UNIVERSITAT de BARCELONA

Final Degree Project

Biomedical Engineering Degree

AI-based diagnostic algorithm for cellular aging using single-cell fluorescence imaging of the cytoskeleton

Barcelona, 11 June 2025

Author: Laia Colomé Xicoy

Director/Tutor: Núria Gavara Casas

Abstract

Aging is a complex, multifactorial process marked by physiological decline, increased disease risk, and progressive functional deterioration. Fibroblasts, which are essential for maintaining tissue structure and repair, exhibit well-documented age-related changes. However, identifying robust, quantifiable markers of these changes remains challenging. This project develops a protocol for detecting cellular aging using single-cell fluorescence imaging.

Fibroblasts from young and old mice are immunostained for nuclear and cytoskeletal markers and imaged via multichannel epifluorescence microscopy. A computational pipeline is developed to segment cells and extract features describing morphology, intensity, texture, and cytoskeletal organization. Features that pass a first statistical filter are used to train and validate interpretable machine learning models later. Logistic regression is chosen for its balance between interpretability and performance, achieving an accuracy of 0.97 on the evaluation set.

The final protocol provides a reproducible method for detecting fibroblast aging solely from microscopy images. It enables both diagnostic through binary classification and mechanistic insight into age-related cellular changes, laying the groundwork for future applications in aging clocks, guiding biomarker discovery, and the screening of therapies aimed at modulating age-related cellular changes.

Resum

L'envelliment és un procés complex i multifactorial caracteritzat per un declivi fisiològic, un augment del risc de malalties i un deteriorament funcional progressiu. Els fibroblasts, essencials per mantenir l'estructura i la reparació dels teixits, mostren canvis relacionats amb l'edat àmpliament documentats. Tot i això, identificar marcadors robustos i quantificables d'aquests canvis continua sent un repte. Aquest projecte desenvolupa un protocol per detectar l'envelliment cel·lular mitjançant imatges de fluorescència de cèl·lules individuals.

Els fibroblasts de ratolins joves i vells són immunotenyits per a marcadors nuclears i citoesquelètics, i se n'obtenen imatges mitjançant microscòpia d'epifluorescència multicanal. Es desenvolupa un mètode computacional per segmentar les cèl·lules i extreure característiques que descriuen la morfologia, la intensitat, la textura i l'organització del citoesquelet. Les característiques que superen un primer filtratge estadístic s'utilitzen després per entrenar i validar models d'aprenentatge automàtic interpretables. Es selecciona la regressió logística per la seva combinació d'interpretabilitat i rendiment, assolint una precisió del 0.97 en el conjunt d'avaluació.

El protocol final ofereix un mètode reproductible per detectar l'envelliment dels fibroblasts només a partir d'imatges de microscòpia. Permet tant el diagnòstic a partir d'una classificació binària com la comprensió mecanística dels canvis cel·lulars relacionats amb l'edat, i estableix les bases per a aplicacions futures com els rellotges d'envelliment, el descobriment de biomarcadors i l'avaluació de teràpies orientades a modular aquests canvis.

PARAULES CLAU: ENVELLIMENT CEL·LULAR, FIBROBLASTS, MICROSCÒPIA DE FLUORESCÈNCIA, CITOESQUELET, APRENENTATGE AUTOMÀTIC, SEGMENTACIÓ D'IMATGES, EXTRACCIÓ DE CARACTERÍSTIQUES, BIOMARCADORS.

Acknowledgements

Glossary of abbreviations

αSMA: Alpha Smooth Muscle Actin
AFM: Atomic Force Microscopy
AI: Artificial Intelligence
ANOVA: Analysis of Variance
AUC: Area Under the Curve
BB: Blocking Buffer
BH: Benjamini-Hochberg
BSA: Bovine Serum Albumin
CSV: Comma-Separated Values
DAPI: 4'6-Diamidino-2-Phenylindole
DMEM: Dulbecco's Modified Eagle Medium
DT: Decision Tree
ECM: Extracellular Matrix
F-ACTIN: Filamentous Actin
FBS: Fetal Bovine Serum
FDR: False Discovery Rate
GLCM: Gray Level Co-occurrence Matrix
K-NN: K-Nearest Neighbors
LR: Logistic Regression
ML: Machine Learning
MMP: Matrix Metalloproteinases
NPV: Negative Predictive Value
PBS: Phosphate-Buffered Saline
PCA: Principal Component Analysis
PFA: Paraformaldehyde
PS: Penicillin-Streptomycin
RF: Random Forest
RFECV: Recursive Feature Elimination with Cross Validation
RNA-seq: RNA sequencing
ROC: Receiver Operating Characteristic
SASP: Senescence-Associated Secretory Phenotype
SVM: Support Vector Machine
TGF: Transforming Growth Factor

List of figures

List of tables

PAGE

Table of contents

1   INTRODUCTION

Aging is a crucial biological process that impacts all living beings, resulting in extensive effects at the cellular, tissue, and systemic levels. This project examines aging from a cellular perspective, with a particular focus on fibroblasts, which are crucial for tissue maintenance and repair, and how their characteristics change with age.

1.1   Motivation and aim of the project

Aging is a complex, multifactorial, and time-dependent process that affects organisms at every biological level, from molecular to systemic, ultimately leading to increased vulnerability to disease and a decline in regenerative capacity[1,2]. As life expectancy continues to rise globally, understanding the mechanisms of aging has become a central focus in biomedical research. In this broader context, cellular aging, the gradual decline in function of individual cells, emerges as a key contributor to the age-related deterioration of tissue structure and performance[3], with a growing need to identify reliable and readily employable indicators of cellular aging[4].

Among the various cell types involved in tissue maintenance, fibroblasts are notably impacted by aging. As key contributors to the structural and functional integrity of connective tissues, fibroblasts support numerous physiological processes essential for tissue homeostasis. However, aging is associated with widespread changes in fibroblast behavior and function, which can disrupt tissue equilibrium and contribute to the onset of age-related pathologies[5]. This makes fibroblasts an excellent model for assessing cellular hallmarks of aging and investigating mechanisms underlying age-related tissue degeneration.

In this context, this project aims to develop a quantitative and interpretable pipeline to characterize age-related phenotypic changes in fibroblasts, focusing on morphology and cytoskeletal organization at the single-cell level. By preparing, acquiring, and analyzing fluorescence images of primary lung fibroblasts isolated from young and old mice, this study aims to identify cellular features that vary with age and can serve as reliable and potentially predictive indicators of the aging process.

This approach provides insights into fibroblast aging, guides identification of potential biomarkers, and enhances understanding of structural cellular changes related to aging. It may also assist in screening treatments to reverse or modulate age-related alterations and, while still early, paves the way for tools to estimate cellular age from visual features, advancing the field of aging clocks.

1.2   Objectives

To achieve the project's aims, the work is structured around the following specific objectives:

- Develop an image analysis pipeline from scratch, including single-cell segmentation and the extraction of quantitative features.
- Train and validate a machine learning classification model to distinguish between young and aged fibroblasts based on the extracted features, prioritizing model interpretability.
- Gain biologically meaningful insights into aging phenotypes by analyzing and interpreting the most informative features that contribute to the classification.

1.3   Scope and limitations

This project encompasses all the objectives outlined in the previous section, including sample preparation through immunostaining assays, image acquisition, analysis, and classification of pulmonary fibroblasts from young and old mice.

However, it relies on several critical preparatory steps that have not been carried out directly by the author. Dissection of lungs for fibroblast isolation has been performed by a licensed technician, due to animal handling regulations. At the same time, a lab researcher has managed cell expansion and maintenance to accommodate project timelines.

The project also presents some inherent limitations. These relate primarily to time constraints, sample availability, and biological relevance.

1. Time-intensive processes

- *Fibroblast culture development:* Establishing cultures at proper confluency and optimal conditions often takes longer than expected, as growth rates differ by batch.
- *Epifluorescence image acquisition:* A manually operated microscope is shared among users, requiring extended sessions in a dark room and manual setting adjustments, which, along with scheduling conflicts, may delay imaging.
- *Computational workload:* Training segmentation software, processing hundreds of images, and extracting features are computationally intensive tasks needing significant processing power. These operations can take several hours and may limit or prevent other simultaneous tasks from being completed.

2. Limited biological material

- Primary fibroblasts are isolated from a small number of mice, each providing a finite number of viable cells. Tissue quality, dissection efficiency, and cell viability introduce variability that can reduce the final pool of usable cells. Additional losses during staining, mounting, or imaging may further limit the number of analyzable cells and replicates, reducing statistical power.

3. Biological relevance

- *In vitro culture:* Fibroblasts are cultured as 2D monolayers on glass, which does not fully replicate the 3D architecture, mechanical forces, and cellular context of their native tissue environment. This discrepancy can influence cellular behavior and phenotype.
- *Species differences:* While murine fibroblasts serve as a valuable model system, they differ from human cells in key aspects such as gene regulation, cellular behavior, and responses to stimuli. As a result, findings obtained in mouse models may not fully translate to human biology and should be interpreted with appropriate caution regarding their broader biological relevance.

1.4   Location and span of the project

This project is carried out in the Biophysics and Bioengineering Unit at the University of Barcelona, which hosts the Mechanobiology of the Cytoskeleton Research Group. The group studies the

mechanical behavior of cells and tissues using physical and engineering principles, with a particular focus on the pulmonary system. Ongoing projects include characterizing the lung extracellular matrix, studying cancer-associated fibroblasts (CAFs), and developing biophysical biomarkers for drug screening in lung diseases[6].

The current project benefits from the unit's established expertise and infrastructure. This includes access to specialized equipment, biological materials, and technical guidance, ensuring that all phases of the project can be carried out successfully.

The project timeline runs from February to June 2025, following a 240-hour internship completed the previous summer in the same laboratory. This internship was crucial for acquiring technical skills, developing laboratory independence, and gaining familiarity with the protocols necessary to conduct the experimental aspects of the study. Considering weekends and holidays, the project spans approximately 133 days, requiring efficient planning and execution across the experimental and computational phases to meet its objectives within the available timeframe.

2   BACKGROUND

The background section outlines the current knowledge relevant to the project, focusing on the biological role of fibroblasts in tissue maintenance and repair, the structural and functional changes they undergo with aging, and the main methods used to study these processes (with a focus on the emergence in recent years of data-driven computational approaches).

Although all fibroblast subsets arise from a common mesenchymal origin, they display substantial heterogeneity due to differences in anatomical location, function and cellular state[5]. This project centers on pulmonary fibroblasts, in line with the host laboratory's research focus; however, findings are expected to have broader relevance across tissue types. Accordingly, the background draws on examples from various organ systems, not just the lung.

2.1   Fibroblasts and their main functions

Fibroblasts are among the most abundant cell types in the body, present in virtually all organ systems, alongside endothelial and blood cells. They are the most widespread in connective tissues[7]. Typically spindle-shaped, they exist in both dense connective tissues like tendons, ligaments, and fasciae, as well as loose connective tissues, including the dermis, mucosal layers, and interstitial spaces between muscle and fat[8].

Fibroblasts play a crucial role in maintaining tissue structure and function, primarily through their production and dynamic remodeling of the extracellular matrix (ECM)[9], a complex network of proteins that includes collagen, elastin, fibronectin, and proteoglycans. This ECM, collectively referred to as the matrisome, provides mechanical support, elasticity, and a scaffold for cell signaling and migration[10].

Beyond shaping tissue architecture and modulating stiffness, fibroblasts play an active role in tissue homeostasis, development, and repair. They secrete and respond to a wide range of growth factors and cytokines, facilitating communication with immune, epithelial, and endothelial cells[11]. As mechanosensitive signaling hubs, fibroblasts detect changes in their microenvironment and adjust their behavior accordingly, influencing inflammation, angiogenesis, and cell differentiation[12]. The broad range of fibroblast functions is summarized in Figure 1.
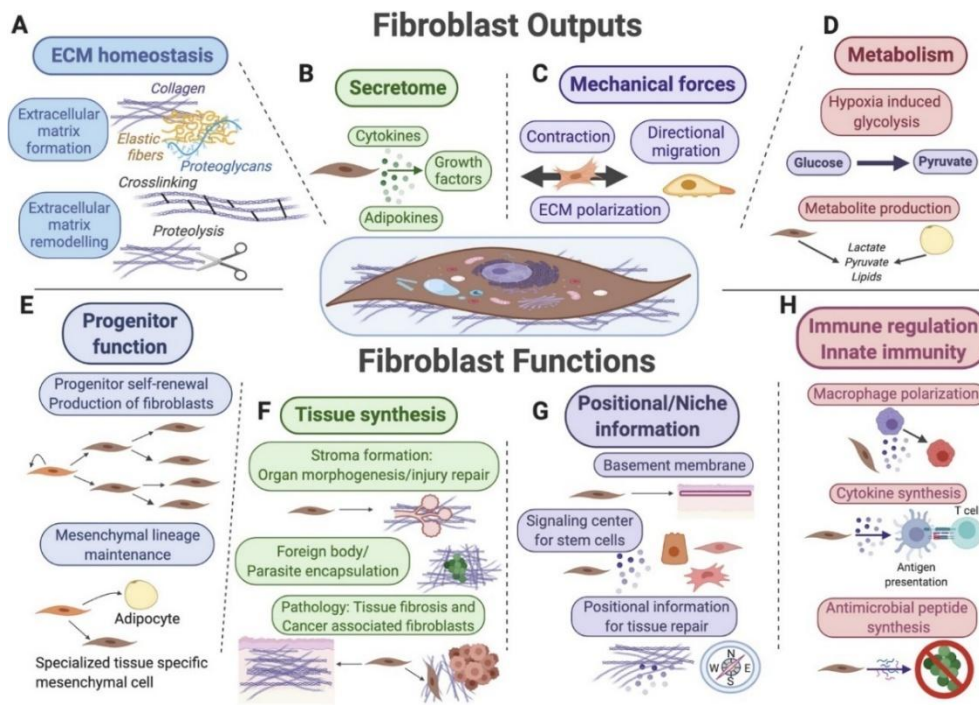
Figure 1 · Overview of fibroblast functions and outputs.
Fibroblasts contribute to a wide range of physiological processes through diverse functional outputs. These include ECM secretion and remodelling (A), secretion of signalling factors for surrounding cells (B), generation of mechanical forces (C), metabolic regulation (D), progenitor cells function for lineage maintenance (E), tissue synthesis and repair (F), positional and niche signalling (G), and roles in immune regulation and innate immunity (H). Together, these functions highlight the central role of fibroblasts in tissue structure, signaling, and homeostasis [8].

One of their best-characterized roles is in wound healing, which progresses through four overlapping phases: hemostasis, inflammation, proliferation, and remodeling, as illustrated in Figure 2. While fibroblasts are not directly involved in hemostasis, they play a central role in the later stages: migrating into wounds, depositing ECM, and differentiating into contractile myofibroblasts (characterized by the expression of α-smooth muscle actin (α-SMA)) that assist in wound closure. In the final phase, they contribute to remodeling the ECM through enzymes such as matrix metalloproteinases (MMPs), ultimately returning to a quiescent state as the tissue heals[7,13].

However, dysregulation of fibroblast activity can lead to pathology. Persistent activation, especially of myofibroblasts, drives fibrosis, characterized by excessive ECM deposition that stiffens tissues, as seen in pulmonary fibrosis or keloids. Conversely, reduced or senescent fibroblast function may impair healing and contribute to chronic wounds[13]. Even regular repair rarely restores full tissue strength; for example, healed skin typically recovers only about 80% of its original tensile strength[15].

Given their central role in tissue architecture, mechanical sensing, and regeneration, fibroblasts are increasingly recognized as key players in age-related functional decline of tissue and as promising targets for regenerative medicine and anti-fibrotic therapies[5].
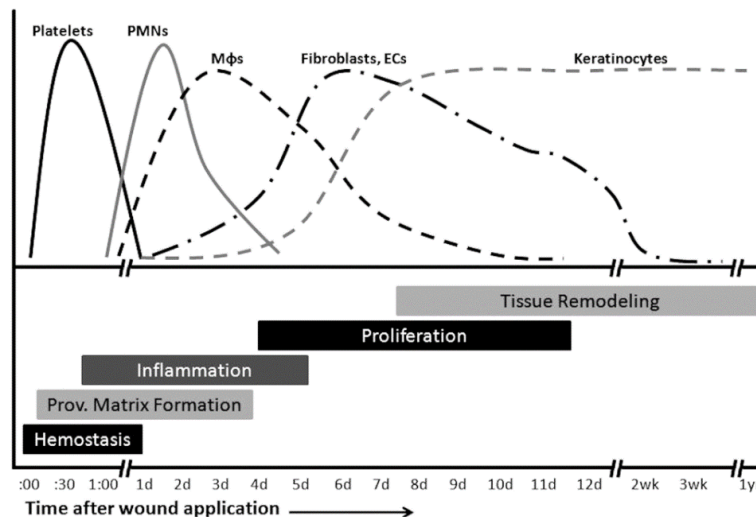
Figure 2 *· Cellular dynamics and phases of wound healing over time.*
Sequential phases of wound healing: hemostasis, inflammation, proliferation, and tissue remodeling, along with the temporal involvement of key cell types. Platelets and polymorphonuclear cells (PMNs) are active early during the inflammatory phase, followed by macrophages (MΦs), fibroblasts, endothelial cells (ECs), and keratinocytes. Fibroblasts contribute to the inflammatory phase and subsequent phases by releasing cytokines, coordinating immune responses, and later supporting extracellular matrix deposition and tissue repair during the proliferative and remodeling phases[14].

2.2   Fibroblasts and ageing

As organisms age, fibroblasts undergo a range of molecular and functional changes that impair their ability to maintain tissue homeostasis. These changes include cellular senescence, altered morphology and mechanics, disrupted signaling, and reduced regenerative capacity. Together, they contribute significantly to the decline in tissue repair and function observed in aged individuals. Below, some of the most well-characterized age-related changes in fibroblasts are described.

Replicative cellular senescence

Fibroblasts undergo a limited number of divisions before entering cellular senescence, which is an irreversible growth arrest during which cells remain metabolically active but can no longer proliferate[16]. A subtype of senescence, known as replicative senescence, is characteristic of fibroblasts and arises from the progressive erosion of telomeres during cell divisions and the activation of the DNA damage response pathway[17].

In vivo studies show that fibroblasts from older individuals proliferate more slowly and accumulate more in a senescent state. Although this process helps prevent cancer and aids temporary wound stabilization, chronic accumulation of senescent fibroblasts impairs tissue regeneration and causes structural decline[5]. For instance, studies indicate that in aging skin, the number of active fibroblasts sharply declines, impairing their ability to renew the dermal matrix and leading to thinner skin, reduced elasticity, and delayed wound healing[18].

Morphological and mechanical changes

Aged fibroblasts show distinct morphological changes, appearing larger and more flattened with irregular nuclear shapes due to chromatin remodeling and nuclear envelope instability.[19,20] These

changes occur alongside cytoskeletal reorganization. Studies report increased levels of all three major cytoskeletal components: Filamentous actin (F-actin), tubulin, and vimentin. In older fibroblasts, F-actin and tubulin appear shorter and thinner, while vimentin forms longer, thicker, and more bundled fibers. This shift increases cytoskeletal tension and stiffness, reducing mechanosensitivity and limiting the cell's ability to respond to physical cues.[21]

Research shows that aged dermal fibroblasts can be 60% stiffer than young ones[22]. This increased stiffness is closely linked to functional impairments, including reduced migration and compromised wound repair. In the aging lung, such fibroblast stiffening is further associated with excessive ECM deposition and decreased tissue compliance, contributing to fibrotic remodeling characteristic of diseases like idiopathic pulmonary fibrosis[23].

Secretory phenotype and inflammaging

Aged and senescent fibroblasts often adopt a pro-inflammatory secretory profile known as the Senescence-Associated Secretory Phenotype (SASP). This includes the release of inflammatory cytokines, such as tumor necrosis factor, MMPs, and growth factors, as well as ECM-degrading enzymes, which can induce inflammation, alter the tissue microenvironment, and affect neighboring cells[24]. While SASP can play beneficial roles in acute responses, such as wound healing, by recruiting immune cells, its chronic expression contributes to a persistent, low-grade inflammation known as inflammaging. This state disrupts immune balance and is linked to age-related diseases, including fibrosis, osteoarthritis, and certain cancers[25,26].

In aged tissues, the accumulation of SASP-expressing fibroblasts, coupled with reduced immune clearance, can drive progressive tissue damage and impaired regeneration. SASP factors can also spread senescence to nearby cells and impair immune function, reinforcing a cycle of inflammation, ECM degradation, and tissue dysfunction[5].

Reduced migration and wound healing capacity

Efficient tissue repair requires fibroblasts to migrate into the wound site, where they deposit new ECM and coordinate with other cell types to facilitate wound healing. However, research indicates that fibroblasts from aged individuals migrate shorter distances, in fewer numbers, and respond less effectively to migratory cues such as growth factors[27]. These impairments have been linked to increased cell stiffness and cytoskeletal changes that physically restrict cell movement, as well as the altered expression of proteins involved in motility[21].

As a result, wounds in aged tissues often remain in the inflammatory phase for more extended periods, heal more slowly, and are at greater risk of developing into chronic, non-healing ulcers[28,29]. In some cases, delayed fibroblast migration may also contribute to excessive fibrosis driven by prolonged inflammation[30].

Altered differentiation and ECM production

Aging fibroblasts exhibit impaired differentiation into myofibroblasts, specialized cells marked by α-SMA expression that are crucial for wound contraction and ECM remodeling[31]. Interestingly, α-SMA expression during aging appears to be context-dependent: it tends to decline during skin wound healing[32] but may increase in fibrotic conditions, such as pulmonary fibrosis[33]. This indicates a loss of proper regulatory control over fibroblast activation rather than a uniform decline.

These changes are further influenced by altered responsiveness to growth factors, particularly transforming growth factor-beta (TGF-β), whose age-related dysregulation promotes fibrosis and disorganized ECM remodeling[34]. Aged fibroblasts also produce less collagen and have a reduced ability to assemble and organize ECM components. The collagen produced becomes increasingly cross-linked and rigid, leading to diminished tissue elasticity and function[35].

2.3 State of the art: Approaches to studying fibroblast aging

In recent years, there has been a significant increase in studies utilizing computational methods, particularly machine learning (ML) and deep learning (DL), to investigate changes associated with fibroblast aging. While fibroblast biology has traditionally been examined through molecular, imaging, and mechanical techniques, the recent integration of these methods with data-driven computational tools marks a new development.

Most research using this combined strategy has emerged since 2021, indicating a trend toward high-throughput, quantitative, and interpretable analyses of cellular aging. This review categorizes the latest state-of-the-art methods for studying fibroblast aging based on their main experimental techniques.

2.3.1 Imaging-based classification

High-content fluorescence imaging is a powerful tool in fibroblast aging research, enabling quantitative analysis of morphological and cytoskeletal changes at the single-cell level. Several studies utilize image-based ML to identify structural hallmarks of aging and senescence. For example, Welter et al. track senescence progression in human lung fibroblasts by measuring features such as increased cell and nuclear size, reduced circularity, and more cytoplasmic protrusions[36]. Similarly, Duran et al. focus on nuclear morphology and chromatin texture to develop classifiers that predict senescence in both cultured cells and tissue sections[37].

Building on these approaches, Neri et al. introduce the FAST pipeline, an automated method that combines β-galactosidase staining (a standard marker for detecting cellular senescence), EdU incorporation (to assess cell proliferation), and nuclear area measurements (related to enlarged morphology) to calculate a per-cell senescence score using random forest models[38]. This system is well-suited for large-scale studies and drug screening. Similarly, Hillsley et al. develop a DL model that detects α-SMA-positive stress fibers, enabling automated identification of myofibroblast activation, a phenotype associated with aging and fibrosis[39].

2.3.2 Transcriptomic and molecular signatures

Beyond morphology, transcriptomic studies provide complementary insights into the molecular programs that drive fibroblast aging. Fleischer et al. perform RNA sequencing (RNA-seq) on human dermal fibroblasts from donors aged 1 to 94 years to determine whether aging signatures are encoded within the transcriptome. Their machine learning model achieves high accuracy and even detects premature aging in patients with progeria, a genetic disorder causing early-onset aging[40].

At single-cell resolution, Yu et al. develop CellBiAge, a pipeline that predicts cell age from single-cell RNA-seq data using binarized gene expression and ML[41]. Although initially applied to brain tissue, the framework is also adaptable to fibroblast datasets. These studies show that fibroblasts

display not only morphological signs of aging but also transcriptional changes, indicating alterations in cell function, identity, and regenerative capacity.

2.3.3 Atomic force microscopy

Atomic force microscopy (AFM) offers a biophysical perspective on fibroblast aging by enabling nanoscale measurements of mechanical properties, such as stiffness, elasticity, and adhesion, at the single-cell level. Multiple studies have reported that aged or senescent fibroblasts exhibit increased stiffness compared to young fibroblasts, a change generally attributed to cytoskeletal remodeling, particularly due to an increased F-actin content[42,43].

While AFM has been extensively used to characterize the mechanical phenotypes of aging cells[44], its integration with classification workflows, particularly ML-based age or senescence prediction, is still relatively limited in the published literature. Although some promising studies exist in related contexts (such as cancer vs. normal cells), clear examples applying AFM feature sets for classifying fibroblast age or senescence status remain scarce.

3 Market analysis: The longevity ecosystem

The longevity ecosystem is a rapidly expanding field focused on extending healthspan, understood as the number of years lived in good health, through innovations spanning health technology, life sciences, healthcare, and wellness.
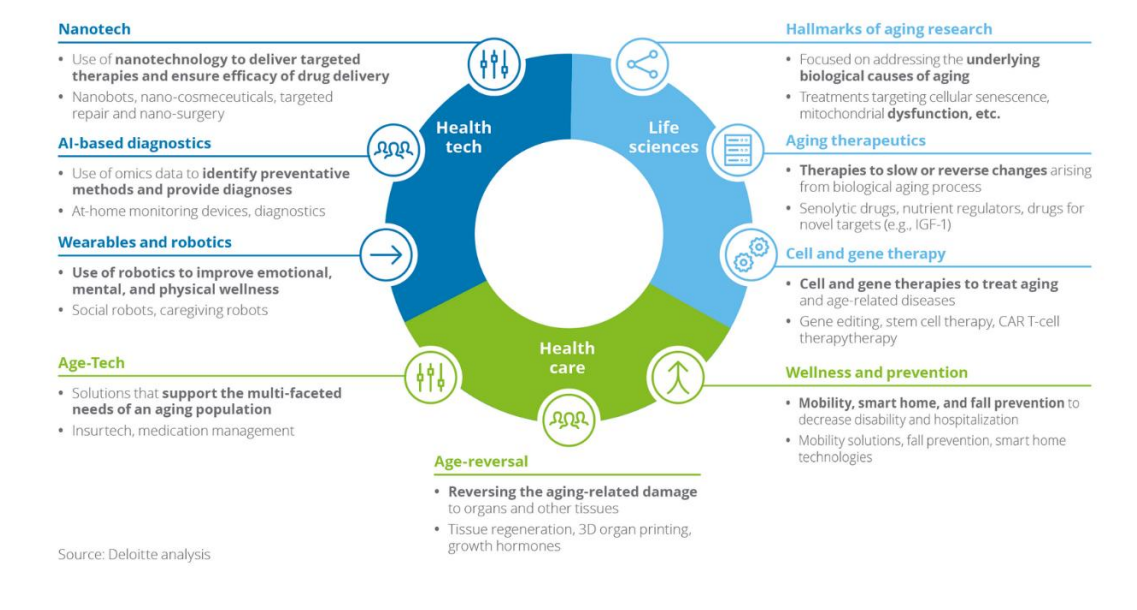
Figure 3 · Overview of the Longevity Ecosystem.
The ecosystem encompasses health tech, life sciences, and care innovations focused on understanding, delaying, or reversing the aging process[45]. (Adapted from Deloitte Analysis)

It includes advances in artificial intelligence (AI) based diagnostics, wearables, and robotics that enable early detection of age-related decline and promote independent living. In parallel, biomedical therapies are being developed to target core biological mechanisms of aging, such as cellular senescence and mitochondrial dysfunction. The ecosystem also encompasses regenerative and preventive solutions, including tissue repair technologies and age-tech tools,

designed to improve quality of life and mitigate the functional decline associated with aging, as illustrated in Figure 3.

Although still emerging, the longevity field is already attracting growing interest and support from investors, universities, and governments. By 2020, the top 50 companies focused on longevity had raised over $1 billion in funding, a figure that continues to increase[45]. This underscores the field's strong potential and promising future.

3.1   Context and strategic relevance

Within this ecosystem, aging research has become a central priority in biomedicine and biotechnology, driven by increasing life expectancy and the growing impact of age-related diseases. As global populations age, healthcare systems face increasing pressure from chronic conditions, tissue degeneration, and a decline in regenerative capacity. Therefore, understanding the biological mechanisms that contribute to aging is essential for developing new approaches to prevent or delay diseases and to improve quality of life in later years[2].

This reflects the central idea of geroscience, which hypothesizes that targeting the biological processes underlying aging itself, rather than treating individual age-related diseases, may prevent, delay, or reduce the severity of multiple conditions simultaneously. This approach also has the potential to extend healthspan[46].

At the same time, public health priorities are shifting toward prevention, early detection, and personalized intervention. This has led to growing interest in identifying measurable indicators of biological aging that go beyond chronological age. Such indicators are essential not only to improve understanding of the aging process but also to tailor medical strategies to individual needs and evaluate the effectiveness of emerging therapies[46,47].

In this context, aging is no longer seen as an inevitable decline, but as a process that can be studied, monitored, and potentially modified. Tools capable of capturing and measuring changes at the cellular and tissue level play a key role in this transformation[48].

3.2   Current players and efforts

These shifts have given rise to a new wave of research and innovation aimed at developing tools to monitor, understand, and intervene in the biological processes of aging. A range of international institutions and companies are now driving this transition through focused programs that target aging at its cellular and molecular roots.

At the forefront of this transformation are research centers such as the Buck Institute for Research on Aging (USA) and the Max Planck Institute for Biology of Aging (Germany)[49,50]. These organizations have played a crucial role in advancing the understanding of the molecular and cellular aspects of aging, including cellular senescence, mitochondrial dysfunction, and impaired intercellular communication. Their research has laid the groundwork for identifying measurable indicators of aging, which can be used to track biological decline and assess treatment options.

In parallel, biotechnology companies are investing in strategies to reverse or slow down the aging process. Calico Life Sciences (a Google subsidiary) and Altos Labs (supported by major technology

investors)[51,52] are conducting long-term research into cellular rejuvenation, epigenetic reprogramming, and tissue regeneration.

3.3   Market opportunities and trends

The aging and longevity market is growing rapidly, with strong interest in tools that can measure biological age more accurately, rather than relying solely on chronological age. One of the most promising areas is the development of aging biomarkers, features that reflect how cells or tissues are aging, often before symptoms appear[53]. These are essential for early diagnosis, personalized treatments, and evaluating the effects of anti-aging interventions.

A significant trend is the use of AI and ML to analyze large datasets from imaging, transcriptomics, and other high-throughput sources[47,54]. This has enabled the rise of aging clocks, predictive models that estimate biological age based on measurable inputs[54]. Examples include ImAge[55], which uses gene expression data, and Google's eyeAge[56], which predicts age from retinal images. Therefore, these tools show how computational methods are transforming aging research into practical diagnostic and clinical tools.

3.4   Target sectors

The growing interest in aging research and longevity technologies is creating opportunities across several key sectors:

- Pharmaceuticals are increasingly focused not only on treating age-related diseases such as fibrosis, osteoarthritis, neurodegeneration, and metabolic disorders but also on targeting the underlying biological mechanisms of aging itself. As aging is now recognized as a major driver of chronic conditions, drug development is shifting toward modulating processes such as cellular senescence[57], chronic inflammation[58], and mitochondrial dysfunction[59], among others. This opens new avenues for therapies that aim to extend healthspan and improve resilience in later life.

- The cosmetics industry, particularly in anti-aging skincare, is driven by consumer demand for products that maintain skin health and appearance. Companies are seeking technologies that assess biological aging at the cellular level to support efficacy claims and differentiate their products[60]. Tools that measure cell morphology, senescence markers, or extracellular matrix production provide objective ways to evaluate the impact of a product. These approaches also open the door to more personalized skincare, where treatments are tailored based on an individual's biological skin profile rather than age alone[61].

- Healthcare is increasingly interested in tools that detect biological aging and early signs of decline before clinical symptoms emerge. The identification of aging markers can inform preventive care, personalize risk assessments, and track responses to interventions. These tools support a shift toward proactive, individualized healthcare, helping to reduce the burden of advanced age-related diseases on healthcare systems[48,62].

4   Concept Engineering

This section identifies the project's specific needs. Based on these requirements, each workflow step is examined in detail, considering alternative approaches and their suitability for the project. This process leads to a comprehensive pipeline that supports the study of age-related changes in fibroblasts at the single-cell level.

4.1   Target outcome

The primary outcome of this project is the development of a comprehensive analysis pipeline that leads to a supervised, interpretable machine learning model capable of classifying pulmonary fibroblasts as either young or aged based solely on morphological, cytoskeletal, and nuclear features extracted from multichannel fluorescence microscopy images. The model is expected to:

- Identify and rank the most predictive cellular features.
- Demonstrate generalization across biological replicates and sexes.
- Support interpretability by revealing how specific cellular structures change with age, offering insights that may guide future biomarker discovery.

This output serves as both a functional classifier and a biological interpretation tool, prioritizing feature explainability over raw predictive power.

4.2   Conceptual workflow design

The development of the project is organized into three primary stages: building the dataset, processing images, and analyzing data. These stages outline the project's conceptual journey.

It is essential to note that some steps in this workflow are based on protocols that have been extensively optimized by the host research group over the years, while others require specific adaptation and development within the scope of this project and are described in greater detail.

4.2.1   Dataset construction

A key step is generating the dataset, which serves as the foundation for all subsequent analysis. Constructing the dataset involves two main stages: sample collection, encompassing all procedures up to the establishment of fibroblast cultures, and image acquisition, which includes immunofluorescence staining and microscopy required to capture high-quality images for analysis.

4.2.1.1   Sample collection

Lung fibroblasts can be obtained from two primary sources: human tissue and animal models. While human samples provide direct relevance to human aging, they often pose ethical, logistical, and standardization challenges. Animal models, particularly mice, offer greater experimental control over variables such as genetics, environment, and treatment exposure. Mice are especially useful due to their genetic similarity to humans, susceptibility to many of the same diseases, well-characterized biology, and short reproductive cycles. In this project, mouse models are selected because they are already well-established in the hosting research group, making them the most practical and feasible option.

Primary fibroblasts isolated from young and old mice are preferred over immortalized commercial cell lines because they retain native, age-specific characteristics and better reflect in vivo biology. In contrast, immortalized commercial cell lines bypass natural aging processes, offering less control over donor age and thereby limiting their suitability for aging studies. While primary cultures introduce more variability due to biological and processing differences, this is considered an acceptable trade-off for preserving physiological relevance in age-related research. A summary of the different fibroblast sourcing strategies is provided in Table 1.

DESCRIPTION	⬆ ADVANTAGES	⬇ DISADVANTAGES	
Use freshly isolated human fibroblasts.	Direct relevance to human aging	Ethical and logistical complexity, limited availability, variability across donors	✖
Isolate primary murine fibroblasts.	A good balance of relevance and feasibility, enabling control over age, sex, and environment.	Species differences compared to humans	✔
Commercial fibroblast cell lines	Readily available; standardized	Typically immortalized and no control over the age variable; reduced physiological relevance.	✖

Table 1 · *Comparison of fibroblasts source options considered.*
Summary of the advantages and disadvantages of using freshly isolated human fibroblasts, primary murine fibroblasts, and commercial cell lines, with a focus on their relevance to aging research, practical feasibility, and biological limitations.

The animals in this project are part of a larger research initiative and are not sacrificed exclusively for this study. As such, the project must adapt to the available population of mice. Cell isolation, culture, and seeding follow standardized protocols established by the research group, ensuring methodological consistency and aligning the experimental procedures with prior validated workflows in the laboratory.

4.2.1.2 Image acquisition

To characterize fibroblasts, particularly in the context of age-associated changes, researchers commonly employ a wide range of techniques, including molecular profiling (such as transcriptomics), biomechanical analysis (such as atomic force microscopy, or AFM), and advanced imaging methods (such as fluorescence microscopy). As outlined in the background section, each of these approaches offers unique insights into cellular function and aging.

This project builds on the host research group's expertise, aiming to derive quantitative features that characterize cell structure, particularly in cytoskeletal organization, nuclear morphology, and cell shape. Multichannel fluorescence microscopy is selected as the most suitable method for achieving this goal.

Different fluorescence microscopy techniques are available for studying cytoskeletal architecture, each with distinct trade-offs in terms of resolution, depth, speed, and complexity[63]. This project selects epifluorescence microscopy, as it is optimized and available in the lab. The research team has extensive prior experience with the instrument, and optimal combinations of fluorophores and filters for each marker are validated in-house. This setup enables efficient project progression without the need for new imaging infrastructure or lengthy calibration times.

DESCRIPTION	PRINCIPLE	▲ ADVANTAGES	▼ DISADVANTAGES	
Wide-field fluorescence microscopy	Uniform illumination of the entire sample; all emitted light (in-focus and out-of-focus) is collected	Simple, fast, cost-effective; suitable for thin samples and high-throughput imaging	Limited optical sectioning; reduced image clarity in thick or densely stained samples; High background due to out-of-focus fluorescence light.	✔
Confocal fluorescence microscopy	Uses a focused laser beam and a pinhole to eliminate out-of-focus light, enabling optical sectioning	High resolution and contrast allow 3D imaging and more precise visualization of cytoskeletal structures.	Slower acquisition, more expensive equipment, higher photobleaching risk	✘
Total Internal Reflection Fluorescence (TIRF) microscopy	Creates an evanescent wave that excites fluorophores within ~100–200 nm of the glass surface	Very high signal-to-noise ratio; ideal for imaging near the basal membrane (e.g., focal adhesions, cortical actin)	Limited to near-surface structures; cannot image deeper cytoskeletal components	✘
Super-resolution microscopy (STED, SIM, STORM, PALM)	Overcomes the diffraction limit using structured light patterns, depletion, or stochastic activation	Ultra-high spatial resolution (20–100 nm); reveals fine cytoskeletal details not visible with conventional methods	Technically complex, slow imaging speed, limited field of view; not suitable for high-throughput analysis	✘

Table 2 · Overview of fluorescence microscopy techniques commonly used for cytoskeletal imaging.
Comparison of wide-field, confocal, TIRF, and super-resolution microscopy methods, summarizing their underlying principles, strengths, and limitations in the context of cytoskeletal visualization [63–65].

One limitation of epifluorescence microscopy (as seen in Table 2) is the significant background fluorescence caused by out-of-focus light. To address this, a background estimation step is implemented post-acquisition, where the background signal is computationally subtracted from the original image. This correction ensures that feature extraction relies on the biological signal rather than imaging artifacts, making it crucial before the quantitative analysis.

4.2.1.2.1 Imaging markers for fibroblast characterization

Cytoskeletal remodeling and nuclear alterations are hallmarks of fibroblast aging, making it logical to select biomarkers that reflect changes in cellular architecture. The aim is to visualize components indicating age-related differences in morphology and cytoskeletal organization. Therefore, a panel of four immunofluorescent markers is utilized to label the nucleus, actin cytoskeleton, contractile fibers, and intermediate filaments, allowing for comprehensive structural characterization of fibroblasts across age groups. The selected markers (described in Table 3) prioritize three main criteria:

- Biological relevance to aging: Each marker targets a structural component affecting cellular aging or senescence.
- Technical feasibility: Staining protocols are validated within the host lab, ensuring reproducibility and efficiency.
- Complementarity: The four markers provide a comprehensive view of cellular structure, including nuclear morphology, cytoskeletal organization, and activation state.

Although microtubule-related markers (like α-tubulin) or mitochondrial markers (like TOM20) could have been considered, they are ultimately not included, as they do not align as well with the three main criteria guiding this decision. In conclusion, two separate immunostaining protocols will be used: one combining α-SMA, phalloidin, and DAPI, and the other substituting vimentin for α-SMA. This approach is necessary due to technical limitations in simultaneously multiplexing all four

markers, and it ensures the reliable visualization of both actin- and intermediate filament-related features while maintaining image quality and signal specificity.

MARKER	DESCRIPTION
4',6-diamidino-2-phenylindole (DAPI)	It is a blue-fluorescent stain that firmly attaches to A-T-rich regions of DNA, facilitating precise observation of cell nuclei[66]. This dye is crucial for distinguishing individual cells and assessing nuclear characteristics, such as size, shape, and chromatin texture, which often change with age. Typically, aged or senescent fibroblasts exhibit enlarged or irregular nuclei and disrupted chromatin organization, indicating senescence or a diminished ability to proliferate[67].
Phalloidin	It binds specifically to F-actin, allowing visualization of the actin cytoskeleton. F-actin forms structures like stress fibers, which support cell shape, force generation, and migration. Aging shifts the actin balance toward disorganized and expanded F-actin networks[68] and stress fiber formation[69], leading to a stiffer, less motile phenotype. In contrast, younger fibroblasts exhibit more dynamic actin structures, such as lamellipodia and filopodia[70]. Phalloidin staining effectively reveals these age-related differences in cytoskeletal organization.
α-SMA	It is a contractile actin isoform expressed by activated fibroblasts (myofibroblasts) during wound healing and fibrosis. Quiescent fibroblasts express little to no α-SMA, but upon activation, it integrates into stress fiber-like structures[71,72]. Staining for α-SMA helps assess whether aged fibroblasts exhibit impaired activation, even in the absence of injury.
Vimentin	It is an intermediate filament protein in fibroblasts that forms a cytoskeletal network from the nucleus to the cell periphery. It plays key roles in maintaining cell shape, positioning organelles, resisting mechanical stress, and facilitating migration, wound healing, and mechanotransduction. It has been seen that aged fibroblasts exhibit increased vimentin abundance and thicker, longer vimentin fibers, and their organization and bundling can potentially contribute to increased stiffness and impaired motility[21]. Vimentin staining allows for the visualization of this intermediate filament network.

Table 3 · Fluorescent markers used for fibroblast characterization.
Summary of the function and relevance of DAPI, phalloidin, α-SMA, and vimentin in visualizing nuclear, cytoskeletal, and activation-related changes associated with fibroblast aging.

4.2.2   Image processing

Before quantitative analysis can be performed, the raw fluorescence microscopy images must undergo a series of processing steps. For clarity, a brief introduction to the software tools used throughout the image processing and data analysis stages is provided on Table 4.

	SOFTWARE TOOLS	DESCRIPTION
	ImageJ/Fiji	Java-based open-source software widely used in biological and medical imaging. Offers tools for visualization, cropping, intensity adjustments, and measurements. Its plugin-based architecture enables flexible and reproducible workflows[73,74].
	Ilastik	Open-source tool for interactive image segmentation using supervised machine learning. Supports pixel classification and object tracking through intuitive brush-based annotations, producing semantic segmentation maps[75].
	CellProfiler	Modular, open-source software for high-throughput biological image analysis. Allows users to build custom pipelines for preprocessing, segmentation, and feature extraction, supporting scalable and reproducible workflows[76].

PROGRAMMING LANGUAGES AND ENVIRONMENTS		DESCRIPTION
	Matlab	High-level programming language and environment for data analysis, image processing, and visualization. Offers specialized libraries and interactive tools for building custom workflows in scientific research.
	R & RStudio	High-level language for statistical computing and data analysis. Paired with RStudio, it offers a user-friendly environment for scripting, plotting, and managing analytical workflows.
	Python & Jupyter Notebook	Versatile, high-level programming language used in scientific computing and image analysis; libraries such as OpenCV and scikit-image support automation and custom scripts. Jupyter Notebook provides an interactive, reproducible coding environment.

Table 4 *· Software tools, programming languages, and environments used in image analysis and data processing.*
Summary of the main platforms employed for image preprocessing, segmentation, feature extraction, and data analysis throughout the project.

4.2.2.1   Segmentation of individual cells

Once the immunofluorescence images are acquired, the first step in the image processing workflow is to split the multichannel images into separate grayscale images for each fluorescent marker. This task is performed using a custom MATLAB script developed within the host laboratory, ensuring compatibility with the microscope output format and downstream processing pipeline.

Before channel separation, cells are cropped semi-automatically using a batch process in ImageJ/Fiji. This decision anticipates common image quality issues, such as uneven staining, partial substrate detachment (which is frequent in aged fibroblast cultures), and overlap between neighboring cells. Thus, cells affected by these limitations can be excluded from analysis while retaining all morphologically distinct or biologically relevant phenotypes that meet basic quality criteria.

Segmentation strategy

Before outlining the image processing pipeline, various segmentation strategies frequently detailed in the literature are examined to determine their suitability for the generated dataset[39,77–80]. A summary comparison of the three approaches considered is provided in Table 5.

Threshold-based segmentation in CellProfiler

An initial approach involves building a comprehensive segmentation workflow in CellProfiler using threshold-based methods. Although it produces visually accurate segmentations, it is sensitive to slight variations in signal intensity. Consequently, a single set of thresholding parameters cannot be consistently applied across the dataset, and manually adjusting parameters for each image is time-consuming and not scalable for large-scale processing.

Hybrid Segmentation Workflow Using Ilastik and CellProfiler

To improve robustness and scalability, a hybrid approach is used by combining Ilastik with CellProfiler. Ilastik performs ML-based image processing through its pixel classification workflow,

where users annotate a subset of images to train a classifier that assigns each pixel a probability of belonging to a class such as nucleus, cytoplasm, or background. This approach is practical in handling signal variability, uneven illumination, and subtle texture differences by using local image features, including intensity, edges, and texture.

The probability maps generated by Ilastik are then used as input in CellProfiler to guide object detection and support accurate identification of nuclei and whole cell structures. Relying on probability thresholds instead of fixed intensity cutoffs allows for more consistent and automated batch processing. CellProfiler produces binary masks that label each pixel as part of a specific object, generating one mask for nuclei and another for entire cell bodies. These masks serve as the foundation for quantitative feature extraction in later stages of the pipeline.

Before generating probability maps, an initial evaluation is conducted to determine whether preprocessing the input images improves segmentation. This preprocessing includes image normalization and background subtraction and is applied only to Ilastik inputs, while CellProfiler continues to work with the original raw images. Results show that this step improves segmentation accuracy by approximately 20%, and it is therefore integrated into the workflow.

Deep Learning based segmentation

Another strategy could involve the use of deep learning-based segmentation models such as U-Net or Cellpose. These methods are becoming the gold standard in biomedical image segmentation due to their ability to learn complex spatial features and handle challenging conditions. However, deep learning models are not implemented in this project due to time constraints, limited access to annotated training data, and the fact that the combined Ilastik and CellProfiler approach already meets the project's accuracy and performance needs.

DESCRIPTION	⬆ ADVANTAGES	⬇ DISADVANTAGES	
Threshold-based segmentation in CellProfiler	Simple to implement and does not require training data	Highly sensitive to noise, uneven staining, and signal variability; requires extensive manual parameter tuning	✖
Hybrid segmentation workflow using Ilastik and CellProfiler	Improved segmentation accuracy; handles variable signal intensity well; minimal user intervention once trained	Requires initial annotation and classifier training; some tuning of thresholds is still needed in CellProfiler	✔
Deep learning-based segmentation	High segmentation accuracy; handles complex morphologies effectively, and challenging imaging conditions	Requires annotated training datasets; computationally intensive; not feasible under current time constraints; and defined project priorities	✖

Table 5 · *Comparison of segmentation strategies evaluated in the project.*
Summary of the three image segmentation approaches considered: thresholding in CellProfiler, a hybrid workflow using Ilastik and CellProfiler, and deep learning-based models; highlighting their main advantages and limitations.

4.2.2.2   Feature extraction

Two main strategies are considered for this task: using CellProfiler with pre-configured modules for feature extraction, and developing custom Python scripts that utilize libraries such as *scikit-image*, *mahotas*, and *OpenCV*. CellProfiler offers a straightforward and scalable solution, but it has limited transparency in feature computation and less flexibility in parameter settings. Given the project's

need for precise measurement interpretation, the Python-based approach is selected for complete control over feature definitions and tailored analysis of aging-related phenotypes.

An alternative approach would be to utilize deep learning models to directly extract latent features from images, thereby eliminating the need for explicit definitions. Although this captures complex spatial patterns, it is not adopted in this project due to the limited dataset size and the need for biological interpretability, which is challenging with high-dimensional representations. A comparison of these feature extraction strategies is summarized in Table 6.

DESCRIPTION	⬆ ADVANTAGES	⬇ DISADVANTAGES	
Feature extraction using CellProfiler	Easy to implement; scalable for large datasets; widely used in cell biology workflows	Limited control over how features are computed; less transparency in feature definitions	✖
Feature extraction using custom Python scripts	Complete control over feature definitions; high interpretability; tailored to specific biological questions	Requires more development time and programming effort; less standardized.	✔
Deep learning–based feature extraction (e.g., CNN embeddings)	Can capture complex or previously unknown patterns; no need for manual feature design	Requires larger annotated datasets; results are difficult to interpret biologically; high computational cost	✖

Table 6 *· Comparison of feature extraction strategies.*
Overview of the three feature extraction approaches considered: CellProfiler-based extraction, custom Python scripting, and deep learning methods. The table highlights their advantages and limitations in terms of scalability, interpretability, flexibility, and computational requirements.

4.2.3   Data analysis

The final stage involves analyzing the extracted features to identify patterns associated with cellular aging and assess their potential for classification. This analysis includes two main components: statistical testing to identify significant and biologically relevant features, and a machine learning phase to determine how well these features distinguish between young and aged fibroblasts.

4.2.3.1   Statistical analysis

Statistical analyses are conducted using R to facilitate both exploratory data analysis and feature filtering. Analysis of variance (ANOVA) identifies features with significant differences between young and aged fibroblasts. It is favored over simpler tests, such as the t-test, as the dataset comprises both male and female samples. Thus, there is a need to consider sex as a factor and isolate age-related differences, leading to more robust and biologically relevant features.

In parallel, correlation analysis identifies variables that are highly correlated with each other. Reducing redundancy among these features helps prevent the overrepresentation of similar information, improving both model performance and interpretability. While dimensionality reduction methods, such as principal component analysis (PCA), are available, they generate abstract components that lack clear biological meaning. The chosen approach prioritizes retaining interpretable features while minimizing redundancy.
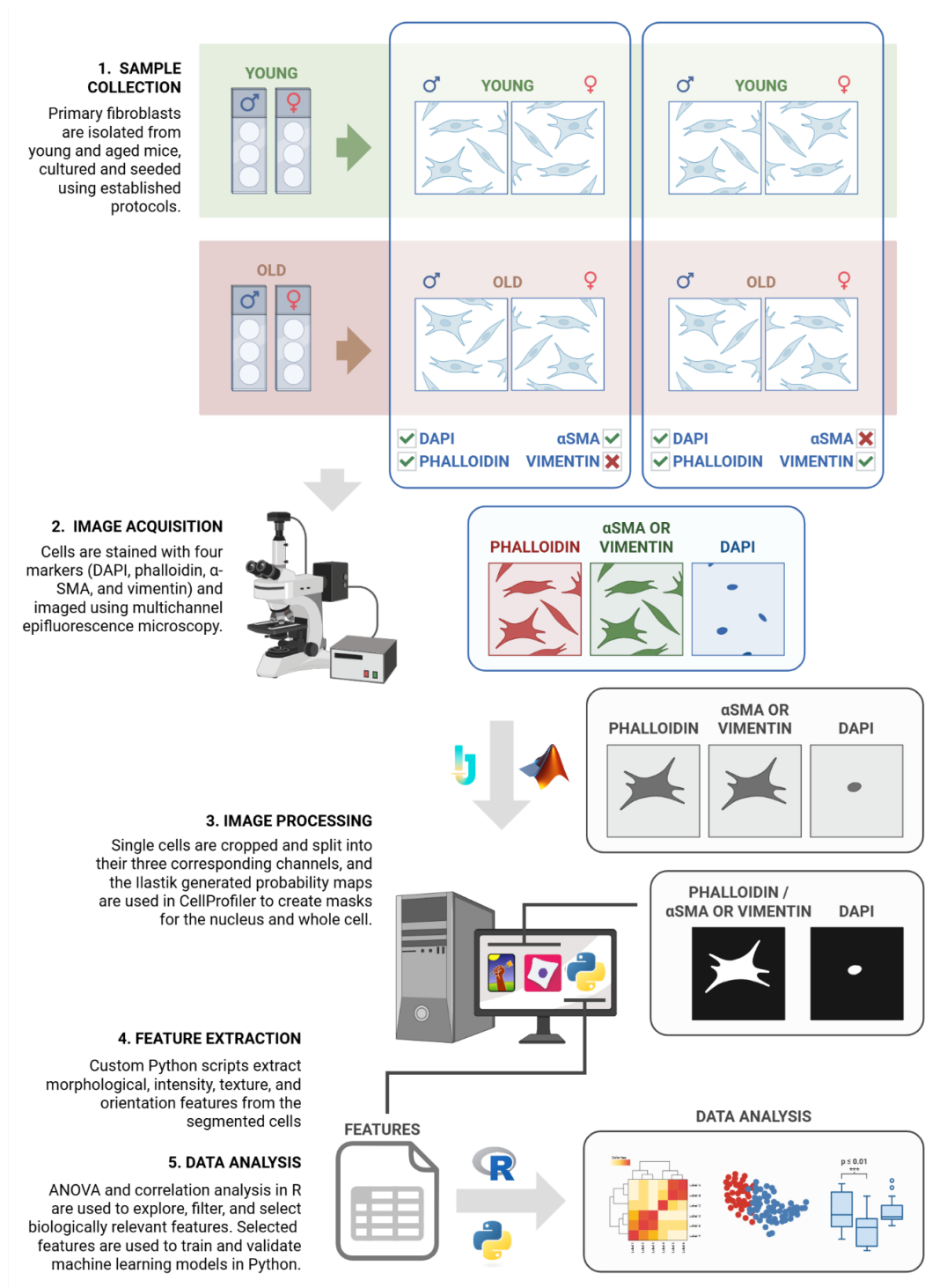
Figure 4 · *Overview of the final pipeline developed.*
Summary of each significant step in the workflow, from fibroblast isolation and imaging to feature extraction and data analysis, outlining how biological samples are transformed into interpretable, quantitative insights about cellular aging.

4.2.3.2 Machine learning

The subset of selected features is then used to train and evaluate supervised ML models using Python. Several algorithms will be explored, including[81,82]:

- Decision tree (DT): splits the data based on feature values to form a set of if-then rules that are highly interpretable.
- k-Nearest neighbors (k-NN): classifies new samples based on the majority class among their closest feature-space neighbors.
- Random Forest (RF): builds an ensemble of randomized decision trees and aggregates their predictions to improve accuracy and reduce overfitting.
- Support Vector Machine (SVM): attempts to find the optimal boundary (or hyperplane) that separates classes in a high-dimensional feature space.
- Logistic Regression (LR): models the probability of class membership by applying a logistic (S-shaped) function to a linear combination of input features.

These models offer different trade-offs in terms of complexity, interpretability, and robustness. Given the aims of this project, particular attention will be paid to models that provide insights into which features contribute most to classification, since interpretability is key for linking quantitative patterns back to biological mechanisms related to aging. Although DL models could also be applied, this option is not pursued, as simpler, more interpretable approaches are better suited to the scale and objectives of this project.

4.3 Overview of the final pipeline

Figure 4 summarizes the key decisions made throughout the conceptual design process and outlines the final pipeline that will be followed in the project.

5 Detail Engineering

This section outlines the practical implementation of the pipeline defined in the conceptual phase, including dataset construction, image processing, and data analysis. It covers isolating and staining fibroblasts, image acquisition, segmentation, feature extraction, as well as training and evaluating predictive models. Finally, the analysis results are discussed in terms of their biological relevance for understanding cellular aging.

5.1 Primary fibroblast isolation and culture

Primary pulmonary fibroblasts are obtained from a total of twelve C57BL/6 mice, stratified by age and sex. The sample includes three young males and three young females at 5 weeks of age, as well as three old males and three old females at 74 weeks of age.

Lungs are extracted by a qualified technician following animal handling protocols. Fibroblasts are then isolated using the crawl-out method within the host laboratory. Once obtained, fibroblasts are cultured in T25 or T75 flasks, based on their count and growth rate. The culture medium is Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin (PS). Before each medium change, cells are washed with phosphate-buffered saline (PBS) 1x containing 1% PS to remove debris and impurities, thereby reducing the risk of contamination. Cultures are kept under standard conditions (37°C, 5% CO_2) until the desired confluence is reached for subsequent steps. During the isolation and culture process, some group-specific differences are observed:

- Young female fibroblasts take longer to adhere and proliferate than those from males. This delay may result from excessive carbon dioxide exposure during euthanasia or difficulties in tissue handling, as in some instances, lung fragments from young females detached before fibroblasts had the opportunity to migrate out.
- Cultures from older animals show a greater tendency for tissue fragments to detach from the wells, possibly indicating age-related changes in tissue integrity or reduced adhesion. Additionally, fewer fibroblasts are recovered from old females compared to males, although the underlying causes remain unclear.

Since the fibroblasts are not isolated exclusively for this project, and given that cell growth is significantly lower in old samples compared to young ones, it is not possible to include all isolated animals in the whole experimental workflow. Notably, old male #1 and old female #2 are excluded from further processing due to insufficient cell yield. Their cells are instead cryopreserved for potential use in future experiments.

5.2 Cell seeding and immunostaining

This section describes the protocol for preparing fibroblast samples for fluorescence imaging, including seeding on glass coverslips and staining to label key cellular structures.

5.2.1 Cell Seeding

Cell seeding is carried out using 13 mm glass coverslips placed inside 40 mm diameter Petri dishes. Each dish contains three coverslips. A total of 10 dishes are prepared for staining with α-SMA, phalloidin, and DAPI, and 10 dishes for staining with vimentin, phalloidin, and DAPI. Coverslips are first immersed in 70% ethanol for approximately 20 minutes to sterilize them. They are then rinsed with PBS 1x and conditioned by adding the cell culture medium DMEM supplemented with 10% FBS and 1% PS.

Cells undergo a trypsinization step to detach them from the culture surface. Before this, they are washed with PBS 1x containing 1% PS. For T75 flasks, 1 mL of 0.25% trypsin is used, while 0.5 mL of the same solution is applied for T25 flasks. Trypsin is neutralized with cell culture medium (5 mL for T75 and 3 mL for T25), and cells are incubated at 37ºC for approximately 2 minutes. The suspension is then centrifuged at 1200 rpm for 5 minutes at 23ºC. After centrifugation, the supernatant is discarded, and the pellet is resuspended in 1 mL of cell culture medium.

To determine the number of cells needed for seeding, cell concentration is measured using a Neubauer chamber. Cells are mixed with Trypan Blue in a 1:2 ratio for viability assessment. Counting is performed across the four large quadrants in each chamber, and the average is calculated between both chambers. The following formula is used to calculate cell concentration:

$$\frac{cells}{mL} = \left(\frac{total\ cells\ in\ chamber}{4}\right) \cdot dilution\ factor \cdot 10^4$$

To achieve a seeding density of $3000\ cells/cm^2$, and knowing that each Petri dish has an area of $9.2\ cm^2$, the target number of cells per dish is calculated as $3000\frac{cells}{cm^2} \times 9.2cm^2 = 27,600\ cells$. From the measured cell concentration, the required volume of cell suspension is calculated and used for seeding. After plating, the cells are incubated for approximately 24 hours to allow

attachment to the coverslips. Following this period, fixation is performed using 4% paraformaldehyde (PFA).

5.2.2 Immunostaining protocol

Since the samples are stored at 4°C in PBS until all conditions are seeded and ready for staining, a PBS 1x wash is performed at room temperature before beginning the staining protocol. This step allows the samples to equilibrate to room temperature before proceeding. A summary of the complete immunostaining protocol, including the mounting stage, is provided in Figure 5.

Permeabilization: To permeabilize the cell membranes and allow antibodies to access intracellular structures, a 0.2% Triton X-100 detergent solution in PBS 1x is applied. Each Petri dish containing three coverslips receives 2 mL of this solution. The samples are placed on an orbital rotator at 80 rpm and incubated at room temperature for 20 minutes. After permeabilization, the cells are washed three times with PBS 1x to remove any remaining detergent.

Blocking: To reduce background and prevent nonspecific antibody binding, a blocking buffer (BB) made of 10% FBS in PBS 1x is applied to each Petri dish, completely covering the coverslips. Each dish is placed on an orbital rotator at 80 rpm and incubated at room temperature for 45 minutes.

Primary Antibody: While the blocking step is in progress, the primary antibody solutions are prepared at the following concentrations: 1:1000 for the mouse monoclonal anti-vimentin antibody and 1:500 for the rabbit recombinant monoclonal anti-αSMA antibody. After blocking, 1 mL of the corresponding antibody solution is added to each Petri dish. The dishes are then placed on an orbital rotator at 80 rpm and incubated overnight at 4°C.

Secondary Antibody: After three washes with BB, 1 mL of the appropriate secondary antibody solution is added to each Petri dish, ensuring the coverslips are fully covered. Anti-rabbit antibodies are used for αSMA and anti-mouse antibodies for vimentin, both prepared at a concentration of 1:500. The dishes are then incubated at 37°C for 2 hours in complete darkness on an orbital rotator set to 80 rpm to promote binding between the primary and secondary antibodies. Following incubation, the samples are washed three times with PBS 1x.

Phalloidin: Phalloidin staining is performed using a directly conjugated phalloidin antibody, eliminating the need for a separate secondary antibody step. The staining solution is prepared by diluting the conjugated antibody at 1:1000 in PBS 1x supplemented with 1% bovine serum albumin (BSA). Each Petri dish receives 1 mL of this solution. The dishes are then incubated in the dark at room temperature for 45 minutes. After incubation, the samples are washed three times with PBS 1x to remove excess unbound stain.

DAPI: For nuclear staining, DAPI is applied at a concentration of 2 drops per mL. One mL of the solution is added to each Petri dish. The dishes are then incubated on an orbital rotator at 80 rpm for 20 minutes at room temperature, in complete darkness. Following incubation, the samples are washed three times with PBS 1x to remove excess reagent.
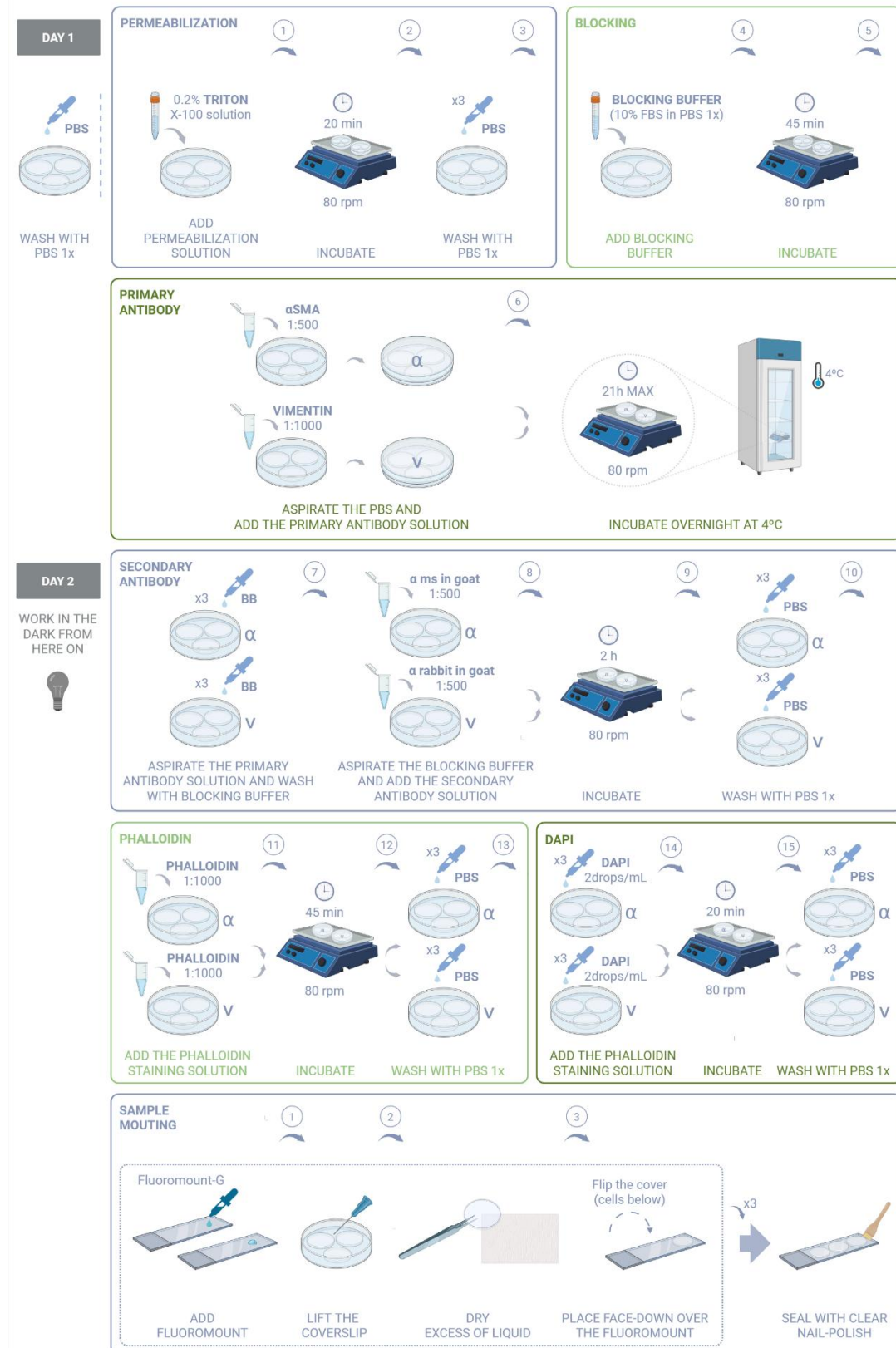
Figure 5 · *Schematic overview of the immunostaining and mounting protocol.*
Step-by-step schematic of the two-day protocol used to label fibroblast cytoskeletal and nuclear structures. It includes permeabilization, blocking, incubation with primary and secondary antibodies, and staining with phalloidin and DAPI.

Mounting: For mounting, clean microscope slides are prepared by placing three drops of the mounting medium Fluoromount-G™ on each slide corresponding to a given condition. Three coverslips per condition are carefully lifted using tweezers, and any excess liquid is gently removed from the edges using absorbent paper. The coverslips are then positioned face down onto the mounting medium and sealed in place using clear nail polish. Once mounted, the samples are stored at 4°C in complete darkness to preserve fluorescence.

5.3   Image Acquisition

Epifluorescence imaging is performed using a Leica SP5 inverted microscope equipped with a CCD camera (C9100, Hamamatsu Photonics) and a Nikon CFI Plan Fluor 20x oil immersion objective. Fluorescent signals are captured through TXRED, FITC, and UV-2A channels, corresponding to the different fluorophores used in the staining protocol.

Image acquisition is performed using Micro-Manager software, which also exports the images in tagged image file format (TIFF) as composite files for analysis. Cells are imaged across various regions of each coverslip to ensure representative sampling. Around 150 images are collected per condition, with each image typically containing about five cells. All visible phenotypes are included, except for cells that appear damaged, detached, or excessively clustered, which are excluded from analysis.

Some samples show suboptimal cell yield or quality. Notably, young male #2 and old male #3 produce fewer usable cells than expected. In the case of young male #2, many cells appear compromised, likely due to technical artifacts during isolation, staining, or mounting. While the exact cause is unclear, these limitations are taken into account during image selection and downstream analysis to ensure data quality.

5.4   Image Processing

This section describes the computational workflow applied to convert raw multichannel fluorescence images into quantitative data suitable for analysis.

5.4.1   Single-Cell Cropping and Channel Splitting

The image analysis workflow is designed to operate at the single-cell level, as mentioned during the concept engineering phase and supported by inspection of the raw images. Several experimental limitations affect full-field consistency. Some cells detach during staining, and slight unevenness in certain coverslips introduces focal distortions, often seen as a blurred ring near the center. Occasional background staining artifacts are also present. These issues reinforce the decision to isolate and analyze individual cells, which simplifies downstream processing.

To achieve this, individual cells are cropped from the images using bounding boxes that are slightly larger than each target cell, ensuring full inclusion with minimal interference. Cropping is performed in batches using a custom macro in ImageJ/Fiji, which displays images sequentially and enables the selection of bounding boxes. Cells are chosen by visual inspection to ensure quality. Detached, severely blurred, or unusually small cells are excluded, while all visible morphological phenotypes are retained to capture within-group variability.

The cropped images are processed using a custom MATLAB script developed in the host lab (*find_all_cells_UBB.m*). This script corrects uneven illumination and separates the composite TIFF files into individual grayscale channels for phalloidin, DAPI, and either αSMA or vimentin, depending on the staining protocol.

5.4.2   Segmentation workflow

This section outlines the workflow used to segment individual cells and nuclei from fluorescence images. It combines image preprocessing, pixel classification, and object identification.

5.4.2.1   Image preprocessing with ImageJ

First, the individual grayscale images are preprocessed in ImageJ/Fiji to improve their quality and facilitate segmentation. These adjustments produce more consistent and visually clear images for posterior Ilastik's pixel classification. Two main steps are applied:

- Background subtraction is performed using the "Subtract Background…" module with a rolling ball radius of 50 pixels, which removes uneven illumination or low-frequency staining variations while preserving cellular structures of interest.
- This is followed by contrast enhancement using the "Enhance Contrast" module, with 2% saturation and normalization enabled. The saturation setting excludes the lowest and highest 2% of pixel intensities to reduce the influence of outliers, while normalization scales pixel values to the [0,1] range, improving contrast uniformly across images.

5.4.2.2   Pixel classification with Ilastik

With the preprocessed single-cell images, probability maps are generated using Ilastik's Pixel Classification workflow. This method performs semantic segmentation by assigning a class label to each pixel based on categories such as nucleus, cytoplasm, or background. Manual annotations are created by painting over representative images to mark each class (Figure 6). A separate model is trained for each fluorescence channel: DAPI, phalloidin, and either αSMA or vimentin. Each model uses approximately 100 annotated cells to define the regions of interest and background.

During training, Ilastik calculates a range of image features including intensity, edges, texture, and neighborhood statistics at multiple spatial scales. It automatically selects features that minimize classification error, prioritizing accuracy even when computational demands are higher. This choice supports higher mask quality and better segmentation. Special care is taken to include fine cytoskeletal structures like filopodia and lamellipodia in the annotations. In images where neighboring cells are close together, thicker background labels are added between them to help the model separate adjacent structures and reduce segmentation errors.

Once the training is completed, the models are applied in batch mode to the complete set of single-cell images. This process generates one probability map per channel and cell, where each pixel is assigned a probability of belonging to the labeled structure.
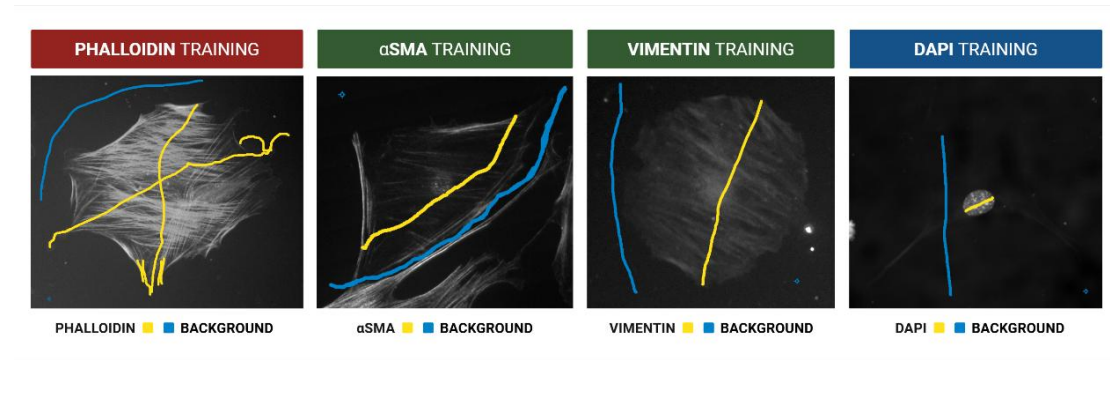
Figure 6 · *Pixel classification training examples for each fluorescence channel.*
Representative annotations used to train Ilastik's pixel classification model for phalloidin, αSMA, vimentin, and DAPI channels.
Brushstrokes in yellow indicate the structure of interest, while blue marks background regions. These manual labels guide the
model in learning to distinguish relevant cellular structures from surrounding areas during the generation of probability maps.

5.4.3   Object identification with CellProfiler

The generated maps are then input into a custom object identification pipeline created with CellProfiler (Table 11). This pipeline utilizes its modular framework to outline and connect specific processing steps. Each module represents a distinct image operation, enabling the development of a customized workflow that detects the objects of interest and produces binary masks for both the nucleus and the entire cell. A detail of the pipeline's modules is available in ANNEX: PIPELINE CELLPROFILER, while the full segmentation pipeline can be found in ANNEX: GITHUB.

The final binary masks are applied to the original raw cropped images rather than the preprocessed ones used for probability map generation. This ensures that feature extraction reflects the actual signal characteristics of each channel, unaffected by preprocessing.
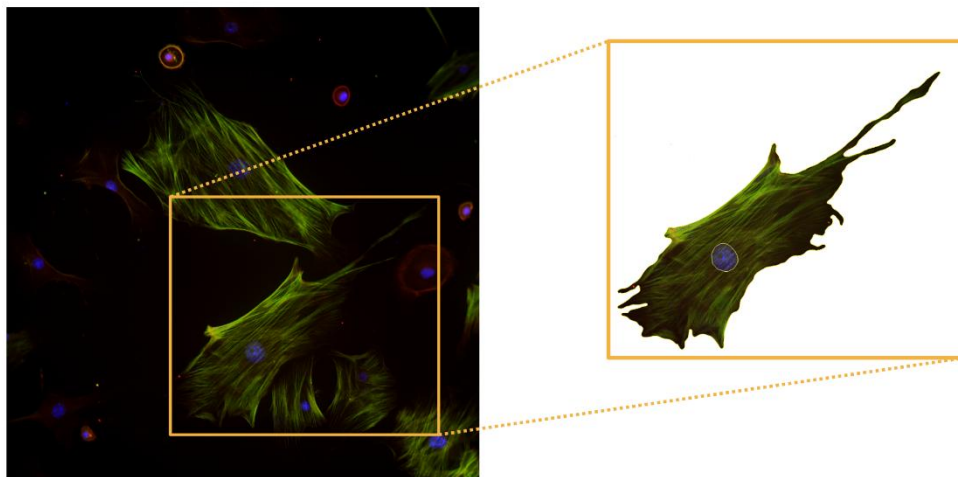
Figure 7 · *Example of single-cell segmentation.*
On the left, a multichannel fluorescence image of fibroblasts is shown, with DAPI (blue), phalloidin (red), and αSMA (green). On the right, a magnified view highlights a segmented single cell, showing both the cell boundary and nuclear mask obtained through the customized pipeline.

To illustrate the pipeline's output, Figure 7 shows a single fibroblast extracted from a field of view, with its corresponding cell and nuclear masks overlaid. This reflects the effectiveness of the combined workflow in isolating individual cells and preserving fine morphological features.

This workflow results in a semi-automated pipeline, where most steps are executed automatically across batches. However, limited user interaction is still required in specific cases, such as manually selecting the correct nucleus when automatic filtering fails, or adjusting thresholding parameters when default values do not produce accurate masks.
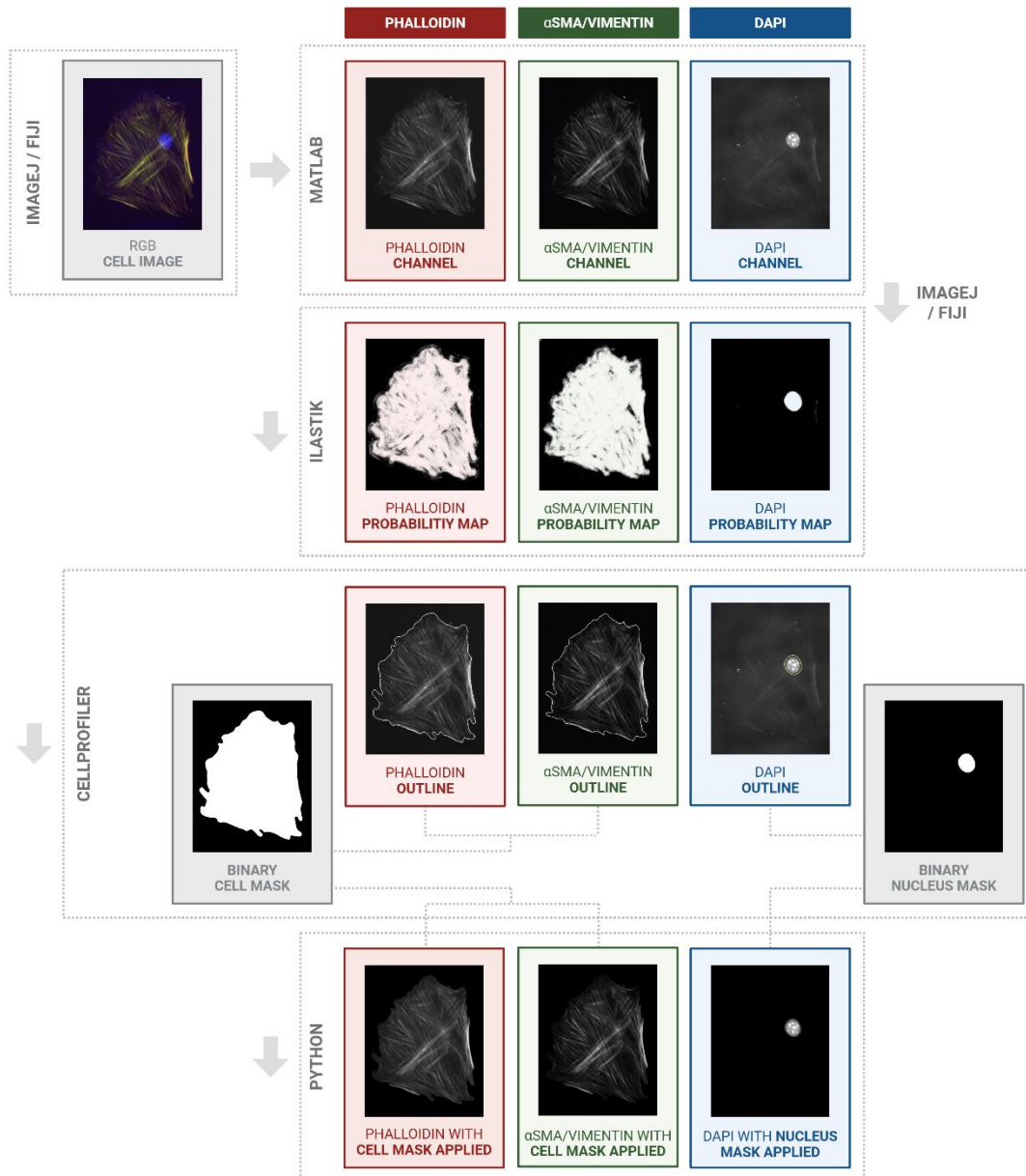
Figure 8 · Visual summary of the segmentation pipeline.
The diagram outlines the complete workflow from raw image preprocessing to the generation of binary masks. RGB images are split into individual channels using MATLAB, probability maps are generated with Ilastik, and final binary masks for the cell and nucleus are created in CellProfiler. For visualization purposes, these masks are shown overlaid on the original grayscale images, although this step is performed later during Python-based feature extraction.

Within the chosen solution, one potential improvement to the segmentation pipeline involves replacing manual nucleus selection with automated heuristics. When multiple nuclei are detected, the pipeline could automatically prioritize the object closest to the image center, which is most likely to correspond to the target cell. This adjustment would reduce user input and enhance the overall workflow efficiency.

5.4.4 Overview of the segmentation workflow

Figure 8 provides a visual summary of the full segmentation pipeline, which combines the key steps described in the previous sections. It illustrates how MATLAB, ImageJ/Fiji, Ilastik, and CellProfiler are used together to generate single-cell and nuclear masks for downstream analysis.

5.4.5 Quantitative feature extraction

Once the cell and nucleus masks are obtained through the segmentation pipeline, a custom Python script is used to extract a wide range of quantitative features from the original grayscale images. This enables the characterization of single-cell morphology, intensity profiles, cytoskeletal organization, and texture, resulting in a detailed quantitative description of each cell.

From this point forward, only samples stained with αSMA are included, as time constraints did not allow for complete processing of the vimentin-stained set.

Intensity correction

Before any image quantification step, each fluorescence channel undergoes an intensity correction to standardize signal levels and compensate for potential inconsistencies due to acquisition settings. The correction follows these steps:

- *Mask-based pixel selection:* For each cell, pixels inside the corresponding binary mask are selected. For the DAPI channel, the nucleus mask is used. For phalloidin and αSMA/vimentin channels, the whole cell mask is applied.
- *Percentile-based baseline estimation:* Within the selected pixels, the 10th percentile intensity value is computed and used as a representative estimate of the local background or low-signal region within the structure of interest.
- *Mean intensity of low-end pixels:* All pixels in the original image (not just within the mask) that are equal to or below this 10th percentile value are extracted. The mean of these pixels is then computed, representing the estimated baseline correction factor for the channel.
- *Normalization:* The original image is divided by the computed baseline value, thereby scaling the image such that the lower-intensity range becomes consistent across cells.

This correction is applied independently to each channel and each image to standardize signal dynamics, allowing downstream intensity and texture features to be computed more consistently and improving feature robustness and comparability across the dataset.

Feature extraction

Quantitative features are grouped into several families that together provide a detailed characterization of single-cell morphology and internal organization. Morphological features describe the shape, size, and geometric complexity of cells and nuclei. Intensity features capture

the distribution and variation of fluorescence signals within each compartment. Texture features quantify spatial patterns in signal variation, reflecting structural organization. Orientation features focus on cytoskeletal arrangement, particularly in actin and αSMA-stained images. A detailed list of all the features extracted, including explanations, library dependencies, and implementation parameters, is provided in ANNEX: EXTRACTED FEATURES, while the complete extraction code is available in ANNEX: GITHUB REPOSITORY.

All feature calculations are performed independently for each cell, and the results are aggregated into a single table. Each row in this table corresponds to a cell, and each column represents a specific feature. The final output is stored in comma-separated values (CSV) format and used for subsequent analysis steps.

Due to the nature of these feature families, a high degree of correlation between related features is expected. This redundancy will be addressed during the data analysis stage to improve interpretability and model performance.

To validate the accuracy of the custom Python pipeline, several features are cross-checked against equivalent available outputs from CellProfiler. This parallel measurement approach confirms consistency in feature extraction and ensures that key cellular descriptors are accurately captured.

5.5   Analysis and results

The feature extraction process generates a high-dimensional dataset, where each cell is characterized by numerous quantitative variables, presenting several analytical challenges. Such complexity makes the data difficult to manage and interpret. Additionally, having more variables than samples increases the risk of overfitting and can compromise the performance and robustness of ML models. Many features are also conceptually or mathematically related, leading to high collinearity (already anticipated during the feature design stage) that can reduce model robustness. To address this, the initial analysis focuses on reducing the number of variables while preserving relevant biological information. The goals are to simplify the dataset, avoid overfitting during model training, and retain only the most informative and independent features.

An essential consideration in the analysis is the inclusion of both male and female samples. While age-related differences are the primary focus, strong sex-specific effects are not expected based on prior knowledge. Treating sex as a separate factor would significantly reduce the sample size for each group, limiting statistical power, especially in a dataset of this scale. Instead, the whole dataset is used to maximize analytical strength, with the understanding that this may introduce some sex-related variability. The analysis is designed to account for this potential noise while prioritizing the identification of robust age-associated features that generalize across all samples.

5.5.1   Class balancing and train/test split

Before statistical analysis or model training, the dataset is balanced to ensure fair comparisons and reduce bias from uneven sample sizes. Balancing is performed across experimental conditions (young male, young female, old male, old female), with an effort to distribute samples as evenly as possible across individual donors.

Each cell is assigned to a group defined by age, sex, and donor. The group with the fewest cells sets the target sample size, and a round-robin sampling strategy distributes cell selection across

donors within each group. Slight variations in donor representation are allowed to preserve equal sizes across experimental conditions and maximize the number of usable cells, since donors with very few cells would otherwise limit the dataset. This approach helps control inter-animal variability and supports reliable downstream comparisons.

Once balanced, the dataset is split into training and test sets using stratified sampling to maintain group representation. 20% of cells from each group are allocated to testing, and the remaining 80% are used for training and feature selection. This split is done before any statistical filtering to prevent data leakage. Applying feature selection only to the training set ensures that the test set remains a valid, independent benchmark.

Additionally, one donor from each experimental condition is excluded entirely and reserved for a final hold-out set. This step provides a rigorous test of whether the selected features and models generalize to unseen individuals, rather than capturing donor-specific patterns, thereby strengthening the robustness of the analysis.

The final distribution of single-cell samples across training, test, and hold-out sets, including individual donor contributions within each group, is summarized in Table 7.

TRAIN / TEST [1052]												HOLD-OUT [234]			
956												234			
YOUNG [536]								OLD [516]				YOUNG		OLD	
478								478				117		117	
YF [278]				YM [258]				OF [239]		OM [277]		YF	YM	OF	OM
239				239				239		239		117		117	
YF1 [143]		YF3 [135]		YM1 [103]		YM3 [155]		OF1 [239]		OM3 [239]		YF2 [94]	YM2 [27]	OF3 [91]	OM2 [26]
119		119		103		136		239		239		94	23	91	26
TRAIN	TEST	TRAIN	TEST	TRAIN	TEST	TRAIN	TEST	TRAIN	TEST	TRAIN	TEST	HOLD-OUT	HOLD-OUT	HOLD-OUT	HOLD-OUT
92	27	98	21	81	22	110	26	191	48	191	48	94	23	91	26

Table 7 · Distribution of samples across training, testing, and hold-out sets.
Summary of the number of single-cell samples selected per animal and condition across each dataset split. Groups are defined by age and sex: young female (YF), young male (YM), old female (OF), and old male (OM). Within each group, samples are evenly distributed across individual donors whenever possible.

5.5.2 Statistical filtering: selection of age-related features

After balancing the dataset and splitting it, a statistical filtering step is performed exclusively on the training set to identify the most relevant features associated with age, while explicitly accounting

for and controlling potential sex-related effects. This selective filtering ensures that only those variables with an independent age-related signal are retained for downstream analysis.

This step utilizes ANOVA. The process is carried out in two stages, both of which include multiple testing correction using the Benjamini–Hochberg (BH) procedure to control the false discovery rate (FDR). A significance threshold of α = 0.05 is applied throughout. Before applying the ANOVA tests, features with near-zero variance (standard deviation < 1e-6) across the training set are removed, since they are unlikely to carry helpful information and help prevent some associated errors during the statistical analysis.

Step 1: Interaction effect filtering

The first step assesses whether a combined effect between age and sex influences each feature. For this, a two-way ANOVA model with an interaction term is fitted:

$$feature \sim age \times sex$$

This model estimates the individual effects of age and sex, as well as their interaction. The p-value associated with the interaction term is extracted and adjusted using the BH method. Features showing significant adjusted p-value (≤ 0.05) are discarded at this stage, as they may reflect inconsistent or sex-specific age effects. Only features without significant interaction are retained, ensuring that the selected features exhibit consistent age-related patterns across both sexes.

Step 2: Main effect filtering

In the second step, the retained features from step 1 are tested using a reduced additive two-way ANOVA model without interaction:

$$feature \sim age + sex$$

This model estimates the main effects of age and sex independently. From this model, two p-values are obtained, one for the effect of age and another for the effect of sex. Both p-values are corrected for multiple comparisons using the BH method. Features are selected only if they show a significant age effect (adjusted p-value ≤ 0.05) and a non-significant sex effect (adjusted p-value > 0.05).

This filtering ensures that the selected features are explicitly associated with age and not confounded by sex-related variation. In addition, the F-statistic for the age effect is stored for each selected feature. This statistic will later be used as an estimate for effect strength in ranking variables or guiding further interpretation.

5.5.3   Hierarchical clustering: correlation-based feature reduction

A Pearson correlation matrix is computed between all features that pass the ANOVA filtering stage. This matrix quantifies the linear relationship between pairs of variables, helping identify features that carry overlapping information. To reduce redundancy, correlation values are transformed into distances using $distance = 1 - |correlation|$, which treats both positive and negative correlations equally by focusing on the strength rather than the direction of the association.
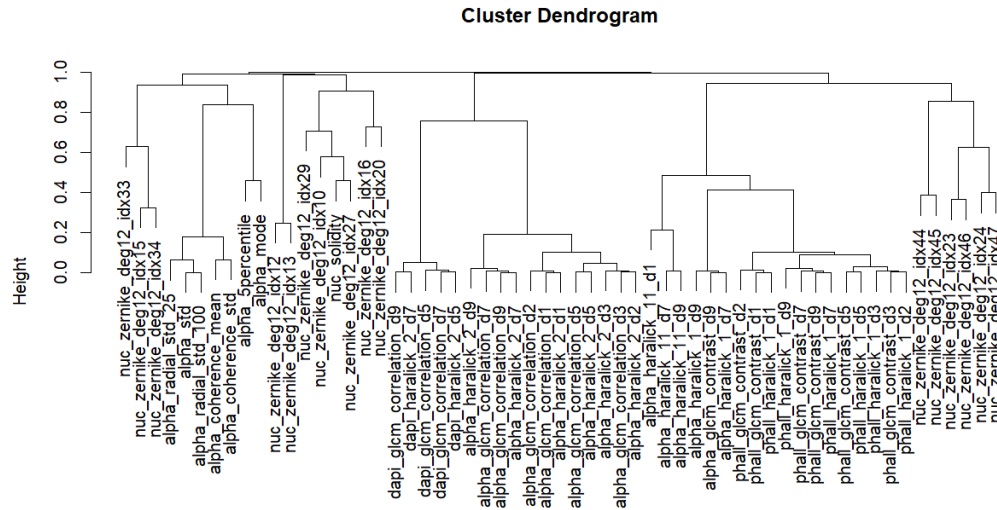
Figure 9 · *Hierarchical clustering of selected features.*
Dendrogram showing the similarity structure among the final set of selected features based on their pairwise correlations. Features are grouped using hierarchical clustering with complete linkage, revealing clusters of highly related variables that often correspond to the same feature family or channel.
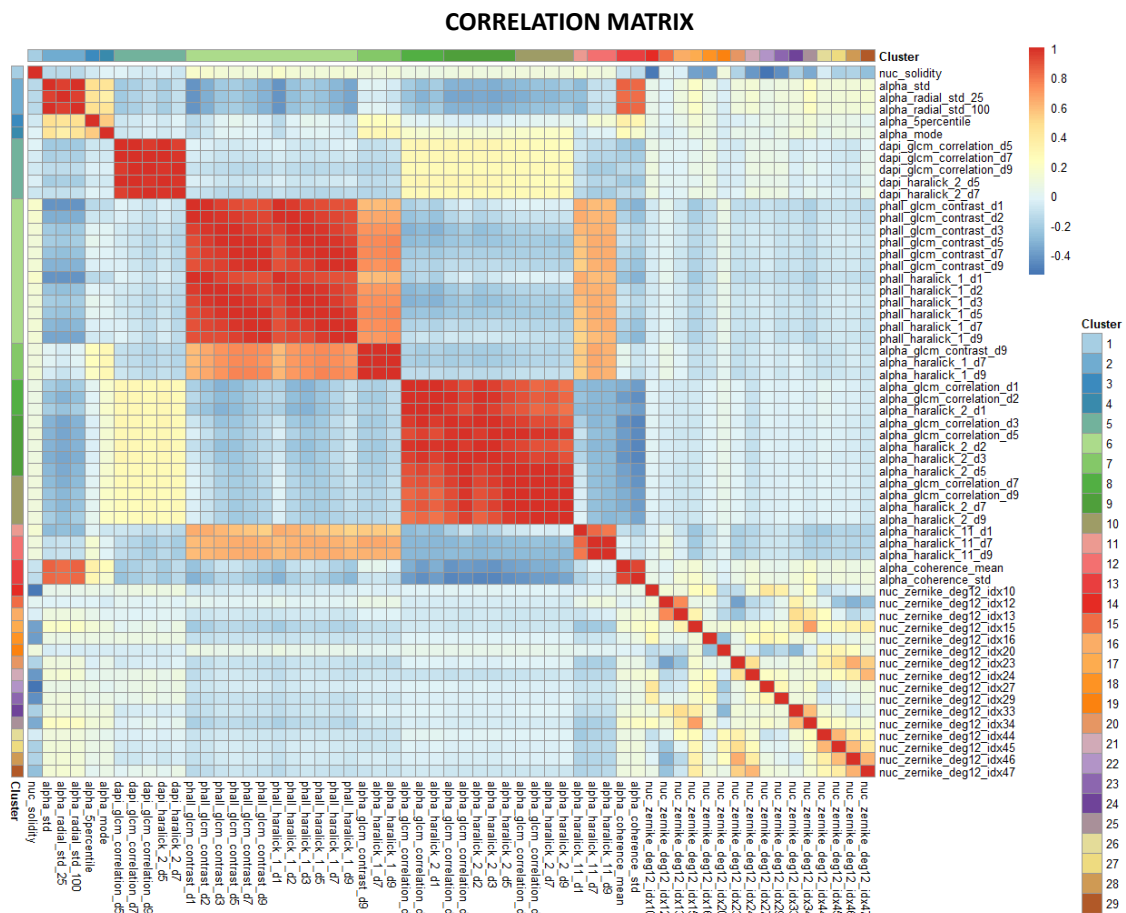
Figure 10 · *Feature correlation matrix and clustering.*
Heatmap showing pairwise Pearson correlations between the selected features, organized by hierarchical clustering. Blocks of high correlation (in red) reveal groups of redundant features, which are further delineated by cluster assignments indicated along the axes.

Hierarchical clustering is then applied to this distance matrix. This method builds a tree-like structure, or dendrogram, by successively merging the most similar features based on their pairwise distances. Cutting the dendrogram at a height of 0.2 creates clusters in which every feature has an absolute correlation of at least 0.8 with the others in the same cluster. From each cluster, only one feature, the one with the highest F-statistic from the previous ANOVA, is retained. This results in a compact set of non-redundant and informative features optimized for downstream analysis. Figure 9 and Figure 10 summarize the results of this correlation-based feature reduction.

5.5.4   Machine learning

Following the feature selection process, the final set of non-redundant and age-informative features is used to train and evaluate several supervised classification algorithms. The aim is to determine whether these features can effectively distinguish between young and old fibroblasts based on their structural profiles. Five commonly used classifiers are considered: Decision tree (DT), k-Nearest neighbors (k-NN), Random forest (RF), Support vector machine (SVM), and Logistic regression (LR). This section outlines the preprocessing steps, the hyperparameter optimization strategy, and the evaluation protocol applied to each model.

5.5.4.1   Preprocessing and pipeline construction

Selected features are standardized when required by the classification algorithm. Standard scaling adjusts each feature to have a zero mean and unit variance, which is essential for distance- and margin-based models, such as k-NN, SVM, and LR. In contrast, DT and RF classifiers are scale-invariant and do not require normalization, as they operate through threshold-based feature splits.

Each classifier is integrated into a pipeline that includes a scaling step only when necessary. During posterior cross-validation and Bayesian optimization steps, the scaler is fitted exclusively on the training data of each fold and subsequently applied to the corresponding validation set. This approach prevents data leakage and ensures a fair evaluation of model performance.

5.5.4.2   Hyperparameter optimization

Each classifier undergoes hyperparameter tuning using Bayesian optimization combined with five-fold stratified cross-validation. The optimization metric is the Area Under the Curve (AUC). Stratification ensures that each fold preserves the original ratio of young and old fibroblasts, maintaining class balance and reducing sampling bias.

In this setup, cross-validation operates within the Bayesian optimization loop. For each hyperparameter set proposed by the optimizer, the model is trained and validated across five stratified folds. In each round, four folds are used for training and one for validation, rotating so each fold is used once for validation. The average AUC across all folds is used as the performance score for that configuration.

Bayesian optimization uses these scores to update a probabilistic model of the hyperparameter space. Instead of searching randomly or exhaustively, it predicts which configurations are likely to perform well based on past results. It selects new candidates by balancing exploration of uncertain regions and exploitation of promising areas. With each iteration, it refines its model to improve future predictions.

This approach provides a reliable estimate of model performance while enabling an efficient search for optimal parameters. Each classifier is tuned using 30 iterations of Bayesian optimization. Detailed descriptions of the best models are provided in ANNEX: BEST MODELS.

5.5.4.3 Evaluation protocol

After identifying optimal hyperparameters through Bayesian optimization, each classifier is retrained on the whole training set using its best configuration. Performance is then evaluated across three data partitions, each capturing a different aspect of generalization.
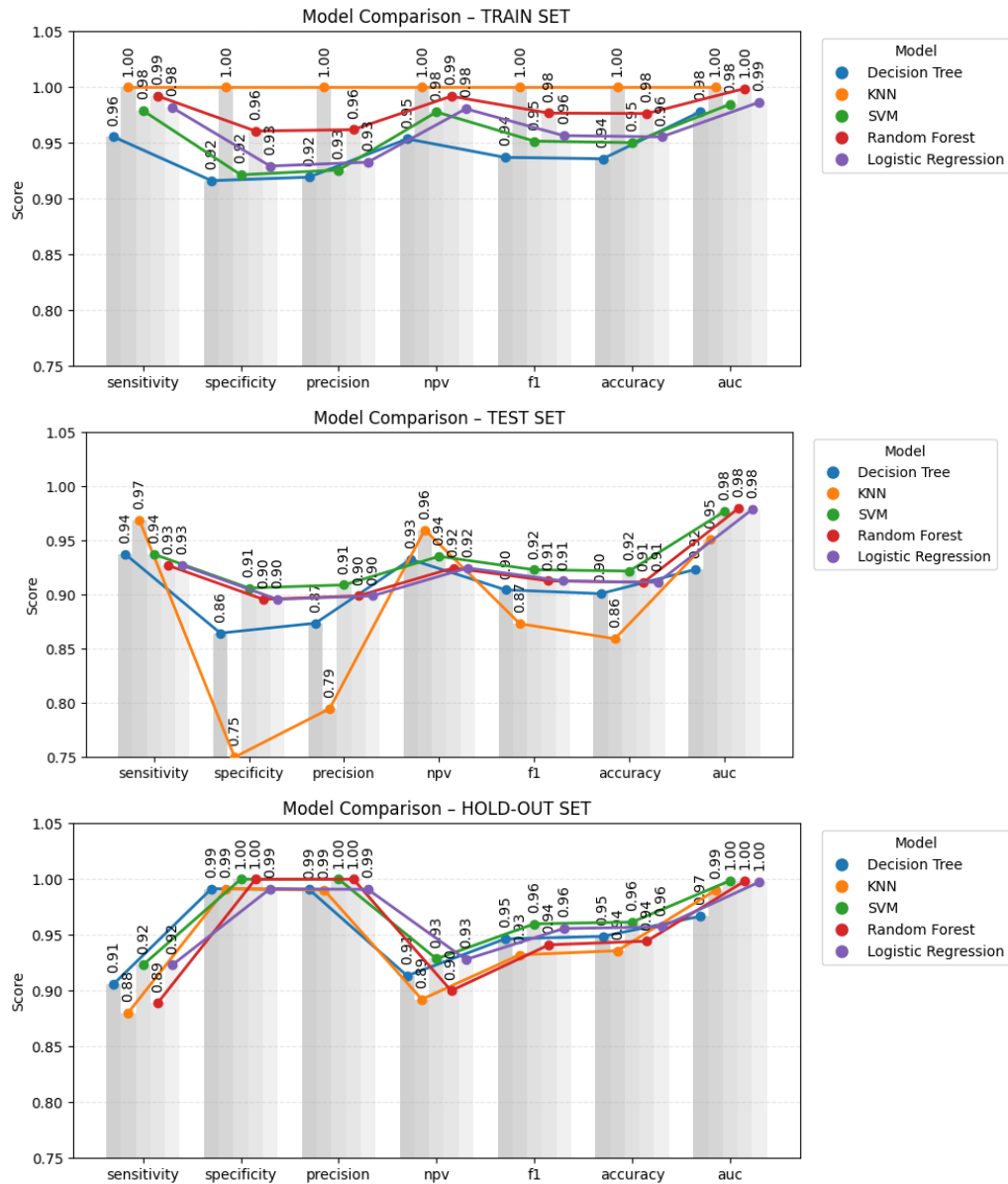
Figure 11 · Model comparison across training, test, and hold-out sets.
Performance of five classification models (Decision Tree, K-Nearest Neighbors, Support Vector Machine, Random Forest, and Logistic regression) evaluated on the training set (top), test set (middle), and validation set (bottom). Metrics include sensitivity, specificity, precision, negative predictive value (NPV), F1 score, accuracy, and area under the curve (AUC).

The training set assesses model fit and helps detect overfitting. The test set, which includes cells not used during training or tuning but originating from the same donor pool, evaluates generalization to new data under similar conditions. The hold-out set, composed of cells from entirely unseen donors, provides the most stringent assessment by testing the model's ability to generalize across biological variability.

Each classifier is evaluated using multiple metrics, including accuracy, precision, recall, F1 score, specificity, sensitivity, and AUC. These results are summarized in Figure 11, while supporting visualizations, such as Receiver Operating Characteristic (ROC) curves and confusion matrices, are provided in ANNEX: MODEL COMPARISON. To verify the effectiveness of the feature filtering process, additional models are trained to predict sex using the same features (ANNEX: SEX CLASSIFICATION INABILITY). All models fail to do so, confirming that sex-related information has been effectively removed and that selected features are specific to age.
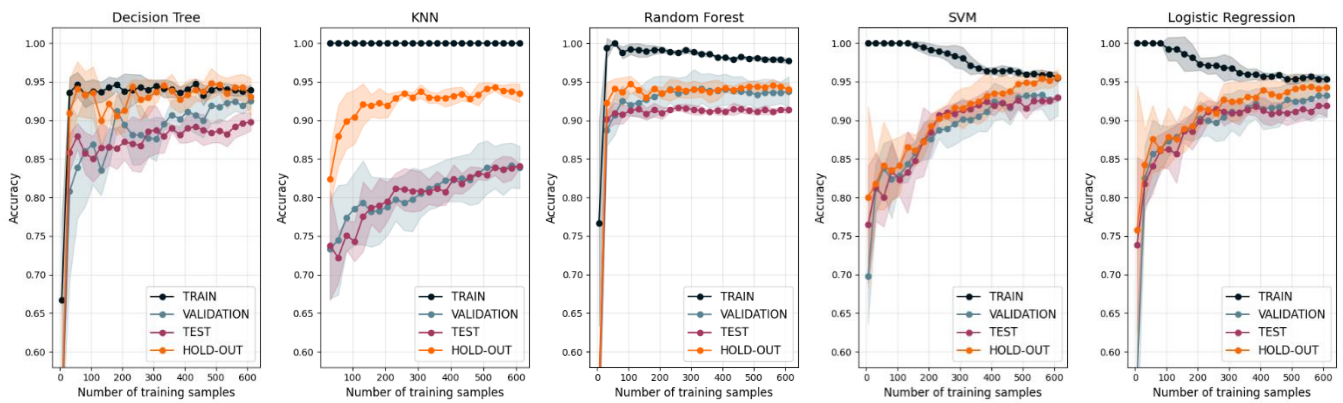
Figure 12 · *Learning curves for classifier performance across data subsets.*
Accuracy as a function of training set size for each classification model, evaluated on the training set, internal cross-validation, test set, and hold-out set.
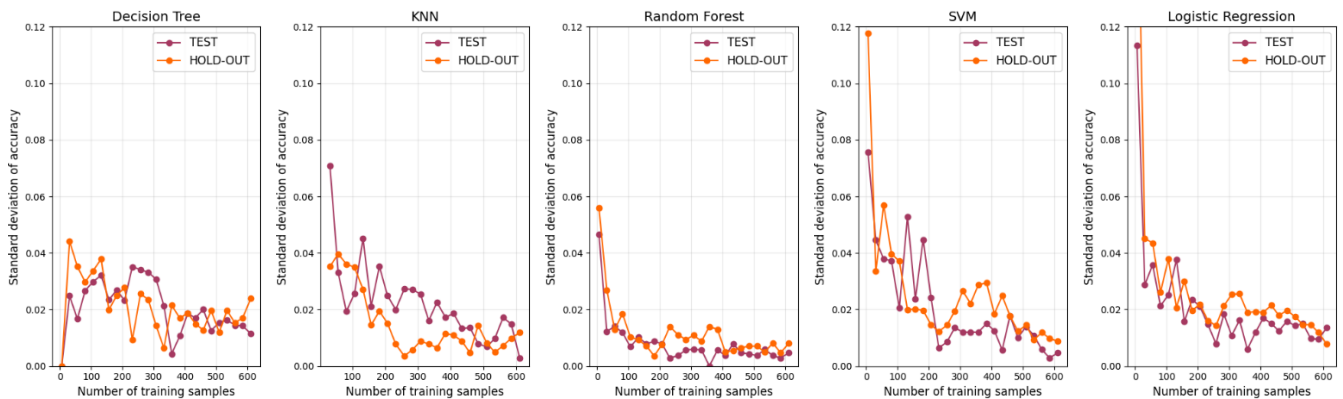
Figure 13 · *Model stability as a function of training set size*
Standard deviation of classification accuracy across training sizes for each evaluation set.

To evaluate how model performance scales with data volume, learning curves are generated using the optimized classifiers. These plots (Figure 12) show accuracy as a function of training set size across four evaluation settings: the training set, internal cross-validation, the test set, and the hold-

out set reserved from entirely unseen donors. Although models are already configured with their optimal hyperparameters, internal cross-validation is included to assess how performance might vary with the training data, even under fixed configurations. Note that the number of samples on the x-axis represents the number of training samples used within each cross-validation fold, not the total training set size.

The curves help identify tendencies towards underfitting or overfitting. A persistent gap between training and validation suggests high variance, while uniformly low scores may indicate high bias. Ideally, training and validation curves should converge as more data is introduced.

To further assess stability, the standard deviation of accuracy is also tracked across all evaluation sets and plotted separately (Figure 13). This variability, also represented by the shaded regions in the learning curves, provides insight into model robustness. A decreasing standard deviation with increasing sample size suggests consistent generalization, while persistent fluctuation indicates sensitivity to data subsets or insufficient training volume.

5.5.4.4　Model selection rationale

All five classifiers evaluated demonstrate strong predictive performance. This high baseline aligns with prior biological intuition. During image acquisition, consistent differences between young and old fibroblasts were already visibly appreciable under standard fluorescence microscopy, suggesting that aging leaves detectable phenotypic signatures.

Given this strong overall performance, the goal of model selection shifts from simply identifying the most accurate classifier to comparing models in terms of generalization, bias–variance tradeoff, interpretability, and robustness to biological variability.

Model comparison and selection

The comparison plots in Figure 11 reveal that on the external test set, both sensitivity and negative predictive value (NPV) show a modest decline compared to the internal test set. Upon closer inspection, nearly all misclassified samples in the external split correspond to cells from a single donor: young male #2. This concentration of errors suggests that the performance drop is not due to a systematic issue in model training or generalization failure, but instead to potential technical variability introduced during the imaging or sample preparation of this specific donor. This hypothesis aligns with prior observations made during image acquisition, where subtle inconsistencies in this donor's dataset were noted. Despite this localized effect, the remaining metrics (particularly accuracy, F1 score, and AUC) are consistently high across all models in the hold-out set, indicating that the classifiers seem to generalize well across independent donors.

The choice of model depends on the intended purpose. In this project, the objective is not only to classify cell age accurately but also to identify which features best reflect aging. The model must generalize reliably with new donors and remain stable across different training subsets, while also offering interpretable outputs that support biological interpretation. Considering the dimensions of generalization, interpretability, and model stability, the subsequent ranking is articulated from most to least favorable:

1.　LR offers the best overall balance for this task. Its accuracy improves steadily with increasing sample size, and the gap between training and evaluation sets narrows

predictably, a sign of good generalization. It also shows low standard deviation across test and hold-out sets as more data are added, demonstrating stability. Importantly, the model is highly interpretable, assigning weights to individual features that directly inform biological insights at both global and single-cell levels.

2. SVM shows a learning curve pattern similar to LR, with early overfitting that diminishes as data grows. The evaluation accuracies increase consistently, and the standard deviation steadily declines with more data. Although less interpretable, its linear nature still allows for partial transparency through analysis of its weight vector, offering some insight into feature relevance.

3. DT is the most transparent model, offering intuitive, rule-based decision paths. Its learning curve shows a rapid initial improvement followed by a slower upward trend, indicating limited but continued benefit from additional data. However, it struggles to generalize, and it has a higher standard deviation, indicating less stability. It's useful for exploration but suboptimal for robust classification.

4. RF achieves high accuracy early but shows a persistent gap between training and evaluation sets, indicating a tendency to overfit. Its standard deviation is relatively low, suggesting stable predictions, but its ensemble structure limits interpretability. Although feature importance can be extracted, individual predictions are opaque due to the combined influence of multiple decision trees, failing to meet the standards for transparency.

5. k-NN performs worst in this context. The model memorizes training data, resulting in near-perfect training accuracy but comparatively poor generalization. The accuracy on test and hold-out sets remains low, and the standard deviation is high, indicating instability. Its non-parametric nature and lack of internal feature weighting make it both unpredictable and uninterpretable in this setting.

Based on this evaluation, LR is selected as the final model. It maintains excellent predictive performance across all evaluation sets, demonstrates strong cross-donor generalization, and enables direct interpretation of feature relevance.

Final model and feature selection

To enhance interpretability and reduce feature redundancy, a refined selection step follows the initial model optimization. Starting from the complete set of features selected during the pipeline search, recursive feature elimination with cross-validation (RFECV) is applied using the same LR configuration. This method ranks features by their contribution to classification performance and iteratively removes the least informative ones. The optimal subset consists of 10 features, which preserve the model's predictive power while simplifying its structure.

A new LR model is trained using only this reduced feature set and evaluated on the training, test, and hold-out sets. Performance remains consistent across all data, confirming that the essential discriminative patterns related to aging are retained. A final accuracy of 0.97 is obtained in the hold-out set, highlighting the model's strong generalization capability to previously unseen donor samples.

Global feature relevance is visualized through a bar plot showing the magnitude and direction of each coefficient, providing insight into how specific variables influence predictions. Additionally, the

model enables interpretation at the single-cell level: by combining each cell's scaled feature values with the corresponding model coefficients, it becomes possible to identify which morphological traits contribute most to its classification. This approach is closely related to SHAP values for linear models, which similarly decompose individual predictions into feature-wise contributions based on the model's weights. This dual-level interpretability strengthens the model's value for both population-level analysis and individual cell assessment.
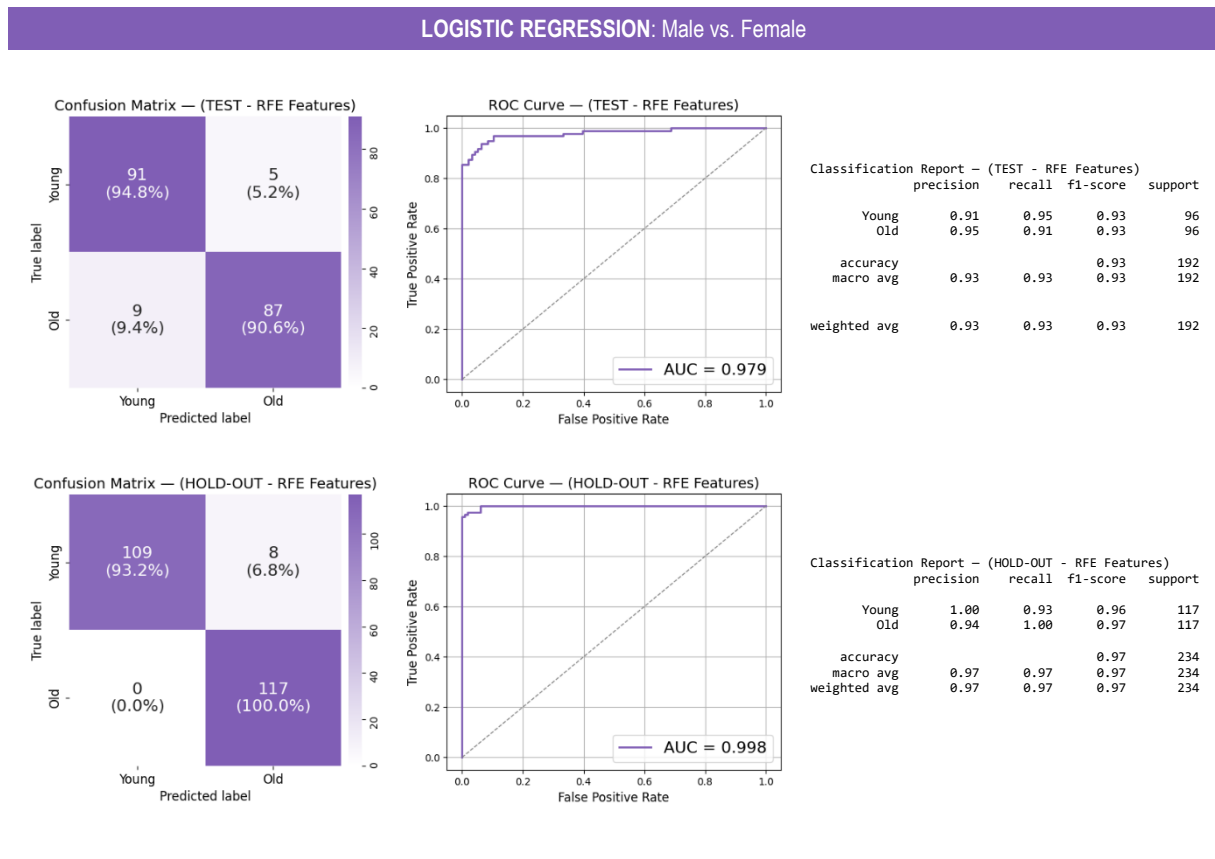
Figure 14 · Final logistic regression model performance after feature selection.
Confusion matrices, ROC curves, and classification reports for the test and hold-out sets of the final logistic regression model trained with features selected via RFECV.
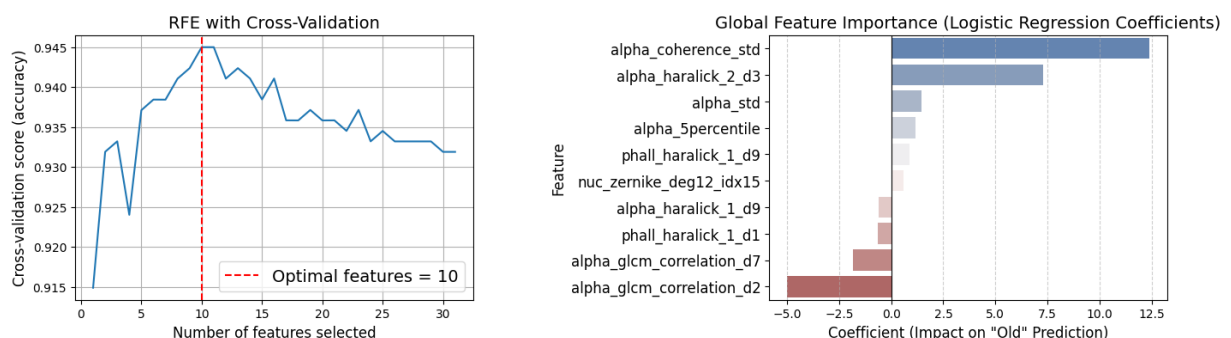
Figure 15 · Feature selection using recursive feature elimination with cross-validation (RFECV).
Cross-validation accuracy as a function of the number of features selected. The optimal number of features is identified as 10, corresponding to the highest average accuracy during cross-validation. This subset is used in the final logistic regression model.

Figure 16 · Global feature importance based on logistic regression coefficients. Bar plot showing the top 10 features selected by RFECV and their corresponding coefficients in the final logistic regression model. Positive values indicate greater contribution to the prediction of "old" fibroblast class, while negative values contribute to "young" classification. Features are ranked by the magnitude of their impact on model predictions.
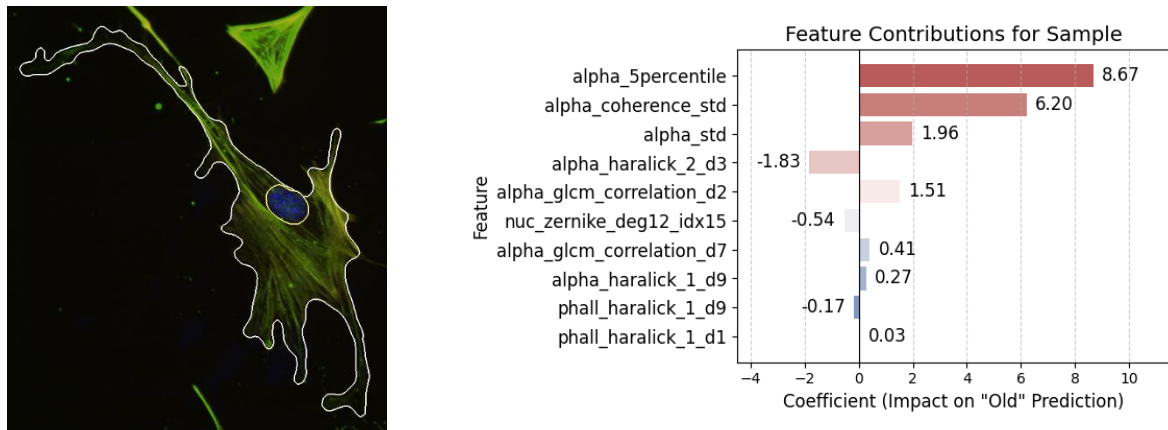
Figure 17 *· Feature contributions for an individual fibroblast classified as "old."*
Bar plot showing the top features contributing to the classification decision for a single cell, based on the coefficients from the final logistic regression model. Positive values increase the likelihood of an "old" prediction, while negative values support a "young" classification. This interpretation is linked to the fluorescence microscopy image of the corresponding cell.

5.5.5   Discussion and biological relevance

Beyond high classification accuracy, the ML framework developed in this study plays a critical role in validating and ranking the features most associated with cellular aging. The LR model, especially when combined with RFE, not only identifies which variables carry predictive value in the context of aging but also enables clear interpretation of their biological relevance, both at the population level and for individual cells.

This interpretability enhances the model's utility beyond simple prediction. By revealing how specific features contribute to classification, the model serves as an analytical tool for identifying and prioritizing the phenotypic traits most strongly associated with aging. It enables researchers to transition from what the model predicts to why it makes those predictions, thereby guiding further biological investigation.

Importantly, the features selected through this modeling process are not arbitrary. Their discriminative power stems from measurable, biologically grounded changes that have been previously assessed through ANOVA. While the correlation filtering step helps reduce redundancy by removing collinear variables, each retained feature remains interpretable and retains its original meaning. The resulting feature set reflects a compact and biologically coherent representation of the aging phenotype, capturing nuclear shape deformation, cytoskeletal remodeling, and changes in spatial organization. The following section discusses each family of features identified as informative for distinguishing cellular age.

Quantitative differences between young and aged fibroblasts

While the features selected by the predictive model are the most informative for this specific classification task, it is important to note that the subsequent biological analysis focuses on all variables that passed the initial statistical significance threshold. These features are grouped and assessed by functional families, as they each show a meaningful association with aging. This

approach ensures that biologically relevant patterns are considered beyond the constraints of model-driven selection, while excluding variables that do not exhibit clear age dependence under the specific experimental conditions of this project. The boxplots showing feature differences between young and old samples, based on their means and standard deviations, are available at ANNEX: BOXPLOTS.

Nuclear morphology

Solidity

Solidity, the ratio of area to convex hull area, measures how convex a shape is. A value of 1 indicates perfect convexity, while lower values reflect indentations or invaginations[83]. In DAPI-stained nuclear masks, aged cells show reduced solidity, consistent with envelope folds, blebs, or irregular nuclear shapes (as illustrated in Figure 18). These deformations align with known age-related changes in the nuclear lamina and chromatin structure, where reduced nuclear stiffness leads to altered nuclear morphology[36].
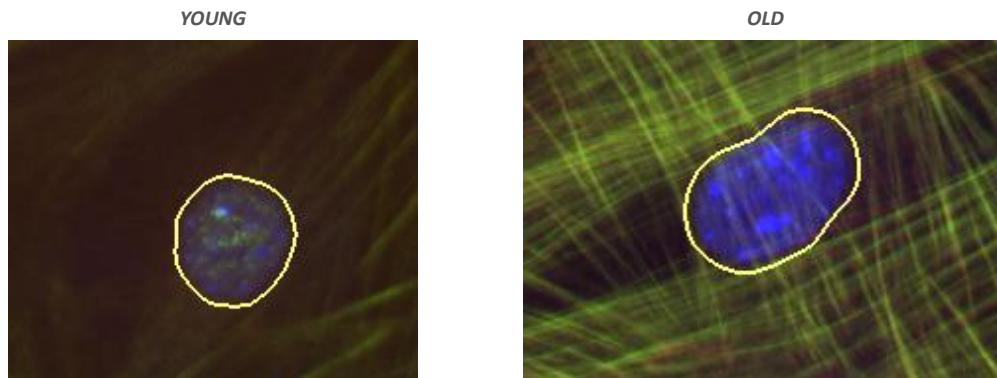
Figure 18 · *Visual representation of age-associated differences in nuclear solidity.*
Comparison of nuclear morphology between young and old fibroblasts, highlighting a reduction in solidity with age. This suggests more irregular and deformed nuclear contours in aged cells, consistent with structural decline.

Zernike moments

Zernike moments, computed up to degree 12, provide a detailed quantification of nuclear shape irregularities. These orthogonal descriptors project a binary shape onto a set of complex polynomials defined over the unit disk. Lower-order moments capture global traits such as circularity, while higher-order moments detect finer features, including wrinkles and lobulations[84].

In aged fibroblasts, numerous higher-order Zernike coefficients show a notable increase, indicating the development of localized deformations in the nuclear envelope. This trend suggests a transition to more intricate boundary geometries.

Zernike moments can be regarded as a more expressive generalization of solidity. Both features quantify aspects of nuclear shape integrity, and their predictive value in the model supports the same biological interpretation: a progressive loss of convexity and regularity in aged nuclei.

α-SMA cytoskeletal intensity and distribution

α-SMA intensity descriptors suggest a complex remodeling of the contractile cytoskeleton in aged fibroblasts. Three key metrics, α_std, α_5percentile, and α_mode are all elevated with age:

- α_std, the standard deviation of background-corrected α-SMA intensity within the cell mask, reflects variability in filament density across the cytoplasm. Its increase suggests greater heterogeneity, with some regions showing more intense α-SMA staining and others less. However, this metric alone does not distinguish whether such variability results from a higher overall number of fibers or from fewer but thicker, more intensely stained bundles. To clarify the nature of this structural remodeling, additional descriptors are considered.
- α_5percentile, the intensity value below which 5% of α-SMA pixels fall, increases with age, suggesting even the dimmest regions show higher expression and pointing to an elevated baseline filament presence.
- α_mode, representing the most frequent intensity bin, also increases, indicating a general upregulation of α-SMA, reflecting more prominent or denser filamentous structures across the cell body.

To determine whether the rise in α_std is due only to overall intensity increases, the coefficient of variation (α_CoV) is analyzed. Defined as the ratio between standard deviation and mean intensity, it is significantly higher in aged fibroblasts. This confirms that α-SMA becomes more heterogeneous both in absolute terms and relative to its average expression.

In addition, radial standard deviation metrics, calculated both within the innermost 25 percent of the cell and across the entire radial extent, show increased variability. This indicates that α-SMA heterogeneity is present across both perinuclear and peripheral regions.

Texture analysis

To quantify the spatial organization and texture of cellular structures, the Gray-Level Co-Occurrence Matrix (GLCM) is used. This method captures how often pairs of pixel intensities occur in a specific spatial relationship within the image. For each pixel, its intensity i is related to that of a neighboring pixel j, located at a defined distance d and angle θ. The resulting matrix $P(i,j|d,\theta)$ summarizes the frequency or normalized probability of such pixel pairs across the image[85].

In this analysis, GLCMs are computed at multiple distances and angles to capture texture features across a range of spatial scales and orientations. This is particularly important in biological structures like actin fibers or chromatin domains, which may exhibit directionality. Unless directional effects are specifically studied, GLCMs are averaged across angles to produce rotation-invariant descriptors.

From each GLCM, statistical features are extracted. Two of the most informative are:

- Contrast, which reflects the intensity differences between adjacent pixels and highlights sharp edges or well-defined patterns.
- Correlation, which quantifies the linear dependence between neighboring pixel values, indicating structural regularity.

These features are part of the broader set of Haralick texture descriptors, a standardized framework for texture analysis. By evaluating them across multiple distances, the analysis captures both local

and extended structural patterns, providing a detailed view of textural complexity and spatial coherence.

DAPI channel

Texture analysis of the DAPI channel reveals age-related alterations in chromatin organization. Haralick correlation features at distances of 5 and 7 pixels (dapi_haralick_2_d5 and dapi_haralick_2_d7) are significantly reduced in aged cells. This decline indicates that chromatin intensity values become less correlated at intermediate spatial scales, reflecting a loss of continuity in chromatin texture. Rather than exhibiting a smooth and diffuse nuclear pattern, aged nuclei show more fragmented and sharply localized high-intensity regions, suggesting a reorganization of DNA into more spatially confined domains during aging.

Phalloidin channel

In the phalloidin channel, Haralick contrast features (Haralick index 1), such as phall_haralick_1_d1 and d9, are significantly lower in aged fibroblasts. A reduction in contrast reflects a decrease in the sharpness of intensity transitions between neighboring pixels. This suggests that actin bundles become less well-defined relative to the surrounding background. Visually, this manifests as a smoother and more uniform F-actin signal, with less distinct fiber boundaries and a more homogeneous cytoplasmic distribution.

Biologically, this pattern points to a general disorganization of the actin cytoskeleton during aging. In younger cells, actin fibers often appear as thick, well-separated structures with substantial signal differences across adjacent pixels. In aged cells, fibers tend to appear thinner, more diffuse, or less spatially distinct, resulting in lower contrast values.

α-SMA channel

In the α-SMA channel, multiple Haralick texture features decrease with aging:

- α_haralick_1 (contrast): Reflects the sharpness of intensity differences. Lower values indicate less distinction between filament edges and the background.
- α_haralick_2 (correlation): Indicates how well intensity in one pixel predicts its neighbor's. Its reduction implies fragmented α-SMA networks with broken continuity across bundles.
- α_haralick_11 (IMC1): This feature quantifies the mutual information between neighboring pixel intensities, capturing how much knowing one pixel's value informs about its neighbor. A lower IMC1 value indicates reduced inter-pixel dependency, reflecting a global decline in structural organization and predictability within the filament network. This suggests that α-SMA becomes more fragmented and spatially disordered with age.

Together, these results indicate that although α-SMA is upregulated (as evidenced by intensity and mode measures), its textural complexity and long-range organization decline with age. This supports the view that aging reorganizes the cytoskeleton into spatially distinct, internally uniform patches, each aligned and contractile, but globally disconnected in intensity and texture.

Cytoskeletal alignment and directionality

The spatial organization of the α-SMA network is further examined using coherence metrics derived from the local structure tensor, which quantify the directional alignment of filaments across the cell.

- α_coherence_mean is significantly higher in aged fibroblasts, indicating that α-SMA fibers tend to be more directionally aligned, reflecting greater local order within filament bundles.
- α_coherence_std is also increased with age, showing higher variability in alignment across different regions. This suggests the presence of both highly ordered patches and disorganized areas within the same cell.

YOUNG   OLD
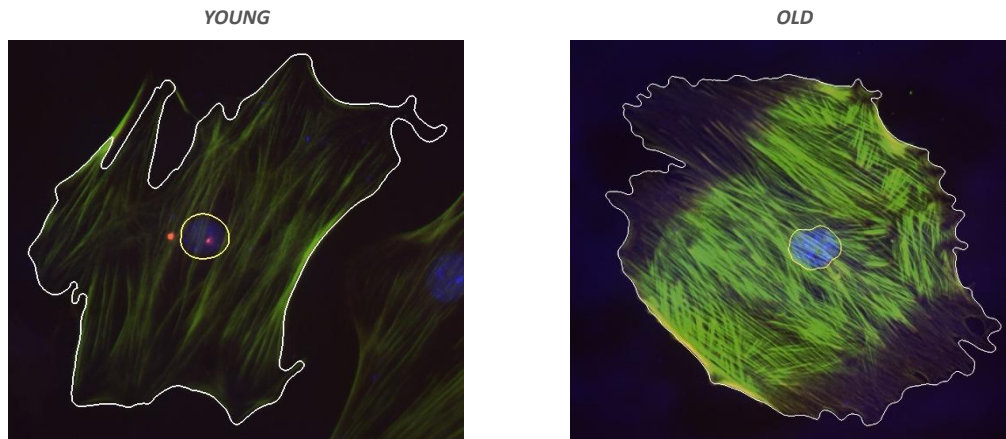
Figure 19 · *Comparison of the cytoskeletal orientation based on aSMA staining.
Visualization of fiber alignment patterns in young and old fibroblasts.*

Together, these metrics reveal a mosaic-like cytoskeletal architecture in aged fibroblasts, where local domains of well-aligned filaments coexist with less organized regions, resulting in a spatially fragmented yet internally coherent pattern (as shown in Figure 19). This heterogeneity helps explain the observed decline in α-SMA Haralick correlation, which reflects reduced predictability of intensity relationships due to misalignment between domains.

Final considerations

While each extracted feature can be associated with plausible biological interpretations based on prior knowledge and visual inspection, these associations remain largely hypothesis-driven. Structural and intensity-based descriptors (such as texture metrics, shape moments, or distribution measures) are computationally well defined, but their precise biological correlates are not always established. As such, the interpretations offered here are biologically grounded and visually consistent but should be viewed as informed hypotheses rather than definitive mechanistic explanations.

Despite this limitation, the results consistently point to underlying biological processes that extend beyond superficial morphological or cytoskeletal differences. Many of the most predictive features likely reflect structural consequences of molecular and regulatory changes associated with aging. For example, the prominence of α-SMA-related features highlights the role of contractile remodeling, influenced by focal adhesion dynamics and age-sensitive signaling pathways. This suggests that the model captures downstream effects of mechanotransduction and cytoskeletal regulation.

In addition, thanks to the single-cell resolution of this study, some intra-population variability has been observed within the same age groups. This cellular heterogeneity suggests the presence of distinct fibroblast subpopulations and aligns with previous findings of the host research group, reinforcing the value of this approach for capturing subtle, cell-level differences.

6   TECHNICAL FEASIBILITY

This section evaluates the technical feasibility of the implemented image-based classification pipeline for aging fibroblasts. It is organized around the four SWOT categories (strengths, weaknesses, opportunities, and threats) to provide a critical and structured overview of the project's technical execution.

S STRENGTHS
- Access to well-equipped imaging and cell culture facilities.
- Established, reliable immunofluorescence protocols.
- Support from experienced researchers.
- Use of well-documented, open-source tools.
- Promotes reproducibility and collaborative development.
- Modular and efficient pipeline structure.
- Moderate computational requirements allow rapid iteration and testing.

W WEAKNESSES
- Manual cropping introduces subjectivity and slows analysis.
- Segmentation requires manual annotation and validation.
- Simplified biological model reduces physiological relevance.
- Small number of cells and donors limits variability.
- Manual imaging introduces inconsistency in focus, exposure, and field selection.
- Lack of automated imaging systems restricts throughput and standardization.

O OPPORTUNITIES
- Pipeline adaptable to other cell types with minimal changes.
- Potential for automation using deep learning segmentation tools (such as Cellpose or StarDist).
- Integration of automated acquisition could increase dataset size and standardization.
- Could evolve into a biological age predictor (imaging-based aging clock).
- Open-source framework supports future collaborations and extensions.

T THREATS
- Rapid development in bioimage analysis tools may render the pipeline outdated.
- Limited translational applicability due to species differences (mouse vs. human).
- Results from in vitro models may not generalize to realistic biological environments.
- Future clinical use would require addressing ethical and regulatory barriers.

Figure 20 · *SWOT diagram of the developed pipeline.*
Summary of the project's internal strengths and weaknesses, as well as external opportunities and threats. The analysis considers technical, biological, and translational aspects relevant to the pipeline's development, applicability, and long-term impact.

6.1   Strengths: Core capabilities

The project is carried out within a well-equipped research unit that provides access to cell culture and fluorescence imaging systems. Established immunofluorescence protocols and support from experienced researchers help ensure consistency and reliability in the experimental procedures, providing a strong technical foundation for the study.

Another key strength is the use of open-source software tools such as ImageJ, Ilastik, CellProfiler, R, and Python. These platforms are well-documented and widely adopted in the scientific community, promoting reproducibility, enabling future adaptations, and facilitating collaboration. Their use also aligns with the principles of open science.

Importantly, the entire pipeline runs with modest computational requirements. Image processing, feature extraction, and model training can be performed efficiently on standard computers, enabling rapid iteration and short processing times. This supports high productivity without the need for specialized hardware, allowing for flexible testing and refinement within limited timeframes.

6.2   Weaknesses: Current limitations

Despite the pipeline's successful implementation, several limitations affect its scalability and generalizability. A key constraint is the manual cropping of individual cells, which introduces subjectivity and slows down analysis. While segmentation is partially automated using Ilastik and CellProfiler, it still relies on manual annotation and validation, reducing reproducibility and complicating large-scale application. Increasing automation is essential for broader deployment, and possible solutions are discussed in the concept engineering section.

Biologically, the experimental model is simplified. Fibroblasts are cultured on rigid glass substrates, which facilitate imaging but do not replicate the mechanical complexity of native tissue, potentially influencing cellular behavior and limiting physiological relevance.

Although classification models perform well across training, test, and validation sets, the dataset remains limited. The number of usable cells is constrained by the availability of biological material, and the small number of donor animals restricts variability, which may impact performance on more diverse samples.

Manual imaging also introduces variability in focus, exposure, and field selection, affecting consistency. Post-acquisition strategies have been applied to mitigate these effects; however, the absence of automated imaging systems remains a significant limitation. Such systems could enhance standardization and throughput but were not accessible within the scope of this project.

6.3   Opportunities: Future potential

Despite current limitations, the project offers clear opportunities for both technical and scientific advancement. The pipeline's modular design supports adaptation to other cell types with minimal modifications, while its single-cell resolution enables detection of cellular heterogeneity often missed by bulk analysis.

Automation is a key area for improvement. Integrating deep learning-based segmentation tools such as Cellpose[86] or StarDist[87] could enhance accuracy and reduce manual input. Likewise, automated image acquisition would improve standardization and substantially increase dataset size.

The approach also holds translational potential. Although currently applied to binary age classification, the pipeline could be extended to estimate biological age quantitatively, contributing to the development of imaging-based aging clocks. Its open-source framework further facilitates collaboration and integration into broader biomedical research.

6.4 Threats: External challenges

Several external threats may limit the long-term feasibility or impact of the pipeline. The field of AI-driven bioimage analysis is evolving rapidly, and newer tools could soon surpass current methods in performance and usability. If the pipeline is not updated accordingly, it risks becoming obsolete.

Additionally, there are significant challenges in translating results obtained from murine models to human biology. Differences in gene regulation, cellular behavior, and aging mechanisms introduce uncertainty when applying the same methodology to clinical samples. Furthermore, results generated in simplified in vitro conditions may not be reproducible in more complex, physiologically realistic settings.

Finally, expansion into human applications would bring about ethical and regulatory considerations, particularly regarding tissue acquisition, patient data management, and diagnostic claims. These external factors must be anticipated in future developments of the pipeline.

7   IMPLEMENTATION SCHEDULE

This section presents the project timeline, structured around key phases and milestones. It includes a Work Breakdown Structure (WBS) to organize tasks, a precedence analysis to define task dependencies, and scheduling tools such as PERT and Gantt charts to support planning and coordination.

7.1   PHASES AND MILESTONES

PHASES AND MILESTONES			
ID	PHASE	DESCRIPTION	MILESTONE
1	Project Setup and Planning	Establish the project's foundation through a focused literature review, definition of objectives, and development of the execution plan and timeline.	Project framework and timeline defined.
2	Sample Preparation and Labeling	Isolate and culture fibroblasts from young and old mice; perform immunofluorescence staining to label cytoskeletal structures.	Labeled cell samples from both age groups ready for imaging.
3	Fluorescence Imaging	Capture multichannel fluorescence images under standardized conditions using epifluorescence microscopy.	Complete and validated image dataset acquired.
4	Image Preprocessing and Segmentation	Develop and apply a reproducible workflow for cleaning, cropping, and segmenting images at the single-cell level.	Segmented single-cell image dataset finalized.
5	Feature Extraction	Extract quantitative features from segmented images, including morphology, intensity, and texture.	Structured feature matrix generated
6	Statistical Analysis	Perform exploratory data analysis and dimensionality reduction to identify relevant patterns and assess data structure.	Reduced feature set and initial insights obtained
7	Model Development and Validation	Train, test, and validate a machine learning model to classify fibroblasts by age based on extracted features.	Model trained and validated with performance metrics reported
8	Feature Interpretation	Interpret the most relevant features driving classification results concerning biological aging processes.	Biologically meaningful conclusions documented
9	Final Reporting	Compile, write, and edit the full final degree report, integrating all results and interpretations.	Complete, reviewed, and formatted final report submitted.

Table 8 · Project phases and corresponding milestones.
Overview of the main phases of the project, each with a brief description and associated milestone, outlining the logical progression from initial planning to final reporting.

The project is divided into distinct phases, each with clearly defined objectives and milestones. This structured approach (illustrated in Table 8) ensures systematic progress, with each phase building logically on the results of the previous one. From initial background research and planning to sample preparation, data acquisition, analysis, and final reporting, the sequence provides a comprehensive roadmap for execution. Milestones act as verification points to track progress and confirm that the project remains on schedule and aligned with its goals.

7.2   WBS

Following the definition of the project's phases and milestones, this section details the Work Breakdown Structure (WBS) (Table 17) and its corresponding dictionary (ANNEX: WBS DICTIONARY). The WBS translates the broader project phases into a structured hierarchy of tasks, grouped into five main categories: project coordination, experimental setup, data collection, feature extraction, and data analysis. This structure mirrors the project's timeline and functional organization.

The WBS Dictionary complements this structure by providing concise definitions for each task. It serves as a practical reference to ensure clarity in execution, support time management, and enable consistent tracking of progress throughout the project.
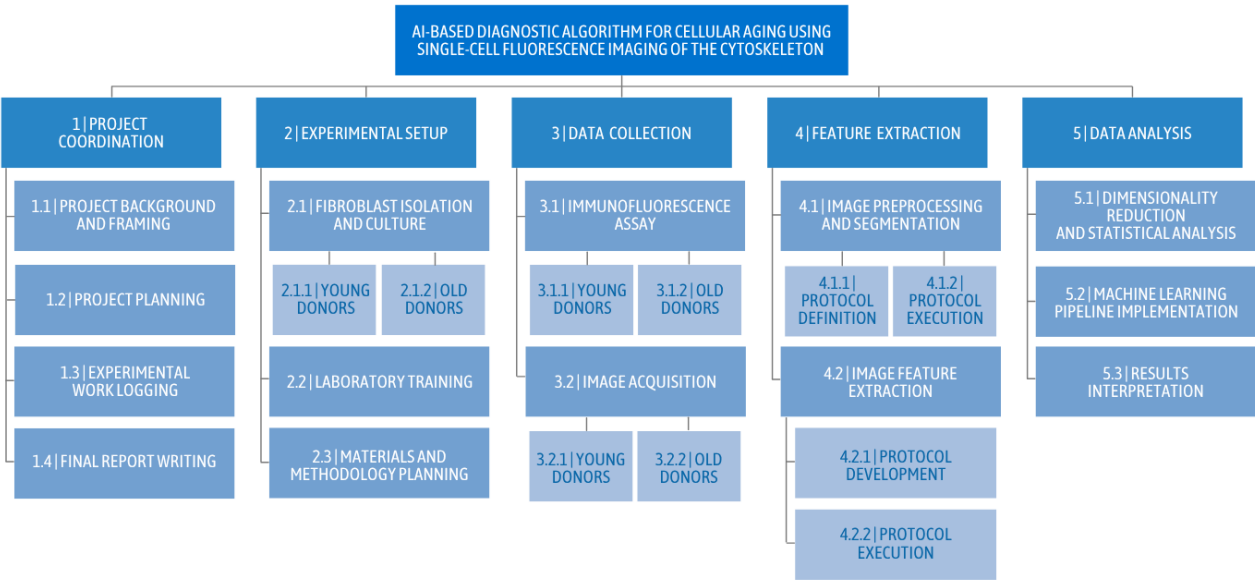
Figure 21 · *Work breakdown structure (WBS)*
Hierarchical decomposition of the project into manageable working packages.

7.3   PRECEDENCE ANALYSIS AND CRITICAL PATH DIAGRAM (PERT-CPM)

To determine the logical order of task execution, a precedence analysis is conducted (Table 9). This identifies dependencies between activities, ensuring that the sequence aligns with methodological requirements and resource availability. The project is planned over a total duration of 133 days.

Building on this analysis, a Program Evaluation and Review Technique (PERT) diagram is developed to map the project's task network. The PERT diagram helps identify parallelizable tasks, estimate realistic timelines, and determine the critical path, that is, the sequence of tasks that directly affects the project's total duration.

ID	TASK	PRECEDENT TASKS				DAYS
	PRECEDENCE ANALYSIS					
A	1.1 | PROJECT BACKGROUND AND FRAMING					20
B	1.2 | PROJECT PLANNING					5
C	1.3 | EXPERIMENTAL WORK LOGGING					47
D	1.4 | FINAL REPORT WRITING	A	C			86
E	2.1.1 | FIBROBLAST ISOLATION AND CULTURE: YOUNG DONORS					15
F	2.1.2 | FIBROBLAST ISOLATION AND CULTURE: YOUNG DONORS					15
G	2.2 | LABORATORY TRAINING					7
H	2.3 | MATERIALS AND METHODOLOGY PLANNING					1
I	3.1.1 | IMMUNOFLUORESCENCE ASSAY: YOUNG DONORS	B	E	G	H	2
J	3.1.2 | IMMUNOFLUORESCENCE ASSAY: OLD DONORS	F	G	H	I	2
K	3.2.1 | IMAGE ACQUISITION: YOUNG DONORS	I				28
L	3.2.2 | IMAGE ACQUISITION: OLD DONORS	J				28
M	4.1.1 | IMAGE PREPROCESSING AND SEGMENTATION: PROTOCOL DEFINITION	K	L			10
N	4.1.2 | IMAGE PREPROCESSING AND SEGMENTATION: PROTOCOL EXECUTION	M				20
O	4.2.1 | IMAGE FEATURE EXTRACTION: PROTOCOL DEVELOPMENT	N				6
P	4.2.2 | IMAGE FEATURE EXTRACTION: PROTOCOL EXECUTION	O				1
Q	5.1 | DIMENSIONALITY REDUCTION AND STATISTICAL ANALYSIS	P				20
R	5.2 | MACHINE LEARNING PIPELINE IMPLEMENTATION	Q				15
S	5.3 | RESULTS INTERPRETATION	R				14

Table 9 *· Precedence analysis.*
Summary of the main project tasks, their order of execution, and estimated durations. It shows how activities depend on one another to guide efficient project scheduling.

7.4   GANTT DIAGRAM

The tasks defined in the precedence analysis are also visualized in a Gantt chart, which offers a clear, time-based representation of the project workflow. This chart outlines the planned start and end dates for each task and highlights overlaps where activities run in parallel. It distinguishes between flexible tasks, which allow for scheduling delays, and critical tasks, whose timing directly affects the overall project timeline. By visualizing dependencies and concurrent processes, the Gantt chart serves as an essential tool for monitoring progress and maintaining adherence to the project schedule.
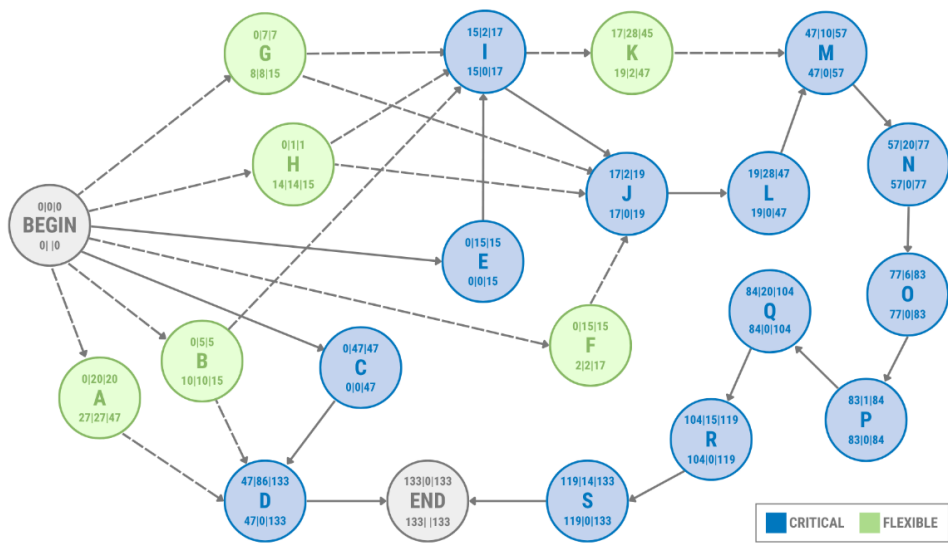
Figure 22 · *Project Evaluation and Review Technique (PERT).*
Project Evaluation and Review Technique (PERT) chart showing the sequence, dependencies, and estimated duration of tasks across the project timeline. It visualizes the critical path and supports planning and coordination.
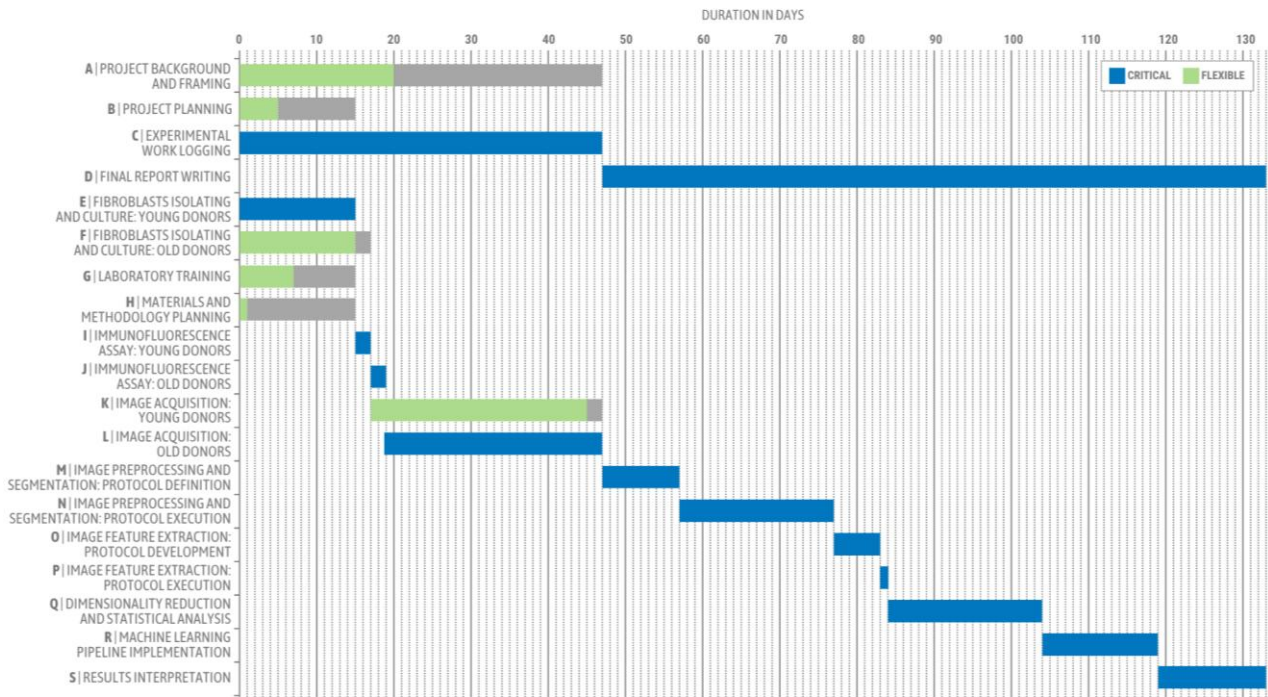
Figure 23 · *Gantt diagram.*
Time-based representation of the project schedule, displaying the start and end dates of each task, their duration, and overlaps. It highlights task dependencies and helps monitor progress throughout the project.

8 ECONOMIC VIABILITY

This section assesses the project's economic feasibility by identifying required resources, linking them to specific project phases, and estimating their costs. Table 10 provides an overview of the projected expenses, including materials and infrastructure needed for the project's implementation.

| Quantity	Material	Description	Cost	Total
Fibroblasts isolation and culture				
Murine models (from The Jackson Laboratory)				
3	C57BL/6J Young Male Mice (5 weeks)	Animal model for isolating young fibroblasts	41,39 € per mouse	124,17 €
3	C57BL/6J Young Female Mice (5 weeks)	Animal model for isolating young fibroblasts	42,33 € per mouse	126,99 €
3	C57BL/6J Aged Male Mice (74 weeks)	Animal model for isolating old fibroblasts	516,51 € per mouse	1549,53 €
3	C57BL/6J Aged Female Mice (74 weeks)	Animal model for isolating old fibroblasts	516,51 € per mouse	1549,53 €
Fibroblasts culture and seeding				
1	Gibco™ High glucose Dulbecco's Modified Eagle's Medium (DMEM)	Cell culture medium used to grow fibroblasts	30,46 € per 500 mL	30.46 €
1	Gibco™ Fetal Bovine Serum (FBS)	Nutrient-rich supplement for cell culture medium	449 € per 500 mL	449 €
1	Sigma-Aldrich Penicillin-Streptomycin (P/S) (10mg/mL)	Antibiotic mix to prevent bacterial contamination	22,10 € per 50 mL	22.10 €
1	Gibco™ Trypsin-EDTA (0.25%), phenol red	Used to detach adherent cells from the flask	21,47 € per 100 mL	21.47 €
1	Trypan Blue (0.4%) solution	Dye used to assess cell viability during counting	20,90 € per 20 mL	20.90 €
45 h	Cell culture facility (laminar hood and incubator)		2.68 € / hour (price for UB users)	120.60 €
360 h	Cell maintenance		0.59 € / hour (price for UB users)	214.40 €
Immunofluorescence assay				
1	Sigma-Aldrich Triton™ X-100 solution	Detergent used for permeabilizing cell membranes	101,00 € per 100 mL	101 €
1	Abcam Recombinant anti-alpha smooth muscle Actin (acetyl E3) + ACTG2 (acetyl E3) antibody [E184] (ab32575)	Detects α-smooth muscle actin	715,00 € per 100 µL	715 €
1	Abcam Monoclonal Anti-Vimentin antibody [VI-10] (ab20346)	Detects vimentin	380,00 € per 100 µg	380 €
1	Abcam Goat Anti-Rabbit IgG H&L (Alexa Fluor® 488) preadsorbed (ab150081)	Fluorescent antibody binding to primary α-SMA	175,00 € per 500 µg	175 €
1	Abcam Goat Anti-Mouse IgG H&L (Alexa Fluor® 488) (ab150113)	Fluorescent antibody binding to primary vimentin	170,00 € per 500 µg	170 €
1	Phalloidin-iFluor 555 Reagent (ab176756)	Fluorescent dye binding to F-actin	250,00 € per 300 tests	250 €
1	NucBlue™ Live ReadyProbes™ Reagent (6 vials of 2.5 mL each)	Fluorescent nuclear stain (DAPI)	186 € per 6 vials	186 €
1	Clear nail polish	Used to seal mounted coverslips	1,69 € per unit	1,69 €
1	Invitrogen by ThermoFisher Scientific Fluoromount-G™ Mounting Medium	Mounting medium preserving the fluorescent signal	93,25 € per 25 mL	93,25 €

1	Sigma-Aldrich Phosphate-Buffered Saline (PBS), tablet, pH 7.2 – 7.6 (1 tablet / 200 mL)	Washing buffer for immunostaining and cell maintenance	118,00 € per 50 tablets	118.€
1	ThermoFisher Scientific paraformaldehyde (PFA) 4% in PBS	Fixative used to preserve cell structure	81,90 € per 250 mL	81.90 €
1	Bovine Serum Albumin (BSA)	Protein used in the blocking buffer to reduce nonspecific binding	418,00 € per 5 g	418 €
	Consumables: Serological pipettes (2 mL, 5 mL, 10 mL, 50 mL), Serological pipette tips (100-1000 µL, 5-200 µL, 0.1-10 µL), 40 mm petri dishes, 13 mm microscope circular glass coverslips, Glass microscope slides, ground edges, frosted end, 75x25 mm, T75 and T25 cell culture flasks with filter, Nitrile gloves, Centrifuge tubes (15 mL, 50 mL), Eppendorf (1.5 mL and 2mL), Kimwipes, lens paper.			
Image acquisition				
50 h	Epifluorescence Microscope (for capturing the immunofluorescence images), accounting for the microscope itself and the oil immersion objective: Nikon CFI Plan Fluor 20× oil immersion objective	Use and capture of images in the epifluorescent microscope	7,5 € / hour	375 €
1	Leica™ Microsystems Immersion Oil for microscopes	Enhances resolution in microscopy with oil objectives	130,00 € per 10 mL	130 €
	µManager software	Control and automation of microscope hardware	Open source	0 €
Feature extraction and data analysis				
	MATLAB software	For data analysis	License price 7€ (students)	7 €
	Ilastik and CellProfiler	For cell segmentation	Open source	0 €
	RStudio (R environment) and Visual Studio Code (Python environment)	For data analysis	Open source	0 €
			TOTAL	7431 €

Table 10 · *Project cost overview.*
Detailed breakdown of materials, reagents, equipment usage, and software tools required for fibroblast isolation, immunofluorescence staining, imaging, and data analysis

The project benefits from access to institutional infrastructure, including imaging and cell culture facilities. Therefore, the prices presented are usage-based estimates calculated based on CCiTUB and IIBB service tariffs[88,89]. Key reagents are included in the cost breakdown due to their significant and direct impact on the experimental outcomes. At the same time, standard consumables are listed but excluded from the total cost calculation because they are commonly available in laboratory settings and represent minor, recurring expenses.

Based on the retributive tables of the University of Barcelona's transparency portal[90] and in accordance with the minimum amounts established by the EPIF, the base monthly salary for a first-year research assistant is €1,515.04. Given the project's defined timeframe of 133 days (approximately 4.5 months), the personnel costs are estimated accordingly at €6,817.68.

Although the murine models are not obtained exclusively for this project, their estimated costs are included to reflect their contribution within the context of a broader research effort. These models account for a significant portion of the total costs due to the need for both young and aged animals of both sexes. This cost difference is attributed to the fact that aged mice must be maintained for much more extended periods, with young mice being 1 to 2 months old and aged mice being 18 to 20 months old.

9 REGULATIONS AND LEGAL ASPECTS

The ethical use of animals in scientific research is guided by the principles of the 3Rs: Replacement, Reduction, and Refinement. These principles ensure that animal use is justified, scientifically necessary, and carried out with the highest standards of care and responsibility.

Replacement refers to methods that avoid or substitute the use of animals, including computer models or the use of less sentient organisms. In this project, replacement is not feasible due to the complexity of the biological question and the nature of aging itself, which involves intricate, systemic processes that cannot be accurately replicated using alternative models.

Reduction involves strategies to minimize the number of animals used while ensuring the validity of scientific outcomes. This project contributes to the reduction by integrating into an existing, previously approved research framework. The animals are not acquired exclusively for this study but are used across multiple experiments, maximizing the data obtained from each individual. Although this project does not specifically analyze sex differences, it includes both male and female samples to maximize the use of available biological material.

Refinement aims to enhance animal welfare by minimizing pain, stress, and discomfort through improved care practices and optimized experimental procedures. All activities follow standardized protocols and incorporate appropriate anesthesia and euthanasia methods to ensure the animals' well-being during their time in the study.

The entire research framework, including this project, complies with all ethical and legal requirements established by the University of Barcelona. The Bioethics Committee of the University of Barcelona (CBUB) is responsible for evaluating research activities that require ethical oversight and for promoting bioethics education among researchers. Given the use of murine models, this study is subject to the supervision of the Animal Experimentation Ethics Committee of the University of Barcelona (CEEA-UB), which ensures that all procedures comply with legal standards and uphold animal welfare.

10 CONCLUSIONS AND FUTURE LINES

This project establishes a complete experimental and computational pipeline for analyzing cellular aging through interpretable, image-based features. From the immunostaining of primary fibroblasts to imaging, quantification, and machine learning-based classification, the workflow demonstrates that morphological and cytoskeletal descriptors capture meaningful age-associated differences at the single-cell level.

The experimental phase generates high-quality, reproducible image data from young and aged donors under controlled conditions. Combined with feature extraction and statistical filtering, this dataset provides a strong foundation for building predictive models. The machine learning framework achieves high classification performance while also enabling the identification and ranking of biologically plausible aging descriptors, particularly those related to nuclear morphology and cytoskeletal organization.

Many of the features retained by the models align with known or hypothesized structural and functional changes associated with aging, supporting their potential for guiding biomarker discovery. This pipeline also opens several avenues for future work. Its modular and adaptable

structure makes it suitable for other cellular systems or experimental conditions, and it provides a foundation for applied applications such as high-content screening, detection of senescent subpopulations, and the development of image-based aging clocks.

Although the current dataset provides solid proof of concept, it remains limited in size, particularly in terms of donor diversity. Expanding the number of biological replicates will be essential to confirm the generalizability of the findings and account for inter-individual variability.

Future work should also focus on increasing the scale of both the experimental and computational components through automation. This would reduce manual variability, improve reproducibility, and allow the workflow to be applied in high-throughput contexts. With a larger and more representative dataset, the model can evolve from binary classification toward continuous or multi-class predictions, thereby better capturing the gradual or intermediate stages of cellular aging.

This approach lays the foundation for several applied directions, including high-content screening of interventions that affect aging, identification of senescent subpopulations, and the development of imaging-based aging clocks that estimate biological age from cellular features. These extensions would enhance both fundamental research and potential translational applications in the fields of aging and regenerative medicine.

11 BIBLIOGRAPHY

1. Guo J, Huang X, Dou L, et al. Aging and aging-related diseases: from molecular mechanisms to interventions and treatments. *Signal Transduction and Targeted Therapy 2022 7:1*. 2022;7(1):1-40. doi:10.1038/s41392-022-01251-0

2. Silva N, Rajado AT, Esteves F, et al. Measuring healthy ageing: current and future tools. *Biogerontology*. 2023;24(6):845-866. doi:10.1007/S10522-023-10041-2/FIGURES/1

3. Ong ALC, Ramasamy TS. Role of Sirtuin1-p53 regulatory axis in aging, cancer and cellular reprogramming. *Ageing Res Rev*. 2018;43:64-80. doi:10.1016/J.ARR.2018.02.004

4. Zhang Y, Gu Z, Xu Y, et al. Global scientific trends in healthy aging in the early 21st century: A data-driven scientometric and visualized analysis. *Heliyon*. 2024;10(1):e23405. doi:10.1016/J.HELIYON.2023.E23405

5. Gauthier V, Kyriazi M, Nefla M, et al. Fibroblast heterogeneity: Keystone of tissue homeostasis and pathology in inflammation and ageing. *Front Immunol*. 2023;14:1137659. doi:10.3389/FIMMU.2023.1137659/XML/NLM

6. Mechanobiology of the cytoskeleton PI: Núria Gavara Casas - Department of Biomedicine - University of Barcelona. Accessed June 10, 2025. https://www.ub.edu/portal/web/dp-biomedicalsciences/group-Gavara

7. Kendall RT, Feghali-Bostwick CA. Fibroblasts in fibrosis: Novel roles and mediators. *Front Pharmacol*. 2014;5 MAY:91491. doi:10.3389/FPHAR.2014.00123/XML/NLM

8. Lendahl U, Muhl L, Betsholtz C. Identification, discrimination and heterogeneity of fibroblasts. *Nature Communications 2022 13:1*. 2022;13(1):1-14. doi:10.1038/s41467-022-30633-9

9. DeLeon-Pennell KY, Barker TH, Lindsey ML. Fibroblasts: The arbiters of extracellular matrix remodeling. *Matrix Biology*. 2020;91-92:1-7. doi:10.1016/J.MATBIO.2020.05.006

10. Hynes RO, Naba A. Overview of the Matrisome—An Inventory of Extracellular Matrix Constituents and Functions. *Cold Spring Harb Perspect Biol*. 2012;4(1):a004903. doi:10.1101/CSHPERSPECT.A004903

11. Plikus M V., Wang X, Sinha S, et al. Fibroblasts: origins, definitions, and functions in health and disease. *Cell*. 2021;184(15):3852. doi:10.1016/J.CELL.2021.06.024

12. D'Urso M, Kurniawan NA. Mechanical and Physical Regulation of Fibroblast–Myofibroblast Transition: From Cellular Mechanoresponse to Tissue Pathology. *Front Bioeng Biotechnol*. 2020;8:609653. doi:10.3389/FBIOE.2020.609653/XML/NLM

13. Cialdai F, Risaliti C, Monici M. Role of fibroblasts in wound healing and tissue remodeling on Earth and in space. *Front Bioeng Biotechnol*. 2022;10:958381. doi:10.3389/FBIOE.2022.958381

14. Pence BD, Woods JA. Exercise, Obesity, and Cutaneous Wound Healing: Evidence from Rodent and Human Studies. *https://home-liebertpub-com.sire.ub.edu/wound*. 2014;3(1):71-79. doi:10.1089/WOUND.2012.0377

15. Ireton JE, Unger JG, Rohrich RJ. The Role of Wound Healing and Its Everyday Application in Plastic Surgery: A Practical Perspective and Systematic Review. *Plast Reconstr Surg Glob Open*. 2013;1(1):e10-e19. doi:10.1097/GOX.0B013E31828FF9F4

16. Dodig S, Čepelak I, Pavić I. Hallmarks of senescence and aging. *Biochem Med (Zagreb)*. 2019;29(3):030501. doi:10.11613/BM.2019.030501

17. Hütter E, Unterluggauer H, Überall F, Schramek H, Jansen-Dürr P. Replicative senescence of human fibroblasts: the role of Ras-dependent signaling and oxidative stress. *Exp Gerontol*. 2002;37(10-11):1165-1174. doi:10.1016/S0531-5565(02)00136-5

18. Zorina A, Zorin V, Kudlay D, Kopnin P. Age-Related Changes in the Fibroblastic Differon of the Dermis: Role in Skin Aging. *Int J Mol Sci*. 2022;23(11):6135. doi:10.3390/IJMS23116135

19. Beck J, Horikawa I, Harris C. Cellular Senescence: Mechanisms, Morphology, and Mouse Models. *Vet Pathol*. 2020;57(6):747-757. doi:10.1177/0300985820943841/ASSET/FFB8FC23-82E5-4B9C-A882-2FF03A4F61E1/ASSETS/IMAGES/LARGE/10.1177_0300985820943841-FIG4.JPG

20. Robin JD, Magdinier F. Physiological and pathological aging affects chromatin dynamics, structure and function at the nuclear edge. *Front Genet*. 2016;7(AUG):213269. doi:10.3389/FGENE.2016.00153/XML/NLM

21. Sliogeryte K, Gavara N. Vimentin Plays a Crucial Role in Fibroblast Ageing by Regulating Biophysical Properties and Cell Migration. *Cells*. 2019;8(10):1164. doi:10.3390/CELLS8101164

22. Schulze C, Wetzel F, Kueper T, et al. Stiffening of Human Skin Fibroblasts with Age. *Biophys J*. 2010;99(8):2434. doi:10.1016/J.BPJ.2010.08.026

23. Pardo A, Selman M. Lung fibroblasts, aging, and idiopathic pulmonary fibrosis. *Ann Am Thorac Soc*. 2016;13:S417-S421. doi:10.1513/ANNALSATS.201605-341AW/SUPPL_FILE/DISCLOSURES.PDF

24. Ohtani N. The roles and mechanisms of senescence-associated secretory phenotype (SASP): can it be controlled by senolysis? *Inflamm Regen*. 2022;42(1):1-8. doi:10.1186/S41232-022-00197-8/FIGURES/3

25.	Torres-Machorro AL, García-Vicente Á, Espina-Ordoñez M, et al. Update of Aging Hallmarks in Idiopathic Pulmonary Fibrosis. *Cells 2025, Vol 14, Page 222*. 2025;14(3):222. doi:10.3390/CELLS14030222

26.	Alqahtani S, Alqahtani T, Venkatesan K, et al. SASP Modulation for Cellular Rejuvenation and Tissue Homeostasis: Therapeutic Strategies and Molecular Insights. *Cells*. 2025;14(8):608. doi:10.3390/CELLS14080608

27.	Kondo H, Yonezawa Y. Changes in the migratory ability of human lung and skin fibroblasts during in vitro aging and in vivo cellular senescence. *Mech Ageing Dev*. 1992;63(3):223-233. doi:10.1016/0047-6374(92)90001-T

28.	Rahal JA, Nassar D. Wound Healing as We Age. *Textbook of Aging Skin*. Published online 2017:1949-1958. doi:10.1007/978-3-662-47398-6_143

29.	Ding X, Kakanj P, Leptin M, Eming SA. Regulation of the Wound Healing Response during Aging. *Journal of Investigative Dermatology*. 2021;141(4):1063-1070. doi:10.1016/J.JID.2020.11.014

30.	Moretti L, Stalfort J, Barker TH, Abebayehu D. The interplay of fibroblasts, the extracellular matrix, and inflammation in scar formation. *J Biol Chem*. 2021;298(2):101530. doi:10.1016/J.JBC.2021.101530

31.	Darby IA, Laverdet B, Bonté F, Desmoulière A. Fibroblasts and myofibroblasts in wound healing. *Clin Cosmet Investig Dermatol*. 2014;7:301-311. doi:10.2147/CCID.S50046

32.	Xia W, Li M, Jiang X, et al. Young fibroblast-derived exosomal microRNA-125b transfers beneficial effects on aged cutaneous wound healing. *J Nanobiotechnology*. 2022;20(1):1-17. doi:10.1186/S12951-022-01348-2/FIGURES/8

33.	Yanai H, Shteinberg A, Porat Z, et al. Cellular senescence-like features of lung fibroblasts derived from idiopathic pulmonary fibrosis patients. *Aging*. 2015;7(9):664-672. doi:10.18632/AGING.100807

34.	Ren LL, Miao H, Wang YN, Liu F, Li P, Zhao YY. TGF-β as A Master Regulator of Aging-Associated Tissue Fibrosis. *Aging Dis*. 2023;14(5):1633. doi:10.14336/AD.2023.0222

35.	Onursal C, Dick E, Angelidis I, Schiller HB, Staab-Weijnitz CA. Collagen Biosynthesis, Processing, and Maturation in Lung Ageing. *Front Med (Lausanne)*. 2021;8:593874. doi:10.3389/FMED.2021.593874/XML/NLM

36.	Welter EM, Benavides S, Archer TK, Kosyk O, Zannas AS. Machine learning-based morphological quantification of replicative senescence in human fibroblasts. *Geroscience*. 2023;46(2):2425. doi:10.1007/S11357-023-01007-W

37.	Duran I. A nuclear morphology-based machine learning algorithm for senescence detection. *Nat Rev Mol Cell Biol*. 2024;25(12):949-949. doi:10.1038/S41580-024-00796-Y;SUBJMETA=1647,48,509,631,80;KWRD=BIOINFORMATICS,SENESCENCE

38.	Neri F, Takajjart SN, Lerner CA, et al. A Fully-Automated Senescence Test (FAST) for the high-throughput quantification of senescence-associated markers. *bioRxiv*. Published online December 23, 2024:2023.12.22.573123. doi:10.1101/2023.12.22.573123

39.	Hillsley A, Santos JE, Rosales AM. A deep learning approach to identify and segment alpha-smooth muscle actin stress fiber positive cells. *Sci Rep*. 2021;11(1):1-11. doi:10.1038/S41598-021-01304-4;SUBJMETA=114,1305,1564,2295,2397,54,61,631;KWRD=BIOMATERIALS+

40.	Fleischer JG, Schulte R, Tsai HH, et al. Predicting age from the transcriptome of human dermal fibroblasts. *Genome Biol*. 2018;19(1). doi:10.1186/S13059-018-1599-6,

41.	Yu D, Li M, Linghu G, et al. CellBiAge: Improved single-cell age classification using data binarization. *Cell Rep*. 2023;42(12). doi:10.1016/j.celrep.2023.113500

42.	Schulze C, Wetzel F, Kueper T, et al. Stiffening of Human Skin Fibroblasts with Age. *Biophysj*. 2010;99:2434-2442. doi:10.1016/j.bpj.2010.08.026

43.	Dulińska-Molak I, Pasikowska M, Pogoda K, Lewandowska M, Eris I, Lekka M. Age-Related Changes in the Mechanical Properties of Human Fibroblasts and Its Prospective Reversal After Anti-Wrinkle Tripeptide Treatment. *Int J Pept Res Ther*. 2013;20(1):77. doi:10.1007/S10989-013-9370-Z

44.	Singam A, Bhattacharya C, Park S. Aging-related changes in the mechanical properties of single cells. *Heliyon*. 2024;10(12):e32974. doi:10.1016/J.HELIYON.2024.E32974/ASSET/77201F4D-A655-4F82-8C9A-E24965CC788D/MAIN.ASSETS/GR1.JPG

45.	The Future of Aging and Longevity | Deloitte US. Accessed June 9, 2025. https://www2.deloitte.com/us/en/pages/life-sciences-and-health-care/articles/longevity-science.html

46.	Moqri M, Herzog C, Poganik JR, et al. Biomarkers of Aging for the Identification and Evaluation of Longevity Interventions. *Cell*. 2023;186(18):3758. doi:10.1016/J.CELL.2023.08.003

47.	Lyu YX, Fu Q, Wilczok D, et al. Longevity biotechnology: bridging AI, biomarkers, geroscience and clinical applications for healthy longevity. *Aging*. 2024;16(20):12955-12976. doi:10.18632/AGING.206135,

48.	Guerville F, De Souto Barreto P, Ader I, et al. Revisiting the Hallmarks of Aging to Identify Markers of Biological Age. *J Prev Alzheimers Dis*. 2020;7(1):56-64. doi:10.14283/JPAD.2019.50

49.	Buck Institute | Research. Accessed June 9, 2025. https://www.buckinstitute.org/research/

50. Max Planck Institute | Research. Accessed June 9, 2025. https://www.age.mpg.de/research

51. Research & Technology - Calico. Accessed June 9, 2025. https://www.calicolabs.com/research-technology/

52. Altos Labs - Science. Accessed June 9, 2025. https://www.altoslabs.com/science

53. Min M, Egli C, Dulai AS, Sivamani RK. Critical review of aging clocks and factors that may influence the pace of aging. *Frontiers in Aging*. 2024;5:1487260. doi:10.3389/FRAGI.2024.1487260/BIBTEX

54. Zhavoronkov A, Mamoshina P. Deep Aging Clocks: The Emergence of AI-Based Biomarkers of Aging and Longevity. *Trends Pharmacol Sci*. 2019;40(8):546-549. doi:10.1016/j.tips.2019.05.004

55. Ninomiya K. ImAge: quantifying epigenetic ageing with single-cell images. *Nature Reviews Cancer 2025*. Published online March 13, 2025:1-1. doi:10.1038/S41568-025-00809-8

56. Ahadi S, Wilson KA, Babenko B, et al. Longitudinal fundus imaging and its genome-wide association analysis provide evidence for a human retinal aging clock. *Elife*. 2023;12:1-28. doi:10.7554/ELIFE.82364

57. Custodero C, Mankowski RT, Lee SA, et al. Evidence-based nutritional and pharmacological interventions targeting chronic low-grade inflammation in middle-age and older adults: A systematic review and meta-analysis. *Ageing Res Rev*. 2018;46:42-59. doi:10.1016/J.ARR.2018.05.004,

58. Lelarge V, Capelle R, Oger F, Mathieu T, Le Calvé B. Senolytics: from pharmacological inhibitors to immunotherapies, a promising future for patients' treatment. *npj Aging*. 2024;10(1):1-8. doi:10.1038/S41514-024-00138-4;SUBJMETA=154,631,80;KWRD=CELL+BIOLOGY,DRUG+DISCOVERY

59. Madreiter-Sokolowski CT, Hiden U, Krstic J, et al. Targeting organ-specific mitochondrial dysfunction to improve biological aging. *Pharmacol Ther*. 2024;262. doi:10.1016/j.pharmthera.2024.108710

60. Klinngam W, Chaiwichien A, Osotprasit S, et al. Longevity cosmeceuticals as the next frontier in cosmetic innovation: a scientific framework for substantiating product claims. *Frontiers in Aging*. 2025;6:1586999. doi:10.3389/FRAGI.2025.1586999

61. Markiewicz E, Idowu OC. Evaluation of Personalized Skincare Through in-silico Gene Interactive Networks and Cellular Responses to UVR and Oxidative Stress. *Clin Cosmet Investig Dermatol*. 2022;15:2221. doi:10.2147/CCID.S383790

62. Kivimäki M, Frank P, Pentti J, et al. Proteomic organ-specific ageing signatures and 20-year risk of age-related diseases: the Whitehall II observational cohort study. *Lancet Digit Health*. 2025;7(3):e195-e204. doi:10.1016/j.landig.2025.01.006

63. McKayed KK, Simpson JC. Actin in Action: Imaging Approaches to Study Cytoskeleton Structure and Function. *Cells*. 2013;2(4):715. doi:10.3390/CELLS2040715

64. Combs CA. Fluorescence Microscopy: A Concise Guide to Current Imaging Methods. *Current protocols in neuroscience / editorial board, Jacqueline N Crawley . [et al]*. 2010;0 2(SUPPL.50):Unit2.1. doi:10.1002/0471142301.NS0201S50

65. Wang X, Lai Y. Three basic types of fluorescence microscopy and recent improvement. *E3S Web of Conferences*. 2021;290:01031. doi:10.1051/E3SCONF/202129001031

66. Kapuscinski J. DAPI: A DMA-Specific fluorescent probe. *Biotechnic and Histochemistry*. 1995;70(5):220-233. doi:10.3109/10520299509108199;WEBSITE:WEBSITE:TFOPB;PAGEGROUP:STRING:PUBLICATION

67. Heckenbach I, Mkrtchyan G V., Ezra M Ben, et al. Nuclear morphology is a deep learning biomarker of cellular senescence. *Nature Aging 2022 2:8*. 2022;2(8):742-755. doi:10.1038/s43587-022-00263-3

68. Zhang T, Yuan X, Jiang M, et al. Proteomic analysis reveals the aging-related pathways contribute to pulmonary fibrogenesis. *Aging*. 2023;15(24):15382-15401. doi:10.18632/AGING.205355

69. Chantachotikul P, Liu S, Furukawa K, Deguchi S. AP2A1 is upregulated upon replicative senescence of human fibroblasts to strengthen focal adhesions via integrin β1 translocation along stress fibers. Published online August 20, 2023. doi:10.1101/2023.08.19.553998

70. Phillips TA, Marcotti S, Cox S, Parsons M. Imaging actin organisation and dynamics in 3D. *J Cell Sci*. 2024;137(2). doi:10.1242/JCS.261389,

71. Salminen A. The plasticity of fibroblasts: A forgotten player in the aging process. *Ageing Res Rev*. 2023;89:101995. doi:10.1016/J.ARR.2023.101995

72. Hinz B, Celetta G, Tomasek JJ, Gabbiani G, Chaponnier C. Alpha-Smooth Muscle Actin Expression Upregulates Fibroblast Contractile Activity. *Mol Biol Cell*. 2001;12(9):2730. doi:10.1091/MBC.12.9.2730

73. Schroeder AB, Dobson ETA, Rueden CT, Tomancak P, Jug F, Eliceiri KW. The ImageJ ecosystem: Open-source software for image visualization, processing, and analysis. *Protein Sci*. 2020;30(1):234. doi:10.1002/PRO.3993

74. Schindelin J, Arganda-Carreras I, Frise E, et al. Fiji: An open-source platform for biological-image analysis. *Nat Methods*. 2012;9(7):676-682. doi:10.1038/NMETH.2019;SUBJMETA=1647,245,631,794;KWRD=IMAGING,SOFTWARE

75.   Berg S, Kutra D, Kroeger T, et al. ilastik: interactive machine learning for (bio)image analysis. *Nat Methods*. 2019;16(12):1226-1232. doi:10.1038/S41592-019-0582-9;SUBJMETA=114,1305,1564,631,794;KWRD=IMAGE+PROCESSING,MACHINE+LEARNING,SOFTWARE

76.   Carpenter AE, Jones TR, Lamprecht MR, et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol*. 2006;7(10):R100. doi:10.1186/GB-2006-7-10-R100

77.   Welter EM, Kosyk O, Zannas AS. An open access, machine learning pipeline for high-throughput quantification of cell morphology. *STAR Protoc*. 2023;4(1). doi:10.1016/j.xpro.2022.101947

78.   Khang A, Barmore A, Tseropoulos G, Bera K, Batan D, Anseth KS. Automated prediction of fibroblast phenotypes using mathematical descriptors of cellular features. *Nature Communications 2025 16:1*. 2025;16(1):1-17. doi:10.1038/s41467-025-58082-0

79.   Shamhan M, Idris AS, Toha SF, Daud MF, Idris IM, Malik H. An automated approach for fibroblast cell confluency characterisation and sample handling using AIoT for bio-research and bio-manufacturing. *Cogent Eng*. 2023;10(1). doi:10.1080/23311916.2023.2240087

80.   Wong KS, Zhong X, Low CSL, Kanchanawong P. Self-supervised classification of subcellular morphometric phenotypes reveals extracellular matrix-specific morphological responses. *Sci Rep*. 2022;12(1):1-14. doi:10.1038/S41598-022-19472-2;SUBJMETA=114,1305,1314,1386,1564,631;KWRD=CLASSIFICATION+AND+TAXONOMY,COMPUTATIONAL+BIOLOGY+AND+BIOINFORMATICS,DATA+PROCESSING,IMAGE+PROCESSING,MACHINE+LEARNING

81.   Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Comput Sci*. 2021;2(3):160. doi:10.1007/S42979-021-00592-X

82.   Machine Learning Algorithms | Microsoft Azure. Accessed June 9, 2025. https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-are-machine-learning-algorithms

83.   Janssen AFJ, Breusegem SY, Larrieu D. Current Methods and Pipelines for Image-Based Quantitation of Nuclear Shape and Nuclear Envelope Abnormalities. *Cells*. 2022;11(3):347. doi:10.3390/CELLS11030347

84.   Alizadeh E, Lyons SM, Castle JM, Prasad A. Measuring systematic changes in invasive cancer cell shape using Zernike moments. *Integrative Biology*. 2016;8(11):1183-1193. doi:10.1039/C6IB00100A

85.   Chaddad A, Tanougast C. Texture Analysis of Abnormal Cell Images for Predicting the Continuum of Colorectal Cancer. *Anal Cell Pathol (Amst)*. 2017;2017:8428102. doi:10.1155/2017/8428102

86.   Stringer C, Wang T, Michaelos M, Pachitariu M. Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods*. 2021;18(1):100-106. doi:10.1038/S41592-020-01018-X;SUBJMETA=114,631,80;KWRD=CELL+BIOLOGY,COMPUTATIONAL+BIOLOGY+AND+BIOINFORMATICS

87.   Weigert M, Schmidt U, Haase R, Sugawara K, Myers G. Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy. *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. Published online March 2020:3655-3662. doi:10.1109/WACV45572.2020.9093435<SPAN

88.   Scientific, technical and other specialized services - University of Barcelona. Accessed June 9, 2025. https://web.ub.edu/en/scientific-technical-and-other-specialized-services

89.   Institute of Biomedical Research of Barcelona | It's time for research, time for life, ...it's time for CSIC. Accessed June 9, 2025. https://www.iibb.csic.es/en/services

90.   Taules retributives PDI i PTGAS - Transparency portal - University of Barcelona. Accessed June 9, 2025. https://web.ub.edu/en/web/transparencia/taules-retributives-pdi-i-ptgas

12 ANNEXES

The following annexes provide supplementary materials that support and expand upon the main content of the report, including technical details, and extended data.

12.1 ANNEX: GITHUB REPOSITORY

All the code and data used in the project are available in the following public repository: https://github.com/lacxy05/AI-based-aging-diagnostic-algorithm.git. It includes:

- **Data files** generated during the feature extraction and statistical filtering steps
- The image segmentation pipeline implemented in **CellProfiler**
- **ImageJ** macros for single-cell cropping and discarding low-quality images
- **Python** notebooks for feature extraction and machine learning.
- **R** scripts for statistical filtering based on ANOVA and correlation analysis

12.2 ANNEX: PIPELINE CELLPROFILER

PIPELINE CELLPROFILER	
PREPROCESSING OF INPUT PROBABILITY MAPS	
ColorToGray (phalloidin, αSMA/vimentin, DAPI)	Converts RGB probability maps into grayscale images for each fluorescence channel.
RemoveHoles (phalloidin, αSMA/vimentin, DAPI)	Fills small holes to produce continuous structures in the probability maps.
MedianFilter (DAPI)	Reduces high-frequency noise while preserving key features.
Morph (fill) (phalloidin, αSMA/vimentin)	Applies morphological operations to fill gaps within cellular structures.
EnhanceOrSuppressFeatures (phalloidin, αSMA/vimentin)	Emphasizes linear and edge-like features to define cytoskeletal organization better.
MedianFilter (phalloidin, αSMA/vimentin)	Further reduces noise and smooths structures.
OBJECT IDENTIFICATION: NUCLEUS	
IdentifyPrimaryObjects	Detects nuclei using adaptive Sauvola thresholding.
Option A: FilterObjects	Automatically retains the most prominent nucleus if multiple nuclei are detected and generates a binary mask.
Option B: EditObjectsManually	Allows manual selection of the correct nucleus when automatic filtering is not sufficient.
OBJECT IDENTIFICATION: CELL	
IdentifySecondaryObjects (phalloidin, αSMA/vimentin)	Expands from the nucleus to segment the entire cell using watershed and adaptive thresholding (Sauvola).
CombineObjects	Merges the two masks obtained from phalloidin and αSMA/vimentin channels into a single cell mask.
ConvertObjectToImage	Converts segmented objects into a binary image format.
Morph (fill)	Applies final morphological operations to smooth and complete object outlines.
RemoveHoles	Eliminates residual internal holes to finalize segmentation.
ConvertImageToObject	Converts the cleaned binary mask image back into object format for downstream analysis.

Table 11 *· Overview of the CellProfiler segmentation pipeline.*
Summary of the modules used in the custom CellProfiler workflow for object identification. The pipeline processes Ilastik-generated probability maps to output binary masks for both the nucleus and the entire cell.

12.3 ANNEX: EXTRACTED FEATURES

CELL	NUCLEUS	FEATURES	DESCRIPTION
		MORPHOLOGICAL FEATURES	
		SIZE AND AREA DESCRIPTORS	
•	•	AREA	Total pixel count within the segmented region.
•	•	CONVEX AREA	Area of the convex hull that encloses the object.
•	•	EQUIVALENT DIAMETER	Diameter of a circle with the same area as the object.
•	•	FERET DIAMETER (MAXIMUM)	Maximum caliper diameter, i.e., the most significant distance between any two points along the boundary.
•	•	BOUNDING BOX WIDTH	Width of the smallest rectangle enclosing the región.
•	•	BOUNDING BOX HEIGHT	Height of the smallest rectangle enclosing the region
		SHAPE AND GEOMETRY	
•	•	PERIMETER	Total length of the outer boundary.
•	•	ECCENTRICITY	Elongation ratio: distance between ellipse foci / major axis.
•	•	SOLIDITY	Ratio of area to the convex hull area. Values near 1 indicate a solid shape with few indentations.
•	•	EXTENT	Ratio of object area to the bounding box area.
•		ASPECT RATIO	Ratio of the major to minor axis length of the fitted ellipse.
		CIRCULARITY AND STRUCTURAL COMPACTNESS	
•	•	CIRCULARITY	$\frac{4\pi \cdot area}{perimeter^2}$. Reflects roundness. Values near 1 indicate a perfect circle.
•		CONVEXITY	Estimated as perimeter² / area. Larger values reflect irregular contours.
		CELL-NUCLEUS SPATIAL RELATIONS	
	•	CENTROID DISTANCE	Euclidean distance between the centroids of the cell and its nucleus.
	•	NORMALIZED CENTROID DISTANCE	Centroid distance normalized by the cell's major axis length.
	•	NUCLEUS TO CELL AREA RATIO	Ratio of nuclear area to total cell area.
		ORIENTATION	
•	•	ORIENTATION	Angle between the major axis of the ellipse and the horizontal axis.
		SHAPE INVARIANCE AND COMPLEXITY	
•	•	FRACTAL DIMENSION	Estimated using a box-counting method, which describes shape complexity.
•	•	HU-MOMENTS (1-7)	Scale- and rotation-invariant descriptors of shape.
	•	ZERNIKE MOMENTS (DEGREE 12)	Advanced invariant shape descriptors, capturing symmetry at multiple spatial frequencies.

Table 12 · Morphological features extracted from segmented cells and nucle obectsi.
Summarizes the quantitative descriptors used to characterize cell and nuclear geometry.

DAPI	PHALLOIDIN	αSMA	FEATURES	DESCRIPTION
			INTENSITY FEATURES	
			BASIC STATISTICAL MEASURES	
•	•	•	MEAN	Average pixel intensity inside the mask.
•	•	•	MEDIAN	The middle value of the intensity distribution (less sensitive to outliers than the mean)
•	•	•	STANDARD DEVIATION	Dispersion of intensity values around the mean.
•	•	•	MINIMUM	The lowest intensity value observed within the mask.
•	•	•	MAXIMUM	The highest intensity value observed within the mask.
			DISTRIBUTION SHAPE DESCRIPTORS	
•	•	•	SKEWNESS	Degree of asymmetry in the distribution. Positive skew indicates a long right tail.
•	•	•	KURTOSIS	Measure of peakness or flatness of the intensity distribution.
•	•	•	COEFFICIENT OF VARIATION (COV)	Ratio of the standard deviation to the mean intensity, providing a normalized measure of dispersion.
			QUANTILES AND ROBUST STATISTICS	
•	•	•	5TH PERCENTILE	Value below which 5% of pixel intensities fall.
•	•	•	95TH PERCENTILE	Value above which 5% of intensities fall.
•	•	•	INTERQUARTILE RANGE (IQR)	Difference between the 75th and 25th percentiles.
•	•	•	MEDIAN ABSOLUTE DEVIATION (MAD)	Median of the absolute deviations from the median (alternative to the standard deviation).
			SIGNAL STRENGTH AND CONTRAST	
•	•	•	INTEGRATED INTENSITY	Sum of all pixel values inside the mask.
•	•	•	ENERGY	Sum of squared intensities, reflecting overall signal power.
•	•	•	MODE	Most frequently occurring pixel intensity bin.
			ENTROPY AND UNIFORMITY	
•	•	•	ENTROPY	Shannon entropy of the normalized intensity histogram, indicating randomness.
•	•	•	UNIFORMITY	Sum of squared probabilities across histogram bins. High uniformity suggests homogeneous intensities.
			NUCLEAR-SPECIFIC METRICS	
•			NUCLEAR DENSITY	Integrated intensity divided by the estimated nuclear volume (assuming a spherical nucleus).
			RADIAL INTENSITY PROFILES	
•	•	•	RADIAL MEAN (AT 25%, 50%, 75%, 100%)	Mean intensity within each region defined by a fractional radius of the maximum.
•	•	•	STANDARD DEVIATION (AT 25%, 50%, 75%, 100%)	Stf deviation within each region defined by a fractional radius of the maximum.

Table 13 · *Intensity features extracted from DAPI, phalloidin, and α-SMA channels.*
Details the statistical, distributional, and spatial descriptors used to quantify fluorescence signal intensity within segmented regions.

DAPI	PHALLOIDIN	αSMA	FEATURES	DESCRIPTION
			TEXTURE FEATURES	
			GRAY-LEVEL CO-OCCURRENCE MATRIX (GLCM) GLCMs measure how often pairs of pixel values occur at a given distance and angle. Texture statistics are computed over six distances (1, 2, 3, 5, 7, 9 pixels) and four angles (0°, 45°, 90°, 135°),	
•	•	•	CONTRAST	Measures local intensity variation. High values indicate strong edges or rough textures.
•	•	•	DISSIMILARITY	Linear intensity difference between neighbors. Highlights the overall difference between neighboring pixel values.
•	•	•	HOMOGENEITY	Measures local uniformity. Higher values indicate smoother textures.
•	•	•	ENERGY	Square root of the sum of squared elements. Also known as Angular Second Moment, this value is high for uniform textures.
•	•	•	CORRELATION	Measures the linear dependency of gray levels in neighboring pixels.
•	•	•	ENTROPY	Shannon entropy of GLCM distribution. Measures the randomness of the texture. High values imply complex or disordered patterns.
			HARALICK FEATURES A comprehensive set of statistical descriptors derived from GLCMs is computed across multiple distances.	
•	•	•	HARALICK DESCRIPTORS (1-13, D=1-9)	Thirteen standard Haralick descriptors (e.g., contrast, entropy, correlation, variance) are calculated and averaged over angles. These features provide a compact summary of complex textures such as filament bundling or chromatin granularity.
			LOCAL BINARY PATTERNS (LPB) LBPs encode local texture by comparing each pixel to its neighbors. The LBP histogram describes micro-texture patterns. These are calculated for two configurations: P=8,R=1 (fine-grain detail) and P=16,R=2 (Broader texture structures)	
•	•	•	MODE	The most frequent local binary pattern. Indicates the dominant structural motif.
•	•	•	VARIANCE	Variability of the LBP histogram. Reflects the diversity of local textures.
•	•	•	ENTROPY	Shannon entropy of the LBP distribution. Captures disorder or irregularity.
•	•	•	ENERGY	Sum of squared LBP histogram values. Higher for more uniform patterns.

Table 14 · Texture features extracted from DAPI, phalloidin, and α-SMA channels.
Summarizes the metrics used to quantify spatial intensity patterns and structural complexity.

PHALLOIDIN	αSMA	FEATURES	DESCRIPTION
		ORIENTATION FEATURES	
		RADIAL SYMMETRY	
•	•	RADIAL SLOPE	Slope of log-log plot of intensity vs. radial distance; reflects intensity decay from center.
•	•	RADIAL VARIABILITY	Std deviation of radial intensity distribution normalized by peak.
		NUCLEUS–FIBER ALIGNMENT	
•	•	NUCLEAR VS. FIBER ALIGNMENT	Cosine of the angular difference between the nucleus orientation and the mean fiber orientation. Values near 1 indicate parallel alignment.
		CIRCULAR ORIENTATION STATISTICS	
•	•	CIRCULAR MEAN	Average orientation angle, using circular statistics.
•	•	CIRCULAR VARIANCE	Dispersion of fiber orientations (0 = aligned, 1 = disordered).
•	•	ORIENTATION RANGE	Angular spread between dominant fiber directions.
•	•	ORIENTATION FWHM	Full width at half-maximum of orientation histogram.
		ANISOTROPY / COHERENCE	
•	•	COHERENCE MEAN	Average orientation coherence from image gradients.
•	•	COHERENCE STD DEV	Standard deviation of coherence values.
		FRANGI	
•	•	FRANGI MEAN	Average response of Frangi filter (fiber detection).
•	•	FRANGI STD DEV	Standard deviation of Frangi response.
		SKELETON NETWORK METRICS	
•	•	SKELETON LENGTH	Total number of skeleton pixels (estimated fiber network length).
•	•	BRANCH POINTS	Number of junctions in the skeleton where more than two fibers meet.
•	•	NETWORK DENSITY	Ratio of branch points to total skeleton length.

Table 15 · Orientation features extracted from phalloidin and α-SMA channels.
Describes metrics that quantify cytoskeletal fiber alignment, anisotropy, and spatial organization.

12.4 ANNEX: BEST MODELS

MODEL	PARAMETERS	BEST
DECISION TREE	CRITERION	ENTROPY
	MAX_DEPTH	4
	MAX_FEATURES	LOG2
	MIN_SAMPLES_LEAF	4
	MIN_SAMPLES_SPLIT	10
KNN	N_NEIGHBORS	21
	P	1
	WEIGHTS	DISTANCE
RANDOM FOREST	CRITERION	ENTROPY
	MAX_DEPTH	7
	MAX_FEATURES	SQRT
	MIN_SAMPLES_LEAF	3
	MIN_SAMPLES_SPLIT	2
	N_ESTIMATORS	200
SVM	C	100.0
	GAMMA	0.05577039432971853
	KERNEL	LINEAR
LOGISTIC REGRESSION	C	5.34024913271107
	PENALTY	L1

Table 16 · *Optimal hyperparameters for the final classification models.*
Summary of the best-performing parameter configurations selected for each classifier.

12.5 ANNEX: MODEL COMPARISON
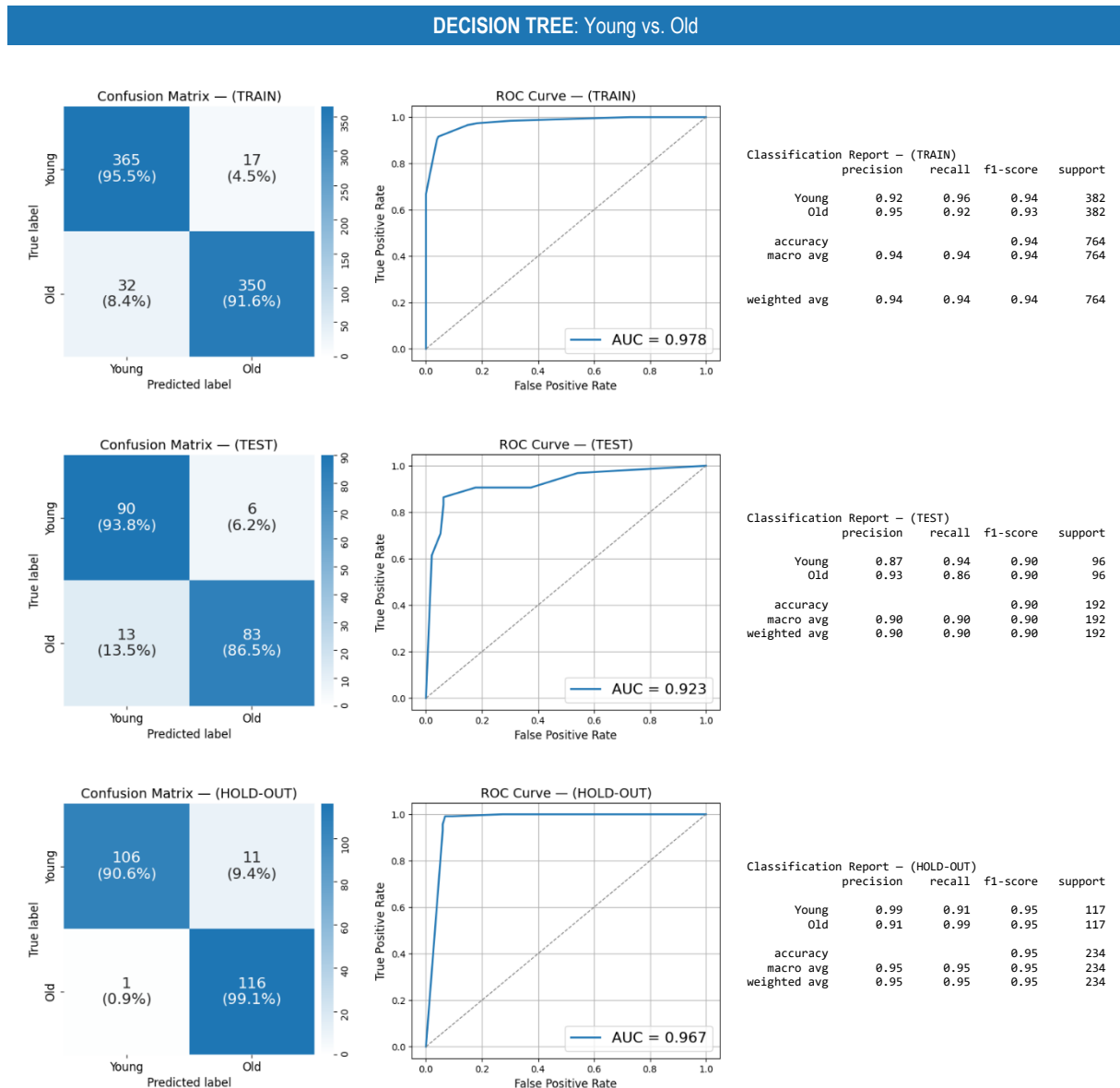
DECISION TREE: Young vs. Old

Figure 24 · *Decision Tree model performance (Aged vs. Young classification).*
Confusion matrices, ROC curves, and classification reports for the DT model trained with Bayesian-optimized hyperparameters. Results are shown for the training, test, and hold-out sets.
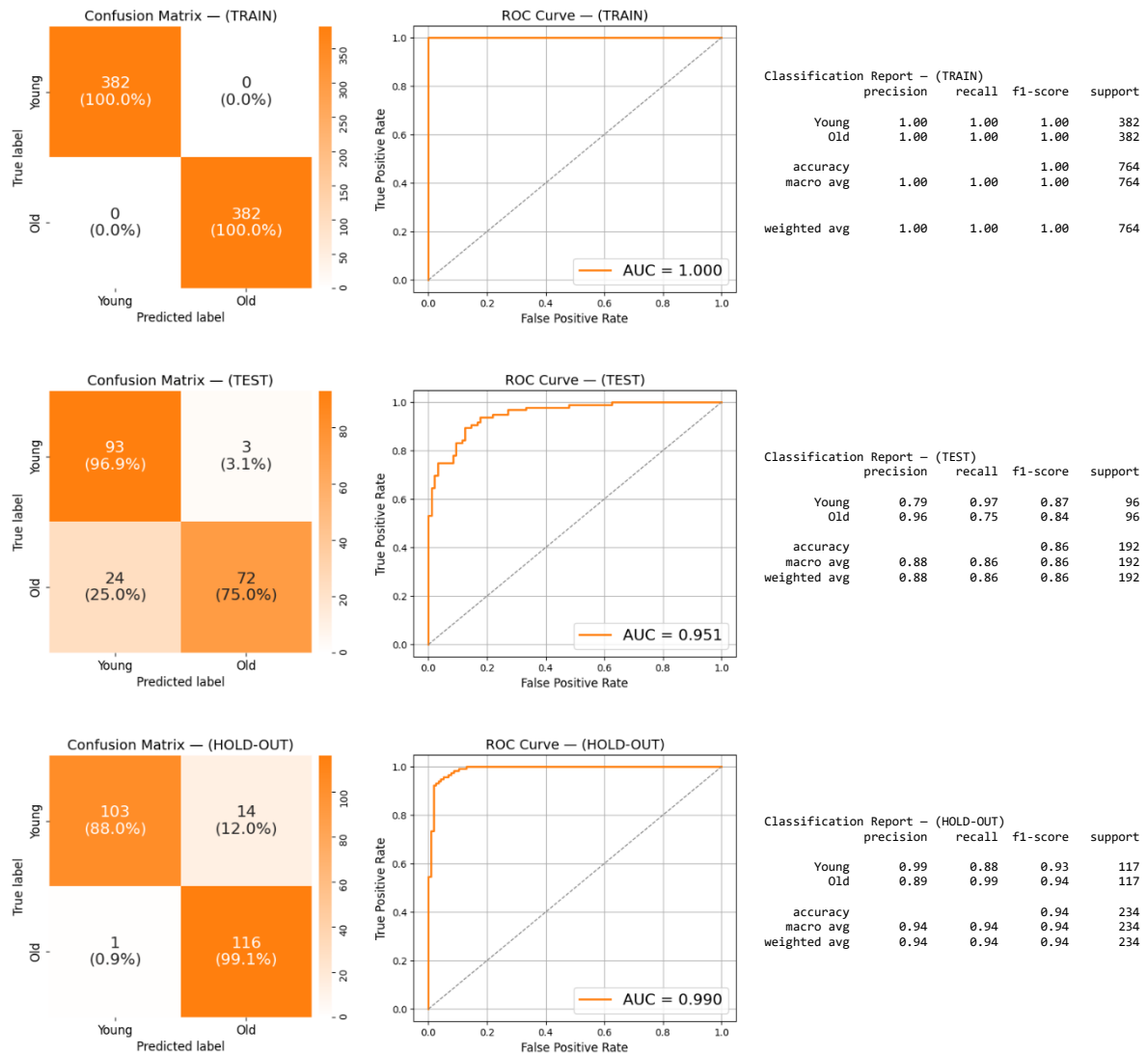
K-NN: Young vs. Old

Figure 25 · k-Nearest Neighbors model performance (Aged vs. Young classification).
Confusion matrices, ROC curves, and classification reports for the k-NN model trained with Bayesian-optimized hyperparameters. Results are shown for the training, test, and hold-out sets.
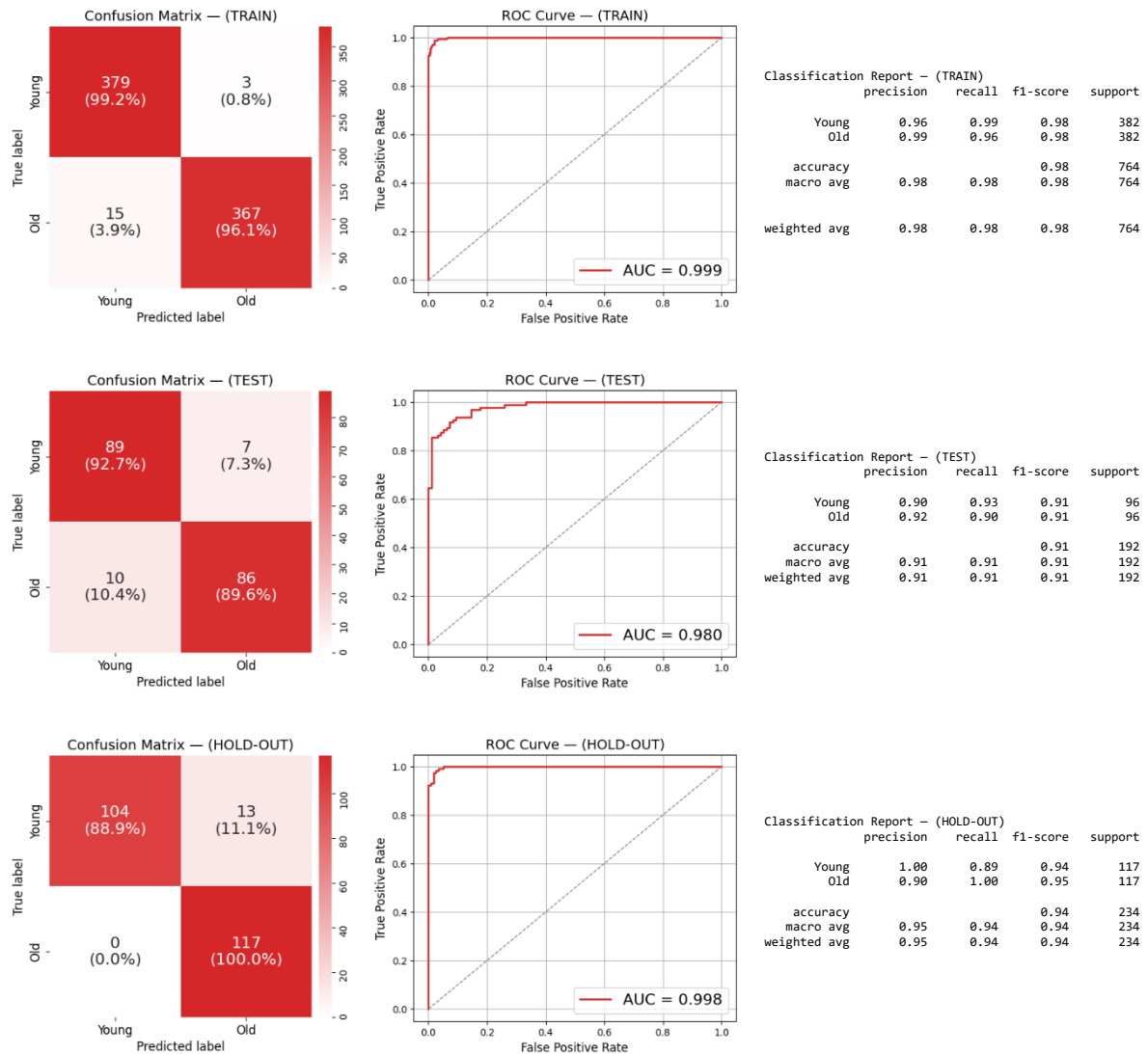
RANDOM FOREST: Young vs. Old

Figure 26 · *Random Forest model performance (Aged vs. Young classification).*
Confusion matrices, ROC curves, and classification reports for the RF model trained with Bayesian-optimized hyperparameters. Results are shown for the training, test, and hold-out sets.
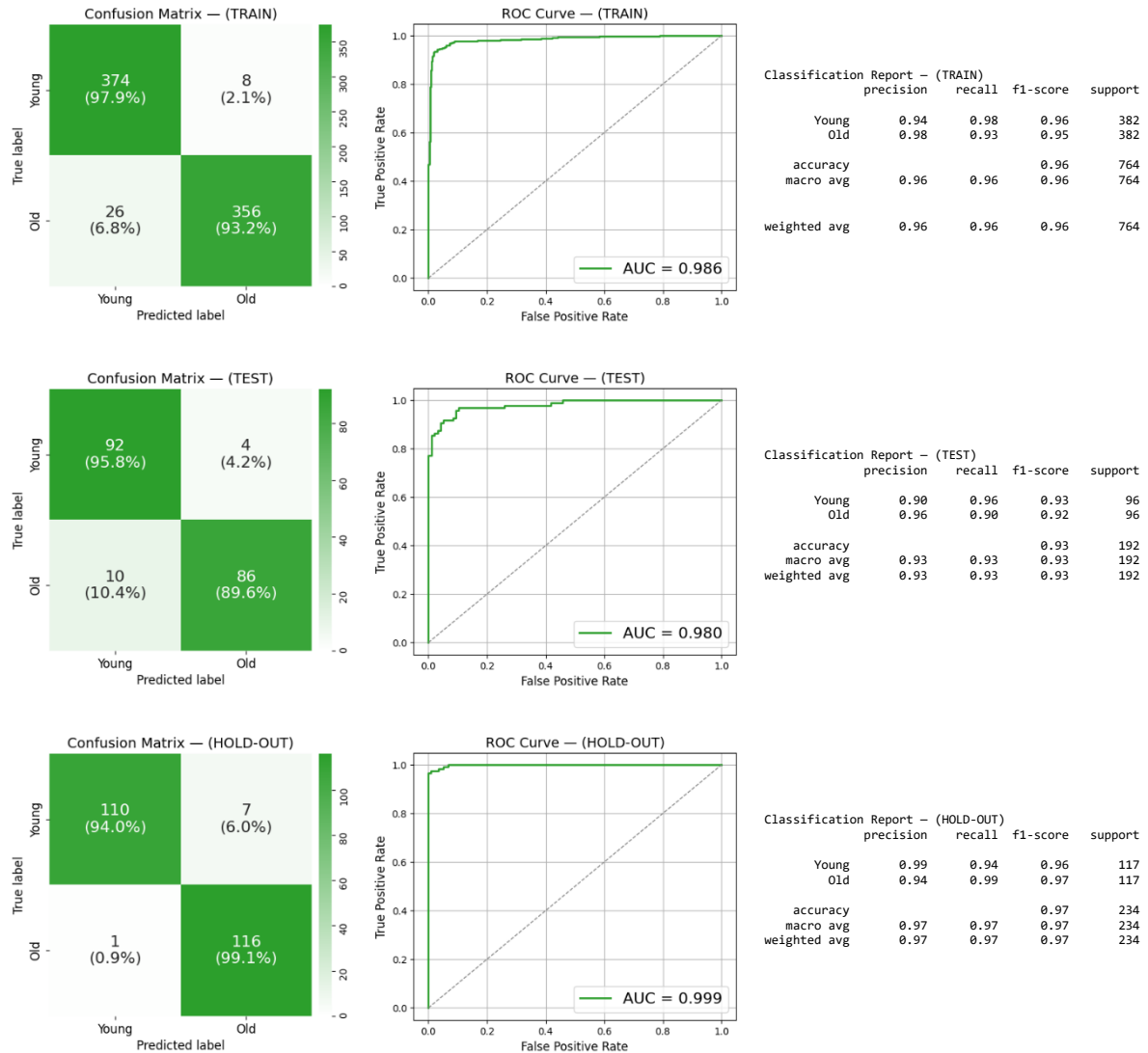
SVM: Young vs. Old

Figure 27 · *Support Vector Machine model performance (Aged vs. Young classification).*
Confusion matrices, ROC curves, and classification reports for the SVM model trained with Bayesian-optimized hyperparameters. Results are shown for the training, test, and hold-out sets.

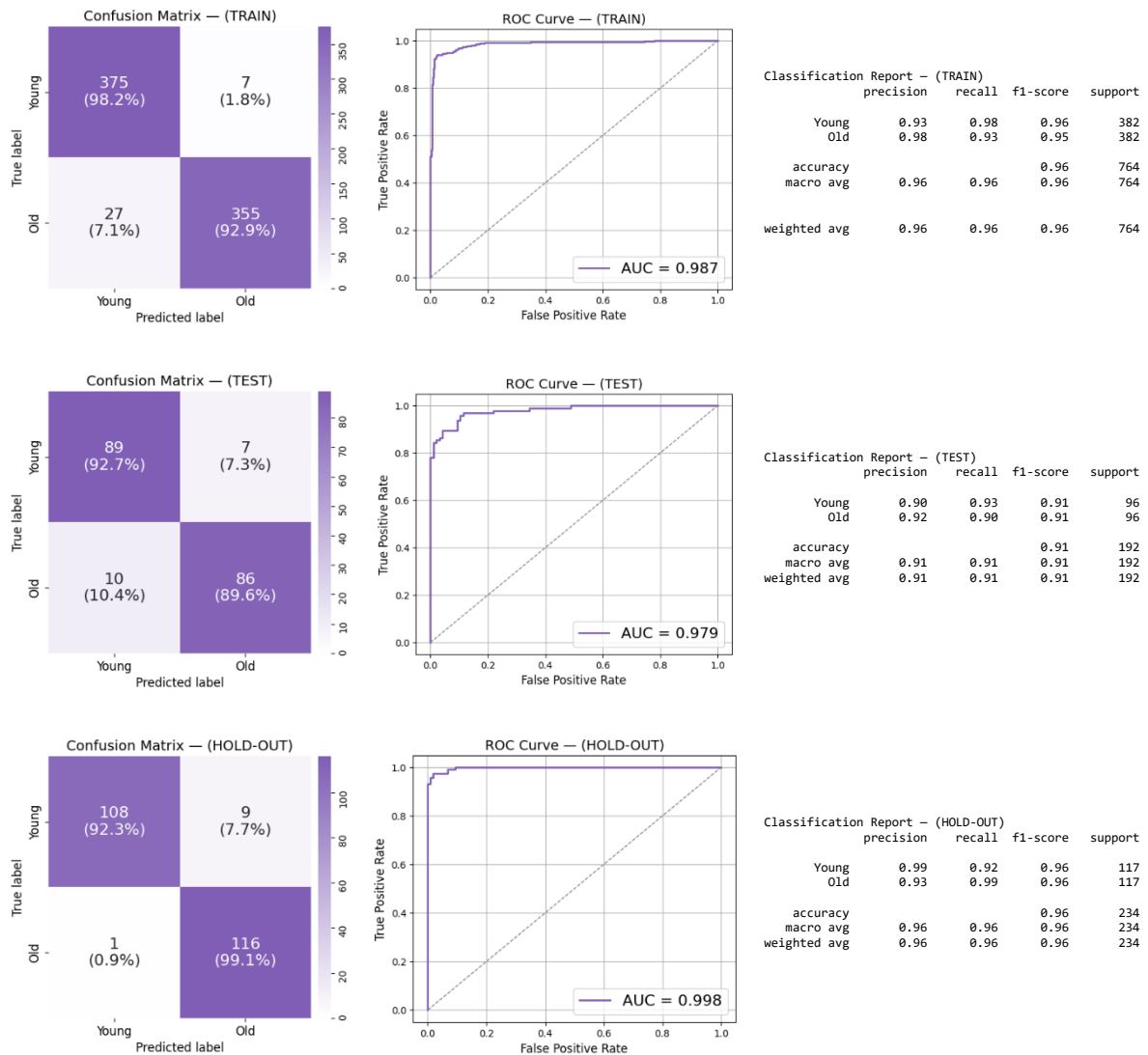**LOGISTIC REGRESSION**: Young vs. Old

Figure 28 *· Logistic Regression model performance (Aged vs. Young classification).*
Confusion matrices, ROC curves, and classification reports for the LR model trained with Bayesian-optimized hyperparameters. Results are shown for the training, test, and hold-out sets.

12.6 ANNEX: SEX CLASSIFICATION INABILITY

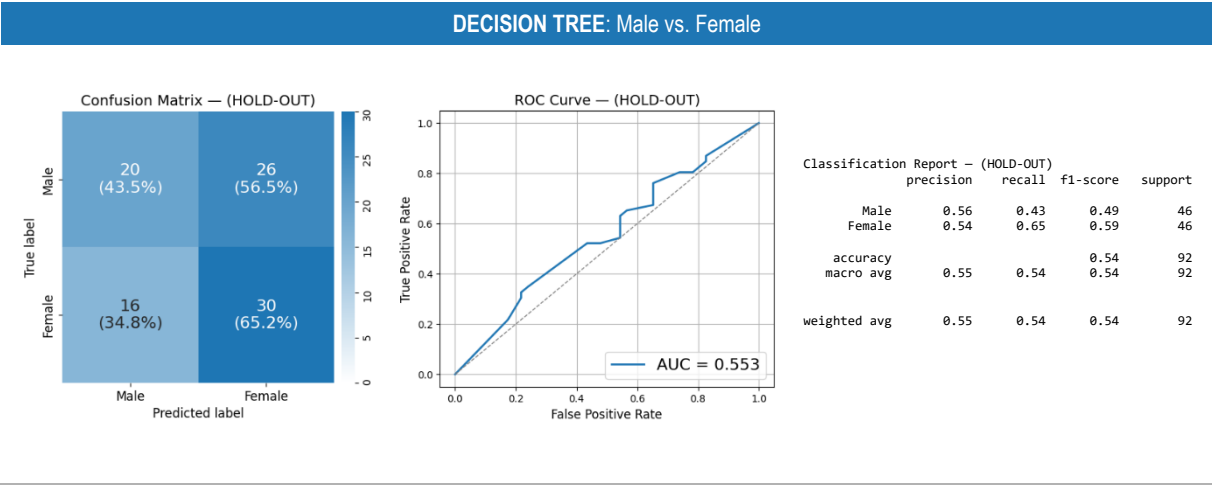**DECISION TREE**: Male vs. Female

Figure 29 · Decision Tree model performance (Male vs. Female classification).
Confusion matrices, ROC curves, and classification reports for the DT model trained to distinguish male from female fibroblasts using the same features extracted for age classification. Results are shown for the test and hold-out sets.

K-NN: Male vs. Female

Figure 30 · k-Nearest Neighbors model performance (Male vs. Female classification).
Confusion matrices, ROC curves, and classification reports for the k-NN model trained to distinguish male from female fibroblasts using the same features extracted for age classification. Results are shown for the test and hold-out sets.

RANDOM FOREST: Male vs. Female

Figure 31 · *Random Forest model performance (Male vs. Female classification).*
Confusion matrices, ROC curves, and classification reports for the RF model trained to distinguish male from female fibroblasts using the same features extracted for age classification. Results are shown for the test and hold-out sets.
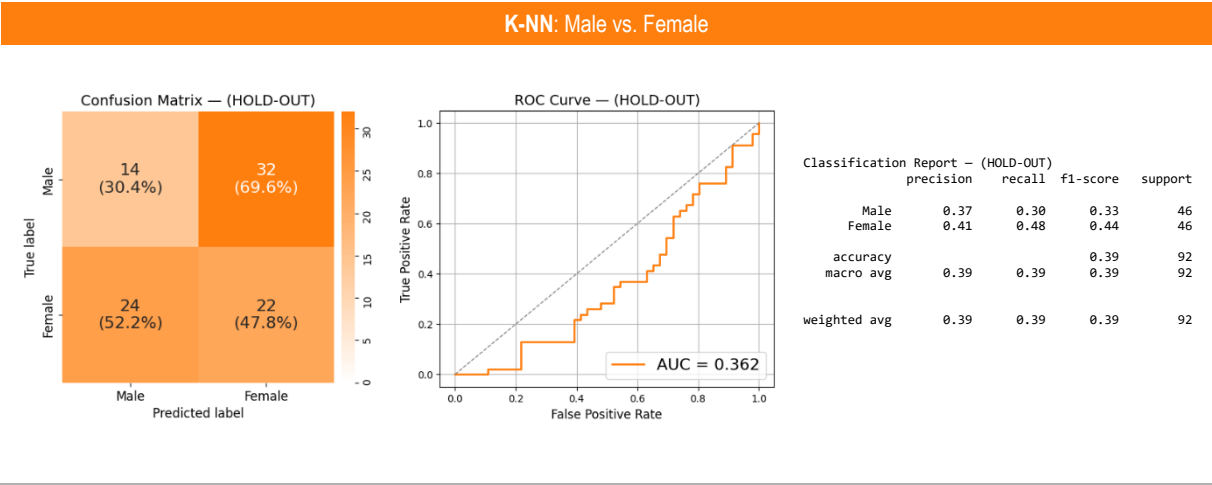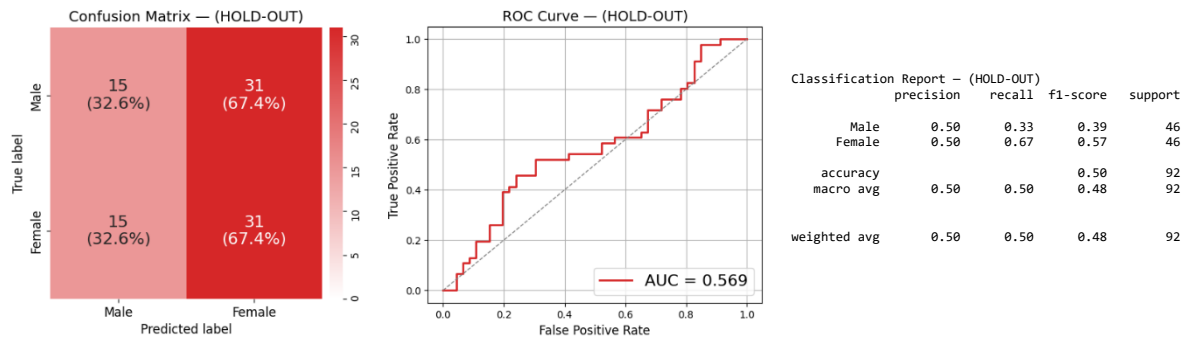
SVM: Male vs. Female

Figure 32 · *Support Vector Machine model performance (Male vs. Female classification).*
Confusion matrices, ROC curves, and classification reports for the SVM model trained to distinguish male from female fibroblasts using the same features extracted for age classification. Results are shown for the test and hold-out sets.
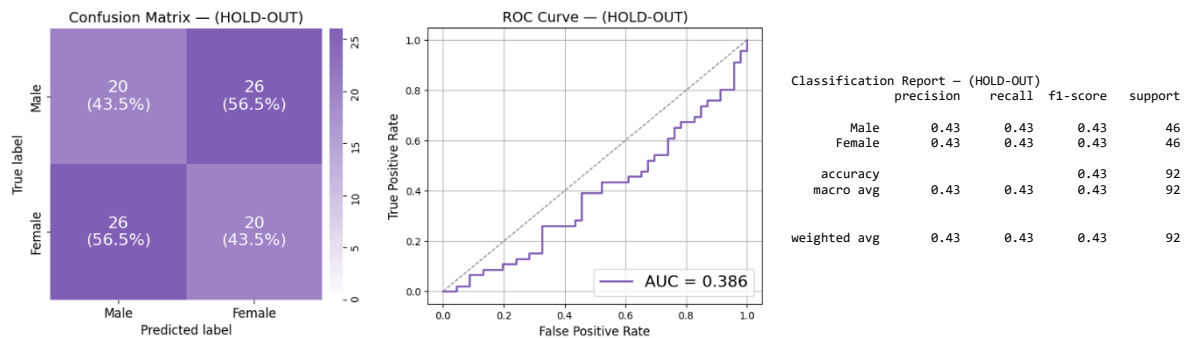
Figure 33 · *Logistic Regression model performance (Male vs. Female classification).*
Confusion matrices, ROC curves, and classification reports for the LR model trained to distinguish male from female fibroblasts using the same features extracted for age classification. Results are shown for the test and hold-out sets.

12.7 ANNEX: BOXPLOTS
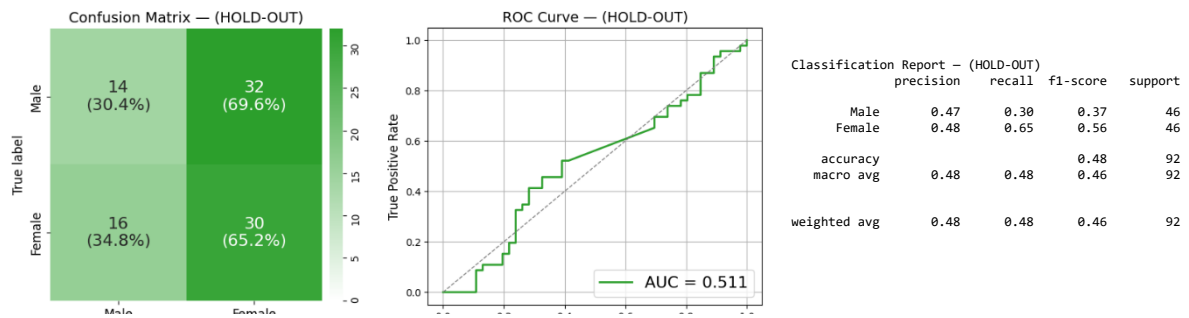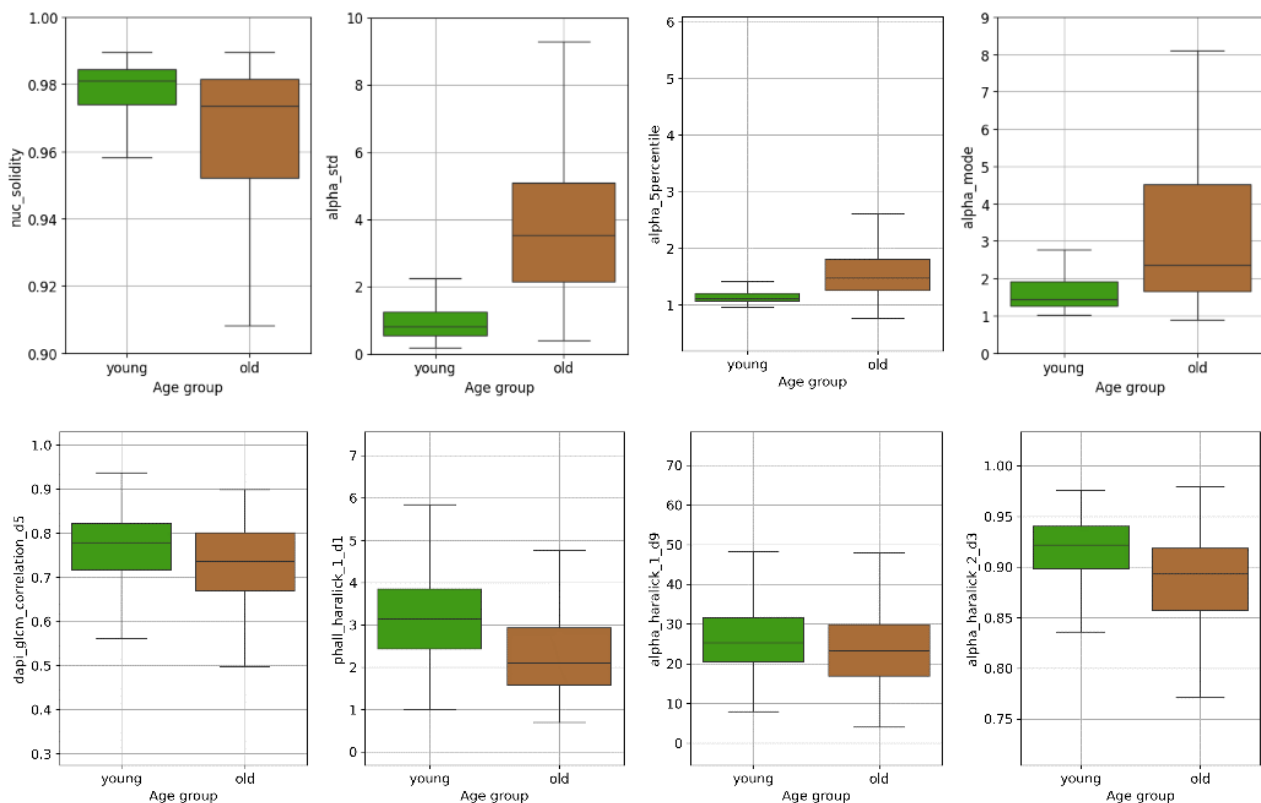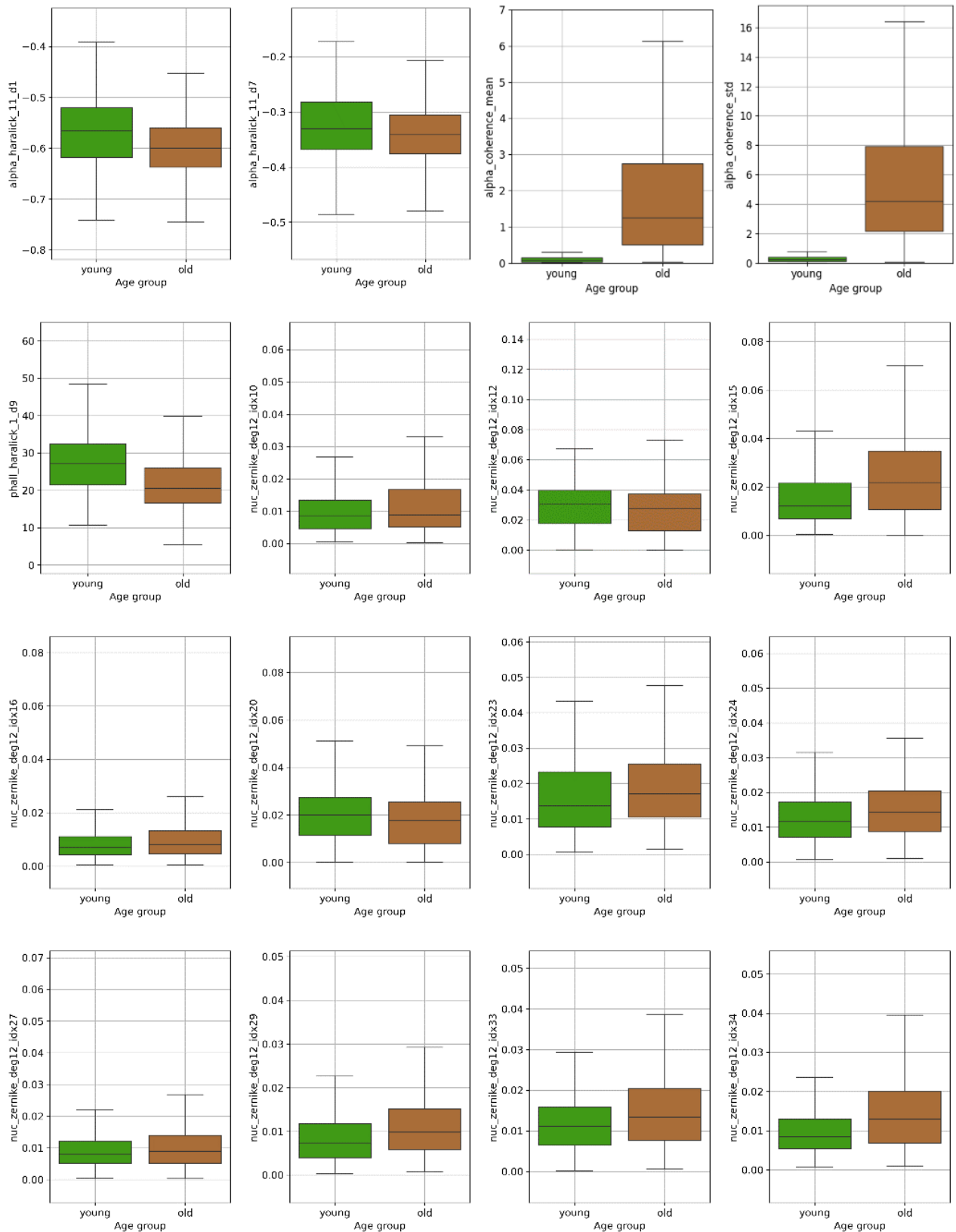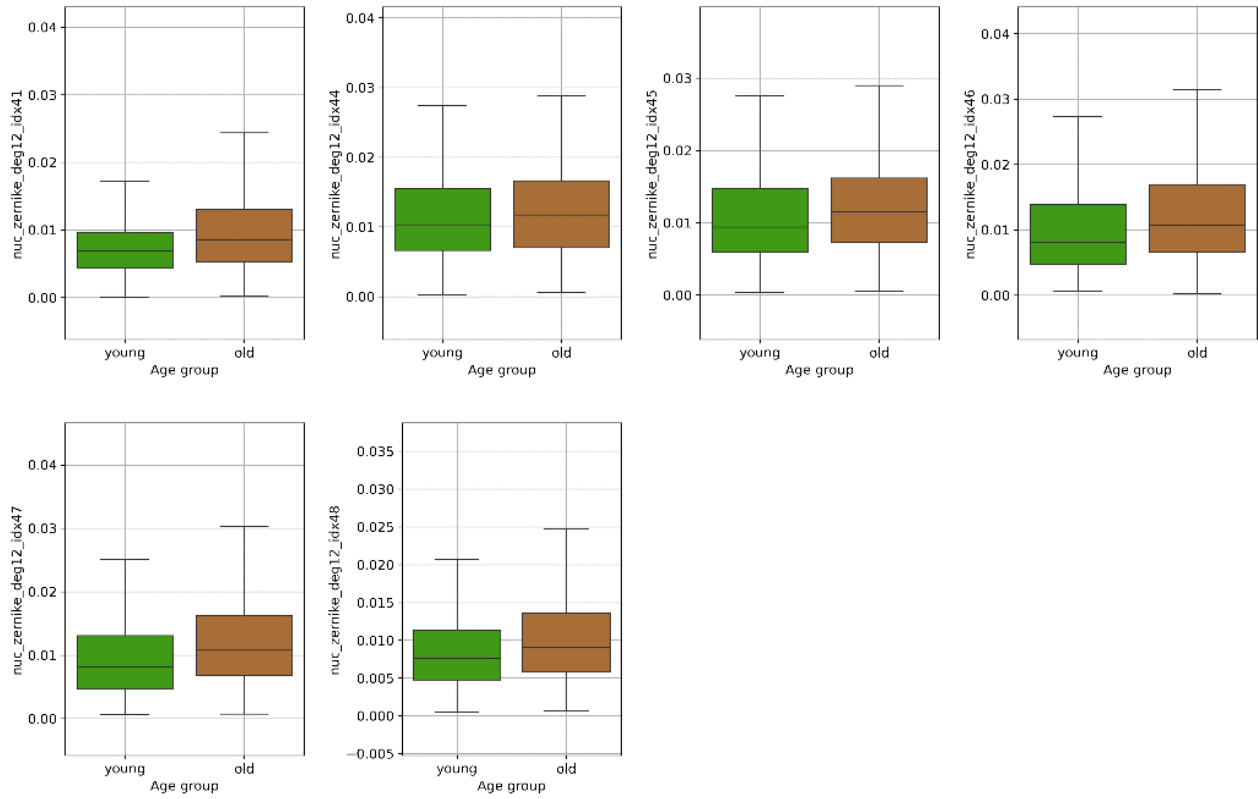
Figure 34 · *Boxplots of selected features after statistical and correlation-based filtering.*
Distribution of all variables that passed the initial statistical filtering step, including the correlation threshold criterion. Each boxplot illustrates how the selected features vary between young and old fibroblasts, highlighting the structural and morphological differences retained for subsequent model training and interpretation.

12.8 ANNEX: WBS DICTIONARY

| 1 | PROJECT COORDINATION	
1.1 | PROJECT BACKGROUND AND FRAMING	
DESCRIPTION	Define the scientific and methodological context of the project through a structured literature review, identification of knowledge gaps, and the conceptual planning of the project's aims, approach, and structure.
ACCEPTANCE CRITERIA	Clear conceptual framework established and documented; curated and thematically organized reference collection prepared.
DERIVABLES	An annotated Mendeley reference library, a written project outline, and defined scope and objectives.
REQUIRED RESOURCES	Internet access, scientific databases, and journal access through CRAI UB.
DURATION	20 days
RISKS	Incomplete contextualization leading to weak alignment between background and objectives.
1.2 | PROJECT PLANNING	
DESCRIPTION	Define the detailed execution plan of the project, including the scheduling of tasks, estimation of durations, identification of dependencies, and structuring of the Work Breakdown Structure (WBS), Gantt chart, and milestones.
ACCEPTANCE CRITERIA	Project schedule fully defined and approved.
DERIVABLES	Work Breakdown Structure (WBS), Gantt chart, PERT chart, phases, and milestones table.
REQUIRED RESOURCES	Project planning software or templates, calendar, and supervisor input.
DURATION	5 days
RISKS	Unrealistic duration estimates, overlooked task dependencies, or planning misalignment with actual workload.
1.3 | EXPERIMENTAL WORK LOGGING	
DESCRIPTION	Maintain a daily record of lab activities and observations.
ACCEPTANCE CRITERIA	Log updated consistently with clear, dated entries, including issues or deviations from the planned protocol.
DERIVABLES	Daily lab activity log.
REQUIRED RESOURCES	Notebook.
DURATION	47 days
RISKS	Incomplete entries, forgotten observations.
1.4 | FINAL REPORT WRITING	
DESCRIPTION	Write, edit, and structure the final degree project report, integrating all project phases into a cohesive document.
ACCEPTANCE CRITERIA	Complete, properly formatted report reviewed by supervisor and ready for submission.
DERIVABLES	Final written report.
REQUIRED RESOURCES	Computer, Microsoft Word access provided by UB, and supervisor feedback.
DURATION	86 days
RISKS	Delays are due to overlapping tasks or limited time for revisions.
2 | EXPERIMENTAL SETUP	
2.1 | FIBROBLAST ISOLATION AND CULTURE	
2.1.1 | YOUNG DONORS	
DESCRIPTION	Isolate, culture, and seed primary fibroblasts from young mice lungs according to established lab protocols.
ACCEPTANCE CRITERIA	Viable cultures established and expanded; sufficient cells obtained to perform a 3000 cells/m² seeding for each donor.
DERIVABLES	Seeded fibroblast samples from young donors, ready for immunostaining.

REQUIRED RESOURCES	Laminar flow hood, reagents (detailed in the Economic Viability Plan), young donor mice, and a licensed animal technician accredited to perform procedures involving live animals, in compliance with institutional and legal regulations.
DURATION	15 days
RISKS	Low cell yield, culture contamination.

2.1.2 | OLD DONORS

DESCRIPTION	Isolate, culture, and seed primary fibroblasts from old mice lungs according to established lab protocols.
ACCEPTANCE CRITERIA	Viable cultures established and expanded; sufficient cells obtained to perform a 3000 cells/m² seeding for each donor.
DERIVABLES	Seeded fibroblast samples from old donors, ready for immunostaining.
REQUIRED RESOURCES	Laminar flow hood, reagents (detailed in the Economic Viability Plan), old donor mice, and a licensed animal technician accredited to perform procedures involving live animals, in compliance with institutional and legal regulations.
DURATION	15 days
RISKS	Low cell yield, culture contamination.

2.2 | LABORATORY TRAINING

DESCRIPTION	Complete hands-on training in cell culture, immunostaining, and fluorescence microscopy protocols.
ACCEPTANCE CRITERIA	Key lab techniques are demonstrated and practiced under supervision.
DERIVABLES	Training completion confirmation.
REQUIRED RESOURCES	Supervisor, laminar flow hood, fluorescence microscope room, training materials.
DURATION	7 days
RISKS	Limited access or limited time for practice.

2.3 | MATERIALS AND METHODOLOGY PLANNING

DESCRIPTION	Prepare for lab work by reviewing protocols, organizing materials, and outlining experimental steps in advance.
ACCEPTANCE CRITERIA	Experimental plan and material checklist reviewed and approved by supervisor.
DERIVABLES	Materials checklist and methods plan.
REQUIRED RESOURCES	Access to lab protocols.
DURATION	1 day
RISKS	Missed items or procedural misunderstandings due to rushed preparation.

3 | DATA COLLECTION

3.1 | IMMUNOFLUORESENCE ASSAY

3.1.1 | YOUNG DONORS

DESCRIPTION	Stain seeded fibroblast samples from young donors using immunofluorescence to visualize cytoskeletal structures.
ACCEPTANCE CRITERIA	Staining is performed following established protocols.
DERIVABLES	Immunolabeled fibroblast samples (young donors).
REQUIRED RESOURCES	Reagents (detailed in the Economic Viability Plan)
DURATION	2 days
RISKS	Cell detachment, uneven or weak staining.

3.1.2 | OLD DONORS

DESCRIPTION	Stain seeded fibroblast samples from old donors using immunofluorescence to visualize cytoskeletal structures.
ACCEPTANCE CRITERIA	Staining is performed following established protocols.

DERIVABLES	Immunolabeled fibroblast samples (old donors).
REQUIRED RESOURCES	Reagents (detailed in the Economic Viability Plan)
DURATION	2 days
RISKS	Cell detachment, uneven or weak staining.

3.2 | IMAGE ACQUISITION

3.2.1 | YOUNG DONORS

DESCRIPTION	Capture high-resolution fluorescent images of immunolabeled fibroblasts from young donors using an epifluorescence microscope.
ACCEPTANCE CRITERIA	At least 150 valid images acquired per condition with appropriate focus, exposure, and staining visibility.
DERIVABLES	Fluorescent image dataset (young donors).
REQUIRED RESOURCES	Epifluorescence microscope, imaging software, data storage system.
DURATION	28 days
RISKS	Low image quality, inconsistent acquisition settings, and equipment malfunction.

3.2.2 | OLD DONORS

DESCRIPTION	Capture high-resolution fluorescent images of immunolabeled fibroblasts from young donors using an epifluorescence microscope.
ACCEPTANCE CRITERIA	At least 150 valid images acquired per condition with appropriate focus, exposure, and staining visibility.
DERIVABLES	Fluorescent image dataset (old donors).
REQUIRED RESOURCES	Epifluorescence microscope, imaging software, data storage system.
DURATION	28 days
RISKS	Low image quality, inconsistent acquisition settings, and equipment malfunction.

4 | FEATURE EXTRACTION

4.1 | IMAGE PREPROCESSING AND SEGMENTATION

4.1.1 | PROTOCOL DEFINITION

DESCRIPTION	Design a reproducible workflow for preprocessing and segmenting fluorescence images using open-source tools.
ACCEPTANCE CRITERIA	The protocol reliably produces interpretable masks across various sample conditions.
DERIVABLES	Segmentation protocol.
REQUIRED RESOURCES	Computer, open-source software (e.g., ImageJ, Ilastik, CellProfiler), and development time.
DURATION	10 days
RISKS	Poor generalization of the protocol may require extensive tuning from image to image, reducing reproducibility.

4.1.2 | PROTOCOL EXECUTION

DESCRIPTION	Apply the defined preprocessing and segmentation pipeline to all acquired images from both donor groups.
ACCEPTANCE CRITERIA	Complete and consistent segmentation output; masks align correctly with the original images.
DERIVABLES	Segmented image dataset.
REQUIRED RESOURCES	Computer, batch processing time.
DURATION	20 days
RISKS	Batch failure, incomplete segmentation, and reliance on semi-automated steps.

4.2 | IMAGE FEATURE EXTRACTION

4.2.1 | PROTOCOL DEVELOPMENT

DESCRIPTION	Design a protocol to extract quantitative features from segmented images.

ACCEPTANCE CRITERIA	The protocol generates structured feature tables that are compatible with downstream analysis.
DERIVABLES	Feature extraction protocol.
REQUIRED RESOURCES	Computer, coding environment, processing time.
DURATION	6 days
RISKS	Using features derived from the same family may lead to irrelevant or redundant data in subsequent analysis steps.

4.2.2 | PROTOCOL EXECUTION	
DESCRIPTION	Apply the developed feature extraction protocol to all segmented images to generate quantitative descriptors.
ACCEPTANCE CRITERIA	Feature matrix successfully generated across all experimental conditions.
DERIVABLES	Complete feature dataset.
REQUIRED RESOURCES	Computer, feature extraction protocol, segmented image files.
DURATION	1 day
RISKS	Missing values (NaNs) and low variability between conditions may affect analysis quality.

5 | DATA ANALYSIS	

5.1 | DIMENSIONALITY REDUCTION AND STATISTICAL ANALYSIS	
DESCRIPTION	Analyze extracted feature data using statistical techniques, with a primary focus on dimensionality reduction to improve interpretability and model performance.
ACCEPTANCE CRITERIA	Dimensionality is successfully reduced while preserving relevant variance and structure.
DERIVABLES	Reduced feature dataset, statistical plots, or summary metrics.
REQUIRED RESOURCES	Computer, R coding environment, relevant libraries.
DURATION	20 days
RISKS	Over-reduction may lead to the loss of meaningful variation; the selection of inappropriate methods could distort the data structure.

5.2 | MACHINE LEARNING PIPELINE IMPLEMENTATION	
DESCRIPTION	Train, test, and validate a classification model using the extracted features to distinguish between experimental conditions.
ACCEPTANCE CRITERIA	The model achieves predefined performance metrics on both the test and validation sets.
DERIVABLES	Trained machine learning model; performance metrics summary.
REQUIRED RESOURCES	Computer, Python environment, scikit-learn library, computing time.
DURATION	15 days
RISKS	Overfitting due to a small dataset can lead to inconsistent accuracy across different data splits.

5.3 | RESULTS INTERPRETATION	
DESCRIPTION	Review statistical outputs and model results to extract biologically meaningful insights related to fibroblast aging.
ACCEPTANCE CRITERIA	Interpretations are logically consistent with the data and supported by relevant literature.
DERIVABLES	Summary of key findings and biological interpretation.
REQUIRED RESOURCES	Analysis results, visualizations, and access to scientific literature provided via CRAI UB.
DURATION	14 days
RISKS	Misidentifying experimental artifacts as biologically meaningful patterns could lead to flawed conclusions.

Table 17 · *Work Breakdown Structure (WBS) dictionary.*
Provides detailed descriptions of each task defined in the Work Breakdown Structure. It includes information on objectives, acceptance criteria, expected deliverables, required resources, estimated durations, and associated risks.