap. To provide a reasonable set of annotations, both the overall and the local portion of the image, as well as small details corresponding to the shape and the edges of each cell (membrane) shown in Figure 3.2b, should be taken into account. Hence, correctly learning or accomplishing the cell detection task on the NMCD would allow models to possess such OD skills.

Another point of view to motivate such a task is the possibility of splitting the first task into two: cell detection over the whole image + cell classification on the single detected cells, building a specialized classifier on top of a more versatile detector.

Each experiment was carried out in an experimental setup composed of an AMD Ryzen Threadripper PRO 3975WX with 32 cores, an Nvidia Quadro RTX 5000 GPU with 16 GB of dedicated RAM and 64 GB of physical RAM.

Cell Recognition

The DETR model, which was pretrained on the COCO dataset [124], was fine-tuned for 50 epochs with our dataset split into training, validation and test sets, as described in Section 3.1.2.

From Figure 3.4 one can assume that the model is clearly capable of detecting cells (or artifacts) and assigning them the correct label, but compared to the objective metrics presented in Table 3.2 it seems that there is no correspondence between qualitative results and quantitative results. These low results can be explained by two motivations.

The first motivation is how restrictive the average precision formula is; even if the classification of a cell is correct, it will not be considered if the IoU between the ground truth (GT) bounding box and the predicted box is lower than the selected IoU threshold. Thus, it is understandable how the difficulty of the human annotator in creating bounding boxes that are consistent can amplify the difference between GTs and predictions.

The second motivation stems from an intrinsic characteristic of the dataset presented in this study: since the sample distribution reflects the cytotypes distribution in the nasal mucosa under normal conditions, the dataset results unbalanced, which is a very frequent condition in medicine; this makes it easier for the model when predicting classes where examples are numerous (Section 3.1.2) and makes its performance very poor when predicting cytotypes with few samples. This, in combination with the stratified splitting of the dataset, results in some minority classes with fewer than 10 examples in the test set, which thus shows a very low AP. As the mAP is the mathematical average between the APs of all classes, these low scores greatly influence the final score. An exception to this principle are eosinophils which can be easily distinguished from other cells and are thus recognizable even if few examples are contained in the dataset.

Class	**Count**	**mAP50**	**mAP50-95**
artifacts	100	0.133	0.062
erythrocytes	5	0	0
eosinophils	54	0.402	0.180
epithelial cells	495	0.738	0.337
epithelial ciliated cells	11	0.099	0.048
lymphocytes	12	0	0
mast cells	2	0	0
metaplastic cells	22	0	0
muciparous cells	50	0.055	0.013
neutrophils	308	0.527	0.204
Total	**1059**	**0.195**	**0.084**

Table 3.2: DETR Coco score on the test set for the task of cell recognition.

Subsequently, the YOLOv8 model, which is also pretrained on the COCO dataset, was fine-tuned for the same number of epochs with the same split as in the previous experiment.

The same experiment was carried our also with Faster R-CNN, pretrained on Imagenet, fine tuned for 50 epochs, with the same splits of the last two experiments.

Table 3.3 shows that the phenomenon described in the previous lines is visible; the classes on which the model performs better (epithelial and neutrophil) are indeed the majority classes among all the cytotypes. We can thus observe a visible correlation between the mAP and representation of those classes in the dataset. Eosinophils, as reported after the DETR experiment, are easily distinguishable from other cytotypes; this is also true for erythrocytes that are blood stains on cytologic slides. YOLO turns out to be better than the other two models not only because it performs better in recognizing cytotypes already identified by the former models but also because it learned to recognize 3 more classes that were not distinguished by DETR and 1 that was not learned by Faster R-CNN.

The solution to this imbalance problem is not trivial; it is true that applying a simple weighted average instead of an arithmetic average will result in higher overall accuracy but will not improve the qualitative results of the system. Instead, these results give us the impression that the system is performing very well, while it is indeed only disregarding the minority classes, among which there are some that are equally, if not more, important at the bdiagnostic level to the more numerous classes.

Cell Detection

To investigate how well the three models can detect the presence of cells and how precise the predicted BBs are (cell detection task), another experiment was carried out.

A duplicated version of the dataset was generated. Because they were not actual cells, all the labels concerning the classes "artifact" and "erythrocyte" were removed and thus considered background. Then, all the remaining labels were renamed and grouped into a single class named

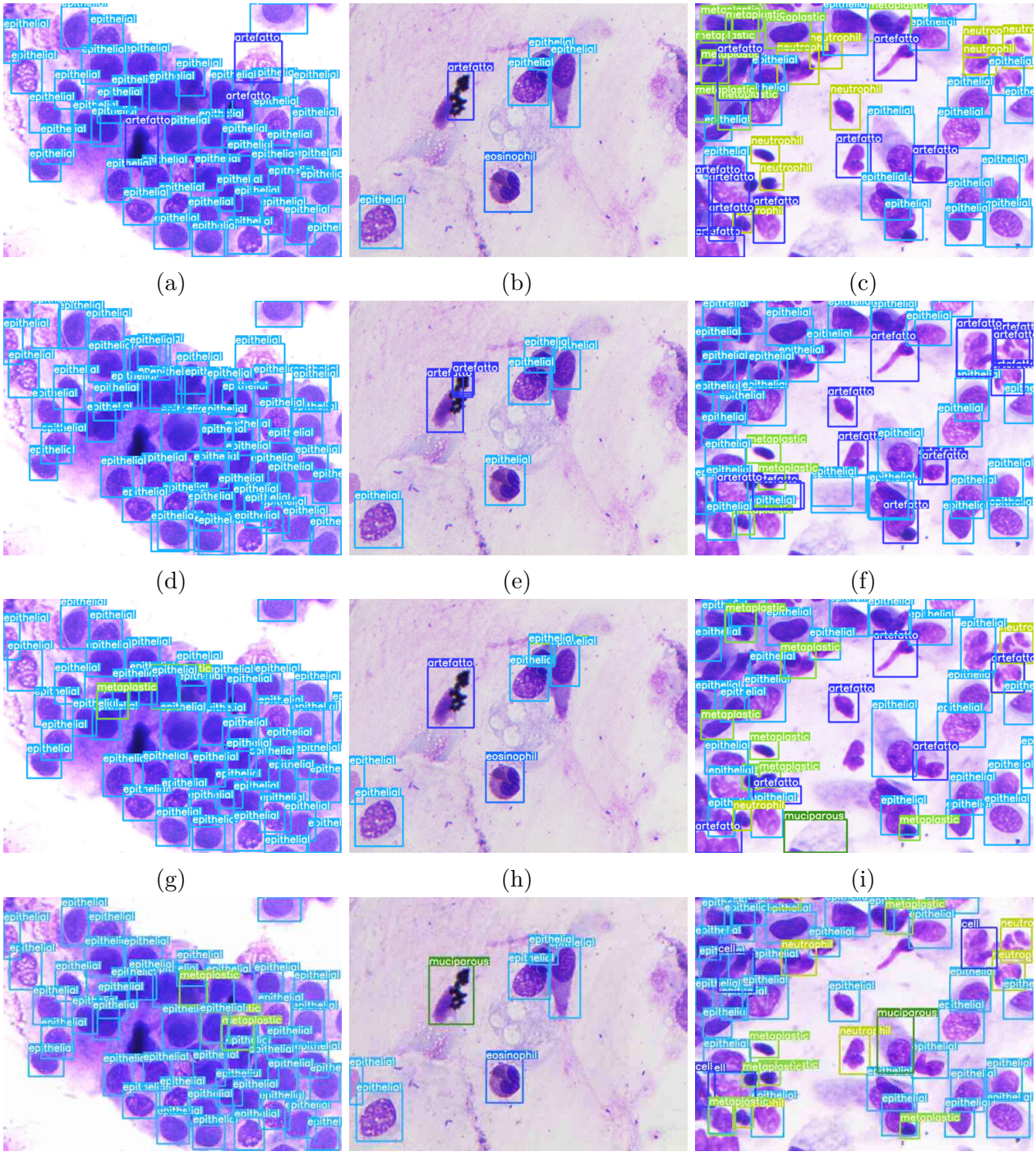


Figure 3.4: Comparison between GTs (a,b,c), DETR predictions (d,e,f), YOLOv8 predictions (g,h,i) and Faster R-CNN (j,k,l) for the task of Cell Recognition.

Class	Count	mAP50	mAP50-95
artifacts	100	0.293	0.171
erythrocytes	5	0.831	0.426
eosinophils	54	0.853	0.528
epithelial cells	495	0.863	0.488
epithelial ciliated cells	11	0.306	0.196
lymphocytes	12	0.223	0.125
mast-cells	2	0	0
metaplastic cells	22	0.051	0.032
muciparous cell	50	0.517	0.251
neutrophils	308	0.843	0.440
Total	**1059**	**0.478**	**0.266**

Table 3.3: YOLOv8 scores on the test set for the task of cell recognition.

Class	Count	mAP50	mAP50-95
artifacts	100	0.123	0.059
erythrocytes	5	0	0
eosinophils	54	0.670	0.321
epithelial cells	495	0.747	0.385
epithelial ciliated cells	11	0.325	0.205
lymphocytes	12	0.112	0.059
mast-cells	2	0	0
metaplastic cells	22	0.004	0.003
muciparous cell	50	0.389	0.154
neutrophils	308	0.680	0.344
Total	**1059**	**0.305**	**0.144**

Table 3.4: Faster R-CNN scores on the test set for the task of cell recognition.

"cell", and their correspondence with the original bounding boxes was maintained.

As previously mentioned, the aim of such experiment is to determine whether models can correctly detect general "cells" objects even within complex dense clusters, independent of their type.

As in the previous experiment, the first model to be tested was DETR which, in this case, was fine-tuned by 50 epochs, followed by YOLO and Faster R-CNN both with the same configurations of the previous experiment.

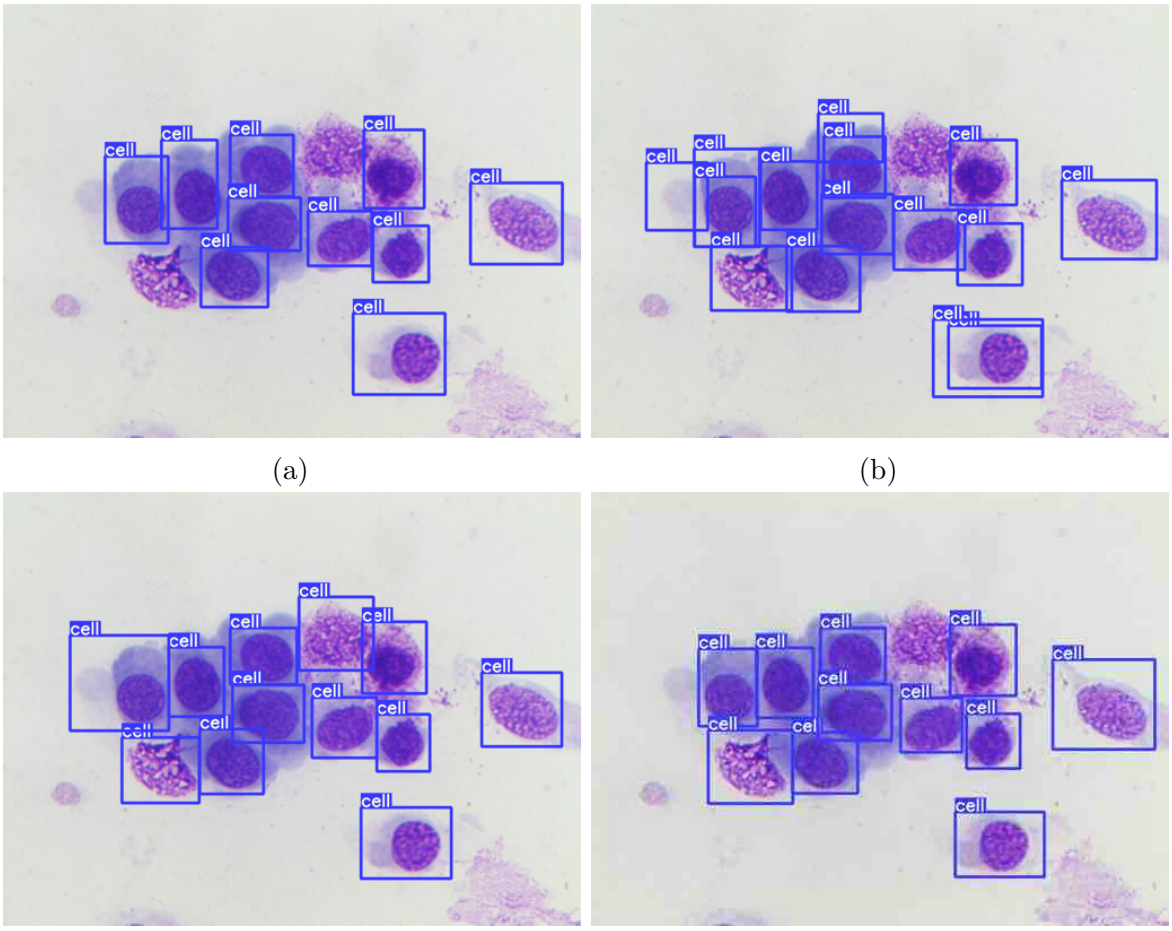
Model	Instances	AP50	AP50-95
DETR	978	0.780	0.339
YOLOv8	978	0.874	0.505
Faster R-CNN	978	0.728	0.357

Table 3.5: DETR, YOLOv8 and Faster R-CNN scores on the test set for the Cell Detection task.

More optimistic yet predictable results can be inferred from Table 3.5: the task of cell

identification is adequately addressed by the all the models, especially YOLO. Since we are not interested in perfect correspondence between the predicted bounding box and the manually annotated bounding box, we can be satisfied by the AP50 obtained by YOLO, which reaches 87%, a result that is even more encouraging according to Figure 3.5c, where annotations are sometimes even more "focused" on the cell than the ground truths.

Although satisfactory, the results are still improvable; even though all models are capable of identifying the majority of cells, they tend to be too inclusive of the learned definition of a cell. This results in some cases where "artifacts" or even "erythrocytes" are classified as cells.



(a) (b)

(c) (d)

Figure 3.5: Comparison between a GT (a) and the respective DETR prediction (b), Yolo prediction (c) and Faster R-CNN (d) for cell detection.

3.1.5 Discussion

Some aspects of the results shown in the previous section are worthy further discussion and analysis to identify the challenges that are characteristic of automatic nasal mucosa cell recognition.

The first aspect to be discussed is how the evaluated models are capable of detecting cells in the pictures of the cytological field while capturing them in their entirety. This was expected since other cytological studies have already applied these (or similar) models in tasks that are

not too dissimilar to the models analyzed in this study [92, 98, 100],in which they obtained good results.

Unfortunately, this is not enough to help clinicians, since the total number of cells has no diagnostic information; the goal is to determine the number of cells for each of the 8 cytotypes ("erythrocyte" and "artifact" are not "cells"). However, as observed in the experimental results, the three models are not yet capable of providing correct information about the number of cells for each cytotype. This is mainly because the proposed dataset has a realistic cell distribution that makes some classes greatly overrepresented while others are nearly not represented at all.

This leads to unbalanced results for the cell classification task: accurate results for the overrepresented/well-represented classes and bordering zero counts for the cells of the minority classes.

The diagnostic significance of a cytotype is independent of its standard quantity within the nasal mucosa but is more dependent on what is the specific task that each cell carries out in the body, with significance proportional to the percentage of variation between the actual distribution and the standard distribution. Thus it becomes crucial to investigate how to balance the dataset to endow the models with a less biased and more objective classification performance.

Methods that could be investigated to solve this type of problem could include the use of a stable diffusion model that allows the dataset to be augmented with artificial examples containing only or mainly minority classes. However, this method would require an expert to confirm that the artificially generated images are medically credible.

Another viable approach would be to divide the task in two steps (detection and classification) so that augmentation can be performed on individual cells, thus increasing the number of examples of a cytotype without affecting the other cells.

Another notable aspect is how the models manage cell clusters; when physiologists are manually counting cells, to avoid losing much time identifying and counting each cell in a cluster, they decide to consider that all the cells in the cluster belong to a specific cytotype (the cytotype to wich the majority of cells in that cluster belong) and add an arbitrary but plausible number of cells to the count for that cytotype. When annotations for the dataset were created, it was specifically chosen to identify single cells even in clusters; this process was very difficult and time-consuming but would allow models to be trained to learn to recognize cells even when clustered.

This can indeed be seen in Figures 3.4d, 3.4g and 3.4j, where it is shown that all models are capable of identifying many clustered cells. Counting cells in a cluster and, more importantly, distinguishing their cytotype instead of assuming that they are all from the same cytotype, can surely guarantee more precise diagnostic results.

3.1.6 Conclusions

In conclusion, this first contribuition paper provides a decisive step toward the integration of artificial intelligence into nasal cytology. By presenting the Nasal Mucosa Cell Dataset (NMCD) and offering baseline experiments with advanced object detection models, the study addresses the pressing need for open resources and reproducible benchmarks in this emerging domain. The results demonstrate that modern detection frameworks can effectively identify and classify nasal mucosa cells, thereby validating the feasibility of automated approaches in this setting.

At the same time, the work highlights important challenges, including the variability of cytological preparations, the imbalance across different cell types, and the need for larger and more diverse datasets to improve robustness and generalization. These limitations underline the importance of further research on interpretability and practical deployment, themes that are directly taken up in the following contributions.

Within the framework of the thesis, this first article provides the foundation: it delivers the dataset and baseline benchmarks upon which the subsequent works expand. The next paper builds on these results by focusing on intrinsically interpretable Deep Learning Models, to try provide some kind of explanation to the phisicians; thus continuing the trajectory from resource creation to clinically relevant AI solutions.

3.2 Prototype Trees for Nasal Cytology

The second contribution of this thesis is represented by the article *Prototype Trees to Develop an Interpretable Automated Pipeline for Nasal Cytology* [1], which focuses on the development of an interpretable and automated pipeline for nasal cytology. While the first paper laid the foundation by creating a dataset and establishing reproducible baselines, this work advances the state of the art by addressing the challenge of interpretability.

The study proposes a divide-and-conquer pipeline that separates the problem into two subtasks: cell detection and cell classification. Detection is handled through a robust object detection model (YOLO), while classification is performed using ProtoTree, an intrinsically interpretable model that provides visual, prototype-based explanations consistent with the reasoning processes of human experts. Furthermore, to enhance robustness and manage uncertainty in decision-making, the method integrates fuzzy logic, evaluating different t-norm operators to optimize decision paths within the ProtoTree framework.

This contribution is significant because it demonstrates that interpretable machine learning methods can be successfully integrated into a practical diagnostic pipeline, delivering both strong performance and transparent reasoning. In the broader context of this thesis, it represents the critical step of moving from dataset creation and benchmarking toward clinically meaningful, interpretable AI systems, setting the stage for application-oriented research and deployment.

3.2.1 Introduction

In recent years, artificial intelligence (AI) algorithms have become integral to numerous aspects of daily life, including healthcare and medicine. Notably, deep learning (DL) models have achieved remarkable success in various clinical tasks, particularly in image-based diagnostics [125]. Among these applications, AI has been employed to enhance cytology: the practice of examining and analyzing cells from tissues or organs to identify abnormalities and detect potential diseases. Advanced AI-driven architectures have enabled the development of automated cytological pipelines, significantly supporting medical professionals in their diagnostic efforts [126].

Explainable Artificial Intelligence (XAI) refers to a set of techniques and methods in AI that aim to make the decision-making processes of machine learning models more transparent and understandable to humans [127]. XAI has gained increasing importance, particularly in fields like healthcare, where AI-powered systems assist in diagnostics, treatment planning, and patient management. In this field, explainability is crucial, because decisions directly impact patient lives, and physicians must understand the reasoning behind AI recommendations to ensure safety, accuracy, and trust [128, 129]. By providing clear insights into AI outputs, XAI fosters collaboration between AI systems and medical professionals, supports regulatory compliance, and addresses ethical concerns, ensuring AI's responsible use in sensitive medical contexts [130].

Despite the growing need for explanations in the healthcare field, among the many models and pipelines proposed for common cytology tasks such as cancer cell identification [131,132], very few

focus on providing explanations to physicians, who then have more difficulty trusting the models and their results. Nasal cytology is a specialized subfield of cytology on the border between rhinology and otolaryngology. It involves microscopic examination of nasal mucosa samples to identify different cell types for diagnosing and monitoring various pathologies. Samples are collected non-invasively using a simple swab, then smeared onto a glass slide for fixation, and stained for analysis. This results in a cost-effective and accessible diagnostic procedure for identifying conditions such as rhinitis and infections. [94] However, these advantages are counterbalanced by the time-consuming and repetitive nature of the procedure, leading to limited adoption by medical practitioners. Research in this area is relatively new, with the first public dataset released only recently [116].

In this paper, we propose a novel automatic pipeline for nasal cytology, aimed at relieving doctors of the tedious task of manually identifying and classifying cells. The pipeline employs a *divide and conquer* approach, splitting the task into two independent subtasks: cell detection and cell classification. For the cell detection task, an out-of-the-box object detection (OD) model (YOLO [114]) was used, while for the cell classification an intrinsically interpretable model (ProtoTree [133]) was chosen, as it is capable of providing physicians with explanations that are consistent with the human reasoning model characteristic of these discernment tasks. Additionally, to enhance interpretability and robustness in decision-making under uncertainty, we incorporated fuzzy logic by evaluating three different fuzzy t-norm operators to determine optimal path selection within the ProtoTree framework.

The document is structured as follows: section 3.2.2 provides an overview of fuzzy system applications in cytology. Section 3.2.3 briefly describes the dataset and the models used in this work. Section 3.2.4 details the experiments conducted to evaluate performance and and optimize hyperparameters for the cell detection and recognition tasks. Finally, Section 5 discusses the findings and presents the study's conclusions.

3.2.2 Related Works

Cytology has become a hot topic in the field of AI in recent years, with the aim of simplifying and speeding up the work of doctors, that lose a lot of time due to the repetitiveness of cell detection. Many works propose the use of AI, that leverage fuzzy logic to handle the uncertainty and ambiguity that characterize cytology.

The most active cytology subfield is the cervical cytology subfield. In [134], the authors apply fuzzy C-means clustering to reduce false positives among centroids detected by a cell nuclei detection clustering algorithm. A comparison between fuzzy C-means and an empirical rule demonstrates the superior effectiveness of the former. Similarly, nucleus segmentation is addressed in [135], where a multi-scale fuzzy clustering algorithm is proposed for segmenting cervical nuclei. This method enables segmentation at different scales, improving detection in clustered cells and achieving a precision of 0.98 and recall of 0.94 in nucleus recognition.

A different task is tackled in [136], where the authors introduce an Improved Fuzzy Deep Learning (IFDL) model to handle uncertainty in Pap smear cell classification, a limitation of standard Deep Convolutional Neural Networks. Their approach integrates a novel layer that combines Deep Belief Networks with a fuzzy weighting system to better manage class ambiguity. Experimental results on the Herlev cell image dataset show a peak accuracy of 99.20% in binary classification.

Both [137] and [138] proposed fuzzy ensembles of Convolutional Neural Networks (CNN) models to classify cervical cytology images. In the first paper, an ensemble of Inception v3, Xception, and DenseNet169 CNNs is used to classify both single-cell and whole-slide images. Their fuzzy rank-based fusion scheme incorporates two nonlinear functions to adjust decision scores based on confidence levels, achieving 98.55% accuracy and 98.52% sensitivity in a two-class setting, and 95.43% accuracy with 98.52% sensitivity in a five-class setting. In the second paper, a similar approach employs the Sugeno Fuzzy Integral to aggregate decision scores from three CNN models.

Nasal cytology, in contrast, is a far less explored subfield due to the limited availability of public data. Only recently has a public dataset been released [116], which was used in this study. Research on AI applications in nasal cytology remains scarce, with one of the few published studies. In [139] presenting a custom CNN trained to classify manually selected 50×50 px cell images into seven classes, yielding promising results.

With this study we try to fill the gap in explainable DL models for nasal cytology by proposing for the first time in this field both a fuzzy and explainable approach. The choice of an explainable, and more specifically an intrinsically interpretable method, is to help physician to trust the system, by giving them the possibility to request explanations that help them understand how the model reasons. In this work we also investigate three detection models to find out which is the best in detecting cells and distinguish between cells and other artifacts in the nasal mucosa.

3.2.3 Materials and Methods

As described in Section 3.2.1, the proposed nasal cytology pipeline is subdivided in two main components: a cell detection and a cell classification component. The workflow of the whole pipeline (fig. 3.6) begins with one (or more) "*rhinocitological fields*", an image of a portion of a cytological slide, that typically contains multiple cells. These images will be the input of the cell detection component where the Object Detection (OD) network will produce bounding boxes (BBs) around all the cells in the image. The BB coordinates are then used to extract individual cell images, which are subsequently processed by the cell classification component.

This classification component assigns each extracted cell to one of eight possible cytotypes. Since the chosen classifier is intrinsically interpretable, no additional post-hoc explanation methods are required. To enhance decision-making under uncertainty, fuzzy t-norm operators

are incorporated to refine path selection in the classification model. When an explanation is requested, the system follows the tree structure of the classifier, providing a reasoning path that closely resembles human decision-making in object recognition.

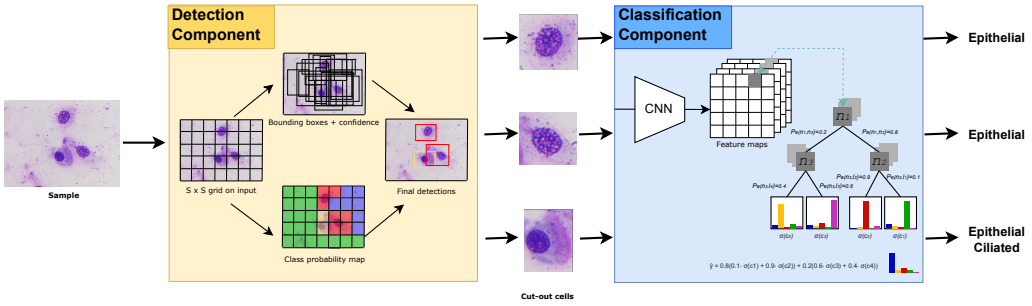


Figure 3.6: Workflow of the proposed pipeline.

Dataset

The dataset used in this study is NMCD [116]. The dataset is composed by 500 pictures of cytological slide portions. Each picture contains at least one cell bounding box (22,7 average BBs per image) that can be labeled with one of the 10 classes (eight of which are cells, the other two are erythrocytes and artifacts). Working with this dataset is not simple since it poses two main challenges both deriving from characteristic proper of the nasal mucosa:

- The dataset is heavily imbalanced. Since the dataset mirrors the distribution of cells in the nasal mucosa, it is heavily imbalanced with very common cytotypes like epithelial and neutrophil cells having more than 3000 samples and rare cytotypes like lymphocytes and mast cells having less than 150 samples each.
- Due to the the high heterogeneity of cells, there is a quite low intra-class similarity and consequently a high inter-class similarity. This is a known problem in the nasal cytology field, where also doctors sometimes struggle to recognize cells and not always agree on the classification.

The dataset comes already split in train (400 images), validation (50 images) and test (50 images) partitions, using the stratified hold-out technique, that guarantees that the 3 partitions keep the original sample distribution.

Metrics

While the classification performances of ProtoTree will be evaluated using common classification metrics like precision, recall, accuracy and f1-score, the performances of the three detection models will be evaluated in terms of Average Precision (AP).

To better grasp some concepts about AP, we introduce a few notes about intersection over union (IoU), which is defined as:

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{3.4}$$

The IoU quantifies the overlap between the prediction and the annotation. An IoU threshold is defined to distinguish between true/false positives/negatives.

For all models, the performance of the evaluators are compared in terms of average precision (3.5).

$$AP = \int_0^1 prec(rec)d(rec) \tag{3.5}$$

With a fixed IoU threshold, the precision and recall can be calculated and then a precision-recall curve is extracted. The AP is then calculated as the area under curve (AUC) for the precision-recall curve. We consider the AP50 which is the AUC of the precision-recall curve with an IoU threshold of 0.5 and the AP50-95 which is the average between APs with IoU thresholds varying from 0.5 to 0.95 with a step of 0.05.

Cell detection component

The cell detection component has a fundamental task within the pipeline: identifying the cells in the image and distinguishing them from other elements like erythrocytes and various artifacts (broken cells, pollen, bacteria, micellar hyphae) present in the nasal mucosa. This step holds much importance since any false positive/negative detected by this component will negatively affect the precision/recall of the subsequent classifier. It is also important to take into consideration the quality of the detections. Every part of the cell is fundamental for a correct classification and so if the detections resulting from this component cut out some parts of the cell, it becomes difficult for the subsequent component to correctly label the sample.

As there are many publicly available object detection models, each with its own merits and shortcomings, it was decided to select three state-of-the-art models to carry out an empirical comparison to evaluate which of these three models performed best on this specific task. The three selected models are Faster R-CNN [123], YOLOv8 [114] and DETR [113]; the same three models tested in section 3.1.4. The former was chosen as it is one of the most widely used for medical contexts and has demonstrated good accuracy when dealing with complex images and small objects, thanks to its two-stage detection pipeline. Yolov8 was chosen for its capability to effectively handle densely packed objects, since cytology images often contain overlapping and closely spaced cells. Moreover, Yolo's use of multi-scale predictions makes it capable of detecting objects of varying sizes, which is useful when dealing with cells whose size can differ significantly. By leveraging transformers, DETR can capture global dependencies within an image, which is beneficial when dealing with nasal cytology, where spatial arrangements and relationships between cells can be crucial for accurate classification. The three models produce annotations in the form of BBs. The bounding boxes' coordinates are than used to cut-out the

detected cells from the original sample image producing one smaller image for each cell. These images are those that are then submitted as input to the classification component.

The three models were tested again since in the first paper each model was only tested once, thus results could be biased by the dataset split.

Cell classification component

Once individual cells have been identified and cut out, they are passed as input to the cell classification component. This component implements an intrinsically interpretable classification tree called Proto Tree [133]. A ProtoTree, as described in [133], combines a convolutional neural network (CNN) with a soft neural binary decision tree structure. An image submitted to a Proto Tree is first forwarded through the backbone CNN, then the resulting feature maps are used as input for the binary tree. The tree is composed of a set of internal nodes and leaf nodes. Each internal node has exactly two child nodes and corresponds to a trainable prototype. These internal nodes perform soft routing, meaning that each input is directed to both children with a fuzzy weight between 0 and 1. The probability of routing a sample to the right child node represents the likelihood of the node's corresponding prototype being present in the image, while the probability of routing it to the left child is the complement of this value. Due to the soft routing mechanism, feature maps traverse all internal nodes and reach each leaf node with an associated probability. Each leaf node corresponds to a class membership function, which is determined during training.

Since all nodes contribute to the final classification, interpretability becomes more complex. To address this, the authors of [133] propose two deterministic strategies to enhance interpretability at test time:

- Greedy path selection, where the sample is routed down the tree by following the most probable path at each decision node.
- Applying the product t-norm 3.6 across all probabilities along each path and selecting the path with the highest resulting value:

$$\top_{\text{prod}}(a, b) = a \cdot b \tag{3.6}$$

In this work, we explore alternative fuzzy t-norms to refine the decision-making process and improve interpretability. Instead of the product t-norm, we investigate three additional fuzzy t-norms to aggregate probabilities along a path before selecting the most probable one:

- the minimun t-norm (3.7);
- the Łukasiewicz t-norm (3.8);
- the Yager t-norm (3.9);

$$\top_{\min}(a,b) = \min\{a,b\} \tag{3.7}$$

$$\top_{\text{Luk}}(a,b) = \max\{0, a+b-1\} \tag{3.8}$$

$$\top_{\text{p}}^{\text{Y}}(x,y) = \begin{cases} \top_{\text{D}(x,y)} & \text{if } p = 0 \\ e^{-(|-log(x)|^p + |-log(y)|^p)^{1/p}} & \text{if } 0 < p < +\infty \\ \top_{\min(x,y)} & \text{if } p = +\infty \end{cases} \tag{3.9}$$

A key observation with the Łukasiewicz t-norm is that it rapidly converges to zero when both probabilities are less than 1. This results in most paths having a value of zero, unintentionally favoring shorter paths, which do not suffer from saturation.

To prevent this saturation effect, we modify the probability value at the root node of the tree. Instead of setting it to 1 (the neutral element of the norm), we assign it the depth of the tree. This ensures that all paths retain meaningful probability values while still maintaining the intrinsic preference for shallower, more interpretable paths.

Regarding interpretability, once the ProtoTree is converted into a deterministic tree, explanations are generated by following the selected path down the tree. At each decision node, the system displays the prototypes it compared against and indicates whether similar patterns were found in the input sample (fig. 3.9). This structured, transparent approach aligns well with human reasoning, making the model's decisions easier for physicians to understand and trust.

3.2.4 Experimental Results

In this section we report the results of all the experiments that were carried out on the single components of the pipeline and on the pipeline as a whole. Each experiment was carried out in an experimental setup composed of an AMD Ryzen Threadripper PRO 3975WX with 32 cores, an Nvidia Quadro RTX 5000 GPU with 16 GB of dedicated RAM and 64 GB of physical RAM.

Cell detection

As described in section 3.2.3, three object detection models were evaluated for the task of cell detection: YOLOv8, DETR, and Faster R-CNN. Each model was trained on the training partition of the dataset for 50 epochs and then evaluated five times on random 80% subsamples of the test set. To focus exclusively on detecting cells, labels were consolidated into a single "cell" class, reducing the number of classes from 10 to 1.

As seen from table 3.6, YOLOv8 outperformed the other two models, demonstrating once again (the first time was in section 3.1.4) superior performance in cell detection. Given its high accuracy, computational efficiency, and lightweight architecture, YOLOv8 was selected as the preferred model for the cell detection component. Its fast inference speed makes it

Model	AP50	AP50-95
Faster R-CNN	0.73 ± 0.012	0.355 ± 0.003
YOLO	0.861 ± 0.048	0.493 ± 0.027
DETR	0.791 ± 0.009	0.344 ± 0.009

Table 3.6: Cell detection performances of the three selected models

particularly suitable for integration into a complex pipeline, where balancing accuracy, efficiency, and interpretability is essential.

Cell classification

The ProtoTree instances responsible for cell classification were trained independently of the cell detection component. This was achieved by training the classification models using cell cut-outs extracted from ground-truth bounding boxes. This process generated a specialized "cell" dataset with over 10,000 samples, split into training (8,708 samples), validation (1,082 samples), and test (1,058 samples) using a stratified hold-out technique.

backbone comparison

Since these components operate independently, any CNN can serve as the backbone, provided an intermediate resizing layer is introduced to match the input feature requirements of the binary tree. The ProtoTree repository includes implementations of ResNet, DenseNet, and VGG as backbone options.

Given YOLOv8's strong performance in cell detection, it was also integrated as a backbone to investigate whether its feature extraction capabilities could enhance ProtoTree's classification performance. In YOLO's architecture, three detection heads extract features from different layers (blocks 15, 18, and 21). A custom parameter was introduced to select which block to use as the feature extractor for ProtoTree.

Backbone	Precision	Recall	F1	Accuracy
VGG-16	0.419	0.424	0.407	0.790
DenseNet 121	0.456	0.373	0.391	0.806
ResNet 50	0.5	0.401	0.412	0.832
Yolo (block 15)	0.467	0.470	0.467	0.835
Yolo (block 18)	0.626	0.493	0.528	0.828
Yolo (block 21)	0.575	0.402	0.435	0.789

Table 3.7: Performances comparison with different backbones.

What stands out immediately by looking at table 3.7 is the difference between the accuracy and other metrics. This large difference can be attributed to the dataset imbalance. This phenomenon can be explained by looking at Table 3.8. The model easily learned to recognize epithelial cells and neutrophils, which have many samples, and eosinophils which are easily distinguishable from other cells, while it has great difficulties with classes containing few samples.

When averaging metrics across classes, the results of the classes with fewer samples make the average drop. The accuracy though, is not averaged across all classes, but calculated on the total number of samples. It thus results in a very high number due to the great influence of the classes with many samples and the near non-influence of those with few samples.

Class	Samples	Precision	Recall	F1
Artifacts	820 (100)	0.632	0.55	0.588
Erythrocytes	37 (5)	1	0.2	0.333
Eosinophils	420 (54)	0.92	0.851	0.885
Epithelial cells	4061 (494)	0.854	0.901	0.877
Epithelial ciliated cells	93 (11)	0.857	0.545	0.667
Lymphocytes	94 (12)	0.181	0.167	0.174
Mast cells	15 (2)	0	0	0
Metaplastic cells	184 (21)	0.181	0.190	0.186
Muciparous cells	403 (50)	0.731	0.6	0.656
Neutrophils	2581 (308)	0.905	0.932	0.918

Table 3.8: Performances of ProtoTree (yolo block18 as backbone) for each of the classes.

Even though the F1-scores appear low, the complexity of the task—combined with class imbalance—makes these results reasonable. In fact, YOLOv8 itself, when used directly for classification, only reached an F1-score of 0.45 at peak performance (prediction confidence threshold 0.338).

A data augmentation pipeline was applied to the train set to solve the imbalance problem. Several image transformations were concatenated with 0.5% probability of being applied to the sample; the applied transformations do not distort nor change the color of the images but only rotate, scale, flip, transpose or add some noise to the image with random intensities/parameters for each transformation. This conservative transformations choice of not changing the shape and color of the cells was dictated by the intention of generating new samples that are as realistic as possible. From each sample from class k, n new samples were generated, where n is the number of samples of the most represented class divided by the number of samples of class k.

We deduce from table 3.9 that simple augmentation techniques that usually solve imbalancing problems do not work well in our case. Further experimentation still need to be performed.

tree depth and pruning The tree depth is another hyperparameter that has to be investigated. Deeper trese usually have higher predictive power, as they allow for more fine-grained decision boundaries. However, excessive depth can lead to overfitting. Moreover, interpretability decreases with tree depth, as a deeper tree contains a larger number of prototypes, making it harder to analyze the decision-making process. Therefore a trade-off point has to be found. Increasing depths were tested, being careful to set the initial depth high enough so that the number of leaves is higher than the number of classes. As expected table 3.10 show an increasing f1-score in proportion to the tree depth. The f1-score will eventually plateau when the tree will have a leaf for each training sample, reaching maximum overfitting.

Class	Samples	Precision	Recall	F1
Artifacts	4100 (100)	0.769	0.5	0.606
Erythrocytes	4070 (5)	0	0	0
Eosinophils	4200 (54)	0.902	0.851	0.876
Epithelial cells	4061 (494)	0.843	0.949	0.893
Epithelial ciliated cells	4092 (11)	0.538	0.636	0.583
Lymphocytes	4042 (12)	0.5	0.25	0.333
Mast cells	4065 (2)	0	0	0
Metaplastic cells	4048 (21)	0	0	0
Muciparous cells	4030 (20)	0.861	0.62	0.721
Neutrophils	5162 (308)	0.897	0.942	0.919

Table 3.9: Performances of ProtoTree (yolo block18 as backbone) for each of the classes, with data augmentation performed on the training set.

Depth	**4**	**5**	**6**	**7**	**8**	**9**
F1-score	0.384	0.435	0.478	0.471	0.526	0.549

Table 3.10: ProtoTree depth comparison.

ProtoTrees, can be pruned to reduce their size, thereby increasing their interpretability, usually at the expense of accuracy. By setting a pruning threshold value, all leaves with maximum class probability lower than the threshold are pruned. If all the leaves of an entire subtree are pruned, then the entire subtree is pruned. In this way, superfluous parts of the tree are removed and only the leaves that discriminate most reliably between cells are kept.

Table 3.11 shows that indeed the f1-score drops a little, but the gain in interpretability of the tree is much larger. A pruning threshold of 0.11 ensures to prune all leaves that have close to zero discriminative power, while a threshold of 0.51 keeps only leaves where a class has absolute majority. The tree size is greatly reduced in this way, in fact the number of prototypes drops by about 90%.

Threshold	Initial F1	F1 pruned	# Prototypes	% pruned
0.11	0.549	0.523	67	87%
0.51	0.549	0.535	48	91%

Table 3.11: Impact of pruning on f1-score and # prototypes.

tree ensemble Similar to classic decision trees, ProtoTrees support ensemble learning, where multiple trees are combined to improve classification accuracy. However, as with many fuzzy and interpretable AI models, this trade-off comes at the cost of increased explanation complexity, potentially reducing interpretability.

To assess the impact of ensemble learning, we experimented with ensembling three and five ProtoTree models with identical parameters. Additionally, we introduced a novel ensembling strategy: combining three ProtoTree models trained on feature maps extracted from different

layers of the YOLO architecture. This approach was inspired by YOLO's multi-scale object detection, where bounding boxes are detected at different depths of the network. The underlying hypothesis is that using trees trained at different depth levels would allow the ensemble to better classify cells of varying sizes, akin to how fuzzy logic handles multi-resolution classification problems by considering multiple levels of abstraction.

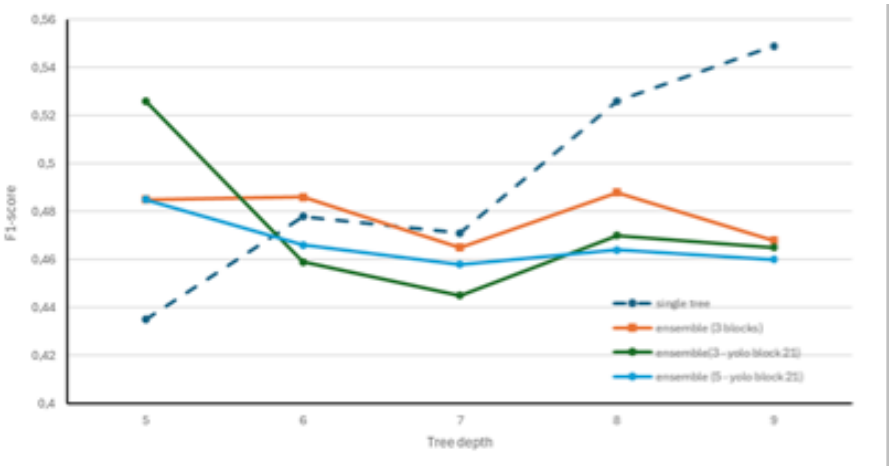


Figure 3.7: Ensembles performances with respect to the tree depth.

Fig. 3.7 illustrates that ensembling enhances performance when applied to low-depth trees. However, as tree depth increases, the benefits of ensembling do not scale proportionally, and deeper single trees eventually outperform ensembles. Moreover, the three-block ensemble did not yield significant improvements, suggesting that the architectural gains seen in YOLO's object detection mechanism may not directly translate to ProtoTree-based classification.

deterministic reasoning As previously discussed, ProtoTrees can be made significantly more interpretable by introducing deterministic reasoning at test time. Instead of allowing a sample to traverse the entire tree through soft routing at each node, we can deterministically select a single path to the leaf node, ensuring a clearer explanation of the model's decision-making process.

We tested six distinct methods to achieve deterministic behavior, as outlined in section 3.2.4). These methods were applied to trees with varying depth, backbone architecture, and pruning thresholds.

Contrary to initial expectations, the soft ProtoTree model performed relatively poorly. This can be attributed to the large number of leaf nodes where the most represented classes tends to have higher membership values. When calculating the weighted sum of membership functions across all leaves, the most frequent class continues to receive high values even when it does not correctly represent the sample.

In contrast, by making the tree deterministic at test time, we select the optimal path that best represents the sample, effectively choosing a single leaf with the highest discriminative power. This adjustment results in improved performance.

When exploring fuzzy t-norms, we introduced three new variants to handle deterministic path selection. Among these, the Łukasiewicz t-norm performed the worst, despite the fix

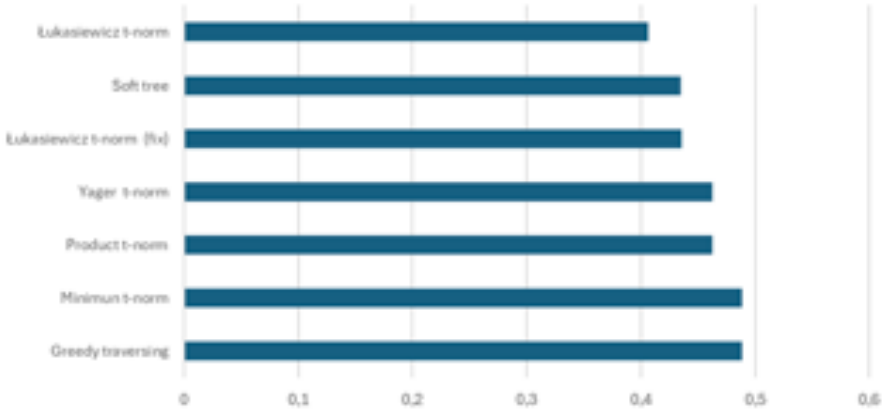


Figure 3.8: Average f1-score of different deterministic reasoning methods, over 7 different trees .

described in section 3.2.3. Its rapid saturation tends to favor shallow paths, which limits its effectiveness.

On the other hand, the minimum t-norm and greedy traversal both exhibited the best performance. Notably, the results of these two methods were identical, as both approaches consistently chose the same leaf path, reflecting a more reliable and robust form of decision-making.

Whole Pipeline Results

Following the various experiments to find the best combination of ProtoTree hyperparameters and configurations, a test was carried out to try out the entire pipeline, concatenating the two components as in fig. 3.6.

The Final ProtoTree uses Yolo (block 18) as backbone, has a depth of 9 and a pruning threshold of 0.51. With this configuration the tree achieves a F1-score of 0.587.

From the fifty images in the test set, five subsets of 35 random images were extracted and than used to test the whole pipeline and evaluate its performances in the task of cell counting.

Class	GT	Pipeline	MAE	CT
Artifacts	74.6 ± 5.57	43.8 ± 2.4	30.8	-
Erythrocytes	2.6 ± 0.8	2.8 ± 1.16	0.2	-
Eosinophils	28.4 ± 11.35	30.2 ± 11.97	1.8	3.5
Epithelial cells	393.2 ± 32.9	508.8 ± 32.15	115.6	100*
Epithelial ciliated cells	7 ± 1.26	2 ± 0.63	5	100*
Lymphocytes	7.2 ± 3.12	8.4 ± 1.49	1.2	3.5
Mast cells	1 ± 0.63	0.6 ± 0.49	0.4	3.5
Metaplastic cells	16.2 ± 4.07	14 ± 2.96	2.2	70
Muciparous cells	41.2 ± 8.79	25.2 ± 2.78	16	70
Neutrophils	187.6 ± 52.32	205.4 ± 56.33	17.8	14

Table 3.12: Whole Pipeline cell counts and MAE for each class.

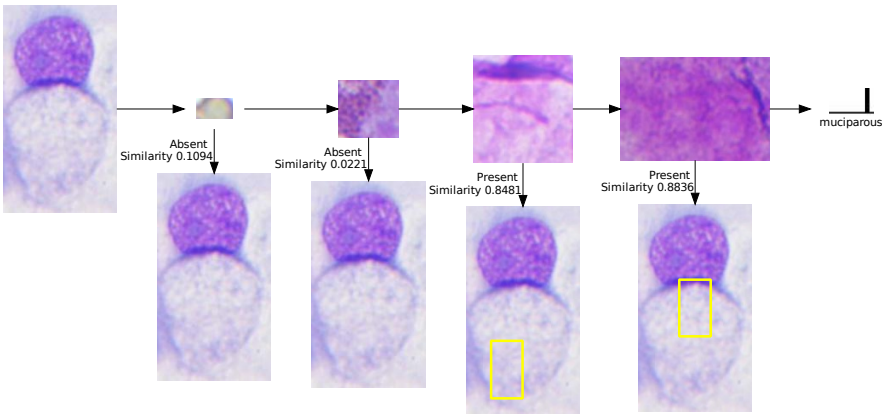


Figure 3.9: Explanation example.

From table 3.12 can be seen how the whole pipeline performs quite good on the cell counting task. For most of the cytotypes the Mean Absolute Error (MAE) is below the Count Threshold (CT). The CT is the number of cell of a specific cytotype required to pass to the next category in the qualitative count.

Fiure 3.9 instead shows an explanation example, where it can be see that a muciparous cell is recognized by the absence of o blood stain and the granularity typical of eosinophils, and by the presence of mucus and the chromatin reinforcement typical of this cells.

3.2.5 Conclusion

In conclusion, this second contribution introduced a novel automatic pipeline for nasal cytology, combining high-performing detection with inherently interpretable classification. By employing YOLO for detection and ProtoTree enriched with fuzzy logic for classification, the study showed that automation can be achieved without sacrificing transparency. The results confirm that it is possible to provide physicians with explanations grounded in human-like reasoning, an essential requirement for building trust in AI-assisted diagnostics.

The study also emphasizes important practical insights. Balancing data distributions and addressing class imbalance remain open challenges that directly impact the robustness of classification. Finally, the integration of fuzzy logic demonstrates how uncertainty can be modeled explicitly, improving interpretability under ambiguous conditions.

Within the framework of this thesis, the second paper marks a decisive transition: it moves beyond dataset provision and benchmarking, and demonstrates how interpretable pipelines can directly support clinical practice.

This sets the stage toward developing application-ready tools capable of embedding these methods into practical and user-friendly platforms for physicians.

3.3 Automated ciliary beat frequency estimation

The third contribution of this thesis is represented by the article DeepCilia: Automated, deep-learning based engine for precise ciliary beat frequency estimation [97]. Unlike the first two works, which focused primarily on the detection and classification of nasal mucosa cells, this study addresses a different yet equally critical aspect of rhinocytology: the measurement of ciliary motility, and specifically the estimation of the ciliary beat frequency (CBF) of ciliated cells.

Ciliated cells form the majority of the nasal epithelium and play a fundamental role in mucociliary clearance, with their rhythmic beating serving as the first line of defense against inhaled pathogens and irritants. Alterations in their beat frequency are closely linked to conditions such as primary ciliary dyskinesia (PCD), chronic rhinosinusitis, and other airway diseases. Moreover, recent studies suggest that the decay of ciliary activity after death can provide valuable information for forensic applications. Despite this importance, existing approaches to CBF estimation often rely on expensive instrumentation, semi-automated protocols, or manual analysis, making them impractical for routine use or deployment outside specialized laboratories.

To overcome these limitations, this work introduces DeepCilia, an end-to-end automated system capable of detecting ciliated striae in microscopy videos and estimating their CBF with high accuracy and efficiency. The architecture combines a YOLOv8-based detection module for robust identification of ciliated striae with a signal processing pipeline leveraging the Short-Time Fourier Transform (STFT) for frequency estimation, even under challenging conditions such as low frame-rate acquisitions or video impurities. The system is designed to be both precise and computationally efficient, achieving sub-second inference times while maintaining an average estimation error below 1.5 Hz.

This contribution is significant because it extends the scope of AI in nasal cytology beyond cell recognition to the quantitative analysis of dynamic cellular behavior, thus opening new avenues in both clinical diagnostics and forensic medicine. Within the broader trajectory of this thesis, it represents the step of broadening the application of interpretable and automated AI methods to cover additional clinically relevant tasks, thereby consolidating the move from theoretical models to practical, usable tools.

3.3.1 Introduction

Technology advancements over recent decades have greatly impacted medical treatments and diagnostic methods. The integration of computer vision in the medical and biomedical fields has led to the availability of additional tools for specialists [5, 140–143]. These tools have enhanced their ability of specialists to acquire, transmit, and analyze digital images and signals, thereby improving their overall performance [6, 144]. Much research has focused on the segmentation and classification of cells from digital images, as in the case of nasal cytology, which is becoming increasingly critical in diagnosing nasal conditions [18,145]. This field focuses on the examination

of nasal mucosa cells with the objective of recognizing changes in the epithelium, which is frequently subjected to acute or chronic irritation and inflammation caused by viruses, bacteria, or fungi.

Among the cell population of the nasal mucosa, ciliated cells make up 80% of the epithelium in the upper airways [146]. These cells are characterised by the presence of cilia that normally beat in a consistent, fluid pattern, easing effective mucociliary clearance. Ciliated cells in humans are usually 15-20 nm in length and beat at a rate of 16 Hz, but this can vary depending on factors such as infections, temperature, age, or inflammation in the upper airway [147]. This cytotype results quite important since its study allows for the diagnosis of primary ciliary dyskinesia (PCD), a rare pathology that refers to the irregular and disorganized movement of cilia and is usually associated with other severe pathologies such as situs inversus, cardiac diseases and male infertility. The diagnosis is typically made through a biopsy or brushing of the ciliary mucosa and the assessment of ciliary function through the measurement of the ciliary beating frequency (CBF).

In addition, [145] demonstrated that the CBF decay of the ciliated cells in a dead body may help to trace the time of death, since a statistically significant relationship between decreasing ciliary movements and increasing postmortem interval was detected even in the presence of putrefactive changes in nasal ultrastructure integrity.

These motivations have fueled the research seeking an architecture that supports such long, complex estimations, and feels necessary in medical and forensics contexts.

As per the available literature, the tools used to measure CBF exhibit practical limitations and often require expensive technology or manual, constant detection of ciliated cells by physicians to obtain reliable results. Consequently, domain experts without specialized laboratories must rely on qualitative evaluation of ciliated epithelium. Thus, is important to invest in research for the development of an efficient and cost-effective CBF measurement system.

Assessing CBF is a challenging task because high oscillatory motion cannot be observed directly under the naked eye. Over the years, various techniques have been developed to tackle this challenge. Initially, methods relying on photoelectric effects [148], photodiodes [149], and stroboscopes [150,151] were proposed. Later, high-speed video cameras allowed for more accurate CBF analysis [152,153], but the process remained semiautomated with CBF determined manually. In 2003, Sisson et al. introduced the first automated technique for measuring CBF, named the Sisson-Ammons Video Analysis (SAVA) [154]. This method is based on the comparison of both analog and digital CBF measurements of the same region of interest (RoI) while varying the temperature, yielding fairly equal results. They also effectuated another measurement with an innovative analysis algorithm called whole-field analysis (WFA), where the entire image is analyzed for each frame. Both the methods have pros and cons: while the RoI method does not need a great deal of computational power to achieve its results, it can usually introduce a selection bias because “the human eye is naturally drawn toward faster beating cells”. On the other hand WFA does not require any human interaction, but it is computationally expensive

if algorithms for motion detection are not implemented to ease the task of detecting motile cells. The study presented in [155] details a CBF measurement technique using the fast fourier transform (FFT). The authors use an iPhone camera to capture images, which are then analyzed on a desktop system while emphasizing the importance of accurate RoI selection, stating that it is "challenging and crucial". Consequently, the authors suggested the need for improvement in RoI selection strategies.

In earlier studies [156,157], CBF was determined by analyzing the variations in pixel intensity. However, the results thus obtained were limited by the specific experimental conditions for video acquisition. In [158, 159], the authors proved the ability to determine the motile areas and CBF using optical flow (OF). OF is described as the representation of the movement of the imaging surface in 2-D images as a result of the actual movement of objects in 3-D space [160]. Nevertheless, it has been emphasized that the accuracy of CBF estimation is highly dependent on the quality of the image and the frequency resolution, which can be affected by factors such as the sampling frequency and the number of frames captured [157, 159–161].

In this study, we propose Deep Cilia, a completely autonomous CBF decay estimation system, whose novelty lies in the following:

- a hybrid WFA/RoI analysis of the video which is completely handled by deep learning models (YOLOv8) that are trained to detect the upper part of a ciliated cell called the ciliated stria. The Deep Cilia architecture is a hybrid between the WFA and the RoI analysis techniques since it analyzes the whole video frame by frame but only returns the frame sections where the ciliated stria is present.

- the ability to correctly estimate the CBF even for low FPS videos through the variation of the light intensity in the frequency domain, calculated using the Short-Time Fourier Transform (STFT).

3.3.2 Materials and Methods

Dataset

All the experiments performed in this study were carried out on a dataset specifically collected for this study. It consist of 119 videos of ciliated cells acquired by an expert at the Unit of Otolaryngology, Department of Clinical and Experimental Medicine, University of Foggia, Italy. The videos were acquired with the use of a Nikon Eclipse 600 (with 1000x magnification) microscope equipped with a camera model MD6iS (Sony IMX236 Sensor).

Each video exhibits a variable number of ciliated cells varying from 1 to more than 5 (in contexts where many cells are amassed). The average video duration in the dataset is 93 seconds, with an upper bound of 352 seconds and the shortest being only 3 seconds long.

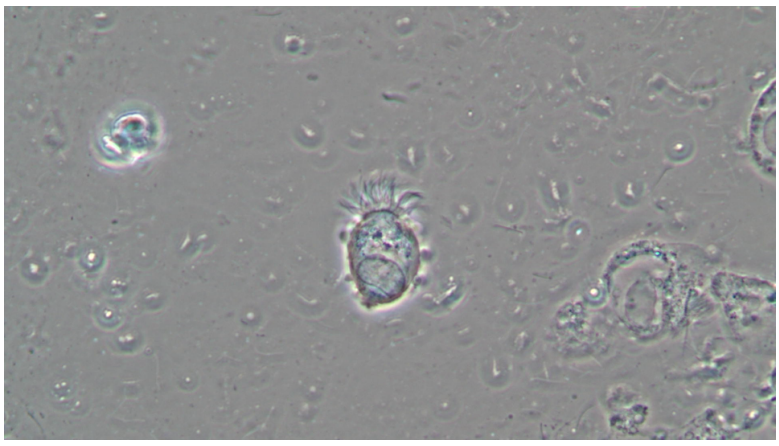
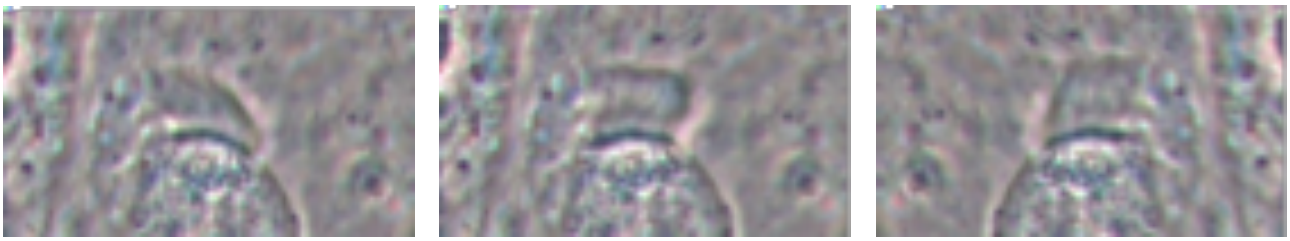


Figure 3.10: Example of frame from one of the dataset's videos.



(a) (b) (c)

Figure 3.11: Manually segmented ciliated strias, used as ground truths when training YOLO.

As previously underlined, the cardinality of ciliated cells recorded in each video can vary. From the 119 videos a total of 246 clearly visible strias (Figure 3.11) were reported, labeled and manually extracted by the authors on each frame.

In addition to the number of cells and video duration, the videos differ from each other in terms of the following:

- Frames per Second (FPS) [$\mu = 23.63$ FPS, $\delta = \pm$ 2.54]
- Steadiness of the camera

Such characteristics, can generally influence the results of the experiments: video recordings with a framerate lower than double the cell CBF (folding frequency for that particular cell) can alter the estimation of the signals' frequency [162]. Similarly a video reporting a few seconds of trembling can irrevocably damage the frequency estimation calculations in that time window (camera shakes will be detected and overlap the metachronal movement of ciliated cells).

To train and evaluate the Ciliated Stria Detection component, all the videos were viewed to allow an expert to manually annotate the ciliated strias of each cell. Then the dataset was partitioned into training, validation and test set containing 89 (185), 18 (36), and 12 (25) videos (cells) respectively.

Preprocessing and Augmentation

Camera lens-equipped microscopes can differ in color calibration, light, resolution, and framerate. For this reason, a standardizing preprocessing phase was defined. Every video in the dataset was first converted to grayscale, and then resized to a resolution of 640x640. These operations make the training phase both consistent and RAMwise light in every part of the analysis thanks to the reduction in both image size and number of color channels (from 3 RGB color channels to one grayscale channel).

This same preprocessing will also be applied to every video that will be submitted to DeepCilia for analysis.

In preparation for the CNN training, the number of images was increased through a data augmentation step. In particular, the images underwent minor modifications between shearing (both vertical and horizontal) and random rotations in the range of +/-180 degrees. This technique has been proven to be an effective method to also prevent overfitting [163, 164], consequently giving the model the chance to better generalize the entities even in different conditions and backgrounds from the ones shown in the training set.

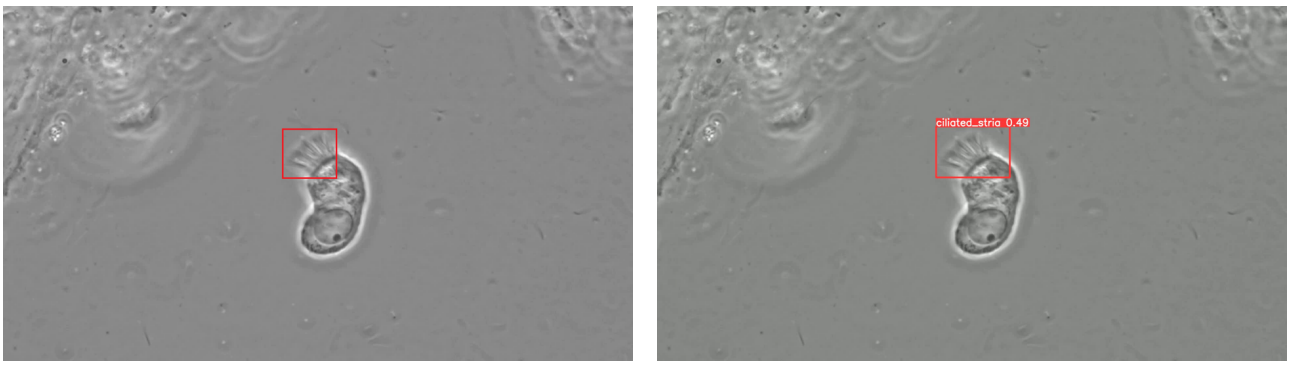
Stria identification

In the current architectural design, DeepCilia leverages a state-of-the-art CNN such as YOLOv8 [165], which is widely regarded for its good performance in object detection.

After training for 200 epochs on the training set, described in Section 3.3.2, YOLOv8 could reliably detect instances of ciliated strias from each frame and return, for each instance, a bounding box of coordinates to track the ciliated stria (Figure 3.12). This tracking allows the engine to not lose track of where the ciliated stria is thus making it robust to small camera movements and translation of the cells in the video.

In order to test YOLOv8 on the task of detection of Ciliated Striae, we submitted to the CNN model a testing set consisting of 5350 images containing 7785 manually-labeled Striae, extracted from 12 videos of the dataset discussed in Section 2.1. For each and every image in the testing set, YOLOv8 yields a bounding box for any plausible Stria detection, and the relative coordinates will be compared to the ground truth in order to obtain the metrics of mAP, Precision and Recall. Different Intersection-Over-Union (IoU) thresholds have been tested, with the best results landing on a precision score of 81% when the IoU is set to 0.5. YOLOv8 also realized a mAP50% of 52% in this testing setting.

Since ciliated cells do not move particularly fast, DeepCilia does not submit to YOLO all the frames of the video but only one per second. This greatly reduces the inference time while still obtaining a precise identification of the ciliated strias.



(a) (b)

Figure 3.12: Comparison between the manually annotated ground truths (a) and the predicted bounding boxes (b).

CBF Estimation

Digital signals (which in this study are frame-by-frame light intensity variations of a video pixel) are characterized by a discrete set of samples that represent the amplitude and time of the signal. These signals may contain noise or unwanted components that need to be attenuated or completely removed in order to extract the desired information. For example, videos recorded with generic cameras can suffer from a variety of issues, such as noise, blur, artifacts, and distortion, which can degrade the analysis. To extract meaningful components of the digital signal, a complex filtering pipeline that elaborates the signal through the selective pass/block of certain band frequencies is needed. Such a pipeline is described in the following from Section 3.3.2 to Section 3.3.2.

Pixel Intensity Through-Time Filtering

The DeepCilia CBF estimation component is based on the analysis of the light variation of each pixel throughout the video. The bounding boxes returned by the Stria identification component that encompass each detected ciliated stria are used to cut windows from the original video V and from it create many "smaller" videos G_k (one for each detection) that will have as resolution the shape of the bounding box.

For each of these new videos we calculate a function called pixel intensity through-time (PITT) which is defined as a function of a single pixel $p(x, y)$ over the whole video such that $PITT(x, y)$ returns a 1-D array of length N where N is the cardinality of the frames of the videos. Each entry $PITT_{i,x,y}$ of the 1-D array is a value in the $[0-255]$ interval (videos are in grayscale, see 3.3.2) representing the light intensity of pixel $p(x, y)$ at the $i-th$ moment (Figure 3.13).

Pixel Wise Filtering

Each of the k grayscale subvideos G_k (and consequently every pixel $G_{k(x,y)}$ composing them) will be processed using a sequence of filters, namely:

- *Frames resizing through interpolation*: Every frame $G_{k(i)}$ of video G_k is resized to become

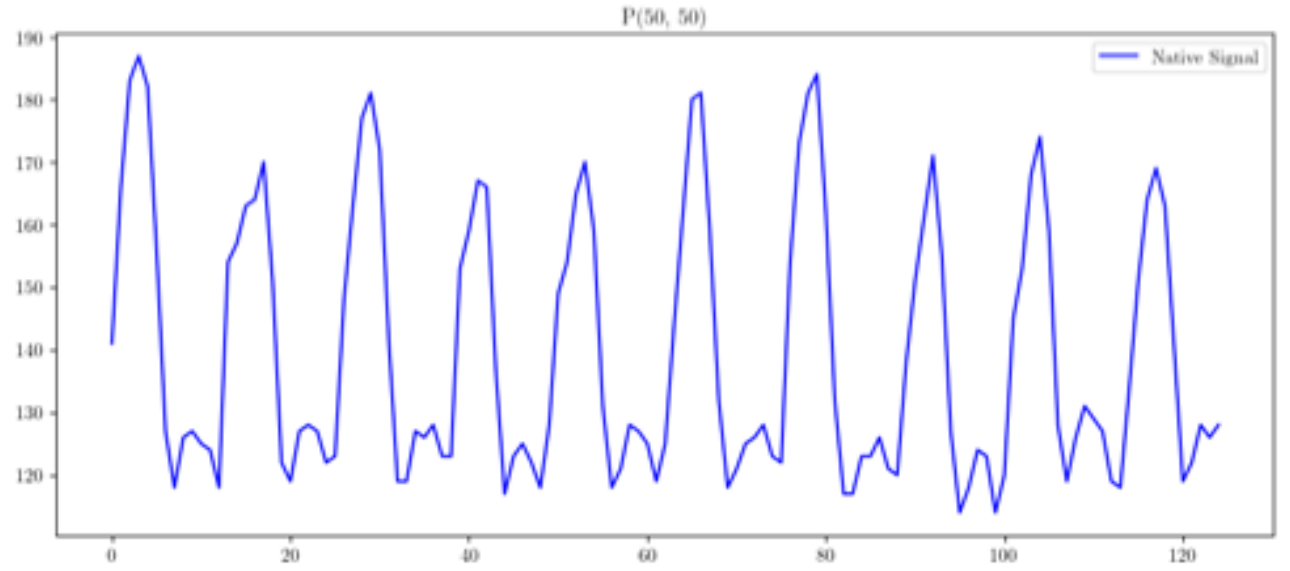


Figure 3.13: 5-second long PITT array of one video pixel.

a square with side length equal to $min(100$, smaller side of $G_k)$, thus returning a new square video called SG_k. This has been proven effective for faster computation and allows videos of the same size to be stacked to be further processed in parallel.

- *Removal of the DC-Offset*: In the context of camera recordings, DC-Offset can be referred to as a baseline signal that is present in the image sensor's output caused by imperfections in the sensor itself or readout circuitry. DC-Offset can be found as a uniform gray or black level in the captured image which reduces the available dynamic range of the sensor and leads to artifacts. A simple filter is applied to mitigate the effects of DC-Offset. This filter, given an SG_k, calculates the average value of each pixel of the video $P(SG_{k,x,y})$ (3.10) and then subtracts it from the values of the pixel itself in every frame (3.11):

$$A_{(SG_{k,x,y})} = \frac{1}{N}\sum_{i=1}^{N} P_i(SG_{k,x,y}) \tag{3.10}$$

$$P_i^*(SG_{k,x,y}) = P_i(SG_{k,x,y}) - A_{(SG_{k,x,y})} \qquad 1 \leq i \leq N \tag{3.11}$$

The above high-pass filter effectively removes the $0\ Hz$ frequency from the frequency spectrum of a signal (Figure 3.14). Figure 3.16b shows that applying this filter returns an image with less contour definition (e.g. the black stria clearly visible in the lower part of Figure 3.16a is no longer present). The intuition behind this filtering technique is that every portion of the image that is constant in position and luminosity shall be removed.

- *First-Order Difference Filter*: a type of digital filter used to compute adjacent samples of a signal. It is a high-pass filter, and is used to detect edges or peaks, to analyze transient behaviours such as sudden changes or perturbations. Consecutive positive (resp. negative) elicitations in the digital signal will be amplified. These results are useful to detect minor camera shakes or image noise, with the downside of injecting a small quantity of noise in

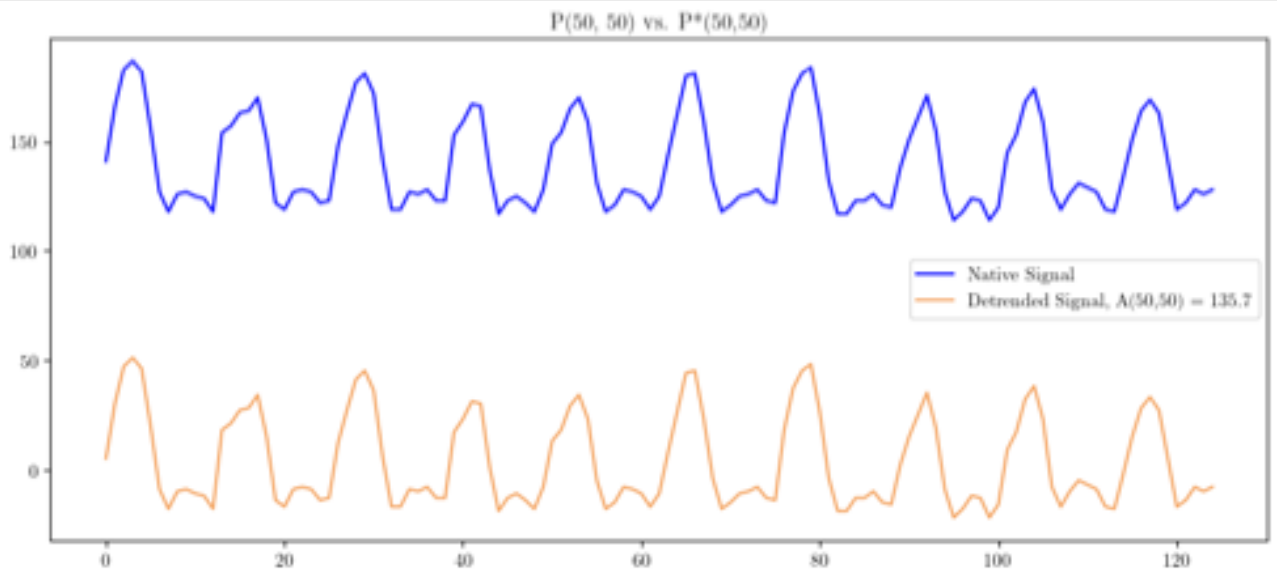


Figure 3.14: Signal from Figure 3.13 compared to the detrended signal. The plot for P*(50,50) is lower because the average value of the pixel throughout the video has been subtracted.

the signal. This filter also operates pixelwise, and it is implemented as a finite impulse response (FIR) digital filter:

$$EQ_i(SG_{k,x,y}) = P_i^*(SG_{k,x,y}) - P_{i-1}^*(SG_{k,x,y}) \qquad 2 \leq i \leq N \tag{3.12}$$

It can be shown that the variance of the noise for each sample obtained by this process is equal to $2\sigma^2$. Suppose that for the i-th frame:

$$P_i^*(SG_{k,x,y}) = C_i^*(SG_{k,x,y}) + N_i^*(SG_{k,x,y}) \tag{3.13}$$

where $C_i^*(SG_{k,x,y})$ is the light intensity value given by the beating ciliated cell in the i-th frame for pixel $p(x,y)$ and $N_i^*(SG_{k,x,y})$ is an additive white Gaussian noise of variance σ^2 in the i-th frame for $p(x,y)$, in the bandwidth which we are currently focusing.

Performing a first-order difference filter as in (3.12) will give:

$$\begin{aligned} &P_i^*(SG_{k,x,y}) - P_{i-1}^*(SG_{k,x,y}) = \\ &\quad (C_i^*(SG_{k,x,y}) + N_i^*(SG_{k,x,y})) - (C_{i-1}^*(SG_{k,x,y}) + N_{i-1}^*(SG_{k,x,y})) = \\ &\quad C_i^*(SG_{k,x,y}) - C_{i-1}^*(SG_{k,x,y}) + N_i^*(SG_{k,x,y}) - N_{i-1}^*(SG_{k,x,y}) \end{aligned} \tag{3.14}$$

Focusing on $N_i^*(SG_{k,x,y}) - N_{i-1}^*(SG_{k,x,y})$, one can note that after the filtering phase the variance has doubled, and now has a value of $2\sigma^2$.

Figure 3.15 shows that higher peaks in the detrended signal (in orange) exhibit a smaller amplitude in the first-difference filtered signal (light blue), and conversely, small peaks are

now much higher. This process, at the cost of the aforementioned noise injection, allows peaks with different amplitudes and the same bandwidth to reduce the amplitude gap itself.

Figure 3.16 shows the effect of this filtering technique on a detrended image. In this case, the most agitated parts of the frame (the cilias) remain visible while static parts are canceled thus making the process of frequency estimation more robust and focused only on the pixels that truly count.

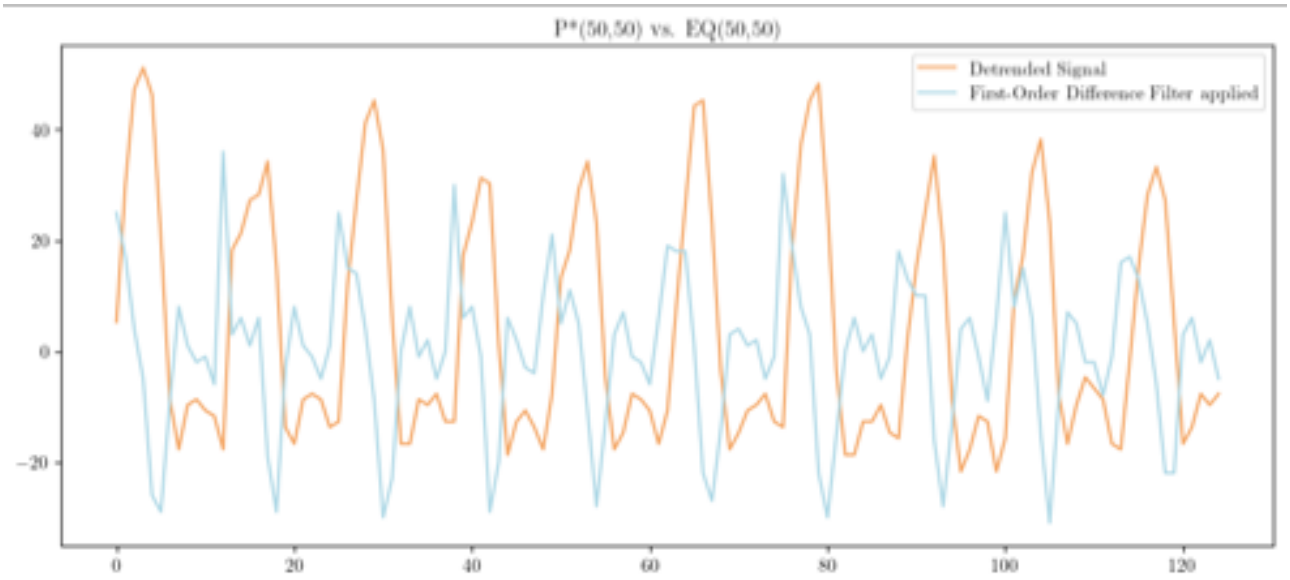
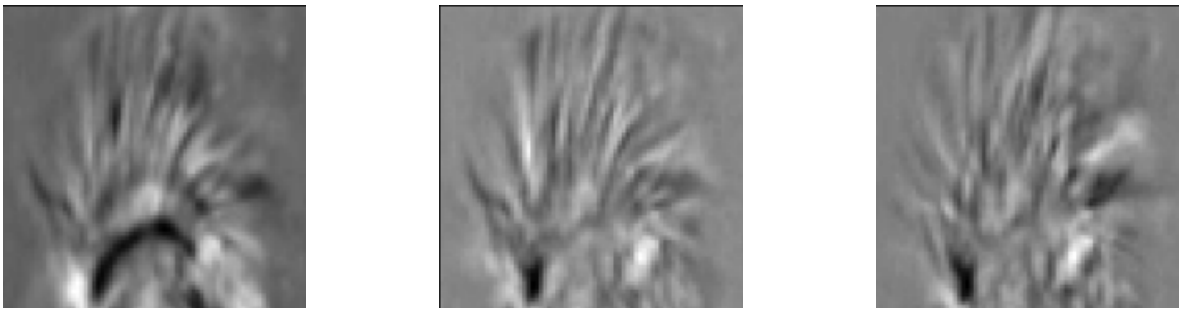


Figure 3.15: First-order difference filter (in light blue) shows smaller peaks than the Detrended signal (in orange), bringing the whole light wave closer to the smaller peaks always preceding the higher ones.



(a) (b) (c)

Figure 3.16: Detection of a ciliated stria made by YOLOv8 (a); same frame but without the camera DC-offset (b); same frame but with first-difference filtering applied: static parts are darker, thus making the process of CBF estimation simpler and more focused on pixels drawing a ciliated stria (c).

Simple Exponential Smoothing

Once every $EQ(SG_{k,x,y})$ PITT for each of the k ciliated cells detected by YOLO has been processed, the next step is to generate two distinct digital signal sequences for each PITT.

Given the values $0 < \alpha < 1$ and $\beta = 1 - \alpha$ (with $\alpha = 0.3$), we define:

- FORWARD PASS:

$$F_i(SG_{k,x,y}) = \begin{cases} EQ_1(SG_{k,x,y}), & \text{if } i = 1 \\ \beta * EQ_i(SG_{k,x,y}) + \alpha * F_{i-1}(SG_{k,x,y}), & \text{if } 1 < i \leq N \end{cases} \tag{3.15}$$

- BACKWARD PASS:

$$B_i(SG_{k,x,y}) = \begin{cases} EQ_N(SG_{k,x,y}), & \text{if } i = N \\ \beta * EQ_i(SG_{k,x,y}) + \alpha * B_{i+1}(SG_{k,x,y}), & \text{if } 1 \leq i < N \end{cases} \tag{3.16}$$

The F (respectively B) sequence is called *forward* (*backward*) *exponential smoothing* (FES-/BES). The properties of FES can be summarized in its ability to stand for values being influenced by what has happened in the immediate past, thus giving the impression that a lag within the signal has been injected. Conversely, BES results in a sequence of values that foresees the real trend of the signal (Figure 3.17).

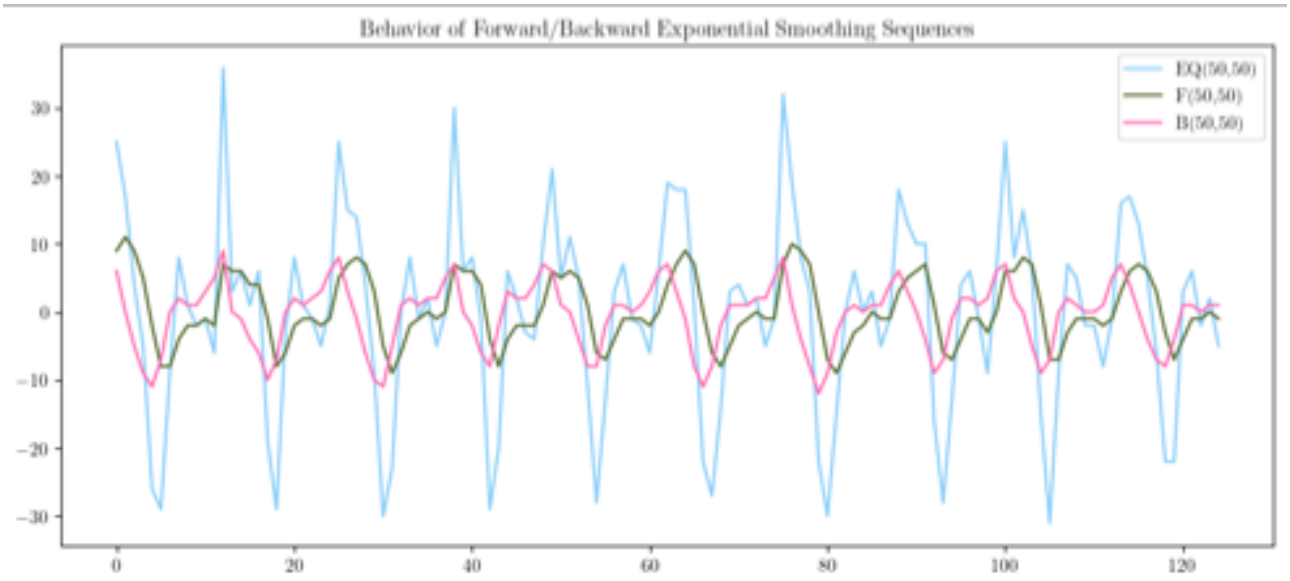


Figure 3.17: FES and BES w.r.t. the detrended signal.

Next, the arithmetical average between the FES and BES is calculated:

$$R_i(SG_{k,x,y}) = \frac{F_i(SG_{k,x,y}) + B_i(SG_{k,x,y})}{2}, \text{ for } 1 \leq i \leq N \tag{3.17}$$

Over $R(SG_{k,x,y})$, the two injected lags from F and B cancel themselves out when the two overlapped signal cross their paths over similar features of the original signals (Figure 3.18).

This exponential averaging function acts as a smoothing factor of small disturbances in the signal. It serves as a low-pass filter where the number of high frequencies flattened is proportional to the value of α. The significant, yet precise dampening factor of the *average exponential smoothing* counterbalances the noise injected by the *first-order difference*. This makes the signal less prone to disturbances while still replicating the sinusoidal behavior of the

original chirped signal. This in turn, allows for a more robust signal analysis without altering the phase (Figure 3.18).

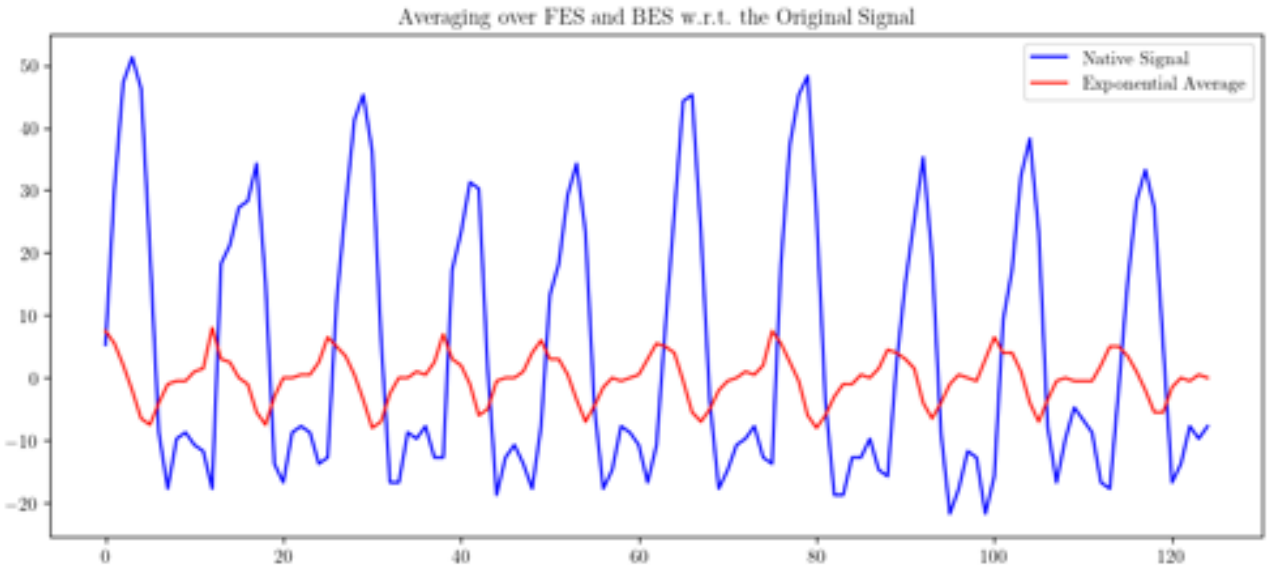


Figure 3.18: Averaged exponential smoothed signal (red) and the detrended signal (blue).

Short Time Fourier Transform

The DeepCilia pipeline ends with the application of the *Short-Time Fourier Transform* (STFT) on every signal $R(SG_{k,x,y})$. This generates an array of results with cardinality $m = \frac{L - windowsize}{hopsize} + 1$ for each of the $R(SG_{k,x,y})$ where L is the length of the signal (windowsize = 128, hopsize = 0).

From each of the m windows, the principal frequency is extracted by calculating the mode between the frequency values of all pixels. This allows DeepCilia to plot the CBF decay through the entire length of the video of the ciliated cell.

3.3.3 Results

The pipeline proposed and explained in Section 3.3.2 was validated by submitting 20 videos of ciliated strias manually cut-out from 20 different videos of the dataset.

To evaluate the DeepCilia pipeline in different working contexts, between these 20 cutout videos, some were purposely extracted from video with very different lengths (15 seconds to 5 minutes) and different quality (cells very close to each other, minor camera shakes, slight changes of focus).

The prediction of the CBF decay obtained by the pipeline was compared with a ground truth CBF that was estimated by the authors by playing in slow motion the 20 videos, keeping track of the full ciliary motion to count a single beat. The final CBF ground truth estimate of each part was selected by majority vote between the manual counts made by the authors.

The results were evaluated and compared in terms of root mean square error (RMSE) and of mean bias error (MBE). This is because the former is useful to detect where and when the

algorithm's estimated CBFs are showing large gaps from the ground truth, while the latter is useful to understand if the pipeline tends to over/underestimate the true ciliary beat.

Table 1 reports the results achieved, while Figure 3.19 reports two example plots of ground truth CBF (blue) against the CBF estimated by DeepCilia (red) of two different ciliated cells in the two longest videos of the dataset.

Threshold	**Initial F1**	**F1 pruned**	**# Prototypes**	**% pruned**
0.11	0.549	0.523	67	87%
0.51	0.549	0.535	48	91%

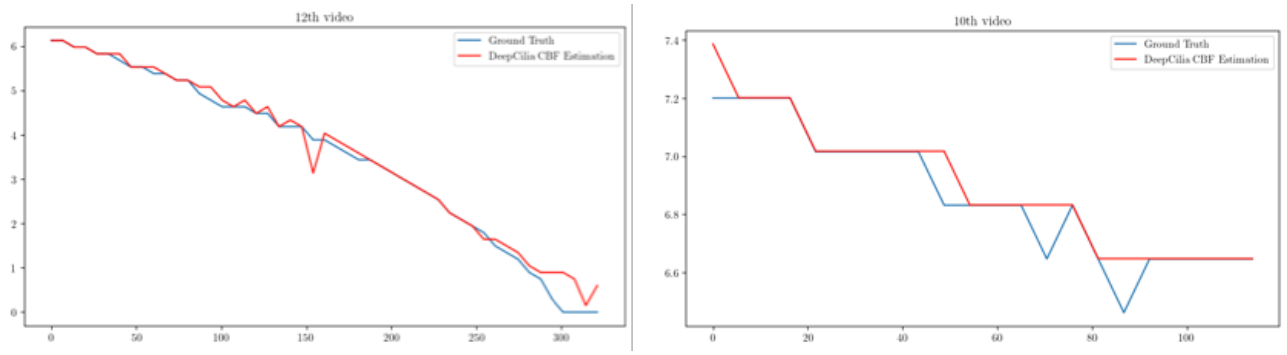
Table 3.13: Impact of pruning on f1-score and # prototypes.

Video Id	**Length (s)**	**Fps**	**RMSE (Hz)**	**MBE (Hz)**
1	21.48	25	0.099	0.060
2	38.60	25	0.174	0.112
3	28.80	25	0.008	0.008
4	61.64	31	0.109	0.052
5	69.78	31	0.154	0.101
6	29.76	25	0.147	0.118
7	60.76	24	0.259	0.225
8	121.38	24	0.172	0.113
9	71.78	24	0.111	0.067
10	119.53	24	0.079	0.035
11	82.74	24	0.077	0.033
12	328.43	19	0.252	0.134
13	67.92	24	0.175	0.113
14	38.44	25	0.127	0.088
15	28.99	24	0.127	0.208
16	17.32	25	0.075	0.038
17	24.36	25	0.089	0.049
18	31.04	25	0.113	0.065
19	64.38	19	0.238	0.168
20	46.05	19	0.117	0.098
Avg	-	-	**0.135**	**0.094**

Table 3.14: CBF estimation algorithm performance scores

After a thorough analysis of these 20 videos in the test set, the videos on which DeepCilia reports a brief gap ($RMSE > \tilde{0}.15$) with respect to the ground truth are only the ones where huge camera jolts are recorded either repeatedly or for an extended period of time (more than 1 second in any video window). This shows the overall resilience of the system to small camera shakes and video impurities.

The focus of this study is also to achieve good performances in terms of time efficiency. Keeping in mind this objective, the architecture of DeepCilia has been designed to be GPU-based to allow parallelization of the different filtering and STFT operations that have to be performed



(a) (b)

Figure 3.19: GT CBF (in blue) against estimated CBF (in red) of video 12 (a) and video 10 (b).

over every single pixel of each video. This optimization guarantees precise results in a matter of seconds. Specifically, the CBF estimation on each of the 20 videos of the test dataset was performed in less than two seconds.

DeepCilia has been implemented with Python (ver. 3.8.12), using the OpenCV library for image manipulations and tensor-computing enabled Pytorch framework for both YOLOv8 deployment, frames parallel preprocessing and STFT computation.

All the experiments were carried out on a machine configured with an Intel-7700HQ CPU, NVIDIA 1050 GPU and 16 GB RAM.

3.3.4 Discussion

The evaluation performed on the 20 videos experimental sample revealed its ability to accurately estimate the ciliary beat frequency decay over time, in fact the average RMSE of only $0.135Hz$ between the estimation and the ground truth CBF across all 20 videos indicates how the system can precisely track changes in the ciliary beat frequency over time. The heterogeneity of the test dataset and error independence to the video lenght and Fps, allow to demonstrate the robustness of DeepCilia to real-word variation in video quality; moreover as reported in the previous section, the system resilience to video impurities like small camera shake can be shown by noting that larger error occurred only in videos where huge camera jolts are recorded repeatedly or for an extended period of time. Also, by looking at the Mean Bias Error values, it can be asserted that estimated values are indeed centered around the true value and not systematically over or underestimated.

These promising results can be attributed to the proposed image filtering a STFT pipeline where the combination of filters enhanced relevant signals while suppressing noise and artifacts allowing the STFT to faithfully extract the principal frequency of the ciliary beat even from lower quality videos, thus proving the pipeline to be capable of accurately estimate the CBF. Moreover the good performances of YOLOv8 in handling the stria detection confirmed the intuition of proposing an hybrid WFA/RoI analysis approach that was able to focus computations only on relevant regions while still analyzing the whole frame, avoiding the selection bias of traditional

RoI methods.

3.3.5 Conclusions

In conclusion, this last contribution introduced DeepCilia, a fully automated and efficient pipeline for the estimation of ciliary beat frequency from microscopy videos of nasal cytology samples. By combining a YOLOv8-based detector with signal analysis through Short-Time Fourier Transform, the system demonstrated that it is possible to achieve accurate, fast, and reproducible measurements of ciliary activity without requiring specialized hardware or manual intervention. This represents a decisive advance in the automation of rhinocytology, extending the scope of artificial intelligence from static cell classification to the dynamic characterization of cellular function.

Beyond its methodological contributions, DeepCilia highlights the broader potential of AI-driven approaches to tackle diverse diagnostic tasks in rhinology, ranging from inflammatory profiling to functional assessment of the mucociliary system. At the same time, it underscores the importance of usability: for these methods to reach clinical practice, they must be delivered within tools that are accessible, transparent, and seamlessly integrated into existing diagnostic workflows.

Within the framework of this thesis, the third paper consolidates the progression from dataset creation to interpretable pipelines and advanced functional analysis. The natural next step is the translation of these research outcomes into a practical platform that clinicians can directly use. For this reason, the following chapter 4 describes the design and development of a dedicated web application, which integrates the proposed models into an accessible interface, enabling physicians to explore their diagnostic potential in real-world conditions.

Chapter 4

WebApp Developing and Documentation

4.1 Technical Overview

4.1.1 Introduction

The web application developed for this PhD project is designed as a specialized tool to support clinicians in the analysis of nasal cytology images, a diagnostic technique widely used to evaluate nasal inflammatory patterns and guide personalized treatment in rhinology. Traditional cytology workflows require manual inspection of microscope slides, with clinicians counting and classifying cells one by one. This is not only time-consuming but also prone to human error, and it places a significant cognitive burden on medical staff. The primary aim of this application is to automate cell detection and classification through a machine learning–based approach, providing clinicians with a rapid, reliable, and reproducible cytotype count.

The platform goes beyond automated image analysis by integrating clinical data management features. In addition to processing cytology images, the system allows clinicians to track patients longitudinally, storing anamnesis, diagnoses, and visit histories in a centralized and secure database. This dual functionality ensures that the application is not only a research demonstration of image analysis algorithms but also a practical clinical support tool that can be readily adopted in hospital workflows.

The software is designed as a fully web-based application to maximize accessibility: users can log in from any computer on the hospital network without local installation, and all computationally intensive operations, such as machine learning inference, are executed on a centralized server. The system provides an intuitive, clinician-friendly interface that simplifies the process of uploading and reviewing cytology images while offering the possibility to correct automated predictions if needed, thus maintaining clinical control and reliability.

4.1.2 System Architecture

The system is implemented as a modular client–server architecture, with a strict separation between the user interface, application logic, and data storage. This separation not only improves maintainability but also enables the platform to scale as new features and algorithms are integrated over time.

At the frontend, the user interacts with a web interface developed in Angular, a widely used TypeScript framework for building robust, component-based applications. Angular was chosen because of its strong support for enterprise-scale projects, its well-defined architecture (Model–View–ViewModel), and its extensive ecosystem of libraries. By adopting Material Design guidelines, the interface achieves a professional and consistent look while also ensuring accessibility and usability for medical professionals, who may not have advanced technical expertise. The frontend is designed to be lightweight, responsive, and capable of dynamically rendering large datasets, such as cytology image thumbnails and detailed prediction results.

The backend is developed in Python using FastAPI, a high-performance framework for building RESTful APIs. FastAPI was selected for its strong typing support, speed, and automatic generation of interactive API documentation, which is particularly valuable for future developers or researchers who may wish to extend the system. The backend is responsible for receiving requests from the frontend, processing cytology images using deep learning models, managing authentication, and interfacing with the database.

A PostgreSQL relational database underpins the application's data management layer. PostgreSQL was chosen because of its robustness, scalability, and ability to handle complex relationships between entities, which is essential for modeling patients, their clinical visits, cytology images, and diagnostic data. Data integrity is enforced through a well-designed schema, which ensures that clinical records remain consistent even in the presence of concurrent users or network failures.

The entire platform is orchestrated by Nginx, which serves as a reverse proxy, handling client requests, routing traffic between the frontend and backend, and providing load balancing capabilities. Nginx was also configured to manage SSL certificates and enforce secure HTTPS connections, which is essential in a healthcare context to ensure data privacy.

Finally, all components of the system—frontend, backend, and database—are packaged into Docker containers and deployed through Docker Compose on an internal server. This deployment strategy simplifies installation, updates, and reproducibility. By isolating services in separate containers, developers can upgrade or modify one part of the application without impacting others, significantly reducing maintenance complexity.

4.1.3 Technology Stack

The selection of technologies for this project reflects a deliberate balance between performance, maintainability, and ease of integration with machine learning models.

On the frontend, Angular provides a mature ecosystem with built-in support for reactive programming through RxJS, which is particularly useful for real-time updates (e.g., dynamically displaying image processing results as they become available). Angular's CLI tools streamline development, testing, and production builds, reducing time spent on configuration. Material Design was chosen not only for its aesthetic qualities but also because it provides a consistent, accessible user experience for clinicians.

The backend relies on FastAPI, which was preferred over alternatives such as Flask or Django for several reasons. FastAPI's async capabilities allow the application to handle multiple concurrent requests efficiently, a key requirement when multiple clinicians are uploading large image datasets simultaneously. Furthermore, FastAPI's strong typing system improves code reliability, making it easier to maintain and extend in a research setting where algorithms and endpoints may evolve rapidly.

For machine learning inference, the backend integrates a YOLO (You Only Look Once) object detection model trained on nasal cytology images. YOLO was selected because of its ability to perform real-time detection with high accuracy, making it suitable for a clinical workflow where responsiveness is critical. The model is optimized to detect and classify multiple cell types, enabling a detailed cytotype count for each uploaded image.

PostgreSQL serves as the data backbone of the system, offering support for complex relational queries and strong transactional guarantees. The schema was carefully designed to model patients, visits, image metadata, and prediction outputs in a way that is both normalized (to avoid redundancy) and optimized for query performance.

The entire application stack is deployed using Docker Compose, with three distinct containers—one each for the frontend, backend, and database. This containerized approach ensures that the application runs consistently across environments and allows for seamless scaling. Nginx sits at the front of the stack, handling all incoming requests, providing caching for static assets, and serving as a reverse proxy to ensure secure and efficient communication between services.

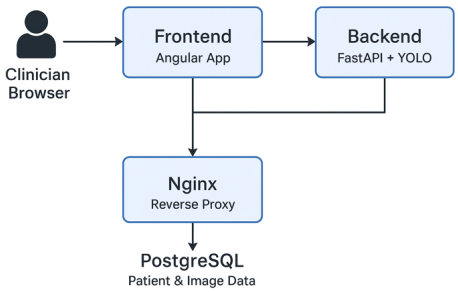


Figure 4.1: Deployement diagram.

4.1.4 Design and Implementation Details

The web application was designed following a user-centered approach, with the primary goal of guiding physicians through the same logical sequence of actions they typically perform during a clinical visit. The interface emphasizes simplicity and intuitiveness, minimizing cognitive load and ensuring that users can interact with the system without prior technical training. Each function—such as patient data entry, image upload, analysis execution, and result visualization—was structured to mirror the natural workflow of nasal cytology assessment. By reproducing the familiar steps of the diagnostic process in a digital environment, the platform allows physicians to focus on clinical interpretation rather than on system operation, ensuring both efficiency and accessibility in everyday practice.

The frontend is organized into reusable Angular components, each responsible for a specific section of the user interface. For example, there are dedicated components for displaying patient records, image galleries, and prediction results. This modular structure makes it straightforward to introduce new functionality or modify the appearance of individual sections without affecting the entire system.

The backend implements a layered architecture, separating the API endpoints, business logic, and data access layers. This approach increases maintainability by ensuring that changes in one layer do not propagate unintended consequences to others. For instance, updates to the YOLO inference module can be made independently of the database logic or API endpoints, as long as data contracts remain consistent.

The YOLO model integration required specific optimizations to meet the performance constraints of a clinical workflow. Images are pre-processed upon upload to ensure uniformity, and the inference pipeline is designed to handle batch processing when multiple images are submitted simultaneously. Prediction outputs are stored alongside their corresponding image metadata, enabling traceability and reproducibility—two key requirements in medical research.

The web platform adopts a server-side authentication system based on JSON Web Tokens (JWT), designed to combine the scalability of stateless authentication with the security and usability of traditional session-based approaches. Each authenticated session is represented by a JWT stored in a secure, HTTP-only cookie, inaccessible to client-side scripts. This prevents token theft via cross-site scripting (XSS) and ensures that authentication behaves transparently for the user. The server validates the token on every request, maintaining a stateless architecture while allowing the session to persist across browser closures. All communication occurs over HTTPS, and the cookie is configured with the Secure, HttpOnly, and SameSite=Strict flags to guarantee data integrity and protection against common web attacks. Each session lasts six hours, after which the JWT automatically expires and the user is redirected to the login page at the next authenticated request. This duration balances usability and security, ensuring convenient workflows without long-term exposure. Overall, this approach provides a robust, scalable, and secure authentication mechanism suited to the platform's medical context, where user identity and data protection are critical.

Data privacy was a central consideration in the system's implementation. Patient data is stored exclusively on internal servers, and all communication between the frontend and backend is secured via HTTPS. The database schema incorporates strict access control, ensuring that sensitive patient information is only accessible to authorized users.

Database Design

The web platform relies on a relational database designed to efficiently manage clinical data, patient information, and analysis results generated by the automated cell count AI model. The database was implemented using SQLAlchemy ORM, which provides a Pythonic abstraction over both PostgreSQL (for production) and SQLite (for development and testing). This dual setup, automatically selected at runtime, ensures scalability and compatibility across different deployment environments. The schema follows a normalized and modular structure, organized around several key entities that reflect the core components of the clinical workflow: users, patients, visits, anamneses, diagnoses, images, and detections.

The User table stores physician credentials and identifying information (name, surname, and email), with passwords securely stored as hashes. Each user is associated with one or more patients, establishing a one-to-many relationship through a foreign key. The Patient entity contains demographic and contact information, including unique identifiers such as tax code, ensuring unambiguous patient tracking across visits.

Every clinical examination is represented by a record in the Visit table, which links to both the Patient and associated medical data. Each visit may include one or more Anamnesis and Diagnosis entries, capturing clinical history, symptomatology, test outcomes, and medical conclusions. These relationships allow for comprehensive longitudinal tracking of the patient's diagnostic process.

Cytological and functional data are managed through the Image, Detection, and DetectionResults tables. The Image table stores binary image data directly within the database for integrity and portability, while the Detection table holds the results of automated cell detection and classification, including bounding box coordinates, confidence scores, and labels. Each detection references a specific image through a foreign key relationship, preserving traceability between raw data and analytical output. The DetectionResults entity groups multiple detections within a single processing batch, associating them with a visit and recording the execution timestamp.

Each visit can refer to more than one Anamnesis/DetectionResults/Diagnosis because it was a design choice not to modify these entities, but rather to create a new instance when the user decides to update them, so as to maintain a history. This solution not only allows us to keep track of all changes made for security reasons, but also to make older versions available to the end user in the future.

To support iterative model refinement and version control, the schema includes UpdatedImage and UpdatedDetection tables. These store processed or re-analyzed images together with their

computed hashes (for deduplication) and corresponding detection results, allowing historical comparison between different algorithmic versions.

The design prioritizes referential integrity, data consistency, and scalability. Foreign key constraints ensure reliable linkage between medical records and analytical results, while the use of indexed columns optimizes query performance for frequent operations such as patient lookup or retrieval of previous visits. Furthermore, the adoption of SQLAlchemy ORM simplifies data access patterns within the backend API, providing transactional safety through session management and automatic schema creation via the create_tables() routine. Overall, the database architecture provides a robust foundation for the platform's analytical and clinical functionalities. It ensures secure data persistence, logical organization of medical records, and seamless integration between user management, patient tracking, and automated diagnostic pipelines.

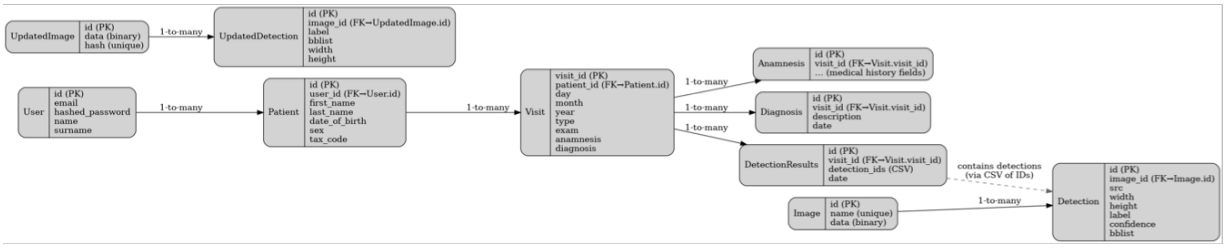


Figure 4.2: Database entity-relation diagram.

4.1.5 Development Workflow

Development was organized around a dual-repository structure, with separate GitHub repositories for the frontend and backend. This separation allows for independent versioning and testing of each component, reducing integration complexity. The repositories follow a feature-branch workflow: new functionality is implemented in isolated branches, reviewed, and then merged into the main branch. This approach ensures stability and facilitates collaborative development.

Testing played a critical role throughout the project. Unit tests were written for key API endpoints to verify that data retrieval, image uploads, and inference results were correctly handled. While automated testing is essential for backend reliability, frontend testing relied on a combination of automated tools and manual reviews to ensure usability and clinical relevance. Clinicians were involved in iterative testing cycles, providing valuable feedback that shaped the interface design and usability improvements.

Deployment is simplified through the use of Docker Compose. Once a new version is ready, updated containers are built and deployed on the internal server with minimal downtime. This workflow ensures that development changes can be propagated to the production environment in a controlled and reproducible manner.

4.1.6 Relation to Research

This application is more than a software project; it is a practical embodiment of the research methodologies developed in this thesis. The integration of YOLO-based object detection for nasal cytology classification demonstrates the viability of applying state-of-the-art computer vision techniques to a highly specialized clinical problem. By providing a user-friendly interface for clinicians, the application bridges the gap between machine learning research and real-world clinical practice.

Furthermore, the system offers a valuable platform for future research. All data, predictions, and user corrections are stored in a structured manner, enabling longitudinal studies, model retraining, and further validation. In this sense, the application not only supports clinical decision-making but also serves as a research tool that can accelerate innovation in digital cytology.

4.2 User Guide

4.2.1 Accessing the Webapp

The developed web application is deployed on a dedicated internal server and is accessible to any user through a standard web browser, without requiring any software installation or configuration on the client side. This deployment strategy was chosen to ensure maximum accessibility: clinicians, researchers, or collaborators can interact with the platform simply by navigating to its URL, regardless of the operating system or device they are using. By consolidating all computation, including image processing, prediction generation, and database management, onto a centralized server, this approach eliminates the need for local computational resources, reduces compatibility issues, and simplifies software maintenance and updates.

The application is reachable at the following address: https://beyond.di.uniba.it (Figure 4.3)

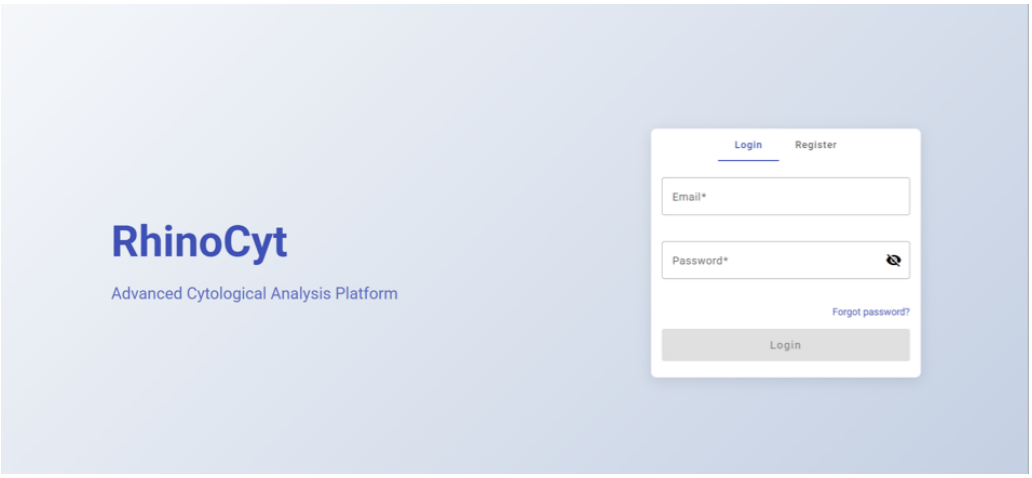


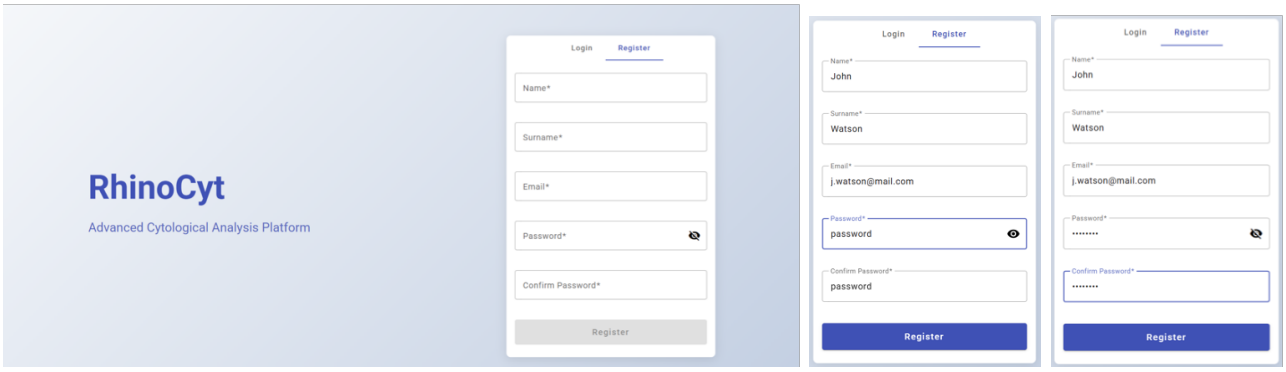
Figure 4.3: Login page.

Although the system is publicly accessible, user registration is mandatory. This requirement serves both administrative and security purposes:

- It allows system administrators to track application usage and monitor potential misuse.
- It enables secure, personalized access, ensuring that each user's uploaded images, diagnostic notes, and patient records are tied to their own account.

To register, new users must provide their name, surname and email and choose a secure password (Figure 4.4). All the fields in the registration form are mandatory, so the "Register" button will not be enabled until all field are filled. The email field of the form always checks the sintax of the text inserted in it, assuring that it follows a valid email Regular Expression (RegEx). If the RegEx is not satisfied, the field will turn red and an error message will be showed under the field itself, and the "Register" button will turn off (Figure 4.5b). The "password" and "confirm password" have their content hidden by deafault for privacy purposes, but it can be revealed (and hidden again) by clicking on the eye icon on the right of the password field (Figure 4.4b). These last two fields require to have identical input; in case this condition is not respected the "Register" button will be turned off. As we said when describing the DB entities, we decided that to each email has to correspond only one account, so when submitting the registration form, the system checks if the email has already be used to register another account; if that happens an error is shown, warning the user that another user has already registered with that email, suggesting him/her to change the email.

If everything goes well the user will be redirected to the main page (Figure 4.6).



(a) (b) (c)

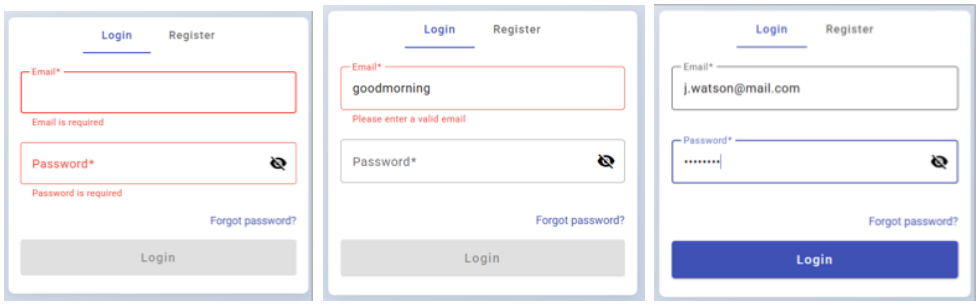
Figure 4.4: Registration page.

If already registered, on their next access, users can log in using their credentials: **email** and **password**. As in the registration form, the login form requires that both field are compiled before allowing the user to press the "Login" button (Figure 4.5a); again, as in the registration form, on the email field is imposed the requirement to satisfy the email RegEx. (Figure 4.5b). When all field are compiled (Figure 4.5c) and the credentials are correct the user will be redirected to the main page (Figure 4.6).

Each session is authenticated on the server side. The platform employs JWT (JSON Web Tokens) for stateless authentication, but instead of manually managing tokens on the client side, the JWT is stored in a secure, HTTP-only cookie, To make the process feel like traditional session-based authentication while maintaining the benefits of JWT. All connections are encrypted via HTTPS, guaranteeing the confidentiality and integrity of transmitted information.

Each authenticated session (unless terminated early by clicking on the logout button 4.9b) last 6 hours. Within this time period the user is authenticated on the server and so even if the browser or the current tab is closed and reopened the website will behave as the closure never happened and let the user perform all the operations that require authentication.

As soon as the 6 hours expire, the user will no more be authenticated so, at the next page update or at the next request requiring authentication the frontend will be notified and bring back the user to the login page, without any way to go back to previous pages that require authentication.



(a) (b) (c)

Figure 4.5: Login form stages: requesting to insert fields (a), not valid email inserted (b), correctly compiled.

From a usability perspective, the system is designed to be platform-agnostic: it runs reliably in modern web browsers such as Google Chrome, Firefox, Safari, and its responsive interface adapts to tablets and mobile devices. However, for optimal use — particularly for reviewing and correcting YOLO predictions on high-resolution cytology images — a desktop browser with a larger screen is strongly recommended.

Because of its web-based nature, users can access the platform from anywhere with network connectivity.

This combination of public availability, mandatory registration, and secure server-side authentication strikes a balance between accessibility and data protection, aligning with the system's purpose as a clinical decision-support tool.

4.2.2 Getting Started

After correctly logging in the user will be redirected to the main page of the web-app.

Overview of homepage

The web-app is designed to guide the physician using it in performing the actions he/she usually does during a visit in the most natural and simple way; so the first page of the web-app is the one in charge of the user the list of its patients.

At first, the page will be empty, since a new user has yet to register all its patients, so a message will be shown to the user suggesting to click the "Add New Patient" button to add a new patient (Figure 4.6a).

This button open a popup form where patient's information are requested (Figure 4.6b), these information are:

- **Name**,
- **Surname**,
- Address,
- City,
- **Date of birth**,
- Phone number,
- Email,
- **Sex**,
- **Tax code**.

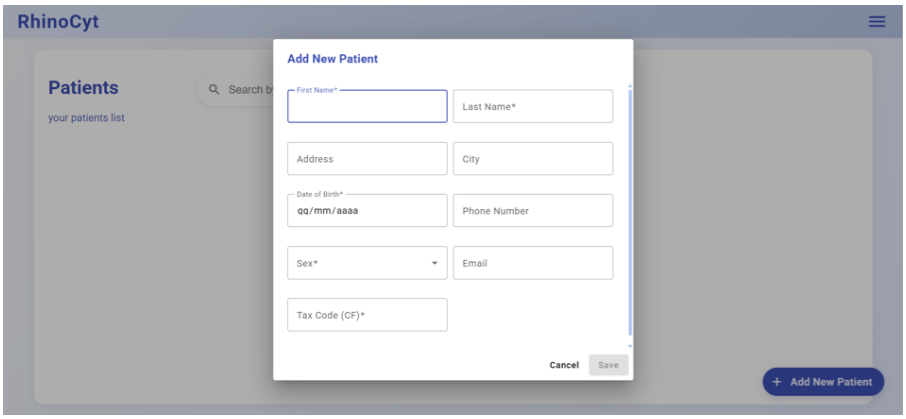
To adhere to the privacy principle of *data minimization* only information in bold are compulsory, the other contact and living information are at discretion of the patient. As soon as the five required fields are correctly compiled, the "Save" button will become active, and clicking it will finalize the patient creation. Clicking the "Cancel" button or clicking outside of the form will close it and all fields are deleted.

When a patient is created his/her card will be added into the main page. A patient card is a small card the shows a male/female avatar in relation to the sex of the patient and then the patient's name, surname and date of birth. Since the main page can become quite crowded with patient cards (Figure 4.6c) a search bar has been added on the top of the page; this bar allows for a dynamic search between the patients showing to the user only the ones that respect the search string criterion (Figure 4.6d).

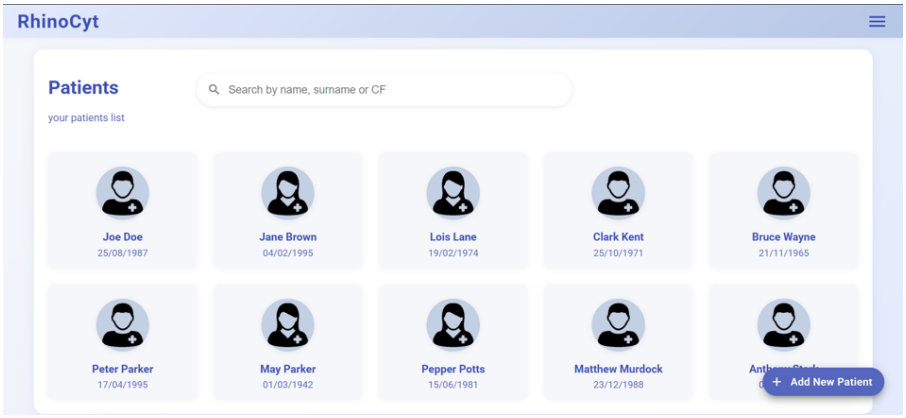
Overview of patient page

By clicking on a patient card the user will be redirected to the patient page that, as the name suggest, shows the patient information alongside the list of the visits recorded for that patient.

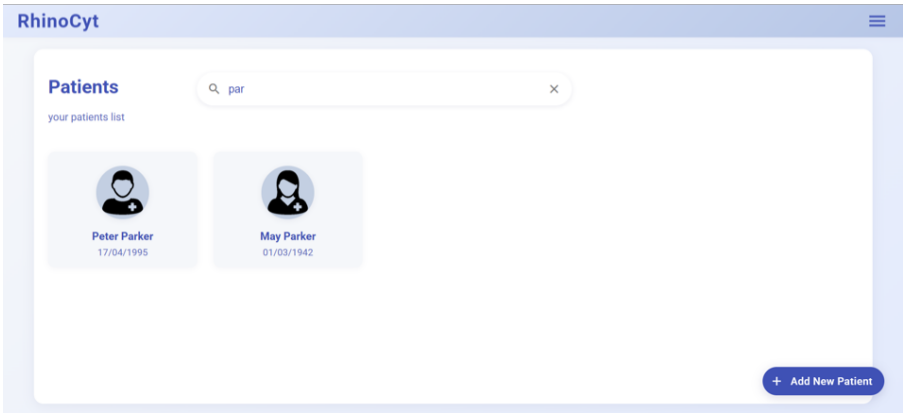
(a)



(b)



(c)

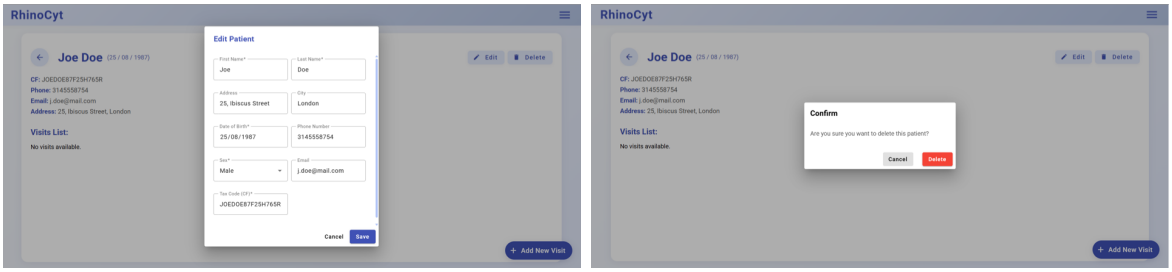


(d)

Figure 4.6: Main page of the webapp at the first start (a), new patient form (b), populated (c) and patient search feature (d).

As before, the first time the page is opened (for each new patient) the visit list will be empty and the page will show only the information that the user registered in the new patient form (Figure 4.7a).

(a)



(b) (c)

Figure 4.7: Patient page reporting patient data and visits list for that patient (a), the edit patient form (b), the delete patient confirmation box (c).

In this page the user can perform four actions:

- **Edit the patient information** - By clicking on the "Edit" button locate at the top-right of the page, a form (similar to the new patient form) will open (Figure 4.7b), the form will show all the patient information in editable text boxes; when finished editing the user can click "Save" to update the patient information or click "Cancel" or outside the form to close the form leaving the patient information to their previous state.

- **Delete the user** If the user wants to delete the patient displayed in the patient page, he/she has to click on the "Delete" button, just to the right of the "Edit" button. Since this is a dangerous action, clicking on this button will open a confirmation form (Figure 4.7c), by clicking again on the red "Delete" button, the system is sure that the user is willingly deleting the patient (along with all of its visits) and the button was not clicked by error. Clicking on "Cancel" or outside the form will close it and leave everything as it was.

- **Register a new visit** - The floating button a the bottom-right of the page allows the user to register a new visit for the current patient. Clicking it will open a small form (Figure 4.8a) asking the user if the visit he/she is currently creating is a "Routine Checkup" or

a "Follow up" visit (Figure 4.8b); after selecting one of these two categories the "Save" button will activate and clicking it will create a new visit. The date of the visit will be inferred by the current date. Clicking on "Cancel" or outside the form will close it and will not create any new visit.

A visit in the visit list will be displayed in a small card, similar to a patient card. A visit card contains the date of the visit (in a format similar to a desk calendar) , the category of the visit and 3 icons: (from left to right) the anamnesis icon, the cell count icon and the diagnosis icon. These 3 icons will initially be grayed out, but can light up if an anamnesis, cell count of diagnosis is present in the respective visit (more in section 4.2.3). A populated visit list will look like in Figure 4.8c.

- **Navigate to a specific visit** - Clicking on a visit card will open the visit page of that specific visit, showing the user the main features of this project (discussed in section 4.2.3).

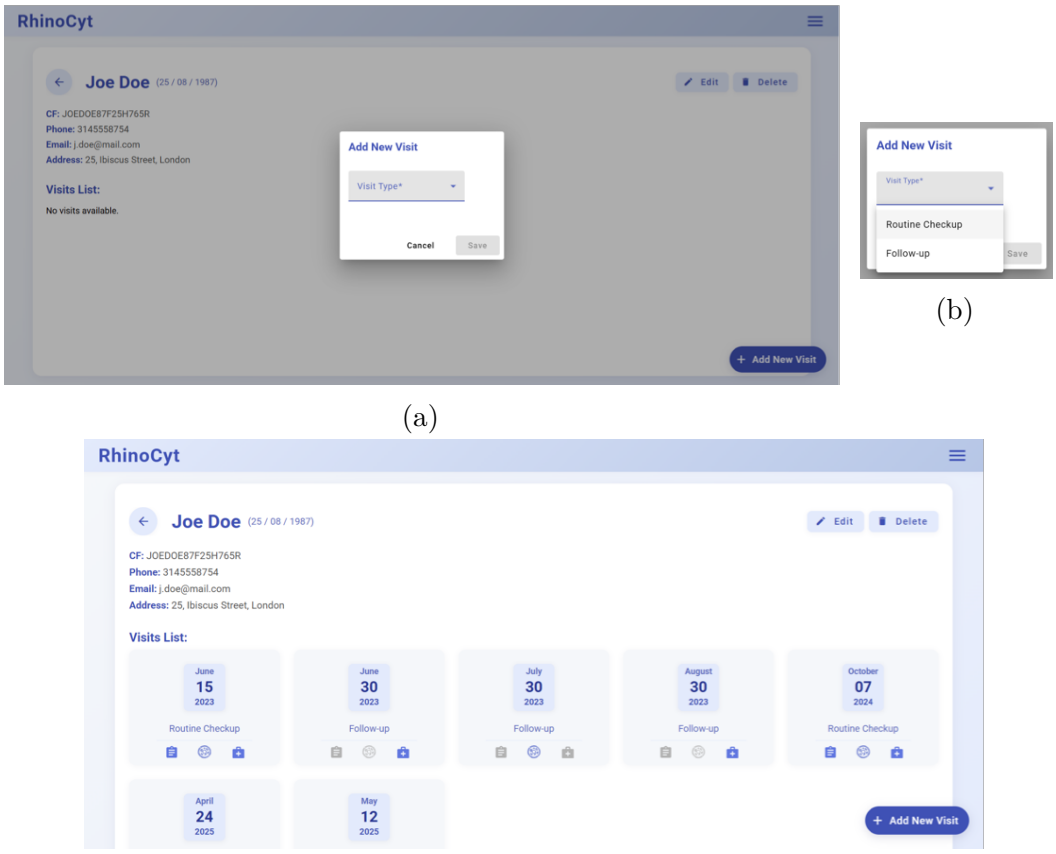


Figure 4.8: Form to add a new visit for the patient(a), the two values of visit type (b) and the populated patient page (c).

Key navigation elements

Before discussing the main features of the web-app it is important to specify what are the navigation elements that allow the user to move between the pages with ease.

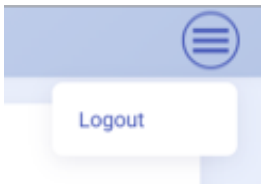
To allow a simple and easy to comprehend navigation, it was deliberately chosen to add only two navigation elements.

The first, and most prominent element, is the back arrow (Figure 4.9a). The back arrow is located in the top left corner of the patient page, the visit page and the cell count page and takes the user back to the previous page (the homepage, the patient page and the visit page respectively).

It is not present on the homepage since to go back to the login page another navigation element is used.

The logout button is the second navigation element. It allows the current user to disconnect to the system and tells the server to delete the authorization token even if the six hours have not passed. The logout operation brings the user to the login page, that is the only page that can be accessed without authorization. The login button can be found by opening the hamburger menu that is located at the right of the top bar (Figure 4.9b), visible in all the pages (except the login page).

(a)



(b)

Figure 4.9: Main navigation elements (circled) (a), hamburger menu, that shows the logout button (b).

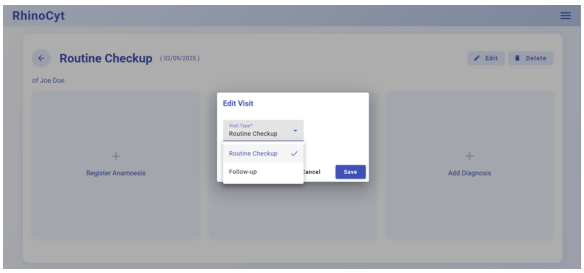
4.2.3 Core Features Walkthrough

In this section the three core features of the web application will be explained. In addition to maintaining a patient list with their relative visits, this application was built to allow otolaryngologists to perform a much faster and accurate cell count, compare the results with the anamnesis to then elaborate a diagnosis. All these 3 functions are accessible from the visit page (Figure 4.10a).

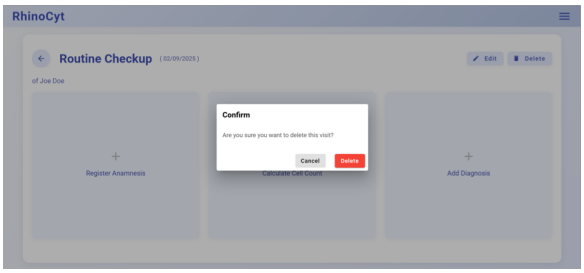
The visit page is mainly composed from 3 cards (one for each of the 3 functions) that initially will be empty, showing a plus icon to encourage the user to click on the card to fill it.

In addition to these cards two buttons are present in the top right corner of the page; similarly to the patient page these 2 buttons allows respectively to edit the visit attributes (that is the category of the visit) and to delete the visit (Figures 4.10b, 4.10c).

(a)



(b)



(c)

Figure 4.10: Visit page (a), edit visit form (b), delete visit confirmation box (c).

Anamnesis

The first of the three cards (both in order of display from left to right and, usually, in order of action execution) is the anamnesis card.

By clicking on the anamnesis card a four pages form will open. This big form comprehends a field for any question that the physician could ask to the patient. For each of the pages there are some required fields, but they are pre-compiled with neutral default values so that the user can freely move between the pages.

The user can move between pages sequentially by clicking the "Next" and "Back" button on the bottom right angle of the form that respectively change the page of the form to the next or the previous one. Alternatively he/she can click on the name of the pages (only already compiled ones) on the top of the form to freely move from one page to the other.

Pages names will initially be accompanied by a white number in a grey circle, showing the uncompleted state of that specific page. Than, when a page is selected the circle will turn blue and the background behind the page name will turn light blue, highlighting the selected page of the form. Already compiled pages will be accompanied by a white pencil icon in a blue circle.

The four pages content is here listed:

1. Familiar anamnesis (Figure 4.11a): in this page are listed all the field relative to pathologies of close relatives like parents and siblings; the physician can annotate here whether the parents/siblings of the patients have an allergy to food or inhalants, if they have any nasal polyposis or bronchial asthma and has space to add some notes on the family medical

history.

2. Symptomatology (Figure 4.11b): this page is dedicated to annotate the symptoms percieved by the patient. The physician can find fields to annotate wheter there is a left, right or bilateral nasal obstruction, hearing loss, tinnitus, and muffled hearing. Is possible to take note about sneezing, olfactory problems, Vertigo syndrome and rhinorrhea. Also check boxes about ocular problems are present.

3. Medical exams (Figure 4.11c): in this page is possible to write down the results of various medical exams that the patient may have undergone:

   - External examination of the nose - with fields about the exterior aspect of the nasal pyramid and nasal valve;
   - Nasal endoscopy - with fields regarding the condition of the nasal septum, the turbinates, the exudates and about the number of polyps in the left/right nasal cavities;
   - Nasopharyngeal examination - with a specific field about Adenoid hypertrophy and a text box regarding eventual allergies;
   - Ear examination - a text box for notes about the ear examination;
   - Rhinomanometric examination - contains number boxed to annotate scores about normal and decongestionate hearing of the left, right and right+left hears.

   the page also includes a free text box for eventual conclusions.

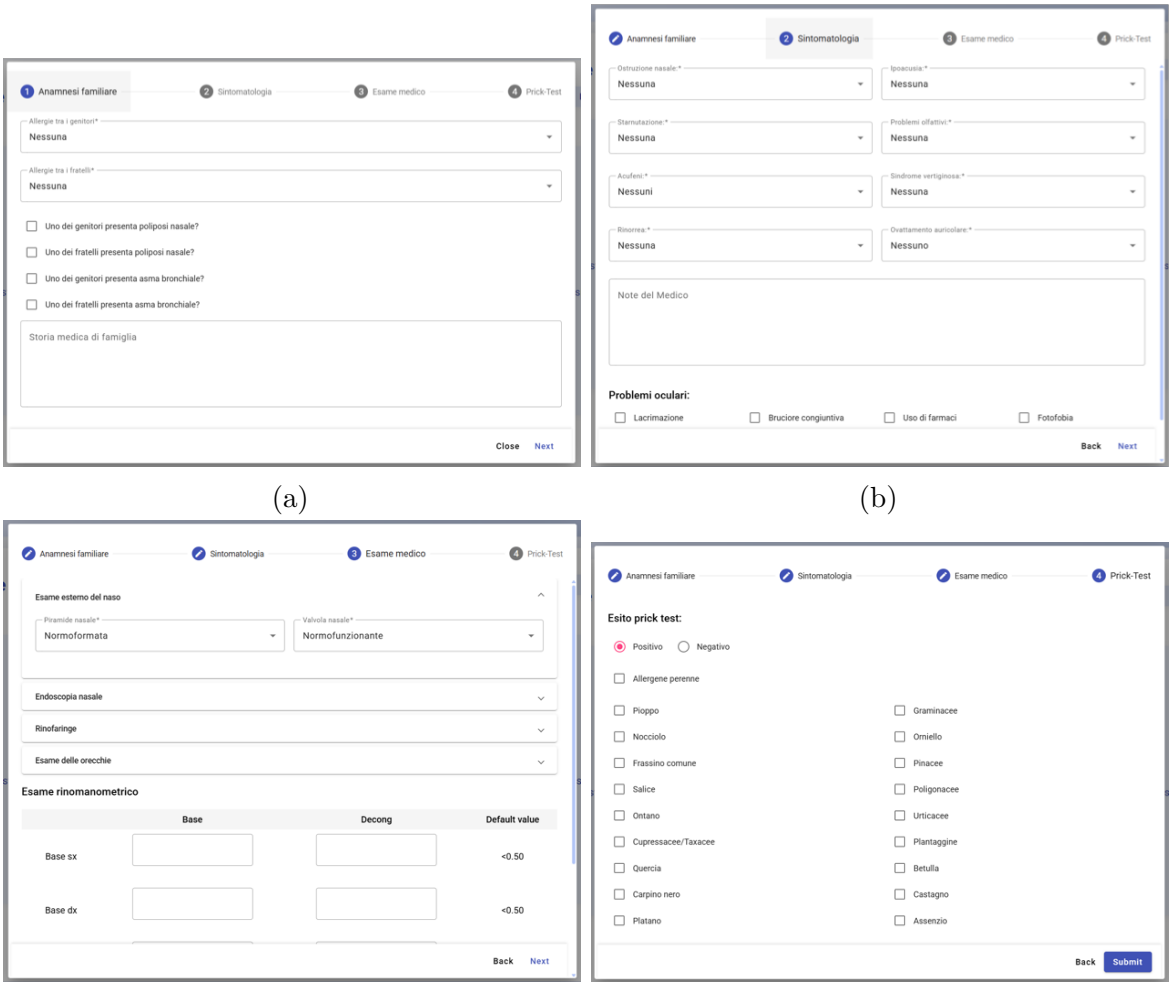
4. Prick test (Figure 4.11d) - allows the medic to annotate whether the prick test had positive or negative outcome and, if positive to check which were the allergens that produced that outcome.

When all four pages are completed, the "Submit" button on the bottom right corner of the last page will light up, clicking it will submit the anamnesis, transforming the anamnesis card from the one showed in Figure 4.10a to the one in Figure 4.12. Now the card shows a summary of all the information present in the anamnesis, the date when the anamnesis was submitted (or was last modified) and an hint to click the card to view the complete anamnesis or edit it.

Also now that an anamnesis is present, the anamnesis icon in this visit card corresponding to this visit on the patient page will light up.

If the user click on the "Close" button on the first page of the form or outside the form, the form will close and the anamnesis will not be registered.

As suggested by the hint on the card, when clicking again the anamnesis card, the anamnesis form will again show up; this time it will be pre-compiled with everything that was inserted by the medic the first time (or at the last update) but it will be completely deactivate and thus is impossible to edit it (Figure 4.13a). The user can navigate the form to look at what is



(a) (b)

(c) (d)

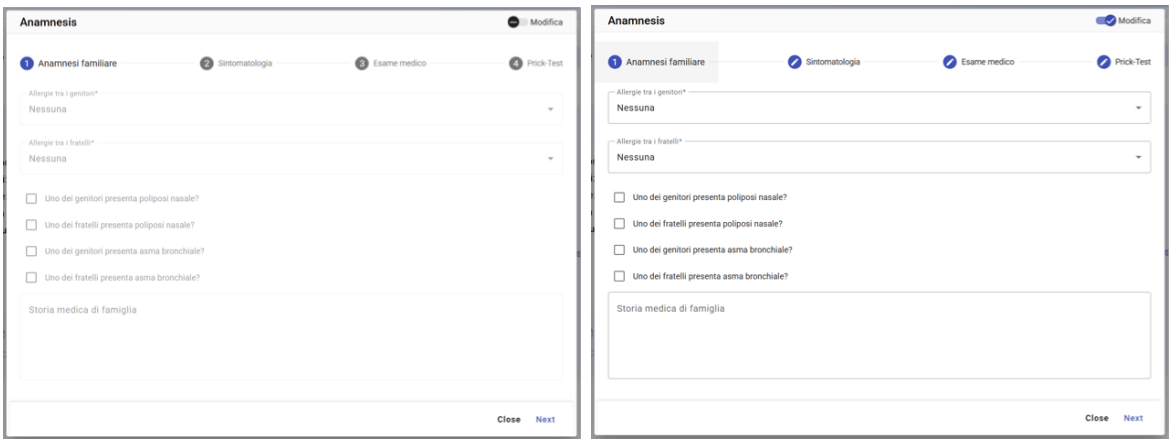
Figure 4.11: The four tabs of the anamnesis form.

Figure 4.12: How the anamnesis card is displayed when an anamnesis is present.

written in the anamnesis with the same modalities as before. To close it, the user can click on the "Close" button on the last page or click outside of the form.

If the user wants to edit/update the anamnesis form for that visit he/she can toggle the "Update" button on the top right corner of the form. Toggling the update mode will make the form active and all the fields editable again (Figure 4.13b). When the user has done editing, it is possible to save the updates by clicking the "Submit" button on the last page of the form. Clicking the "Close" button on the first page of the form or outside the form, will close the form discarding the changes.

If the anamnesis form is updated, also the summary and the date showed in the card will change.



(a) (b)

Figure 4.13: Visualizing a previous compiled anamnesis (a), turning on the "edit" switch allow the user to edit the anamnesis (b).

Cell Count

After compiling the anamnesis, the next step in an hypothetical standard visit is to perform the *Nasal Cytology* diagnostic exam. Until now rhino-cytologist had to take a sample of the nasal mucosa with a swab, transfer it on a cytological glass, then fix and stain it; then the most tedious part would have started: observing the mucosa at the microscope, choosing 50 fields and then count a classify one by one each of the hundreds of cells in each field.

This platform is built specifically to solve this problem and count the cells in an automatic way, in a matter of seconds, thus making life easier for the physicians who want to perform this diagnostic exam.

Of the three cards present in the visit page, the one devoted to cell counting is the middle one. Clicking it will open a dialog at the center of the page that allows the user to upload the images of the fifty cytologic fields (Figure 4.14a). The image can be uploaded by dragging them on the dialog itself from another window of the device were the web-app is executed or by clicking on the Browse button that will open a new file explorer window where the user can select the files he/she wants to upload. When the files have been selected, the dialog will show

the list of their names, so that is possible to double check which files have been chosen (Figure 4.14b). To close the dialog and cancel the upload operation the user can click on the "Cancel" button or outside of the dialog.

When at least one file has been selected, the "Upload" button will activate. The user can click on this button (after double checking that the chosen file are correct) to start the upload and prediction process. This process is displayed to the user through the "upload and prediction" progress bar. This bar will be displayed on the screen throughout the whole upload and prediction process. The bar is composed by two concentric circles: a thick blue circle shows the progress of the upload process (Figure 4.14c); as soon as the upload process ends (usually is really fast) the inner, thinner circle will start to fill up in green (Figure 4.14d); this displays the progress of the prediction process. Inside the circles a number will display the completed percentage of the upload process before and of the prediction process after.

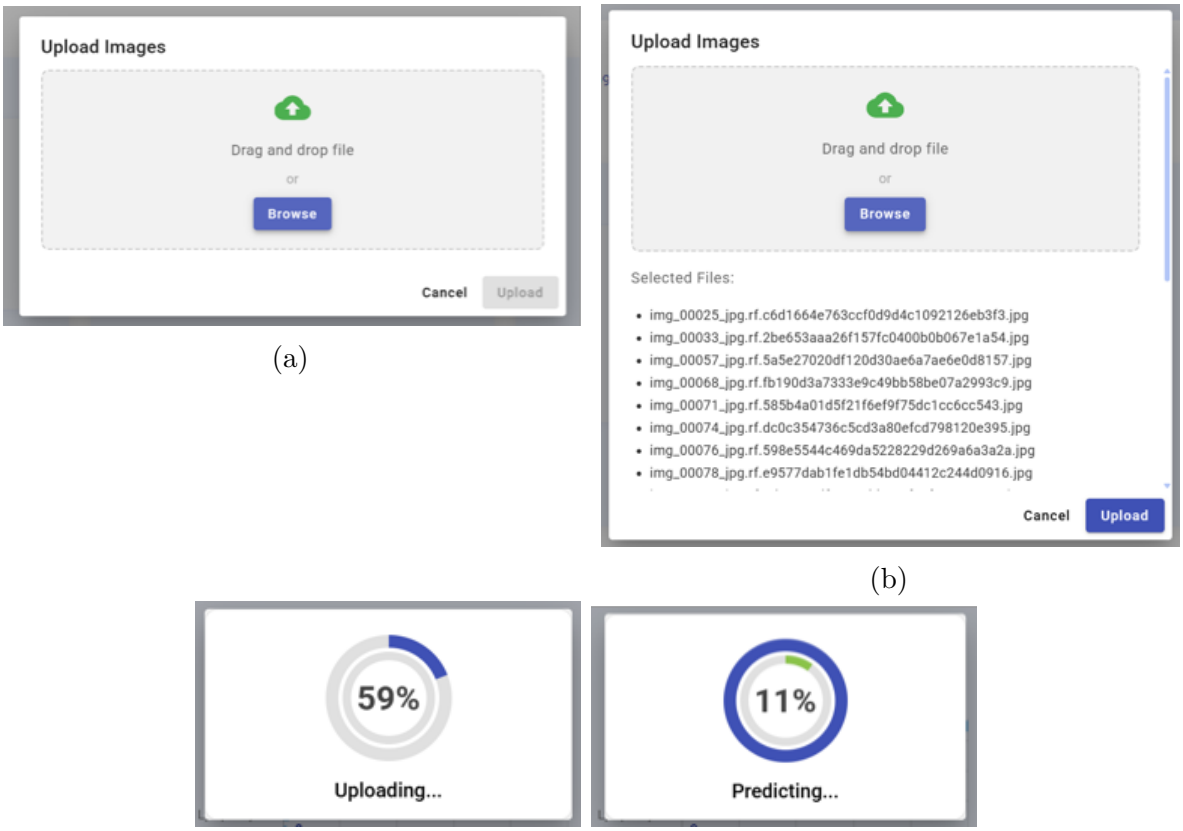


Figure 4.14: Image upload dialog (a), the list of selected images is shown in the dialog (b), the upload progress bar (c), the prediction progress bar (d).

As soon as the prediction process ends, the progress bar will disappear and a new page will be displayed: the **Cell Count page** (Figure 4.15). This page is characterized by two tabs, the "Field Visualization" tab and the "Cell Visualization" tab.

The **Field Visualization** tab (Figure 4.15a) is the active one by default, and is characterized by a two column layout: on the right column the user will see the list of upload images with bounding boxes drawn around all the identified cells; the color of the bounding boxes changes depending on the predicted cytotype of the specific cell. At the bottom right corner of the left

column, the user can find a toggle buttons that allows to deactivate bounding boxes (Figure 4.15b). The smaller right column is dedicated to a deeper analysis of the images; at first, when no image is selected it will only show the chart that display the number of cells for each cytotype.

(a) (b)

Figure 4.15: The cell count page, showed at the end of an analysis (or when clicking an already populated cell card) (a), by clicking on the flag on the bottom right of the gallery, the user can turn on/off the bounding boxes (b).

The physician has the possibility to click on any image on the left column and this will appear in the left column (Figure 4.16a). This allows to have a better look at the specific image and the predictions made on it; on the bottom of the image a chart will also be displayed, showing the number of cells for each cytotype in the selected image (the total count chart is still present, can be seen by scrolling the right column downwards). Other 3 elements are present in the right column when a cell is selected: on the bottom right of the image the is a toggle button to deactivate/activate the bounding boxes (Figure 4.16b), while at the top right of the image are present the "Inspect" button (explained at the end of the section dedicated to the Cell Visualization tab) and a "Close" button (with the "X" icon); clicking this last button will bring the right column back at the status where no image was selected. Clicking on other image in the left gallery, while another one is still displayed will change the displayed image (and relative chart) to the last clicked one.

The **Cell Visualization** tab has a two columns layout, similar to the one of previous tab (Figure 4.17a). Again the left column is dedicated to show all the cells, while the right one is devoted to in depth analysis of the single cell.

In the right column of the Cell Visualization tab there is not a single gallery like in the Field Visualization tab; this time there is a gallery for each of the cytotype. Initially, for an easier navigation, the galleries are compressed in one expansion tab for each cytotype; the user can expand or close as many as he/she wants, whenever he/she wants, by clicking on the expansion tab itself. Expanding one of the tab will show all the cells of that cytotype from all the the uploaded field images (Figure 4.17b).

(a) (b)

Figure 4.16: The detailed visualization of a specific image (a), by clicking on the flag just below the image, the user can turn on/off the bounding boxes on the image (b).

Similarly to the previous tab, when clicking on a cell image, the left column will change. In the upper part of the column will be visualized the field image that contains the selected cell, with only one bounding box drawn, highlighting where the selected cell is located in the image (Figure 4.17c). Clicking on another cell in one of the galleries on the left will update the content of the right column. On the top-right of the image in the left column a "X" button is present; clicking it makes the image disappear and brings the right column to its initial (empty) state.

Just below of the image the is another expansion box that allows to user to correct the system predictions.

Scrolling down in the right column the user can see some examples and counter-examples that could be useful when evaluating the correctness of the prediction (Figure 4.17d).

The AI system at the core of the cell counting part of the web-app, was thought as a continuous learning system. It thus has the capabilities to improve with time, learning from user corrections. Also when designing the system it was decided that the user should be able to make corrections, since the system can always fail, and thus the final decision should be always left to the human, especially in the medical field.

Specifically for these motivations was created the expansion box in Figure 4.18. Expanding it will reveal a text box and a button; the textbox is used to choose between all the different cytotypes (Figure 4.18b), and the "Apply" button is the one that confirm the choice made and applies the updates to the selected cell. Modifications done with this procedure only affect the selected cell (the one highlighted in the figure above and chosen by the user in the left column gallery), to update the classification of another cell the user has to select it first and then follow this procedure again. Selecting a new cytotype without clicking on the apply button does not do anything in term of prediction updates.

Once a prediction update has been successfully completed, the gallery of the newly assigned cytotype will open, with the updated cell highlighted by red borders on the image (Figure 4.18c).

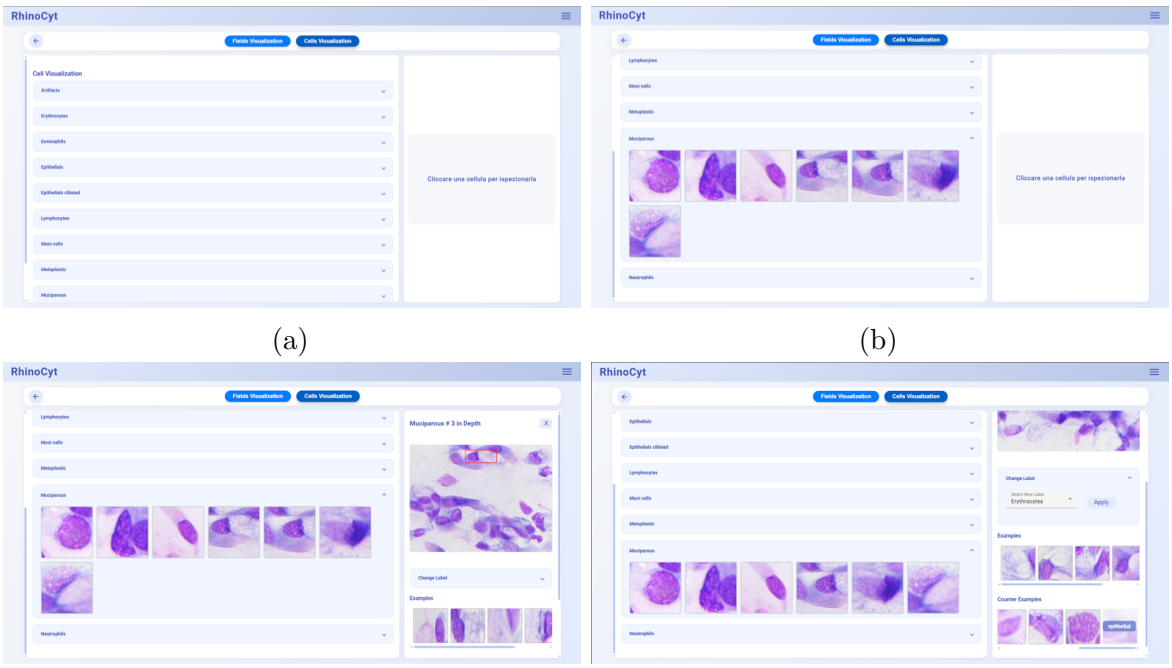


Figure 4.17: The cell count page in "cell visualization" mode (a), the user can see the single cells sorted by cytotype (b) and can inspect each one of them and see the cell in its field (c); to help the cytologist examples of similar cells from the same cytotype and counterexamples from other classes are also shown (d).

The borders of cells images with updated predictions will remain red until the page changes, to let the user know which prediction have already been updated.

Cytotype modifications made by the user will be saved on the DB, so that the system can be later trained also on this new data, and learn to not commit these mistakes anymore, thus becoming better with time and thanks to user corrections.

On the bottom of the change label expansion tab, there are two galleries: the example gallery and the counter example gallery.

The example gallery displays the pictures of the nine most similar cells with respect to the selected cell, from the same cytotype of the selected cell itself. I.e. if the selected cell is currently a muciparous cell, the example gallery will show the nine muciparous cells that resemble the most the selected cell.

The counter-example gallery will instead contain nine more images showing the most similar cells with respect to the selected cells, but from different cycotypes (one for each of the nine remaining cytotypes). Continuing from the previous example, the counter-example gallery will contain the image of the eosinophil cell that resembles the most the selected cell, and the same for the epithelial cells, the artifacts and so on ...

While its easy to understand which is the cytotype of the cell in the example gallery (the same of the selected cell), the same cannot be said about the counter-examples gallery. That is why all the cell images (in both the examples and counter-examples galleries) when hovered with the mouse will become opaque, to allow for the name of the respective cytotype to appear

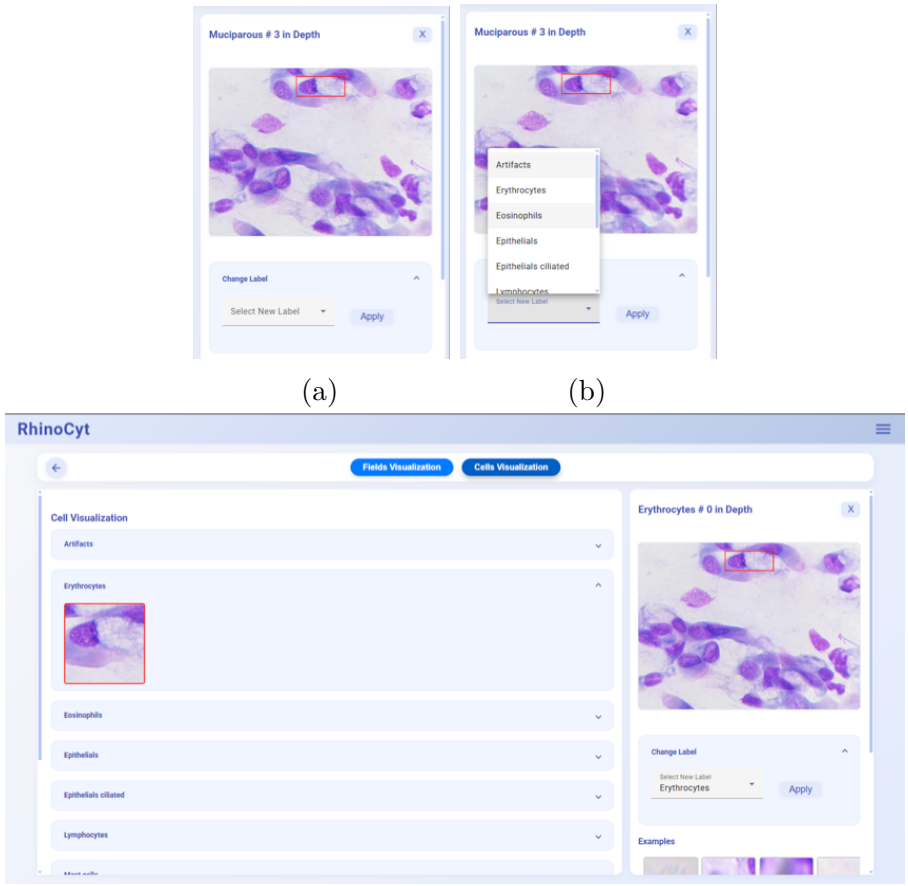


Figure 4.18: Label correction expansion tab. The user can correct the system prediction by opening the "change label" curtain (a), he/she has to choose the correct label from the list and click apply (b); the curtain of the corrected cytotype will be opened with new newly re-labeled cell hightlighted (c).

(Figure 4.17d).

When talking the Image Visualiziation tab, the function of a specific button has been deliberately ignored. What the "Inspect" button (top-right of the selected image, Figure 4.16) does is strictly related to the Cell Visualization tab functioning, that is why its function is explained here. What this button actually does is opening the Cell Visualization tab, but instead of having cells from all the images in the gallery, only cells from the selected field image will be displayed. In this way is easier for the physician to find a specific cell from a specific field image, take a closer look and eventually changing its predicted label.

When a cell count has already been performed for the current visit, the cell count card in the visit page will look like in Figure 4.19) and the corresponding icon on the visit card in the patient page will turn on. When a cell count has been performed the card will display a chart showing the cell counts of the individual cytotypes; plus a button on the bottom that allows to perform a new count for that visit (to face situations like the ones where a some fields where forgotten to be uploaded for the first count).

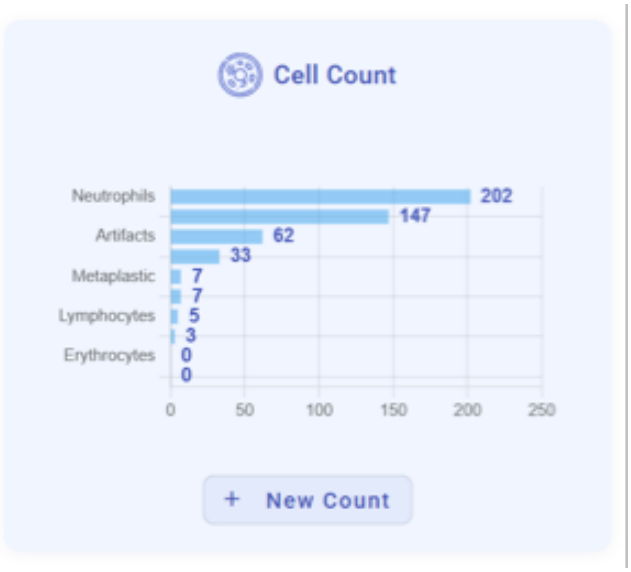


Figure 4.19: The cell count card will display a chart showing the cell count for each cytotype when a cell count exam has been done.

Diagnosis

Theoretically the diagnosis is always the last step in a visit (followed only by an eventual drug prescription) and that is why this is the last card in the visit page.

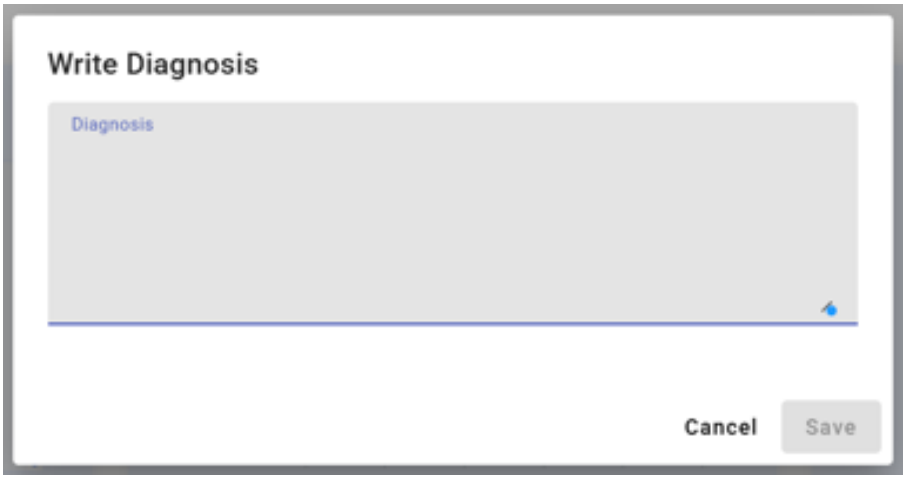


Figure 4.20: The diagnosis dialog allow to insert a new diagnosis or edit an already exsting one.

Since a diagnosis has no standard format it was decided to only use a large textbox to let the medic annotate what were his/her conclusion. So the functioning of the diagnosis card is very similar to the one of the anamnesis card, but simpler. When clicking on the card a form with a texbox will open (Figure 4.20) and allow the user to write his/her conclusion or annotations about the current visit. To save it, the physician has to just click on the "Save" button, that will become active as soon as something is written in the form. To cancel the diagnosis writing and close the form, the user can click on the "Cancel" button or outside of the form, as always.

When a diagnosis is saved, similarly to the anamnesis, the card will show what was written in the form, with the corresponding date and an hint, suggesting that the card can still be clicked to edit the diagnosis (Figure 4.21). Clicking the card will re-open the diagnosis textbox

form, this time pre-compiled with the previous diagnosis; the user can edit it and click the "Save" button. The card will now show the edited diagnosis, with the date on which the update was performed.

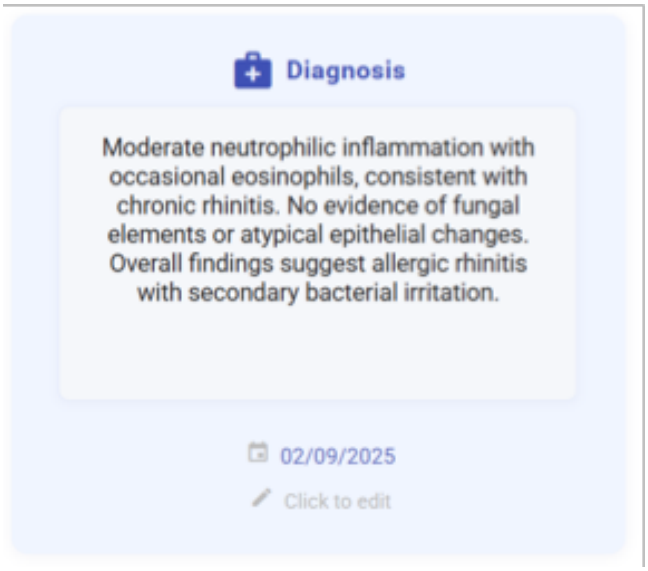


Figure 4.21: How the diagnosis card is displayed when a diagnosis is present.

Figure 4.22: A visit page with all cards populated.

Figure 4.22 shows how the visit page looks when all the three cards are in their compiled version, when the anamnesis has been completed, a cell-count has been performed and a diagnosis has been made.

With the visit page completely populated, the web application user guide can be called complete. Many functions will be added in the future, but as for now, each page, function and detail has been explained in depth.

Chapter 5

Conclusion and Future Works

The work presented in this thesis set out to investigate how artificial intelligence can be effectively applied to nasal cytology, a diagnostic domain that, despite its clinical potential, has remained largely manual and dependent on specialized expertise. By combining medical knowledge, deep learning methodologies, and software engineering principles, the research aimed to design and implement an interpretable, automated pipeline capable of assisting clinicians in the analysis of cytological samples. This objective has been pursued through an incremental and multidisciplinary approach that integrated data collection, model design, interpretability, and deployment into a unified workflow.

The first contribution of this work consisted in the creation of a dedicated dataset for nasal cytology, representing one of the first structured and annotated resources in this field. This dataset enabled the training and benchmarking of modern object-detection and classification models, providing a foundation for reproducible research and future improvements. The dataset design emphasized both quality and diversity, ensuring that the different cytotypes and pathological conditions were adequately represented. This resource was essential to bridge the gap between theoretical model design and its practical application to real medical data.

Building upon this foundation, the thesis investigated a variety of deep learning architectures for the tasks of cell detection and classification. Through comparative experimentation with state-of-the-art models—such as DETR, YOLOv8, and Faster R-CNN—it was possible to identify the architectures most suitable for cytological images characterized by complex backgrounds, overlapping cells, and variable staining. The experiments demonstrated that deep learning approaches can achieve robust and accurate performance in recognizing and classifying nasal cells, thus confirming the feasibility of automated cytological analysis.

Beyond accuracy, this work has emphasized the importance of interpretability and transparency in medical AI. The development of a prototype-based interpretable model (ProtoTree) introduced a novel way to combine the predictive power of convolutional features with the clarity of rule-based reasoning. This approach allowed clinicians to visualize the prototypes influencing each decision, thereby fostering trust in model outputs and providing a tool that can

support—not replace—human expertise. The interpretable design thus contributes to bridging the cognitive gap between artificial intelligence and medical reasoning.

A further step has been the extension of AI analysis beyond static cytological images. The third study explored the estimation of ciliary beat frequency (CBF) in video sequences of ciliated cells. This task, which quantifies an important functional aspect of nasal epithelium physiology, demonstrated the flexibility of computer vision techniques to address dynamic diagnostic parameters. By combining motion-analysis algorithms with prior deep learning components, the proposed pipeline succeeded in estimating CBF values consistent with clinical measurements, opening the way to objective and reproducible assessment of mucociliary function.

While the methodological contributions represent the core of the thesis, their integration into a deployable clinical tool has been equally important. To this end, a web-based application was designed and implemented, transforming research prototypes into a usable digital assistant for physicians. The system was conceived to guide the user through the same logical steps followed during a real cytological examination—patient registration, visit management, image upload, automatic analysis, and diagnosis editing—while maintaining simplicity and usability. The server-side authentication, encrypted communication, and modular architecture ensure data security and scalability, aligning with medical software requirements. Through this web application, the thesis demonstrates that artificial intelligence in medicine can be translated into practical solutions capable of being adopted in real workflows.

Overall, the work contributes to the digital transformation of nasal cytology, addressing three main research challenges: (1) the lack of structured data resources, (2) the need for interpretable and reliable AI models, and (3) the gap between research and clinical usability. Each of these challenges has been met through concrete outputs: a dataset, interpretable models, and an integrated platform. Taken together, they form a complete pipeline that moves from data acquisition to end-user interaction, offering both scientific and practical value.

Several directions can be pursued to extend the contributions of this thesis. A first priority is the expansion of the cytology dataset through multi-center collaboration, enabling broader data diversity and improving model generalization across different laboratory settings. In parallel, clinical validation studies involving physicians and patients are needed to rigorously assess the system's diagnostic performance and usability in real-world conditions. From a technical standpoint, future work could explore self-supervised and multimodal learning approaches to reduce dependence on manual annotations and to integrate additional information such as clinical metadata or temporal sequences. Moreover, enhancing the explainability interface of the web application could improve physician interaction with model outputs, transforming interpretability from a visual feature into a true decision-support mechanism. Finally, the extension of the platform to other cytological domains—such as sputum, bronchial, or Pap smear cytology—would further test the versatility and robustness of the proposed methodologies.

In conclusion, this thesis demonstrates that the combination of deep learning, interpretable modeling, and thoughtful software design can significantly advance the automation and accessi-

bility of nasal cytology. By uniting medical knowledge with computational innovation, it shows a feasible pathway toward reliable, transparent, and user-centered diagnostic support systems. The outcomes achieved not only advance the scientific understanding of AI applications in cytology but also lay the groundwork for the next generation of intelligent tools that can assist clinicians in improving diagnostic accuracy and patient care.

Bibliography

[1] M. G. Camporeale, U. Kaymak, G. Dimauro, Prototype trees to develop an interpretable automated pipeline for nasal cytology, in: 2025 IEEE International Conference on Fuzzy Systems (FUZZ), IEEE, 2025, pp. 1–7.

[2] G. Dimauro, N. Barbaro, M. G. Camporeale, V. Fiore, M. Gelardi, M. Scalera, Deepcilia: Automated, deep-learning based engine for precise ciliary beat frequency estimation, Biomedical Signal Processing and Control 90 (2024) 105808.

[3] M. Camporeale, G. Dimauro, M. Gelardi, G. Iacobellis, M. S. Ladisa, S. Latrofa, N. Lomonte, A nasal cytology dataset for object detection and deep learning, arXiv preprint arXiv:2404.13745 (2024).

[4] G. Dimauro, D. Caivano, P. Di Pilato, A. Dipalma, M. G. Camporeale, A systematic mapping study on research in anemia assessment with non-invasive devices, Applied Sciences 10 (14) (2020) 4804.

[5] G. Dimauro, M. E. Griseta, M. G. Camporeale, F. Clemente, A. Guarini, R. Maglietta, An intelligent non-invasive system for automated diagnosis of anemia exploiting a novel dataset, Artificial Intelligence in Medicine 136 (2023) 102477.

[6] G. Dimauro, M. G. Camporeale, A. Dipalma, A. Guarini, R. Maglietta, Anaemia detection based on sclera and blood vessel colour estimation, Biomedical Signal Processing and Control 81 (2023) 104489.

[7] M. G. Camporeale, L. Colizzi, N. Lomonte, A. Ragone, Development of a desktop application to enable doctors to remotely monitor patients' hematological parameters, in: International Conference on Product-Focused Software Process Improvement, Springer, 2023, pp. 48–59.

[8] L. Simone, M. G. Camporeale, N. Lomonte, G. Dimauro, V. Gervasi, An efficient deep learning approach for arrhythmia classification using 3d temporal svcg, in: 2023 IEEE International Conference on Digital Health (ICDH), IEEE, 2023, pp. 234–239.

[9] V. S. Barletta, M. G. Camporeale, N. Lomonte, M. Scalera, E. Gentile, Deskemo, a multi-platform desktop software for remote control and monitoring of hematological parameters, in: Italian Forum of Ambient Assisted Living, Springer, 2023, pp. 224–232.

[10] M. Camporeale, F. Clemente, G. Dimauro, N. Lomonte, R. Maglietta, C. Pasciolla, D. Sacco, G. M. Zaccaria, Highly reliable personalized noninvasive hemoglobin estimation by using vision transformers and dual fine-tuning, Computers in Biology and Medicine 197 (2025) 111026.

[11] L. Simone, M. G. Camporeale, V. M. Rubino, V. Gervasi, G. Dimauro, Interpretable early detection of parkinson's disease through speech analysis, in: International Conference on Artificial Intelligence in Medicine, Springer, 2025, pp. 373–378.

[12] A. Elliott, The culture of AI: Everyday life and the digital revolution, Routledge, 2019.

[13] K.-H. Yu, A. L. Beam, I. S. Kohane, Artificial intelligence in healthcare, Nature biomedical engineering 2 (10) (2018) 719–731.

[14] A. Al Kuwaiti, K. Nazer, A. Al-Reedy, S. Al-Shehri, A. Al-Muhanna, A. V. Subbarayalu, D. Al Muhanna, F. A. Al-Muhanna, A review of the role of artificial intelligence in healthcare, Journal of personalized medicine 13 (6) (2023) 951.

[15] S. A. Alowais, S. S. Alghamdi, N. Alsuhebany, T. Alqahtani, A. I. Alshaya, S. N. Almohareb, A. Aldairem, M. Alrashed, K. Bin Saleh, H. A. Badreldin, et al., Revolutionizing healthcare: the role of artificial intelligence in clinical practice, BMC medical education 23 (1) (2023) 689.

[16] P. Rajpurkar, E. Chen, O. Banerjee, E. J. Topol, Ai in health and medicine, Nature medicine 28 (1) (2022) 31–38.

[17] M. Gelardi, M. L. Fiorella, C. Russo, R. Fiorella, G. Ciprandi, Role of nasal cytology, International Journal of Immunopathology and Pharmacology 23 (1_suppl) (2010) 45–49.

[18] M. Gelardi, et al., Atlas of nasal cytology Stampa for the Differential Diagnosis of Nasal Diseases SECOND EDITION, ITA, 2012.

[19] V. Özdemir, N. Hekim, Birth of industry 5.0: Making sense of big data with artificial intelligence,"the internet of things" and next-generation technology policy, Omics: a journal of integrative biology 22 (1) (2018) 65–76.

[20] A. C. Chang, Intelligence-based medicine: artificial intelligence and human cognition in clinical medicine and healthcare, Academic Press, 2020.

[21] G. Krishnan, S. Singh, M. Pathania, S. Gosavi, S. Abhishek, A. Parchani, M. Dhar, Artificial intelligence in clinical medicine: catalyzing a sustainable global healthcare paradigm, Frontiers in artificial intelligence 6 (2023) 1227091.

[22] W. Van Melle, Mycin: a knowledge-based consultation program for infectious disease diagnosis, International journal of man-machine studies 10 (3) (1978) 313–322.

[23] N. Muthukrishnan, F. Maleki, K. Ovens, C. Reinhold, B. Forghani, R. Forghani, Brief history of artificial intelligence, Neuroimaging Clinics 30 (4) (2020) 393–399.

[24] V. Kaul, S. Enslin, S. A. Gross, History of artificial intelligence in medicine, Gastrointestinal endoscopy 92 (4) (2020) 807–812.

[25] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, B. Scholkopf, Support vector machines, IEEE Intelligent Systems and their applications 13 (4) (1998) 18–28.

[26] L. Breiman, Random forests, Machine learning 45 (1) (2001) 5–32.

[27] J. Pearl, Probabilistic reasoning in intelligent systems: networks of plausible inference, Elsevier, 2014.

[28] J. Jiang, P. Trundle, J. Ren, Medical image analysis with artificial neural networks, Computerized Medical Imaging and Graphics 34 (8) (2010) 617–631.

[29] A. Depeursinge, D. Van de Ville, A. Platon, A. Geissbuhler, P.-A. Poletti, H. Muller, Near-affine-invariant texture learning for lung tissue analysis using isotropic wavelet frames, IEEE Transactions on Information Technology in Biomedicine 16 (4) (2012) 665–675.

[30] T. J. Sejnowski, The deep learning revolution, MIT press, 2018.

[31] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs, jama 316 (22) (2016) 2402–2410.

[32] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, S. Thrun, Dermatologist-level classification of skin cancer with deep neural networks, nature 542 (7639) (2017) 115–118.

[33] X. Feng, J. Yang, A. F. Laine, E. D. Angelini, Discriminative localization in cnns for weakly-supervised segmentation of pulmonary nodules, in: International conference on medical image computing and computer-assisted intervention, Springer, 2017, pp. 568–576.

[34] O. Oren, B. J. Gersh, D. L. Bhatt, Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints, The Lancet Digital Health 2 (9) (2020) e486–e488.

[35] V. Nivethitha, R. A. Daniel, A. Debnath, V. Dwarakanathan, G. Jeer, G. Kavipriya, Diagnostic accuracy of artificial intelligence models for imaging detection of hepatic

steatosis through systematic review and meta analysis, Scientific Reports 15 (1) (2025) 34408.

[36] H. R. Tizhoosh, P. Diamandis, C. J. Campbell, A. Safarpoor, S. Kalra, D. Maleki, A. Riasatian, M. Babaie, Searching images for consensus: can ai remove observer variability in pathology?, The American journal of pathology 191 (10) (2021) 1702–1708.

[37] M. Khosravi, Z. Zare, S. M. Mojtabaeian, R. Izadi, Artificial intelligence and decision-making in healthcare: a thematic analysis of a systematic review of reviews, Health services research and managerial epidemiology 11 (2024) 23333928241234863.

[38] F. Maleki, K. Ovens, R. Gupta, C. Reinhold, A. Spatz, R. Forghani, Generalizability of machine learning models: quantitative evaluation of three methodological pitfalls, Radiology: Artificial Intelligence 5 (1) (2022) e220028.

[39] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, C. Zhong, Interpretable machine learning: Fundamental principles and 10 grand challenges, Statistic Surveys 16 (2022) 1–85.

[40] Y. Wang, Q. Chen, B. Zhang, Image enhancement based on equal area dualistic sub-image histogram equalization method, IEEE transactions on Consumer Electronics 45 (1) (1999) 68–75.

[41] K. Ito, K. Xiong, Gaussian filters for nonlinear filtering problems, IEEE transactions on automatic control 45 (5) (2002) 910–927.

[42] J. Astola, P. Haavisto, Y. Neuvo, Vector median filters, Proceedings of the IEEE 78 (4) (2002) 678–689.

[43] A. Y. Kwentus, Z. Jiang, A. N. Willson, Application of filter sharpening to cascaded integrator-comb decimation filters, IEEE Transactions on signal Processing 45 (2) (1997) 457–467.

[44] I. Sobel, G. M. Feldman, An isotropic 3×3 image gradient operator, in: Stanford Artificial Intelligence Project (SAIL), 1990.

[45] J. M. Prewitt, et al., Object enhancement and extraction, Picture processing and Psychopictorics 10 (1) (1970) 15–19.

[46] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8 (6) (1986) 679–698. doi:10.1109/TPAMI.1986.4767851.

[47] N. Otsu, et al., A threshold selection method from gray-level histograms, Automatica 11 (285-296) (1975) 23–27.

[48] N. Ikonomatakis, K. Plataniotis, M. Zervakis, A. Venetsanopoulos, Region growing and region merging image segmentation, in: Proceedings of 13th International Conference on Digital Signal Processing, Vol. 1, IEEE, 1997, pp. 299–302.

[49] T. F. Chan, L. A. Vese, Active contours without edges, IEEE Transactions on image processing 10 (2) (2001) 266–277.

[50] J. B. Roerdink, A. Meijster, The watershed transform: Definitions, algorithms and parallelization strategies, Fundamenta informaticae 41 (2000) 187–228.

[51] R. M. Haralick, K. Shanmugam, I. H. Dinstein, Textural features for image classification, IEEE Transactions on systems, man, and cybernetics (6) (2007) 610–621.

[52] K. Doi, Computer-aided diagnosis in medical imaging: historical review, current status and future potential, Computerized medical imaging and graphics 31 (4-5) (2007) 198–211.

[53] L. Zhang, R. Sankar, W. Qian, Advances in micro-calcification clusters detection in mammography, Computers in Biology and Medicine 32 (6) (2002) 515–528.

[54] T. H. Baron, G. C. Harewood, A. Rumalla, N. L. Pochron, L. M. Stadheim, G. J. Gores, T. M. Therneau, P. C. De Groen, T. J. Sebo, D. R. Salomao, et al., A prospective comparison of digital image analysis and routine cytology for the identification of malignancy in biliary tract strictures, Clinical Gastroenterology and Hepatology 2 (3) (2004) 214–219.

[55] T. Cover, P. Hart, Nearest neighbor pattern classification, IEEE Transactions on Information Theory 13 (1) (1967) 21–27. doi:10.1109/TIT.1967.1053964.

[56] J. R. Quinlan, Induction of decision trees, Machine learning 1 (1) (1986) 81–106.

[57] M. E. Plissiti, C. Nikou, A. Charchanti, Combining shape, texture and intensity features for cell nuclei extraction in pap smear images, Pattern Recognition Letters 32 (6) (2011) 838–853.

[58] G. Sun, S. Li, Y. Cao, F. Lang, Cervical cancer diagnosis based on random forest, International Journal of Performability Engineering 13 (4) (2017) 446.

[59] T. Ching, D. S. Himmelstein, B. K. Beaulieu-Jones, A. A. Kalinin, B. T. Do, G. P. Way, E. Ferrero, P.-M. Agapow, M. Zietz, M. M. Hoffman, et al., Opportunities and obstacles for deep learning in biology and medicine, Journal of the royal society interface 15 (141) (2018) 20170387.

[60] E. Kondrateva, M. Pominova, E. Popova, M. Sharaev, A. Bernstein, E. Burnaev, Domain shift in computer vision models for mri data analysis: an overview, in: Thirteenth International Conference on Machine Vision, Vol. 11605, SPIE, 2021, pp. 126–133.

[61] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, C. I. Sánchez, A survey on deep learning in medical image analysis, Medical image analysis 42 (2017) 60–88.

[62] F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain., Psychological review 65 (6) (1958) 386.

[63] Y. LeCun, L. Bottou, Y. Bengio, P. Haffner, Gradient-based learning applied to document recognition, Proceedings of the IEEE 86 (11) (2002) 2278–2324.

[64] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[65] F. Long, Microscopy cell nuclei segmentation with enhanced u-net, BMC bioinformatics 21 (1) (2020) 8.

[66] R. Hong, D. Fenyö, Deep learning and its applications in computational pathology, BioMedInformatics 2 (1) (2022) 159–168.

[67] A. A. Ahmed, M. Abouzid, E. Kaczmarek, Deep learning approaches in histopathology, Cancers 14 (21) (2022) 5264.

[68] Y. Gur, M. Moradi, H. Bulu, Y. Guo, C. Compas, T. Syeda-Mahmood, Towards an efficient way of building annotated medical image collections for big data studies, in: International Workshop on Large-Scale Annotation of Biomedical Data and Expert Label Synthesis, Springer, 2017, pp. 87–95.

[69] K. Stacke, G. Eilertsen, J. Unger, C. Lundström, Measuring domain shift for deep learning in histopathology, IEEE journal of biomedical and health informatics 25 (2) (2020) 325–336.

[70] J. J. Wadden, Defining the undefinable: the black box problem in healthcare artificial intelligence, Journal of Medical Ethics 48 (10) (2022) 764–768.

[71] X. Li, H. Xiong, X. Li, X. Wu, X. Zhang, J. Liu, J. Bian, D. Dou, Interpretable deep learning: Interpretation, interpretability, trustworthiness, and beyond, Knowledge and Information Systems 64 (12) (2022) 3197–3234.

[72] D. Minh, H. X. Wang, Y. F. Li, T. N. Nguyen, Explainable artificial intelligence: a comprehensive review, Artificial Intelligence Review 55 (5) (2022) 3503–3568.

[73] European Parliament and Council of the European Union, Artificial intelligence act, regulation (eu) 2024/1689, Official Journal of the European Union (2024).

[74] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, A. Torralba, Learning deep features for discriminative localization, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 2921–2929.

[75] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, D. Batra, Grad-cam: Why did you say that?, arXiv preprint arXiv:1611.07450 (2016).

[76] M. Biehl, B. Hammer, T. Villmann, Prototype-based models in machine learning, Wiley Interdisciplinary Reviews: Cognitive Science 7 (2) (2016) 92–111.

[77] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[78] K. Kaczmarek-Majer, G. Casalino, G. Castellano, M. Dominiak, O. Hryniewicz, O. Kamińska, G. Vessio, N. Díaz-Rodríguez, Plenary: Explaining black-box models in natural language through fuzzy linguistic summaries, Information Sciences 614 (2022) 374–399. doi:10.1016/j.ins.2022.10.010.

[79] M. Kiani, J. Andreu-Perez, H. Hagras, A temporal type-2 fuzzy system for time-dependent explainable artificial intelligence, IEEE Transactions on Artificial Intelligence 4 (3) (2022) 573–586.

[80] G. Casalino, G. Castellano, D. Margherita, A. G. Valerio, G. Vessio, G. Zaza, Exploring the expressive power of large language models in neuro-fuzzy system explainability: A study on eeg-based seizure detection, language 6 (2025) 7.

[81] F. Mumuni, A. Mumuni, Explainable artificial intelligence (xai): from inherent explainability to large language models, arXiv preprint arXiv:2501.09967 (2025).

[82] M. Mesinovic, P. Watkinson, T. Zhu, Explainability in the age of large language models for healthcare, Communications Engineering 4 (1) (2025) 128.

[83] L. Lavitas, O. Redfield, A. Lee, D. Fletcher, M. Eck, S. Janardhanan, Annotation quality framework-accuracy, credibility, and consistency, in: NEURIPS 2021 Workshop for Data Centric AI, Vol. 3, 2021.

[84] A. Althnian, D. AlSaeed, H. Al-Baity, A. Samha, A. B. Dris, N. Alzakari, A. Abou Elwafa, H. Kurdi, Impact of dataset size on classification performance: an empirical evaluation in the medical domain, Applied sciences 11 (2) (2021) 796.

[85] N. Japkowicz, The class imbalance problem: Significance and strategies, in: Proc. of the Int'l Conf. on artificial intelligence, Vol. 56, 2000, pp. 111–117.

[86] A. Hagiwara, S. Fujita, Y. Ohno, S. Aoki, Variability and standardization of quantitative imaging: monoparametric to multiparametric quantification, radiomics, and artificial intelligence, Investigative radiology 55 (9) (2020) 601–616.

[87] L. Perez, J. Wang, The effectiveness of data augmentation in image classification using deep learning, arXiv preprint arXiv:1712.04621 (2017).

[88] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, pp. 248–255. doi:10.1109/CVPR.2009.5206848.

[89] G. Ferrante, A. Licari, S. Fasola, G. L. Marseglia, S. La Grutta, Artificial intelligence in the diagnosis of pediatric allergic diseases, Pediatric Allergy and Immunology 32 (3) (2021) 405–413. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/pai.13419, doi:https://doi.org/10.1111/pai.13419.

[90] D. Buongiorno, G. D. Cascarano, I. De Feudis, A. Brunetti, L. Carnimeo, G. Dimauro, V. Bevilacqua, Deep learning for processing electromyographic signals: A taxonomy-based survey, Neurocomputing 452 (2021) 549–565. doi:https://doi.org/10.1016/j.neucom.2020.06.139.

[91] L. Simone, M. G. Camporeale, N. Lomonte, G. Dimauro, V. Gervasi, An efficient deep learning approach for arrhythmia classification using 3d temporal svcg, in: 2023 IEEE International Conference on Digital Health (ICDH), IEEE, 2023, pp. 234–239.

[92] L. Wong, A. Ccopa, E. Diaz, S. Valcarcel, D. Mauricio, V. Villoslada, Deep learning and transfer learning methods to effectively diagnose cervical cancer from liquid-based cytology pap smear images, International Journal of Online and Biomedical Engineering (iJOE) 19 (04) (2023) 77–93. doi:10.3991/ijoe.v19i04.37437.

[93] N. Wentzensen, B. Lahrmann, M. A. Clarke, W. Kinney, D. Tokugawa, N. Poitras, A. Locke, L. Bartels, A. Krauthoff, J. Walker, R. Zuna, K. K. Grewal, P. E. Goldhoff, J. D. Kingery, P. E. Castle, M. Schiffman, T. S. Lorey, N. Grabe, Accuracy and Efficiency of Deep-Learning–Based Automation of Dual Stain Cytology in Cervical Cancer Screening, JNCI: Journal of the National Cancer Institute 113 (1) (2020) 72–79. arXiv:https://academic.oup.com/jnci/article-pdf/113/1/72/37792039/djaa066_supplementary_data.pdf, doi:10.1093/jnci/djaa066.

[94] M. Gelardi, et al., Atlas of nasal cytology Stampa for the Differential Diagnosis of Nasal Diseases SECOND EDITION, ITA, 2012.

[95] P. Dey, The emerging role of deep learning in cytology, Cytopathology 32 (2) (2020) 154–160. doi:10.1111/cyt.12942.

[96] A. Victória Matias, J. G. Atkinson Amorim, L. A. Buschetto Macarini, A. Cerentini, A. S. Casimiro Onofre, F. B. De Miranda Onofre, F. P. Daltoé, M. R. Stemmer, A. von Wangenheim, What is the state of the art of computer vision-assisted cytology? a systematic literature review, Computerized Medical Imaging and Graphics 91 (2021) 101934. doi:https://doi.org/10.1016/j.compmedimag.2021.101934.

[97] G. Dimauro, N. Barbaro, M. G. Camporeale, V. Fiore, M. Gelardi, M. Scalera, Deepcilia: Automated, deep-learning based engine for precise ciliary beat frequency estimation, Biomedical Signal Processing and Control 90 (2024) 105808.

[98] H. Jiang, Y. Zhou, Y. Lin, R. C. Chan, J. Liu, H. Chen, Deep learning for computational cytology: A survey, Medical Image Analysis 84 (2023) 102691.

[99] S. Mitra, N. Das, S. Dey, S. Chakraborty, M. Nasipuri, M. K. Naskar, Cytology image analysis techniques toward automation: Systematically revisited, ACM Comput. Surv. 54 (3) (2021). doi:10.1145/3447238.

[100] K. Ikeda, N. Sakabe, S. Maruyama, C. Ito, Y. Shimoyama, S. Sato, K. Nagata, Relationship between Liquid-Based Cytology Preservative Solutions and Artificial Intelligence: Liquid-Based Cytology Specimen Cell Detection Using YOLOv5 Deep Convolutional Neural Network, Acta Cytologica 66 (6) (2022) 542–550. doi:10.1159/000526098.

[101] J. Zou, Z. Xue, G. Brown, R. Long, S. Antani, Deep learning for nuclei segmentation and cell classification in cervical liquid based cytology, in: P.-H. Chen, T. M. Deserno (Eds.), Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications, Vol. 11318, International Society for Optics and Photonics, SPIE, 2020, p. 1131811. doi:10.1117/12.2549547.

[102] J. Caputa, D. Łukasik, M. Wielgosz, M. Karwatowski, R. Frączek, P. Russek, K. Wiatr, Fast pre-diagnosis of neoplastic changes in cytology images using machine learning, Applied Sciences 11 (16) (2021). doi:10.3390/app11167181.

[103] Y. Tao, Y. Cai, H. Fu, L. Song, L. Xie, K. Wang, Automated interpretation and analysis of bronchoalveolar lavage fluid, International Journal of Medical Informatics 157 (2022) 104638. doi:https://doi.org/10.1016/j.ijmedinf.2021.104638.

[104] L. Gedefaw, C.-F. Liu, R. K. L. Ip, H.-F. Tse, M. H. Y. Yeung, S. P. Yip, C.-L. Huang, Artificial intelligence-assisted diagnostic cytology and genomic testing for hematologic disorders, Cells 12 (13) (2023).

[105] R. Awan, K. Benes, A. Azam, T.-H. Song, M. Shaban, C. Verrill, Y. W. Tsang, D. Snead, F. Minhas, N. Rajpoot, Deep learning based digital cell profiles for risk stratification of urine cytology images, Cytometry Part A 99 (7) (2021) 732–742. doi:https://doi.org/10.1002/cyto.a.24313.

[106] H. Sadeghi, H.-S. Braun, B. Panti, G. Opsomer, O. Bogado Pascottini, Validation of a deep learning-based image analysis system to diagnose subclinical endometritis in dairy cows, PLOS ONE 17 (1) (2022) 1–14. doi:10.1371/journal.pone.0263409.

[107] B. Murdoch, Privacy and artificial intelligence: challenges for protecting health information in a new era, BMC Medical Ethics 22 (1) (Sep. 2021). doi:10.1186/s12910-021-00687-3.

[108] E. D. McAlpine, L. Pantanowitz, P. M. Michelow, Challenges Developing Deep Learning Algorithms in Cytology, Acta Cytologica 65 (4) (2020) 301–309. arXiv:https://karger.com/acy/article-pdf/65/4/301/3690762/000510991.pdf, doi:10.1159/000510991.

[109] M. M. Rahaman, C. Li, X. Wu, Y. Yao, Z. Hu, T. Jiang, X. Li, S. Qi, A survey for cervical cytopathology image analysis using deep learning, IEEE Access 8 (2020) 61687–61710. doi:10.1109/ACCESS.2020.2983186.

[110] C. Edlund, T. R. Jackson, N. Khalid, N. Bevan, T. Dale, A. Dengel, S. Ahmed, J. Trygg, R. Sjögren, LIVECell—a large-scale dataset for label-free live cell segmentation, Nature Methods 18 (9) (2021) 1038–1045. doi:10.1038/s41592-021-01249-6.

[111] M. Schwendy, R. E. Unger, S. H. Parekh, EVICAN—a balanced dataset for algorithm development in cell and nucleus segmentation, Bioinformatics 36 (12) (2020) 3863–3870. doi:10.1093/bioinformatics/btaa225.

[112] C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, Cellpose: a generalist algorithm for cellular segmentation, Nature Methods 18 (1) (2020) 100–106. doi:10.1038/s41592-020-01018-x.

[113] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, in: European Conference on Computer Vision, Springer, 2020, pp. 213–229.

[114] G. Jocher, A. Chaurasia, J. Qiu, Ultralytics YOLO (Jan. 2023).
URL https://github.com/ultralytics/ultralytics

[115] S. Ren, K. He, R. Girshick, J. Sun, Faster r-cnn: Towards real-time object detection with region proposal networks, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (6) (2016) 1137–1149.

[116] M. G. Camporeale, G. Dimauro, M. Gelardi, G. Iacobellis, M. S. Ladisa, S. Latrofa, N. Lomonte, Nasal mucosa cell dataset (nmcd) (2024). doi:10.21227/0erx-zn98.
URL https://dx.doi.org/10.21227/0erx-zn98

[117] M. Avanzo, M. Porzio, L. Lorenzon, L. Milan, R. Sghedoni, G. Russo, R. Massafra, A. Fanizzi, A. Barucci, V. Ardu, M. Branchini, M. Giannelli, E. Gallio, S. Cilla, S. Tangaro,

A. Lombardi, G. Pirrone, E. De Martin, A. Giuliano, G. Belmonte, S. Russo, O. Rampado, G. Mettivier, Artificial intelligence applications in medical imaging: A review of the medical physics research in italy, Physica Medica 83 (2021) 221–241. doi:https://doi.org/10.1016/j.ejmp.2021.04.010.

[118] A. M. Bulfamante, F. Ferella, A. M. Miller, C. Rosso, C. Pipolo, E. Fuccillo, G. Felisati, A. M. Saibene, Artificial intelligence, machine learning, and deep learning in rhinology: a systematic review, European Archives of Oto-Rhino-Laryngology 280 (2) (2022) 529–542. doi:10.1007/s00405-022-07701-3.

[119] Y. J. Jun, J. Jung, H.-M. Lee, Medical data science in rhinology: Background and implications for clinicians, American Journal of Otolaryngology 41 (6) (2020) 102627. doi:https://doi.org/10.1016/j.amjoto.2020.102627.

[120] C. Trincianti, M. A. Tosca, G. Ciprandi, Updates in the diagnosis and practical management of allergic rhinitis, Expert Review of Clinical Pharmacology 16 (7) (2023) 669–676, pMID: 37314373. doi:10.1080/17512433.2023.2225770.

[121] D. B., N. J., roboflow (Version 1.0) [Software]. Available from https://roboflow.com. computer vision. (2022).

[122] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[123] R. Girshick, J. Donahue, T. Darrell, J. Malik, Region-based convolutional networks for accurate object detection and segmentation, IEEE Transactions on Pattern Analysis and Machine Intelligence 38 (1) (2016) 142–158. doi:10.1109/TPAMI.2015.2437384.

[124] T.-Y. Lin, M. Maire, S. Belongie, L. Bourdev, R. Girshick, J. Hays, P. Perona, D. Ramanan, C. L. Zitnick, P. Dollár, Microsoft coco: Common objects in context (2015). arXiv:1405.0312.

[125] R. Aggarwal, V. Sounderajah, G. Martin, D. S. Ting, A. Karthikesalingam, D. King, H. Ashrafian, A. Darzi, Diagnostic accuracy of deep learning in medical imaging: a systematic review and meta-analysis, NPJ Digital Medicine 4 (1) (2021) 65.

[126] H. Jiang, Y. Zhou, Y. Lin, R. C. Chan, J. Liu, H. Chen, Deep learning for computational cytology: A survey, Medical Image Analysis 84 (2023) 102691.

[127] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, et al., Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai, Information Fusion 58 (2020) 82–115.

[128] D. Saraswat, P. Bhattacharya, A. Verma, V. K. Prasad, S. Tanwar, G. Sharma, P. N. Bokoro, R. Sharma, Explainable ai for healthcare 5.0: opportunities and challenges, IEEE Access 10 (2022) 84486–84517.

[129] N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. L. de Prado, E. Herrera-Viedma, F. Herrera, Connecting the dots in trustworthy artificial intelligence: From ai principles, ethics, and key requirements to responsible ai systems and regulation, Information Fusion 99 (2023) 101896.

[130] J. Amann, A. Blasimme, E. Vayena, D. Frey, V. I. Madai, P. Consortium, Explainability for artificial intelligence in healthcare: a multidisciplinary perspective, BMC Medical Informatics and Decision Making 20 (2020) 1–9.

[131] H. Jiang, Y. Zhou, Y. Lin, R. C. Chan, J. Liu, H. Chen, Deep learning for computational cytology: A survey, Medical Image Analysis 84 (2023) 102691.

[132] N. Wentzensen, B. Lahrmann, M. A. Clarke, W. Kinney, D. Tokugawa, N. Poitras, A. Locke, L. Bartels, A. Krauthoff, J. Walker, et al., Accuracy and efficiency of deep-learning–based automation of dual stain cytology in cervical cancer screening, JNCI: Journal of the National Cancer Institute 113 (1) (2021) 72–79.

[133] M. Nauta, R. Van Bree, C. Seifert, Neural prototype trees for interpretable fine-grained image recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 14933–14943.

[134] M. Plissiti, E. Tripoliti, A. Charchanti, O. Krikoni, D. Fotiadis, Automated detection of cell nuclei in pap stained cervical smear images using fuzzy clustering, in: 4th European Conference of the International Federation for Medical and Biological Engineering: ECIFMBE 2008 23–27 November 2008 Antwerp, Belgium, Springer, 2009, pp. 637–641.

[135] J. Huang, T. Wang, D. Zheng, Y. He, Nucleus segmentation of cervical cytology images based on multi-scale fuzzy clustering algorithm, Bioengineered 11 (1) (2020) 484–501.

[136] M. Benhari, R. Hossseini, An improved fuzzy deep learning (ifdl) model for managing uncertainty in classification of pap-smear cell images, Intelligent Systems with Applications 16 (2022) 200133.

[137] A. Manna, R. Kundu, D. Kaplun, A. Sinitca, R. Sarkar, A fuzzy rank-based ensemble of cnn models for classification of cervical cytology, Scientific Reports 11 (1) (2021) 14538.

[138] R. Kundu, H. Basak, A. Koilada, S. Chattopadhyay, S. Chakraborty, N. Das, Ensemble of cnn classifiers using sugeno fuzzy integral technique for cervical cytology image classification, arXiv preprint arXiv:2108.09460 (2021).

[139] G. Dimauro, G. Ciprandi, F. Deperte, F. Girardi, E. Ladisa, S. Latrofa, M. Gelardi, Nasal cytology with deep learning techniques, International Journal of Medical Informatics 122 (2019) 13–19.

[140] M. Erten, I. Tuncer, P. D. Barua, K. Yildirim, S. Dogan, T. Tuncer, R.-S. Tan, H. Fujita, U. R. Acharya, Automated urine cell image classification model using chaotic mixer deep feature extraction, Journal of Digital Imaging (2023) 1–12.

[141] S. Y. Rubaiat, M. M. Rahman, M. K. Hasan, Important feature selection & accuracy comparisons of different machine learning models for early diabetes detection, in: 2018 International Conference on Innovation in Engineering and Technology (ICIET), IEEE, 2018, pp. 1–6.

[142] G. Casalino, G. Castellano, G. Zaza, Evaluating the robustness of a contact-less mhealth solution for personal and remote monitoring of blood oxygen saturation, Journal of Ambient Intelligence and Humanized Computing (2022) 1–10.

[143] M. K. Hasan, M. H. Aziz, M. I. I. Zarif, M. Hasan, M. Hashem, S. Guha, R. Love, S. Ahamed, Help me: Recommendations for non-invasive hemoglobin level prediction in mobile-phone environment, JMIR Mhealth Uhealth (2019).

[144] M. Valueva, G. Valuev, N. Semyonova, P. Lyakhov, N. Chervyakov, D. Kaplun, D. Bogaevskiy, Construction of residue number system using hardware efficient diagonal function, Electronics 8 (6) (2019) 694.

[145] M. Gelardi, L. Iannuzzi, N. Quaranta, M. Landi, G. Passalacqua, Nasal cytology: practical aspects and clinical relevance, Clinical & Experimental Allergy 46 (6) (2016) 785–792.

[146] G. Caruso, M. Gelardi, G. C. Passali, M. M. De Santi, Nasal scraping in diagnosing ciliary dyskinesia, American journal of rhinology 21 (6) (2007) 702–705.

[147] B. T. Lemieux, J. J. Chen, J. Jing, Z. Chen, B. J. Wong, Measurement of ciliary beat frequency using doppler optical coherence tomography, in: International forum of allergy & rhinology, Vol. 5, Wiley Online Library, 2015, pp. 1048–1054.

[148] J. Yager, T.-M. Chen, M. J. Dulfano, Measurement of frequency of ciliary beats of human respiratory epithelium, Chest 73 (5) (1978) 627–633.

[149] H. Teichtahl, P. Wright, R. Kirsner, Measurement of in vitro ciliary beat frequency: a television-video modification of the transmitted light technique, Medical and Biological Engineering and Computing 24 (1986) 193–196.

[150] J. Gray, The mechanism of ciliary movement.—vi. photographic and stroboscopic analysis of ciliary movement, Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character 107 (751) (1930) 313–332.

[151] A. Coste, M.-C. Millepied, C. Chapelin, P. Reinert, F. Poron, M. Boucherat, E. Escudier, Incidence of primary ciliary dyskinesia in children with recurrent respiratory diseases, Annals of Otology, Rhinology & Laryngology 106 (10) (1997) 854–858.

[152] I. Schipor, J. N. Palmer, A. S. Cohen, N. A. Cohen, Quantification of ciliary beat frequency in sinonasal epithelial cells using differential interference contrast microscopy and high-speed digital video imaging, American journal of rhinology 20 (1) (2006) 124–127.

[153] C. O'Callaghan, K. Sikand, M. A. Chilvers, Analysis of ependymal ciliary beat pattern and beat frequency using high speed imaging: comparison with the photomultiplier and photodiode methods, Cilia 1 (2012) 1–7.

[154] J. H. Sisson, J. Stoner, B. Ammons, T. A. Wyatt, All-digital image capture and whole-field analysis of ciliary beat frequency, Journal of microscopy 211 (2) (2003) 103–111.

[155] J. J. Chen, B. T. Lemieux, B. J. Wong, A low-cost method of ciliary beat frequency measurement using iphone and matlab: rabbit study, Otolaryngology–Head and Neck Surgery 155 (2) (2016) 252–256.

[156] G. Mantovani, M. Pifferi, G. Vozzi, Automated software for analysis of ciliary beat frequency and metachronal wave orientation in primary ciliary dyskinesia, European archives of oto-rhino-laryngology 267 (2010) 897–902.

[157] C. M. Smith, J. Djakow, R. C. Free, P. Djakow, R. Lonnen, G. Williams, P. Pohunek, R. A. Hirst, A. J. Easton, P. W. Andrew, et al., ciliafa: a research tool for automated, high-throughput measurement of ciliary beat frequency using freely available software, Cilia 1 (1) (2012) 1–7.

[158] W. Kim, T. H. Han, H. J. Kim, M. Y. Park, K. S. Kim, R. W. Park, An automated measurement of ciliary beating frequency using a combined optical flow and peak detection, Healthcare informatics research 17 (2) (2011) 111–119.

[159] E. Puybareau, H. Talbot, G. Pelle, B. Louis, J.-F. Papon, A. Coste, L. Najman, A regionalized automated measurement of ciliary beating frequency, in: 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), IEEE, 2015, pp. 528–531.

[160] O. Meste, F. Brau, A. Guyon, Robust estimation of the motile cilia beating frequency, Medical & biological engineering & computing 53 (2015) 1025–1035.

[161] G. Dimauro, A new image quality metric based on human visual system, in: 2012 IEEE International Conference on Virtual Environments Human-Computer Interfaces and Measurement Systems (VECIMS) Proceedings, IEEE, 2012, pp. 69–73.

[162] L. Lévesque, Nyquist sampling theorem: understanding the illusion of a spinning wheel captured with a video camera, Physics Education 49 (6) (2014) 697.

[163] J. Wang, L. Perez, et al., The effectiveness of data augmentation in image classification using deep learning, Convolutional Neural Networks Vis. Recognit 11 (2017) (2017) 1–8.

[164] C. Shorten, T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, Journal of big data 6 (1) (2019) 1–48.

[165] G. Jocher, A. Chaurasia, J. Qiu, YOLO by Ultralytics (Jan. 2023).
URL `https://github.com/ultralytics/ultralytics`