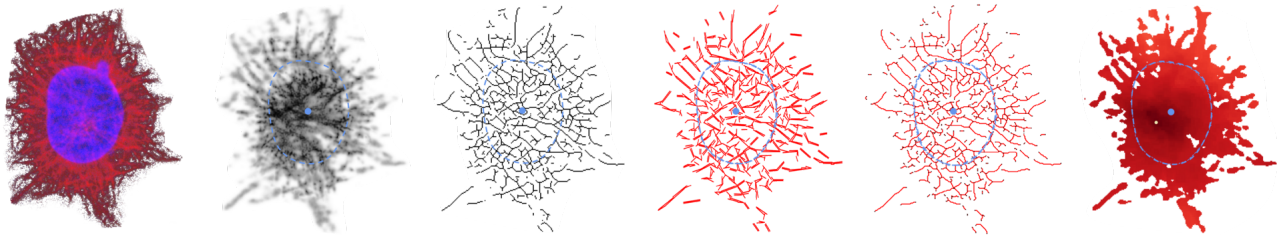
TÉCNICO LISBOA



Unraveling the cytoskeletal architecture of cancer cells: development of a novel computational approach

Diogo João Fróis Vieira

Thesis to obtain the Master of Science Degree in

Biomedical Engineering

Supervisors: Prof. João Miguel Raposo Sanches
Dr.ª Maria Sofia Fernandes

Examination Committee

Chairperson: Prof.ª Maria Margarida Fonseca Rodrigues Diogo
Supervisor: Prof. João Miguel Raposo Sanches
Member of the Committee: Prof.ª Maria Margarida Campos da Silveira

November 2022

I declare that this document is an original work of my own authorship and that it fulfills all the requirements of the Code of Conduct and Good Practices of the Universidade of Lisboa.

Preface

The work presented in this thesis was performed at Instituto de Sistemas e Robótica (ISR), Instituto Superior Técnico (Lisboa, Portugal), during the period October, 2021 to November, 2022, under the supervision of Prof. João Sanches at Instituto Superior Técnico. The thesis was co-supervised by Maria Sofia Fernandes at Instituto de Patologia e Imunologia Molecular da Universidade do Porto (IPATIMUP) / Instituto de Investigação e Inovação em Saúde (i3S). The author of this thesis wants to acknowledge IPATIMUP/i3S for the biological support of this research and providing the cell line and dataset analyzed in this work.

Acknowledgments

Objetivo cumprido. Vou relembrar este trabalho e toda a minha jornada académica como um período de muito empenho, dedicação e paixão. No entanto, foi marcado pelo sacrifício de valores que considero essenciais, com o objetivo de concluir ambos com distinção como sempre desejei.

Sacrifícios. Mesmo assim, houve quem nunca tivesse desistido de mim. Neste parágrafo, quero deixar os mais profundos e sentidos agradecimentos aos que realmente importam. À Mãe, ao Pai, à Inês, à Sofia, às avós Mina e Lina, aos avôs José e Manuel, à Titia, mas também ao Sal e ao Pimenta. Palavras não são suficientes para mostrar o quão grato estou. Antes de qualquer esforço, dedicação e trabalho meu, vem o vosso AMOR E APOIO INCANSÁVEL, a base da razão de estar onde estou agora. Sem vocês, concluir esta dissertação e o curso seria um cenário reservado para os meus sonhos. Hoje, é uma realidade. MUITO OBRIGADO.

Sem laços de sangue, mas também preciosos e insubstituíveis, os meus amigos. Em primeiro lugar, gostaria de agradecer ao Miguel e à Maggy, amigos com quem já não consigo viver sem. Ao Diogo, futuro e brilhante matemático, um agradecimento especial pela ajuda no formalismo matemático

deste trabalho. Aos murtalenses e amigos de longa data, Vasco e Lúcia, um obrigado está longe de ser suficiente para agradecer pelo vosso apoio! À Vera, das pessoas que melhor compreendeu o sofrimento que foi esta jornada, um sincero obrigado à melhor companhia das manhãs!

Aos biomédicos que me acompanharam neste percurso desde o dia 1 e que agora também terminarão os deles brevemente, muito obrigado! Um agradecimento especial em particular ao Marrafa, o meu braço direito e futuro business partner. À Carolina, aloco o mais carinhoso agradecimento pelo apoio diário indispensável e determinante para conseguir chegar ao fim!

Se perguntarem por segredos para o sucesso, arranjem amigos arquitetos como a Bia, a Raminhos, a Maria e o Hugo. Também eles acompanharam a reta final deste percurso e sempre se preocuparam comigo e com a minha dose mensal de croissants, muito obrigado! Ao Gonçalo, ao Sebastião e à Lorrane, não tarda vamos brindar! Muito obrigado pelo vosso apoio e compreensão!

Por fim, um agradecimento muito especial aos que acompanharam de perto o desenvolvimento deste trabalho. À Raquel, que em pouco tempo muito me ensinou e lamento por não o poder fazer mais. Ao professor João e à Sofia, um sincero agradecimento pela disponibilidade, pelos preciosos conselhos e por acreditarem no meu potencial desde o início. A conclusão deste trabalho é uma vitória nossa!

Abstract

Cancer remains a major health problem and one of the leading causes of mortality worldwide. This is mostly due to the development of metastatic disease and the lack of powerful tools to detect invasive cancer cells at an early stage. Therefore, there is an urgent need to unravel new strategies to identify invasive cancer cells. It is well established that the structure and organization of the cytoskeleton are dynamically orchestrated during many cellular processes, including cancer invasion. Thus, in this study, a novel computational approach was developed to characterize the cytoskeletal architecture of cancer cells. Immunofluorescence images were used to devise a pipeline to characterize the structural pattern of α-tubulin, a cytoskeleton component, in cells with E-cadherin mutations leading to loss of cell-cell adhesion and a potential invasive phenotype. Microtubule organization features were evaluated, including morphology, orientation, compactness, and radiality. The strategy involved a preprocessing step for image enhancement followed by feature extraction (processing) of manually segmented microtubules based on three methods: analysis of grayscale pixel intensity distribution, 2D spatial rearrangement of automatically-detected line segments/microtubules and graph theory features. Results of the proposed method applied to cells with E-cadherin mutations have shown that the microtubules of cells with disrupted E-cadherin are shorter, have uniform length patterns and are more compactly distributed as compared with cells with wild-type E-cadherin. This study reveals that cytoskeletal features could provide an efficient strategy to identify cells with invasive potential, ultimately impacting cancer diagnosis and prognosis.

Keywords

Cytoskeleton; Image segmentation; E-Cadherin; Microtubules; Architectural features; Feature extraction

Resumo

O cancro continua a ser um grande problema de saúde e uma das principais causas de mortalidade a nível mundial. Isto deve-se principalmente ao desenvolvimento de doenças metastáticas e à falta de instrumentos capazes de detectar células cancerígenas invasivas numa fase precoce. Por conseguinte, há uma necessidade urgente de desvendar novas estratégias para identificar células cancerígenas invasivas. Está bem estabelecido que a estrutura e organização do citoesqueleto é dinamicamente orquestrada durante muitos processos celulares, incluindo a invasão do cancro. Assim, neste estudo, foi desenvolvida uma nova abordagem computacional para investigar a arquitetura do citoesqueleto de células cancerígenas. Imagens de imunofluorescência foram utilizadas para conceber uma metodologia para caracterizar o padrão estrutural da α-tubulina, uma componente do citoesqueleto, em células com mutações na E-caderina, levando à perda de adesão celular e a um fenótipo potencialmente invasivo. Foram avaliadas características de organização dos microtúbulos, incluindo morfologia, orientação, compactação, radialidade. A estratégia envolveu uma etapa de pré-processamento para melhoramento da imagem seguida de extracção (processamento) de microtúbulos segmentados manualmente com base em três métodos: análise da distribuição de intensidade de pixels em escala de cinzentos, rearranjo espacial 2D de segmentos de linha/microtúbulos detectados automaticamente e características da teoria de grafos. Os resultados mostraram que os microtúbulos de células com E-caderina mutada são mais curtos, têm padrões uniformes de comprimento e são distribuídos de forma mais compacta em comparação com células com E-caderina *wild type*. Este estudo revela que as características do citoesqueleto podem fornecer uma estratégia eficiente para identificar células com potencial invasivo, com possível impacto no diagnóstico e prognóstico do cancro.

Palavras Chave

Citoesqueleto; Segmentação de imagem; E-caderina; Microtúbulos; Características arquitetónicas; Extração de características

Contents

List of Figures

List of Tables

Acronyms

AAI	Average Actin-cytoskeletal Intensity
AMI	Average Microtubule Intensity
CHO	Chinese Hamster Ovary
CNF	Cytoskeleton Network Features
CV	Coefficient of Variation
$CVar$	Circular Variance
DCF	Deconvoluted Cell Features
ECM	Extracellular Matrix
EMT	Epithelial-Mesenchymal Transition
GLCM	Gray-Level Co-occurrence Matrix
LFT	Line Filter Transform
LSD	Line Segment Detector
LSF	Line Segment Features
MIP	Maximum Intensity Projection
MTOC	Microtubule-Organizing Center
NP	Non-Pathogenic
OFT	Orientation Filter Transform
OOP	Orientational Order Parameter
P1	Pathogenic 1
P2	Pathogenic 2
PAD	Partial Actin-cytoskeletal Deviation
ROI	Region Of Interest

TAD Total Actin-cytoskeletal Deviation

WT Wild Type

1

Introduction

Contents

In the 1950s, when the role of the cytoskeleton in cells was not well understood, Crick, F. and Hughes, A. [21] proposed a model for this fundamental intracellular structure - the Mother's Work Basket, "a jumble of beads and buttons of all shapes and sizes, with pins and threads for good measure, all jostling about and held together by colloidal forces". Current knowledge is enough to decipher this metaphor better. The cell is a dynamic and hierarchically ordered system of molecules (beads and buttons) and supramolecular entities (threads). Beads, buttons, pins, and threads... life depends on their spatial and temporal interactions [22]. The cytoskeleton is a complex but delicate 3D weaving piece made of these fabrics, i.e., filamentous cytoskeletal proteins, motor proteins, and their regulatory molecules, that perform a well-orchestrated plethora of crucial functions. This network occupies the whole cell volume and plays pivotal roles in controlling fundamental cellular processes involving cell mechanics, cell division, cell shape, intracellular transport, motility, and adhesion [1].

1.1 Motivation

It is no wonder that when such a ubiquitous structure like the cytoskeleton is destabilized, it may cause impacting damage in several cellular processes in a way that turns the cell into a cancer cell. Cancer is a significant health problem and one of the leading causes of death worldwide [23]. Nowadays, from engineering and biology perspectives, key knowledge has been acquired to study the cytoskeleton in cancer cells [24]. It is well established that the cytoskeleton is subject to a dramatic reorganization program in cancer cells [25, 26]. However, little is known about which cytoskeleton properties are changed, nor about how these changes are quantified. Medical imaging and computational processing of the extracted data have also suffered remarkable advances, but little effort has been put into cancer progression prevention via image processing of the cytoskeleton. Consequently, qualitative and quantitative descriptors are needed to unpeel what "organization" really means when applied to cytoskeletal networks, aiming to understand from which points of view this organization differs between cytoskeletal structures of non-invasive vs. invasive cancer cells.

Understanding and quantifying the structure and organization of the cytoskeleton can help to provide key insights into various biological processes. In particular, cytoskeleton organization of invading and non-invading cancer cells, whose knowledge could be used for cancer prevention and diagnosis.

1.2 Objective and Original Contributions

The overall objective of this work was to perform a qualitative and quantitative description of the cytoskeletal organization, to better understand and quantify microtubule structural differences between non-invasive and invasive cancer cells, in order to predict invasive potential. The proposed approach

relies on a pipeline that includes the (pre)processing of fluorescence microscopy images of cancer cells expressing Wild Type (WT) E-cadherin or mutant E-cadherin. The disruption of E-cadherin, as a result of mutations, compromises the adhesion of epithelial cells, this phenomenon being associated with the progression of aggressive cancers [27–29]. In this study, distinct E-cadherin missense mutations were evaluated, including a non-pathogenic mutation (Non-Pathogenic (NP)) and two pathogenic mutations with increased invasive potential (Pathogenic 1 (P1); Pathogenic 2 (P2)). Importantly, cells were grown in laminin, considering the Extracellular Matrix (ECM) interaction. Upon transient transfection of vectors encoding WT or mutant E-cadherin, cells were labeled for α-tubulin, a main component of the cytoskeleton. α-tubulin features were then investigated to better understand how modifications in the cytoskeletal architecture cooperate to mediate cancer cell invasion. Furthermore, nuclei were also investigated for morphological alterations, as these are important for cancer invasion [7]. For this purpose, image preprocessing was used to enhance and highlight cytoskeletal structures. Individual cell processing involved the extraction of 1D and 2D features from the cytoskeleton highlighted after preprocessing. Single cell processing comprised a feature extraction methodology of three different natures: by computing morphological-, intensity-, textural- and frequency-based features on deconvoluted versions of cytoskeletal structures (Deconvoluted Cell Features (DCF)) and by obtaining line segment- (Line Segment Features (LSF)) and graph-based (Cytoskeleton Network Features (CNF)) features from skeletonized versions after automatic line detection and graph conversion, respectively. The features calculated by these three methods were correlated with ten aspects of the cytoskeleton organization - *orientation*, *bundling*, *parallelism*, *morphology*, *compactness*, *connectivity*, *quantity*, *complexity*, *radiality*, and *cytoskeleton-nucleus interaction* -, allowing a complete characterization of this structure. Overall, this work aimed to comprehensively analyze the cytoskeleton complex network at the individual cell level and globally. Notably, although the proposed framework was developed to characterize α-tubulin microtubules and predict invasive potential with enormous implications in cancer diagnosis and prognosis, this strategy can be applied to other cellular processes. Moreover, it can also be applied to other types of cytoskeletal proteins including actin and intermediate filaments, but also to other morphologically similar structures, such as retinal blood vessels or neuronal networks with impact in other diseases.

1.3 Thesis Outline

The work presented in this Thesis aims to help unveil how cytoskeletal features can be evaluated taking advantage of a novel computational framework. The outline of this document is as follows: in Chapter 2, the introductory chapter, a brief description of the biological background is presented concerning the importance of the cytoskeleton in cellular processes namely cell invasion, the role of E-cadherin in cell-cell interactions and cancer, and the involvement of the nucleus in cancer progression. Chapter

3 presents a review of the state-of-the-art computational methods and engineering tools available to analyze cytoskeletal structures and their implications. At the end of Chapter 3, the aim of the study is introduced. In Chapter 4, the methodology is described in detail, from image acquisition and processing to the proposed approaches aiming to highlight the cytoskeletal network structure followed by an extensive feature selection and extraction. Chapter 5 presents the main results and discussion and Chapter 6 focus on concluding remarks and proposed future work.

2

Biological Background

Contents

In this Chapter, a biological overview is given in the context of the proposed work. First, the importance of the cytoskeleton in cancer progression is defined, followed by a biological description of this structure and of microtubular conformations. Next, the role of the cytoskeleton and the interplay with the adhesion molecule E-cadherin, a marker of cancer progression, is described. The importance of the nucleus in matrix invasion is also addressed.

2.1 The Cytoskeleton and Cancer

Cancer is one of the leading causes of death worldwide [23]. As of this writing, more than 5 million people have died from this major health problem in 2022 [30]. Cancer is a heterogeneous disease that arises from the transformation of normal cells into tumor cells in a multi-stage process [31]. From the primary tumor site, cells can acquire an inappropriate migratory and invasive phenotype and go beyond their usual boundaries to invade adjoining body parts and spread to other distant organs, creating metastases [23, 26]. Metastatic cancer remains the leading cause of cancer-related deaths [26].

Structural and regulatory proteins composing the cytoskeleton are essential for critical cellular processes of both normal and cancer cells. The cytoskeleton aids all kinds of cellular morphogenesis, such as extension, retraction, stabilization, and adhesion [32]. Different cytoskeletal proteins do not work in isolation but are inextricably linked together to promote migration (especially during the avascular phase of tumor growth), invasion, proliferation, and resistance to cellular and environmental stresses, promoting tumor development [33, 34]. During cancer progression, cells need to move and change their shape to pass through spatial cues and physical constraints within the target organ, requiring pronounced temporal and spatial reorganizations of the cytoskeleton by the (dis)assembly of its proteins [1, 34, 35].

Given its multifunctional role, cytoskeleton defects have been associated with many human diseases, including neurodegenerative disorders, muscular dystrophies, cardiac disorders, and cancer [36]. In particular to cancer, structural and regulatory proteins that compose the cytoskeleton have been important and effective targets of potential therapeutics against metastatic dissemination [26]. Despite the continuous improvements in therapeutic approaches targeting cytoskeletal components, they still remain largely unexplored [33].

The cytoskeleton is a complex fibrous regulatory and structural protein network [2, 26]. As shown in Figure 2.1, this structure is composed of polymers intricately linked together with well-orchestrated functions, forming a dynamical network pivotal for the control of fundamental cellular processes [26].

The cytoskeleton is mainly composed of three classes of fibers (Figure 2.2), that differ in their mechanical stiffness, their polarity, their (dis)assembly dynamics and the type of molecular motors with which they associate [35]: *microfilaments*, *intermediate filaments*, and *microtubules*. The plastic structure of the cytoskeleton provided by these polymers has a central role in sensing the mechanical cellu-

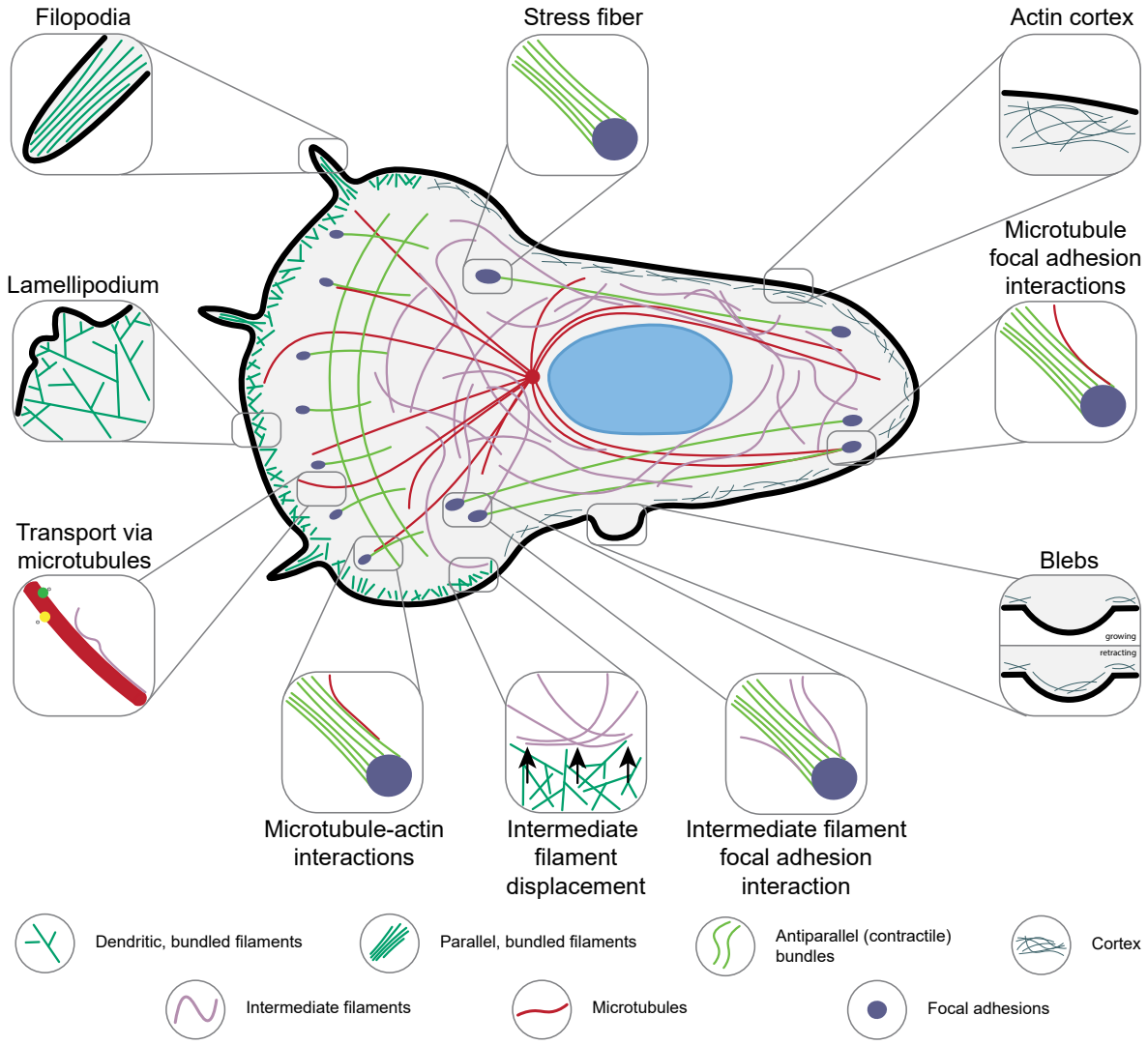


Figure 2.1: Proteins that form the cytoskeleton (actin, microtubules, and intermediate filaments) and their physical interactions form dynamic organizational structures inside the cell [1]. Figure adapted from [1].

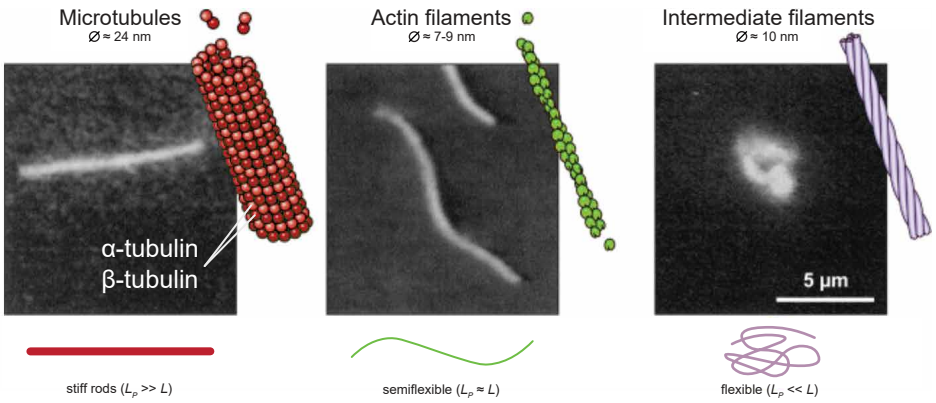


Figure 2.2: Types of fibers that compose the cytoskeleton. These fibers are long chains of multiple copies of one or more small protein subunits [2]. Different physicochemical properties provide different persistent length L_P compared to the total filament length L. Figure adapted from [3].

lar microenvironment, transmitting compressive/tensile stresses accordingly and distributing forces and fluctuations throughout the cell [34, 35]. These changes in the cell mechano-environment allow it to change its shape, resist deformation, respond to externally applied forces and maintain the integrity of intracellular compartments [1, 35, 37]. Mechanically, microtubules are compressive-resisting ”girders”, while microfilaments and intermediate filaments are tension-bearing elements [4].

In particular, *microtubules* are heterodimers of globular proteins α/β-tubulin that self-associate into long, hollow polymer tubes (Figure 2.2, left panel) [34]. They are the stiffest of all three cytoskeletal polymer types and have the most complex assembly/disassembly dynamics [35]. Their stiffness allows them to form approximately linear tracks that might even cross the cell from one end to the other. During interphase, they can form radial arrays that function as central hubs and ”highways” for intracellular trafficking (Figure 2.3 (a)) [35]. On the other hand, microtubules are rearranged to form the mitotic spindle during cell division, separating sister chromatids between daughter cells (Figure 2.3 (b)) [24, 34, 35]. Additionally, microtubules are important for directional migration (Figure 2.3 (c)) and action potential transmission in neurons (Figure 2.3 (d)) [2].

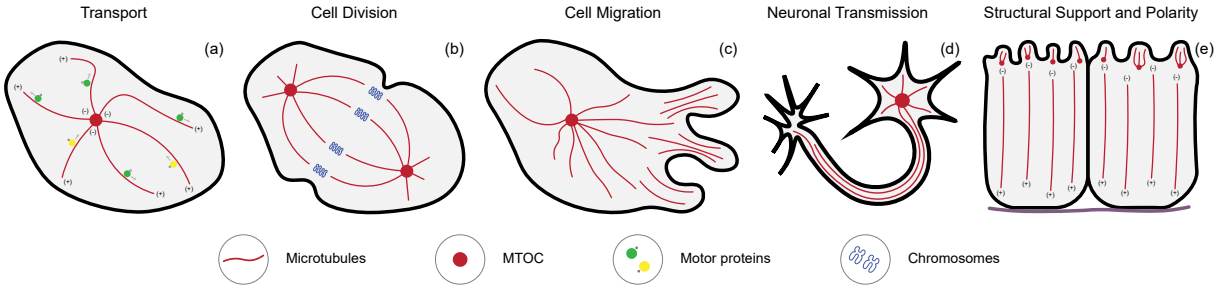


Figure 2.3: The role of microtubules in several important cellular phenomena. While (+) represents the positive polarity (fast-growing ends), (-) represents the negative polarity (slow-growing ends). Figure adapted from [2, 4].

The majority of microtubules nucleate from a Microtubule-Organizing Center (MTOC). A small intercellular organelle called centrosome is commonly considered the major MTOC [38]. Microtubules minus ends are tethered from the centrosome and form radial extensions in interphase animal cells [39, 40]. In nonpolarized cells, most microtubule minus-ends remain anchored to the centrosome next to the nucleus, resulting in a microtubule array with all plus-ends facing the cell cortex (Figure 2.3 (a) and (c)) [39]. Two centrosomes are the MTOCs that nucleate the microtubules of the mitotic spindle (Figure 2.3 (b)). Finally, in polarized cells such as epithelial cells, the centrosome is located below the lumenal domain [39]. In these cells, non-centrosomal microtubules are aligned parallel to each other along the apical ($(-)$ ends) to basal ($(+)$ ends) polarity axis of the cell [40]. This microtubular conformation is essential for secretion and absorption and serves as a barrier (Figure 2.3 (e)) [39, 40]. Basal bodies are the MTOCs that assemble the microtubules of cilia and flagella (Figure 2.3 (e)) [2].

Normal and neoplastic cells often show different patterns of tubulin isotype expression, post - trans-

lational modifications and associated modifying enzymes [41]. Consequences of these alterations affect specific cancer properties, including poor outcome/prognosis and metastatic abilities. Thus, cancer tubulin isotypes and post-translational modification signatures can potentially be used as biomarkers and for therapeutic development [24, 41].

2.2 E-cadherin in Cell-Cell Interactions and Cancer

Carcinomas are cancers arising from epithelial cells [2]. These cells are incorporated in epithelial tissues, sheet-like layers with cells bound tightly together to structurally and functionally cover all internal and external surfaces of organs [2, 5], as shown in Figure 2.4. Carcinomas are the most common cancers in humans, accounting for about 80% of cases, and can arise in any epithelial tissue in any part of the body [42]. Epithelia in different body locations have characteristic morphologies and functions; for instance, the renal epithelium is specialized in filtering, while the lung epithelial specializes in gas exchange [2].

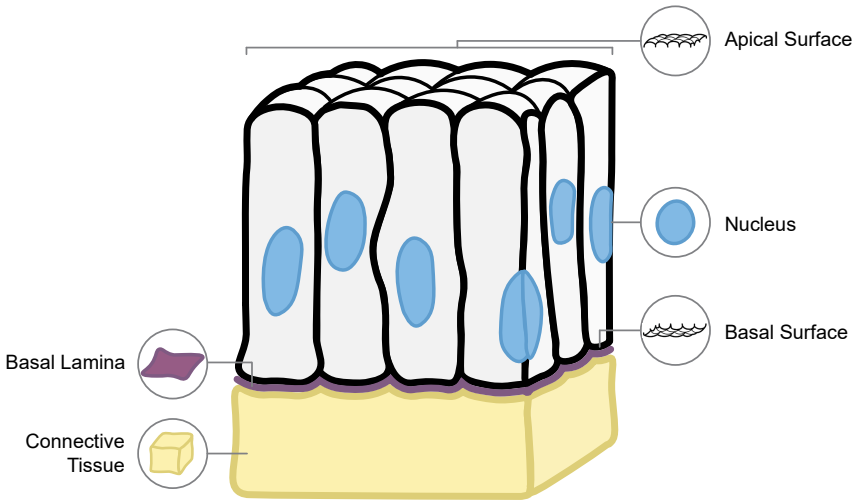


Figure 2.4: Representative image of epithelial tissue. The cell's plasma membrane can be divided into *apical* and *basolateral* domains. The apical side faces the sheet's free surface, while the basal and lateral (collectively, basolateral) surfaces embrace the organism's interior. In particular, the basal surface contacts an underlying specialized ECM, the *basal lamina*, and *connective tissue* [2, 5]. Figure adapted from [2].

The stability of epithelial tissue is guaranteed by firm cell-cell and cell-ECM interactions between adhesion molecules along the basolateral domain in clusters called cell junctions [2]. These connections tightly link adjacent cells and hold them together to the ECM [2].

Among other classes of cell junctions, adherens junctions anchor adjacent cells to each other in circumferential belt-like structures [42], as shown in Figure 2.5. Adherens junctions are protein complexes containing densely packed cadherin molecules that laterally interact with other cadherin domains

of adjacent cells. In particular, *E-cadherin*, or epithelial cadherin, is a significant component of adherens junctions in epithelia [42]. E-cadherin comprises *three* major structural domains (Figure 2.5), whose properties reflect the varied and vital functions this transmembrane protein fulfills: a large *extracellular domain*, a single *transmembrane segment* and a short *cytoplasmatic domain* [28, 43]. The extracellular domain establishes Ca^{2+}-dependent homotypic links with adjacent cells by forming zipper-like structures that contribute to cell-cell contact and adhesion [44]. The transmembrane segment connects the extracellular and intracellular segments. Lastly, inside the cell, the cytoplasmatic domain interacts with a set of proteins called catenins, which bridge E-cadherin and the actin cytoskeleton, essential for stable cell-cell contact, homeostasis and signal transduction [28, 45]. Importantly, outside the cell, through the homophilic association of E-cadherin molecules expressed in neighboring cells, they ensure intercellular adhesion amongst epithelial cells [46]. Inside the cell, the interaction with the catenin complex allows signal transduction and dynamic linkage with the cytoskeleton, coordinating the shape, polarity, and function of the cells in an epithelium [45].

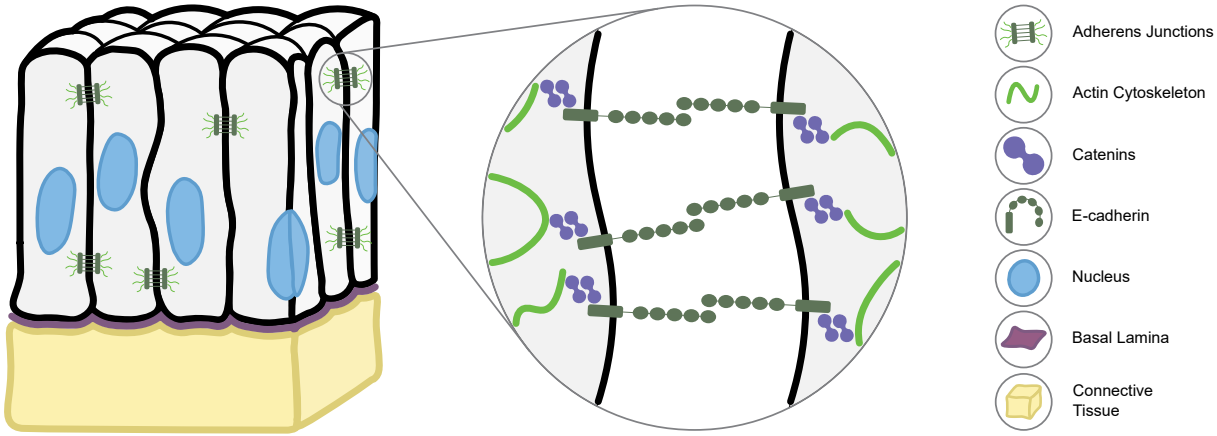


Figure 2.5: E-cadherins are major components of adherens junctions. These proteins are composed of *three* structural domains: An extracellular domain that binds to the extracellular domain of E-cadherin on adjacent cells, promoting cell-cell adhesion; a transmembrane domain; and a cytoplasmatic domain connected to the actin cytoskeleton through catenins. Figure adapted from [2].

A normal epithelium is characterized by stable intercellular junctions, apical-basal polarity and interactions with the basal lamina. However, in response to signals that cells receive from their microenvironment, cells can start losing their epithelial features and appearance [25]. E-cadherin is a prototypical marker whose expression loss is described as a key event in cancer progression [47, 48]. E-cadherin deregulations can be a consequence of several events, such as mutations in the *CDH1* gene, overexpression of transcription repressors, alterations of microRNAs, deregulation of protein trafficking, and aberrant post-translational mechanisms [28, 29, 45]. An abnormal function or loss of expression of E-cadherins compromises cell-cell adhesion through adherens junctions. Consequently, on a bigger scale, this leads to an aberrant tissue architecture [27]. Figure 2.6 depicts how abnormal functioning

E-cadherins can define the overall epithelial appearance [49]. In late stages, individual cells separate from one another, acquiring migratory capacities and increased invasive potential [25, 29, 48].

E-cadherin dysfunction results in unstable adherens junctions and affects the actin cytoskeleton through catenins [44]. Therefore, throughout the process of acquisition of an invading phenotype by a non-invading cell, the cytoskeleton plays a central role by suffering a dramatic reorganization. Even though adherens junctions link to the actin cytoskeleton, studies show that microtubules are anchored to adherens junctions and are required for its efficient assembly and to deliver factors that regulate their dynamics [50, 51]. Several mechanisms contribute to tethering microtubules to adherens junctions during epithelial differentiation and polarization [50, 51]. Therefore, E-cadherin dysfunctions could interfere in microtubule homeostasis and cause reorganization of the microtubule cytoskeleton that can influence the organization of other cytoskeletal proteins [50].

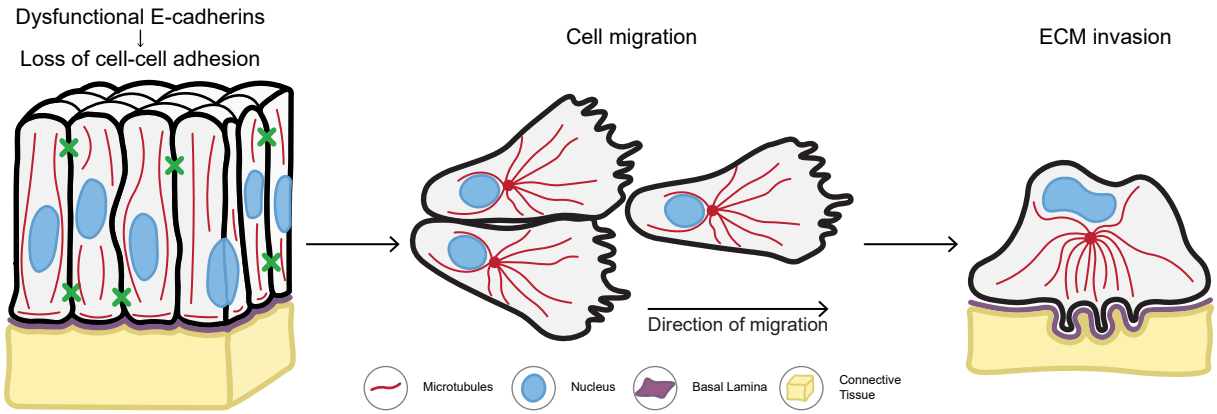


Figure 2.6: Dysfunctional E-cadherins may cause a loss of cell-cell adhesion, causing cytoskeletal reorganization favorable for cancer progression. Figure adapted from [6].

E-cadherin dysfunction is associated with highly invasive and lethal cancers [27–29], including gastric cancer [8, 45]. Indeed, it is well established that inactivating E-cadherin alterations cause diffuse gastric cancer [52].

Cell-cell junctions link neighboring cells but also interact dynamically with a specialized ECM, the basal lamina [2]. This thin, planar assembly lies beneath all epithelial cells and links the cells to the interstitial matrix. The ECM is a complex three-dimensional network of collagen, laminin, fibronectin, and other molecules. It is ubiquitously present in the non-cellular compartment of tissues and is key to provide mechanical support, selective permeability, maintaining normal tissue architecture and modulating intercellular crosstalk [8, 42, 53]. The anchorage of the epithelium to the underlying basal lamina is maintained by other types of cell junctions on the basal surface of epithelial cells [2]. The cytoskeleton also performs a key role in cell-ECM adhesion since these cell junctions link internally with the cytoskeleton [8].

2.3 Nuclear Morphology and Cancer Progression

Abnormalities in nuclear morphology and chromatin organization are hallmarks of many diseases, including cancer [54]. Cancer progression to a metastatic state requires cells to migrate and invade. In particular, cells must overcome barriers like the ECM and neighboring cells and sometimes through pores smaller than the cell itself [7]. Therefore, during these processes, cells either modulate the ECM or deform the nucleus, the latter being rate limiting, as interphase nuclei are stiffer and more viscous than the cytoplasm [7, 55], as shown in Figure 2.7. The nucleus dynamically changes its stiffness and rigidity while the cytoskeleton remodels to enhance cell migration and invasion [7].

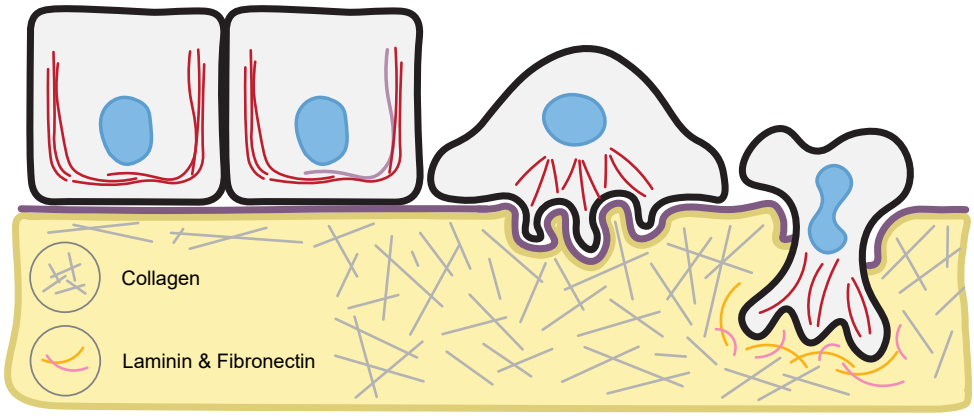


Figure 2.7: Representative image of normal and invading cells. While the cytoplasm, plasma membrane, and the majority of other small organelles may readily overcome constraints, the nucleus is the key limiting component due to its size and rigidity [7]. Figure adapted from [7, 8].

The cytoskeletal networks collectively serve as an important link between the extracellular microenvironment and the nucleus, allowing for nuclear morphology maintenance and signaling [7, 54]. External and internal forces of the cell are transmitted through the cytoskeleton and the nucleoskeleton, and altered mechanical environment can induce nuclear deformation [55]. During cell motility, dynamic, mechanical feedback between the cytoskeleton and the nucleus is needed to coordinate force transmission throughout the cell and organize cell polarity and nucleus-centrosome orientation [7, 55]. In conclusion, defects in the cytoskeleton-nucleus coupling have been linked to changes in nuclear morphology and mechanotransduction pathways, which can lead to various diseases, including cancer [54].

3

Computational Methods Background

Contents

This Chapter presents a review of the state-of-the-art computational methods and engineering tools available to analyze cytoskeletal structures. First, cytoskeleton imaging techniques in cancer research are introduced. Then, a "toolbox" of common image processing techniques aiming to segment cytoskeletal structures is described. Subsequently, three processing methodologies are addressed. Finally, aspects of the cytoskeletal organization from a biological and quantitative point of view are established in light of the processing techniques discussed. At the end of the Chapter, the aim of the study is introduced.

3.1 Bioimaging of the Cytoskeleton in Cancer Research

Early-stage detection of invasive cancer cells remains challenging nowadays. When identified early, cancer is more likely to respond to treatment, resulting in a greater probability of survival with less morbidity and less expensive treatment [23]. Therefore, there is an urgent need to unravel novel and automatic strategies to identify invasive cancer cells.

Imaging is an indispensable tool for discovery and diagnostics in biology, biomedicine and, in particular, cancer research. It is a fast-evolving field, with frequent emergence of novel techniques [56]. Many studies rely on computational image processing techniques, from which qualitative and quantitative studies can be performed. As discussed, the cytoskeleton plays a key role in the control of fundamental cellular processes [26]. Imaging has been used to understand and assess the role of such an essential structure in various cellular mechanisms.

Several fluorescence and electron microscopy methods are used for cytoskeletal filament imaging. Technique-specific features such as contrast properties, spatiotemporal resolution and signal-to-noise ratio limit the range and potential of suitable image processing tools for cytoskeleton analysis [56]. For example, electron microscopy techniques require dead cells and have higher resolutions, which is essential for an accurate cytoskeleton segmentation, but presents low signal-to-noise ratios and contrasts [53]. On the other hand, fluorescence-microscopy-based methods (e.g., widefield and confocal) are faster and allow imaging of live cells, focusing on the filament dynamics and their co-localization with other proteins. However, fluorescence microscopy generally has lower resolutions and artifacts associated with out-of-focus fluorescence [56].

In particular, immunofluorescence microscopy is an imaging technique that uses fluorescent-labeled antibodies to visualize the distribution of (glyco)proteins and other molecules in intracellular structures at the cellular or tissue level [57]. Figure 3.1 depicts two examples of antibody-stained cytoskeletons of HeLa cells (cervical cancer cell line), imaged using confocal microscopy.

Notably, several issues must be overcome to properly evaluate fluorescence images and develop computational methods. For instance, since the amount of acquired radiation is small, and high optical

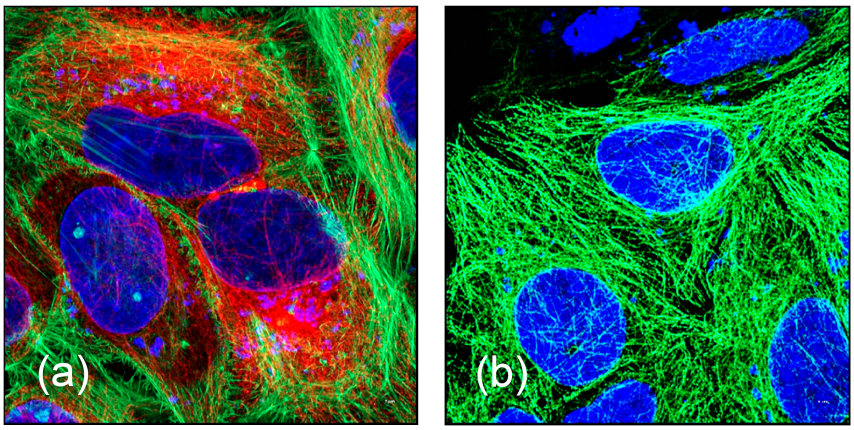


Figure 3.1: Immunofluorescent images of HeLa cells taken using airy scan mode in Zeiss 800 confocal microscope using 63x objective. (a) shows actin (green), tubulin (red) and nucleus (blue); (b) shows tubulin (green) and nucleus (blue) [9].

and electronic amplifications are involved, the resulting fluorescence images are usually corrupted with a signal/pixel-dependent noise described by a Poisson distribution associated with the photon counting process. Unlike other imaging modalities, where the noise associated is additive or algebraically multiplicative, fluorescence images have a wide-sense multiplicative noise, meaning that the time and/or space varying variance is related to the number of fluorescent dyes at each pixel in the image [58, 59]. Another relevant aspect of immunofluorescence microscopy is that it is not able to reach a resolution that permits the precise localization of each cytoskeleton filament due to the tiny lateral widths of the fibers (Figure 2.2), that are far beyond the resolution limit of these techniques [56, 60]. This lack of precision can make adjacent filaments unresolvable within the diffraction limit of light, compromising further quantitative measurements [56].

Of the three types of proteins that are part of the cytoskeleton, actin has been the most commonly studied, given its broad-ranging functions in various cellular processes. Intermediate filaments, on the other hand, are of great importance in the context of carcinogenesis and cancer cell alterations, namely during Epithelial-Mesenchymal Transition (EMT). Targeting microtubules, a number of drugs are currently used for many cancer types as part of therapeutic approaches [26]. It is therefore urgent to develop novel strategies involving the development of computational methods to better understand how the modulation of cytoskeletal proteins impacts cell behavior, including microtubules during cancer invasion.

3.2 Cytoskeleton Preprocessing

In fluorescent images, cytoskeletal structures can be described as networks of curvilinear objects due to the filament-like geometry of cytoskeletal proteins (Figure 2.2 and Figure 3.1). Each filament's curvilinearity can vary depending on the image's cytoskeletal protein. For example, intermediate filaments are

more curly, while microtubules are less curvilinear, given their greater rigidity [53]. However, this characteristic is not exclusive to cytoskeletal structures, as blood vessels or neuronal networks present similar morphological properties [61, 62]. Given these similarities between neurons/vessels and the cytoskeleton, several techniques used in the state-of-the-art for cytoskeleton segmentation have been created to segment neurons or vessels.

Preprocessing of cytoskeletal structures consists of separating curvilinear/filamentous structures (cytoskeletal proteins) from noise, blur effects, background and other non-linear structures. Preprocessing of cytoskeletal structures faces many challenging difficulties, not only due to the *many variables inherent to the scientific essay in question*, but also because of the *natural complex 3D network-like morphology of the cytoskeleton*. Regarding the former, the methodology to be applied is highly dependent on several factors, such as the type of microscopy used (fluorescence, electron) and magnification factor, cell line (epithelial, fibroblasts, vegetal), type of staining and cytoskeletal protein (actin, tubulin, intermediate filaments) under study. Additionally, blurring and imaging-related artifacts are common and might be very question-specific, weakening high-throughput cytoskeleton preprocessing solutions [10]. With so many variables to control for, it is not surprising the wide range of different state-of-the-art tools used to preprocess cytoskeletal structures in images. In contrast, despite the intention to highlight similar structures (cytoskeleton structures), different combinations of variables shorten the range of suitable preprocessing options for a given image. Regarding the latter, a segmentation, tracing, and enhancement methodology-oriented pipeline is required due to the complex network-like scaffold morphology of the cytoskeleton. Ultimately, individual and question-specific solutions must be devised as automation seems to be an important challenge in (high-throughput) cytoskeleton analysis.

Therefore, preprocessing steps for enhancement and segmentation of cytoskeletal structures must be included before its analysis. They can be broadly subdivided into *supervised* and *unsupervised* segmentation methods, depending on the requirement for annotated data. The former methods rely on a ground-truth dataset, including examples of segmentations, usually created by experts with manual annotation [56]. With deep learning or other machine learning algorithms, segmentation models are fit using this dataset. However, deep and machine learning were out-of-scope topics in this work, as unsupervised methods are the main focus. These methods need no training data to segment cytoskeletons.

Many unsupervised methods for cytoskeleton segmentation include two main steps: an *enhancement* step, followed by a *labeling* step. In the enhancement step, image filters denoise and enhance vessel-like structures while suppressing structures that deviate from curvilinear geometries. Then, based on the features enhanced by these filters, filaments from the cytoskeleton are extracted in a labeling step to group pixels into their corresponding filaments [56]. Naturally, more challenging images (e.g., with discontinuous/blurred contours or low signal-to-noise ratio) require more complex enhancement and labeling steps.

In a recent publication by Özdemir, B. and Reski, R. [56], close to 100 publications about ”automated and semi-automated enhancement, segmentation and tracing of cytoskeletal networks in microscopic images” were reviewed. A summary of the publications covered is provided, categorized by bioimaging technique, biopolymer type, key strategies, availability of user-interactive tools, and main segmentation/-tracing/enhancement tasks. However, only a fraction of these works involved unsupervised strategies, as the other fraction involved deep learning and other supervised models [56]. Additionally, another significant partition of unsupervised strategies mentioned in this publication was created for the purpose of cytoskeletal tip tracking and other time-dependent methodologies, which were out of the scope of this work. Notably, many methods were developed primarily for plant cells where the stomatal is set as the reference direction. Thus, they cannot be applied to quantify cytoskeletal structures of living animal cells under usual conditions [56, 63, 64].

Most cytoskeleton preprocessing methodologies in the state-of-the-art are based on *intensity-threshold methods*. These usually encompass denoising/deblurring filters, filament-enhancement filters, and thresholding [56, 64]. The following sections will address common image operations in other investigations' preprocessing pipelines in an attempt to build a toolbox of suitable image preprocessing tools for cytoskeleton enhancement. Figure 3.2 shows examples of cytoskeleton segmentation byproducts from other works, and Table 3.1 comprises the most important contributions in cytoskeleton segmentation in images using intensity-threshold methods [10–12, 65–68].

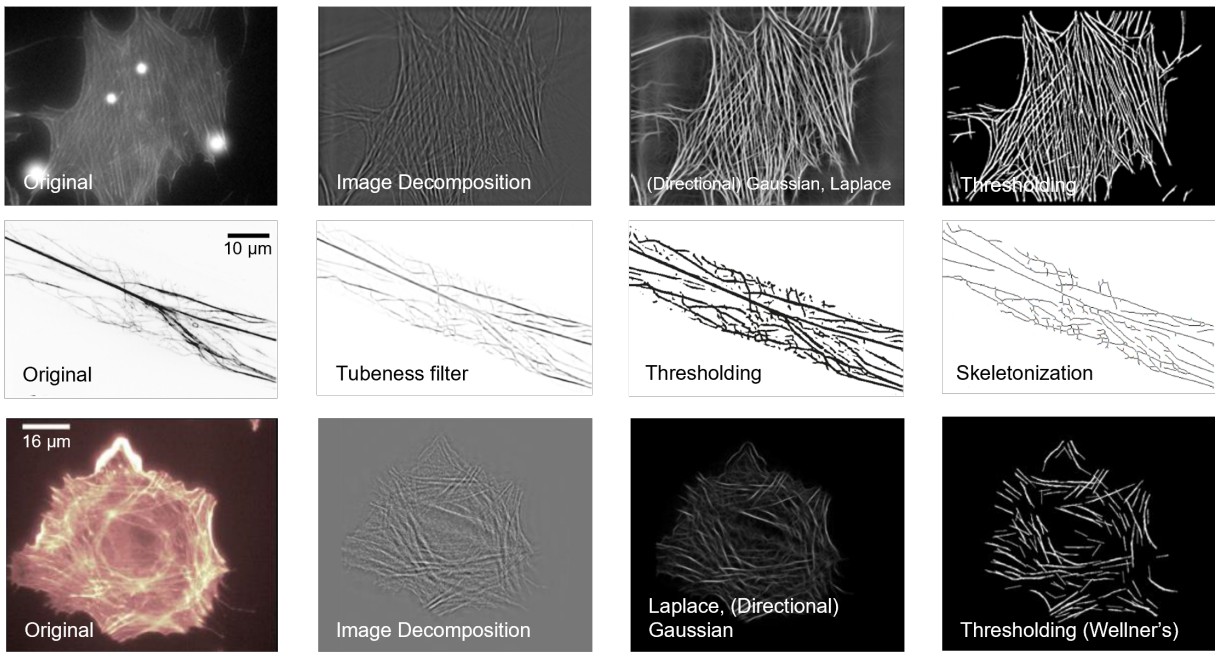


Figure 3.2: Examples of byproducts from other investigations' intensity-threshold processing pipelines. Top row, middle and bottom rows represent the preprocessing of a mouse osteoblast, a plant cell and a human lung adenocarcinoma cell, studied in [10], [11] and [12], respectively.

Table 3.1: Prominent imaging operations employed in cytoskeleton preprocessing in other state-of-the-art works.

Publication	Bioimaging type	Biopolymer type	Cell	Denoising	Line filters	Thresholding	Other operations
Alioscha-Perez, M. et al. (2016) [10]	Fluorescence microscopy	Actin filaments	Mouse osteoblasts	Image decomposition, Gaussian filter	Laplacian filter, directional Gaussian	Wellner's adaptive thresholding	-
Zhang, Z. et al. (2016) [66]	Single-molecule localization mic. (SMLM)	Microtubules	Mouse fibroblasts	-	Line Filter Transform (LFT)/Orientation Filter Transform (OFT)	Otsu	Skeletonization
Breuer, D. et al. (2017) [11]	Confocal microscopy	Actin filaments	Arabidopsis (plant)	-	Hessian-based	Adaptive median	Skeletonization
Faulkner, C. et al. (2017) [65]	Confocal microscopy	Microtubules	Arabidopsis (plant)	Gaussian filter	-	Otsu	-
Rogge, H. et al. (2017) [67]	Widefield fluorescence microscopy	Actin filaments	Mouse podocytes	Gaussian filter	Iterative tracing algorithm	Hysteresis	Morphological openings
Xia, S. et al. (2019) [68]	Stochastic optical reconstruction mic. (STORM)	Actin filaments	Mouse embryonic stem cells	H-minima transform	LFT/OFT	Meyer's watershed algorithm	-
Basu, A. et al. (2021) [12]	Atomic force microscopy (AFM)	Actin filaments	Human lung cancer cells	Image decomposition, Gaussian filter	Laplacian filter, directional Gaussian	Wellner's adaptive thresholding	-

Denoising and Deblurring: Noise removal in cytoskeleton images is usually one of the first image preprocessing steps. Common denoising strategies include Gaussian filters [10, 56, 65, 67, 69] that are good in smoothing images and relatively increase image signal at the expense of overall image detail.

Image deconvolution is an image-processing technique that removes blur or enhances contrast and resolution. Both functions are crucial for accurate segmentation due to the thin filaments that compose the cytoskeleton. This technique has been used in other works [68], but it can be equipment-specific since microscopes produce different imaging and noise properties.

Filament-Enhancers: After denoising, curvilinear/tubular geometries are enhanced via image filters of different natures. Laplacian filters are often used for image sharpening, where the edges of objects are highlighted [10]. Directional Gaussian filters tuned to specific orientations can also highlight filament geometry [10]. A group of filters to be highlighted are the ones based on the multiscale Hessian matrix [70, 71], as used in [11].

Other tools used in the state-of-the-art (as in [66] and [68]) to accentuate filamentous features include two transforms [72]: LFT, where the orientation and intensity of multi-orientational line integrals computed at each pixel are recorded and optimized to enhance linear features; to remove artifacts (as high-density localization hotspots) created by LFT, OFT is employed to operate on the optimized directions and intensities of the LFT map and selectively enhance pixels with the greatest directional coherence within their neighborhood [66].

Further strategies include morphological operations like path openings [67]. The goal of these operations is to suppress non-elongated structures.

Thresholding: In a binary segmentation, a simple labeling step can involve an intensity-threshold of the enhanced image, thus obtaining a binary mask of filaments [56]. Image binarization via Otsu thresholding is a common practice in many image processing pipelines [65, 66]. Likewise, intensity thresholds can be calculated globally according to the overall pixel intensity distribution or locally by adaptively computing a threshold for each fractional region of the image. Adaptive thresholds used in the state-of-the-art

include median thresholding [11, 56]. However, overlapping objects in binary cytoskeletons cannot be correctly labeled based on spatial connectivity, establishing a significant limitation [56].

Hysteresis thresholding is another thresholding technique, that instead of using a single value limit as most thresholders do, uses two values: low and high thresholds. Canny edge detector has been the most used method [64, 67, 73]

Other Operations: Further preprocessing of these binary filament masks may include a skeletonization procedure [11, 66, 69]. This reduces binary objects to 1-pixel-wide representations. This morphology thinning method may give rise to filaments with very short lengths due to noise or other factors. Short filaments are removed mainly using size filters [11]. Thus, a skeletonized version of a cytoskeleton corresponds to a 2D arrangement of 1-pixel-wide filaments, resembling a network.

3.3 Cytoskeleton Processing and Quantification

For subsequent cytoskeleton evaluation, the next step is to extract information from the enhanced cytoskeleton images obtained after preprocessing. The various strategies used in state-of-the-art allow this information extraction, using methods of different natures.

It remains a major challenge to quantitatively describe, analyze and compare a complex network of filaments like the cytoskeleton. Most image processing pipelines in the state-of-the-art focus on extracting features of post-segmented cytoskeleton filaments. A feature is an individual measurable property/characteristic of a phenomenon or object. For example, a cytoskeleton can be simply but incompletely described by the total number of filaments (1D feature) or by each fiber orientation (2D feature). Methods of different natures produce different types of features. For instance, a long filament can be detected by a long straight line or by an edge connected by two far-away nodes. However, further interpretation is needed to correlate them to architectural aspects.

Processing strategies of cytoskeletal structures include grayscale distribution studies, line segments rearrangements (Figure 3.3 (a) and (b)) and graph networks (Figure 3.3 (c) and (d)), as will be addressed in the following sections.

3.3.1 The Cytoskeleton as a Grayscale Distribution

Grayscale analyses have been widely reported in the literature [32, 63, 64, 74, 75]. Morphology can be studied in images in any preprocessing state, such as filament-enhanced images or skeletons. Features such as radius/diameter, area, perimeter, moments, centroid, circularity, compactness, or complexity are commonly extracted in this kind of analysis [32, 64]. Although these features are included in this chapter,

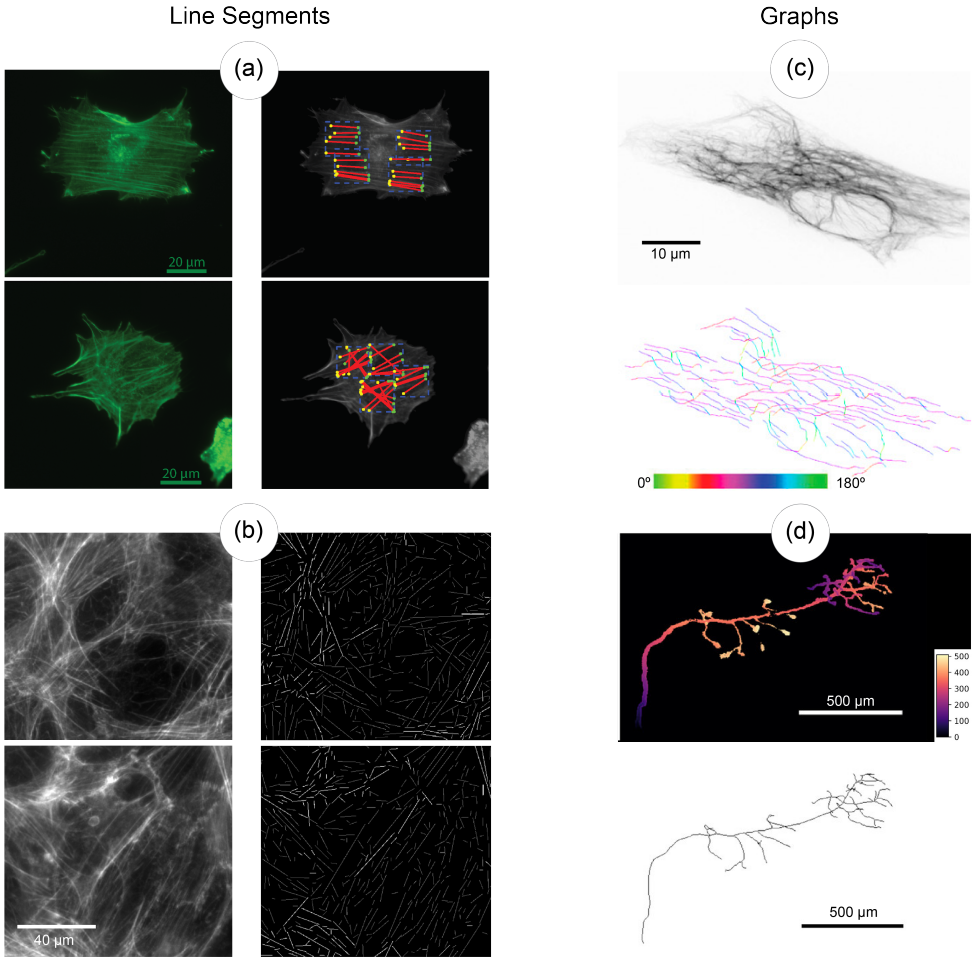


Figure 3.3: Cytoskeletal structures processed as line segment rearrangements and graph networks (nodes and edges). (a) Representation of the actin cytoskeleton of two mouse embryonic fibroblast cells as a set of line segments [13]; (b) Lines extracted from two actin cytoskeletal structures of human ovarian carcinoma cells [14]; (c) Vimentin (intermediate filament) network of a human foreskin fibroblast [15]; (d) Depth projection of a neuron and corresponding network [16].

they are calculated considering only binary values. That is, 1 corresponds to the cytoskeletal content estimated from the image (intensity values higher than 0 were converted to 1), and 0 is the background.

The gray level histogram of an image is a graphical representation of possible gray levels in the image against the number of times it occurs. Other statistical grayscale features can be extracted from this, such as the mean, variance, skewness, and kurtosis of the intensity distribution [32, 75]. More in-depth, the Gray-Level Co-occurrence Matrix (GLCM) is a statistical approach used in texture analysis. It examines the spatial connection among pixels, determining how frequently a specific combination of pixels appears in an image with a given direction and distance [76]. Entropy, correlation, contrast, among others, are extracted from the GLCM and used in texture analysis in other works [32, 74]. However, the meshwork topography of the cytoskeleton limits the utility of these features since many cannot accurately and comprehensively describe its organization.

3.3.2 The Cytoskeleton as a Set of Line Segments

One of the most used methods is based on the hypothesis that cytoskeletal proteins in images can be approximated by straight lines (or line segments), given their filament-shaped geometries with different stiffnesses/deformabilities depending on the protein in question. As previously mentioned, filaments can be interpreted as curvilinear objects, so this approximation aims to represent a filament as one or several spatially connected line segments. The extraction of these line segments can be done automatically through methods such as Hough Transform [19], Line Segment Detector (LSD) [77], or OFT [78]. After processing, a set of lines defined in the 2D space of the image is obtained, where their positions, orientations, and lengths are known [10, 79–81]. Nevertheless, little effort has been put into designing more complex features with this information, aiming to more accurately describe cytoskeleton organization aspects [12, 64] or other morphologically similar structures [82].

The advantage of this approach is that the cytoskeleton is uniquely represented by a set of line segments, reducing the computational burden. However, using this method, information such as filament curvature is harder to extract. Line-merging algorithms may solve this problem by considering a filament as a piecewise composition of quasi-straight segments instead of simple line segments, as made by Alioscha-Perez, M. et al. (2016) [10] with mouse osteoblast cytoskeletons and by Zhang, Z. et al. (2016) [66] with mouse fibroblasts. Such analysis is yet to be performed in human cancer cells.

3.3.3 The Cytoskeleton as a Network

Organization of a filament network obtained after skeletonization is usually quantified with metrics using information directly extracted from the network. These networks are represented by nodes (crossings or endpoints of filaments), and weighted edges capturing cytoskeletal segments' intensity [11]. Number of filaments [16, 68], (average) filament lengths [11, 16, 68], (average) filament curvatures [66], node angles [66, 68], assortativity [11] and junction types [16, 66, 68] are common cytoskeleton quantification metrics extracted from 2D filament networks.

In Asgharzadeh, P. et al. (2018) [83], confocal microscopy images of living cells tagged with a fluorescent protein were converted to networks. From these networks, a comprehensive feature extraction pipeline was applied to characterize the secondary structure of the FtsZ protein, a central component of plant chloroplast cytoskeletal structures. Features extracted included network (enclosed) volume, volume density, diameters, stretch, node thicknesses, pairwise distances, and node-to-surface distances, as well as angles and segment lengths, thicknesses, inhomogeneities and curvatures [83].

Similar studies in animal cells did not go as far, one example being the study by Costigliola, N. et al. (2017) [15], where only orientations and fiber lengths were extracted from cytoskeletal structures of human foreskin fibroblasts. In this publication, a graph-matching algorithm (similar to line-merging algo-

rithms) was used to cluster network branches likely belonging to the same fiber based on co-orientation and spatial proximity [15].

A complete network analysis is yet to be applied to animal cell cytoskeletal structures and correlate network features with cytoskeletal organization and their influence in cancer.

3.3.4 Qualitative and Quantitative Cytoskeleton Characterization

For a comprehensive cytoskeletal evaluation, a number of cytoskeleton features need to be addressed and characterized. For instance, Higaki, T. (2017) [63], considered four descriptors to analyze plant cell cytoskeletal structures: orientation, bundling, parallelism and density. However, other investigators developed features for cytoskeletal quantification that do not fit perfectly in any of the four. Therefore, other architectural features must be created and cytoskeleton organization aspects must be listed, interpreted qualitatively and quantified with features/metrics. Important points of view to grasp the cytoskeleton organization include: *orientation*, *bundling*, *parallelism*, *morphology*, *compactness*, *connectivity*, *quantity*, *complexity*, *radiality*, and *interaction with the nucleus*, as shown in Figure 3.4.

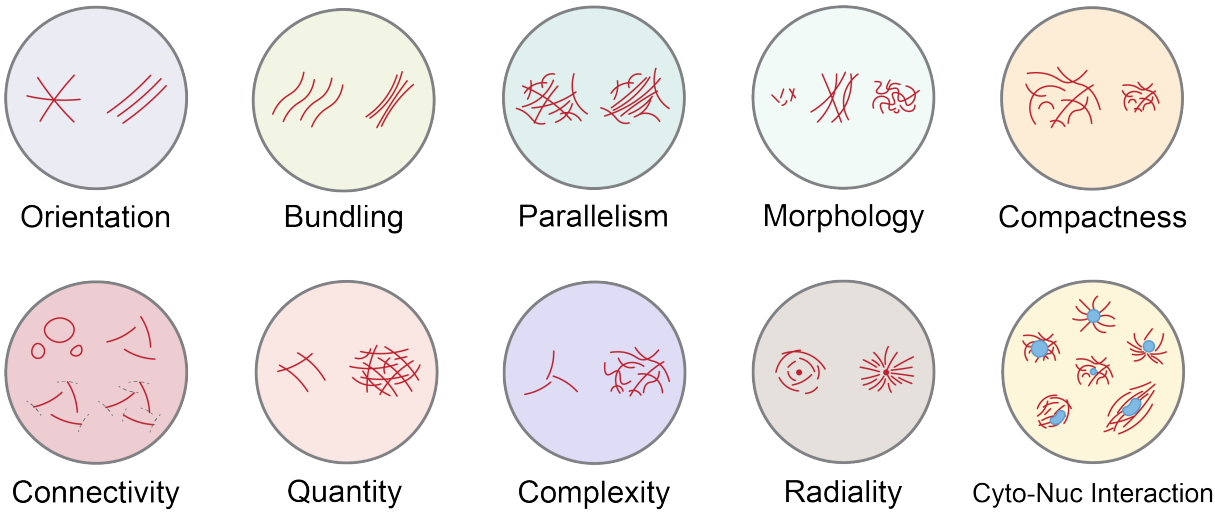


Figure 3.4: Schematic representations of cytoskeleton organization aspects included in the model.

Orientation: Orientation is related to the position of something concerning its surroundings. The orientation of a fiber is typically quantified by its angle with some reference, for example, the horizontal axis or the cell's long axis [63]. Globally, the collection of fiber orientations gives rise to an angle distribution, where its spread increases with the degree of disorganization [12].

Basu, A. et al. (2021) [12] have published the most recent effort on cytoskeletal quantification procedures. In this work, an image quantification tool of the actin cytoskeleton called Statistical Parametrization of Cell Cytoskeleton (SPOCC) was employed to identify intermediate EMT states based on cytoskeletal orientations. As a processing stage, actin geometry was extracted as a series of straight

lines. For alignment quantification, the Orientational Order Parameter (OOP) was calculated, considering the angular distribution of these straight lines. OOP is the maximum eigenvalue of the Mean Order Tensor of a set of vectors representing each straight-line angle. When the angular distribution is narrow, the lines are well-aligned, and OOP is close to 1. The lines are disorganized when the distribution is broad, resulting in OOPs close to 0. The alignment of stress fibers characterizes EMT progression of a given cell to a mesenchymal phenotype with time, so OOP was used to validate the gradual increase of alignment [12].

In Qu, Y. et al. (2020) [82], metrics were designed to quantify the alignment of collagen fibers, which are structures that are morphologically similar to cytoskeleton fibers. These metrics should be able to quantify cytoskeleton orientation as well. Impulsively, orientation could be quantified with line segment angles' spread (or standard deviation), as commonly adopted in statistic indices. However, this indicator depends on the relative angles with respect to the reference axis. Such limitations were overcome by using circular statistics concepts, and Qu, Y. et al. proposed the Circular Variance ($CVar$) to quantify orientation: the higher it is, the broader the angle distribution. They also proposed the homogeneity index (HI) based on limiting entropy as an indicator of the disorder of fiber orientations [82].

The multi-orientational index (MOI) [69] is an indicator of orientation complexity. This index is defined as the entropy of the orientation distribution, calculated by applying a series of opening operations. The value of MOI is higher for isotropic, random filament orientation distributions, and lower for homogeneous orientation distributions.

Bundling: Cytoskeleton bundling is crucial in various cellular processes. Fiber bundling is associated with increased object thicknesses or widths [65]. The bundling level of cytoskeleton filaments has been quantitatively measured by the skewness of the fluorescence intensity distribution [63,84,85]. Skewness is a statistical parameter that quantifies the asymmetry of a distribution. When cytoskeleton filaments form bundles, the fluorescence distribution is expected to extend more toward high fluorescence levels than low levels. This is caused by the stronger fluorescence of labeled cytoskeleton bundles compared to single filaments, thus increasing asymmetry (and skewness) of the fluorescence distribution [84]. The Coefficient of Variation (CV) of the intensity distribution, defined by the ratio of the standard deviation to the mean, is a measure of relative variability and has been used in the literature as a cytoskeleton bundling quantifier [86].

Thickness and bundling are closely-related terms. In Kimori, Y. (2015) [69], cytoskeleton thickness distribution was determined by applying a pattern spectrum to the segmented cytoskeleton. The global thickness of fibers was estimated as the median of the spectrum.

Parallelism: Fiber parallelism is closely related to fiber orientation, but instead provides an index to quantify variations around the average fiber angle [63]. Parallelism is based on fiber angles relative to the

image and not the cell axis; therefore, no reference direction is needed [63]. From another perspective, parallelism can be seen as the degree of local orientation: increased local parallelism is expected to be found in aligned fibers, so their orientations are narrowly spread around the mean orientation [63].

Regarding the design of cytoskeleton parallelism quantification metrics via data processing, Liu, Y. et al. (2018) [73] introduced Partial Actin-cytoskeletal Deviation (PAD) and Total Actin-cytoskeletal Deviation (TAD). After validating this set of features in synthetic images, a preprocessing step, including Canny and Sobel filters and skeletonization, lines were extracted with Hough transform and used to calculate PAD and TAD. PAD is defined as the standard deviation of detected line segment angles in a sub-image of the original one. TAD is calculated globally and is defined as the standard deviation for the mean direction in each sub-image against the average of these means. Lower PADs and TADs point to a narrow angular distribution of line segments correlated with highly aligned cytoskeletal proteins, and vice versa. The authors referred to PAD and TAD as orientation indicators [73]. Nevertheless, according to this work's distinction between orientation and parallelism, these metrics fit better as parallelism quantifiers.

Morphology: Fiber morphology is closely related to the geometry of a given cytoskeletal filament on the image. Fiber length is a commonly extracted morphology descriptor for visualization and quantification via line segment lengths [10, 79] or graph edge distances [11, 56, 66, 83].

To globally assess fiber morphology, average lengths are the most straightforward features. Further, length dispersions were quantified with the CV of filaments/edges lengths in network representations of cytoskeletal structures in Breuer, D. et al. (2017) [11].

Additionally, fiber contortion is related to how the fiber twists or bends in the image. The closer the fiber is to a line segment (i.e., the fiber is approximately straight), the less contortion it presents. On the other hand, the more curls and twists it presents, the higher its contortion level. Aiming to detect arterial contortion abnormalities and their relation with disease, contortion has been quantified with tortuosity, defined as the ratio between the length of the fiber and the linear distance between the two endpoints [87]. This ratio was obtained in [16] with graph representations of cytoskeletal structures. Alternatively, contortion has been quantified in [11] with the CV of edge angles in network representations of plant cytoskeletons [11].

Compactness: Cytoskeleton compactness aims to evaluate how spread fibers are in relation to one another. A cytoskeleton with a given number of fibers is compact if their spatial positions are close to one another, and is sparse if the same number of fibers are well spread on the 2D plane.

Compactness was assessed by Asgharzadeh, P. et al. (2018) [83] in 3D plant cytoskeletons represented as graphs/networks. It was quantified by the normalized difference between the mean distance to the center of gravity and the mean distance to the network surface of all nodes [83].

Connectivity: Specifically to graph representations, the cytoskeleton can be subject to a connectivity-based analysis, where individual filaments are evaluated in terms of other connected filaments to any of the two extremities [83]. An isolated cycle is a filament represented by an edge with one node. An isolated filament is a spatially isolated edge. Further, a junction-to-endpoint branch is a filament where one node connects to another branch, while in junction-to-junction filaments, both nodes connect to two or more other filaments. Connectivity is one of graph theory's basic concepts and may measure a network's resilience. This type of analysis was done in [16].

The algebraic connectivity was computed in [11] to assess cytoskeleton robustness. By obtaining the second-smallest eigenvalue of the Laplacian matrix of the graph, the graph was quantified in terms of how well connected it was. Assortativity was also obtained to assess graph heterogeneity [11].

Quantity: Fiber density indicates the amount of cytoskeleton per unit area [63]. It can be helpful to quantify the degree of cytoskeletal disruption imposed by inhibition of gene function or external stimuli. Density is commonly estimated, considering the pixel grayscale distribution [66, 73].

Fiber density was estimated in [66] by the fraction of cell area occupied by microtubule filaments. In [73], a preprocessing step was applied, including Canny and Sobel filters for edge detection. The Average Actin-cytoskeletal Intensity (AAI) was computed from the grayscale histogram distribution to quantify the amount and intensity of filaments on the preprocessed image.

Complexity: The measurement of the cytoskeletal fractal dimension to characterize complexity has been widely reported in the literature [18, 69, 88, 89]. Fractals are irregular geometric patterns characterized by self-similarity and complexity, with a non-integer value for their dimension. Thus, fractal objects cannot be defined by the traditional Euclidean geometry, so fractal geometry is used instead to quantify their properties [18]. Fractal dimension is a ratio that provides a statistical index of complexity comparing how the detail in a pattern changes with the scale at which it is measured. Both Alhussein, G. et al. (2016) [88] and Revittser, A. et al. (2021) [18] used the box-counting method on cytoskeletal structures, which is a widely used algorithm to calculate the fractal dimension of an object on a grayscale image. The former measured the fractal dimension independently over squares with desired areas through a discrete square grid placed on the image. Then, the fractal dimension of discrete square portions inside the cytoskeleton was obtained, observed and analyzed over time [88]. Finally, the latter computed a whole-image fractal dimension by averaging fractal dimensions of actin in human mesenchymal cells rotated by an angle of 15° until 90°, initially (angle = 0°) aligned with the nucleus major axis [18].

Radiality: The cytoskeletal proteins of a cell can form radial arrays nucleating from the centrosome [38, 39]. In these cases, the centrosome is the point in the cell with a higher degree of radiality (radiality center). However, cytoskeletal proteins may arrange themselves in parallel, not depending on the

centrosome, with a lower radiality [39] such as in polarized epithelial cells. No studies were found that addressed the radiality parameter quantitatively in cytoskeletal proteins of cells.

Cytoskeleton-Nucleus Interaction

Nucleus (Pre)processing: Similar to cytoskeleton features, nuclear architectural features could provide an efficient processing strategy to identify cells with invasive potential at an early stage. But first, a preprocessing step for nuclei segmentation is required. Classical strategies roughly include thresholding (Yen or Otsu) to obtain a binary mask, followed by a watershed algorithm to separate single and clumped nuclei [32]. Other strategies include deep learning models using U-Net [90] or Mask R-CNN [91], networks built for medical image segmentation.

Quantification: The dynamic interplay between the nucleus and the cytoskeleton is crucial in several cellular processes involved in cancer progression. For instance, the nucleus continuously changes its stiffness and rigidity, while the cytoskeleton changes its organization during cell migration and invasion [7, 25]. Thus, it is necessary to design metrics that assess nuclear features and relate them to those of the cytoskeleton.

In 2018, Al-Mamun, A. M. et al. [32] applied a quantitative image analysis pipeline to study the cellular cytoskeleton during *in vitro* tumor growth. Two separate preprocessing pipelines were applied to the cytoskeleton and nucleus of a given cell. A grayscale analysis was applied by exploring the intensity histogram and GLCM-based features (for textural assessment). However, this methodology was applied to analyze ruffling regions in cells and not to relate nucleus features with cytoskeleton organization [32].

Thus, more studies are needed to correlate the state of the nucleus and the organization of the cytoskeleton in any cellular process, particularly in cancer cell progression.

3.4 Objective

It is well established that the cytoskeleton undergoes a dramatic reorganization during many cellular processes, including cancer invasion. Nevertheless, the identification of cytoskeletal features and how the modifications in the cytoskeletal network cooperate to mediate cell invasion are still limited. Thus, the present study aimed to develop a novel computational approach to unravel the cytoskeletal architecture of invasive cancer cells. Specifically, a pipeline was established to characterize α-tubulin, a major component of the cytoskeleton, in cells with or without E-cadherin mutations that impact cell-cell adhesion and cell invasion. The strategy involved the application of methods of image preprocessing, processing, feature extration and feature analysis. Ultimately, from simple immunofluorescence images,

it is expected that this framework will enable the prediction of cell invasive potential, with future potential impact in cancer diagnosis and prognosis.

4

Methodology

Contents

4.1 Overview

The cytoskeleton is a three-dimensional network of filaments dispersed in the cytoplasm with a varying organization over time. The main objective of this work was to characterize the organization of this structure. For that purpose, aspects of organization related to typical patterns of cytoskeletal structures in cells were created, including: orientation, bundling, parallelism, morphology, compactness, connectivity, quantity, complexity, radiality, and interaction with the nucleus. Algorithmic and mathematical methods were then developed to quantify and visualize each aspect of fiber-by-fiber architecture. All this methodology was put into practice to distinguish cytoskeletal structures of non-invasive and invasive cancer cells.

Figure 4.1 illustrates a graphical abstract of the bioimaging pipeline employed in this work. In summary, the cytoskeletal structure of cancer cells expressing WT or mutant E-cadherin, leading to distinct cell-cell adhesion and invasion phenotypes, was evaluated. A dataset of 12 RGB images stained for α-tubulin (red), a major cytoskeleton component, and nuclei (blue) was obtained. Cells were grouped into four experimental conditions (groups), with three images per group: WT cells, transfected with functional E-cadherin, and three groups of cells transfected with E-cadherin carrying different mutations. In particular, these include a non-pathogenic group (NP) and two groups of pathogenic phenotypes (P1 and P2). Importantly, although in the groups NP, P1 and P2 E-cadherin harbors a mutation, the pathogenic mutations P1 and P2 are associated with an invasive phenotype while the NP is not. In the proposed strategy, an image preprocessing framework was applied to this dataset using Python. Blue (nuclei) and red (α-tubulin) channels were deconvoluted and subject to different (pre)processing methods. Automatic nuclei segmentation with a trained machine learning model (Stardist [92]) allowed the extraction of morphological-, intensity-, textural- and frequency-related features from deconvoluted nuclei, as well as their centroids, contours, and areas. On the other hand, preprocessing of deconvoluted cytoskeletal images aimed to enhance cytoskeletal structures by applying three filters: a Gaussian filter (for smoothing), a Sato filter [70] (for curvilinear structure enhancement), and a Hessian filter [71] (for image binarization), followed by skeletonization.

Binary segmentation masks were drawn manually and allowed the analysis of the cytoskeleton and nucleus of individual cells. Single-cell processing comprised a feature extraction methodology of three different natures: one by computing morphological-, intensity-, textural- and frequency-based features on deconvoluted versions of cytoskeletal structures (DCF) and the others by obtaining line segment- (LSF) and graph-based (CNF) features from their skeletonized versions after automatic line detection (with LSD [77]) and graph conversion (with Skan), respectively.

The features calculated by these three methods were correlated with several aspects of the cytoskeletal organization (including *orientation*, *bundling*, *connectivity*, *morphology*, and others), allowing a complete characterization of this structure. Hence, cytoskeletal organization was compared qualita-

tively and quantitatively between groups of three example cells. All cells meeting the inclusion criteria were manually segmented and (pre)processed for a global analysis. Organization aspects of microtubules were compared between cancer cells expressing WT E-cadherin and cancer cells expressing mutant E-cadherin (NP, P1 and P2). Remarkably, this methodology was applied to α-tubulin images, but it can be used with any other cytoskeletal protein.

4.2 Preparation and Acquisition of Biological Data

In this section, the methodology for preparing the biological material that was analyzed in this work is described, namely in cell culture preparation and microscopy for image acquisition.

Cell Culture

In this study, Chinese Hamster Ovary (CHO) cells, which are negative for E-cadherin, were used. For analysis, WT E-cadherin and a panel of E-cadherin variants were selected. Specifically, three variants were evaluated, two pathogenic (P1 and P2) with increased invasive potential and a non-pathogenic (NP) with no invasive potential. NP and P1 are located on the signal peptide domain of E-cadherin while P2 is located on the prodomain. Importantly, the NP mutation does not impact the expression and function of E-cadherin, while the P1 mutation impacts E-cadherin expression and function, and P2 mutation does not impact E-cadherin levels but impacts its function (Seruca's Lab personal communication). The P1 mutation corresponds to a heterozygous germline mutation c.38_46del, leading to the amino acid deletion p.L13_L15del, which was identified from a New Zealand patient as previously described [93]. Due to clinical confidentiality issues, the current Thesis cannot disclose the identity of the NP and P2 mutations. However, this information can be provided if requested.

Briefly, cells were maintained at 37 ºC under 5% CO_2 humidified air in a α-MEM (+) medium (Gibco, Invitrogen) supplemented with 10% fetal bovine serum (Hyclone), 1% penicillin/streptomycin (Gibco, Invitrogen) and 5 µg/mL blasticidin (Gibco, Invitrogen). Cells were cultured on 6-well plates and transiently transfected with vectors encoding either the WT E-cadherin or E-cadherin mutants as previously described [93]. Cells were then cultured on Corning™ BioCoat™ Poly-D-Lysine/Laminin 8-well culture slides (Corning) and grown until 70-80% confluency was reached.

Immunofluorescence Staining, Image Acquisition and Dataset Description

Cells cultured on culture slides were fixed with 4% paraformaldehyde for 20 min. Following a 10 min wash in phosphate buffered saline (PBS), cells were permeabilized with 0.1% Triton X-100 in PBS for 15 min at room temperature. Cells were blocked with 3% bovine serum albumin (BSA) in PBS and

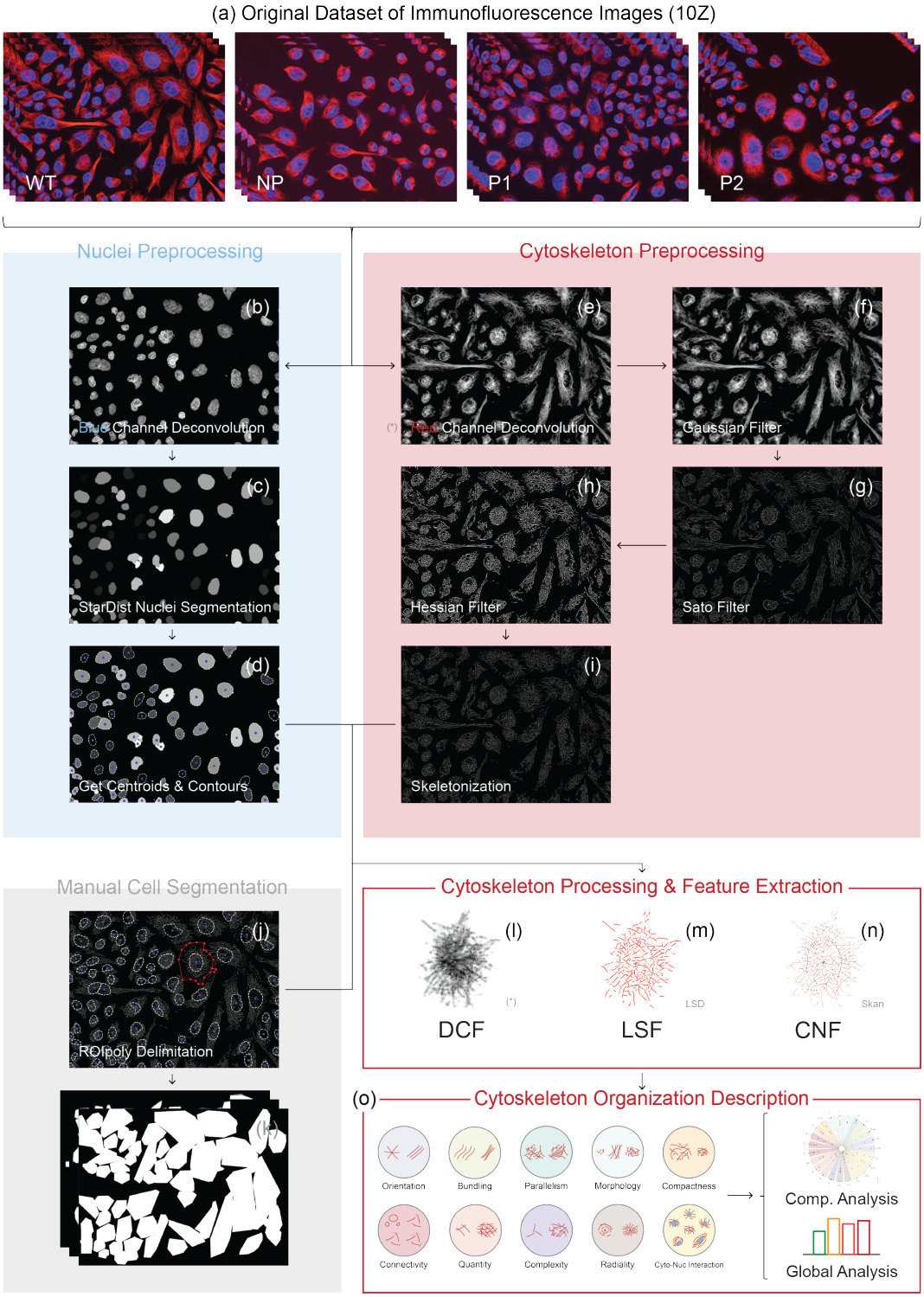


Figure 4.1: Overall pipeline to evaluate the cytoskeletal architecture of cancer cells. The processing pipeline involves the analysis of α-tubulin and DAPI immunofluorescent images (a). Preprocessing methods are subsequently applied to all the images for nuclei (b-d) and cytoskeleton (e-i) segmentation. After manual delimitation of cell boundaries (j-k), features are extracted (l-n) and analyzed comparatively and globally (o). WT, Wild Type; NP, Non-Pathogenic mutation; P1, Pathogenic mutation 1; P2, Pathogenic mutation 2; DCF, Deconvoluted Cell Features; LSF, Line Segment Features; CNF, Cytoskeleton Network Features.

stained overnight at 4 ºC with α-tubulin rabbit primary antibody (Invitrogen, #PA5-16891). Subsequently, cells were incubated with Alexa Fluor 594 goat anti-rabbit (Invitrogen) for 1 h in the dark. For nuclei staining, 4,6-diamidino-2-phenylindole (DAPI) (Vector Laboratories) was used to bind to adenine–thymine-rich regions in DNA, conferring a blue-colored fluorescence to nuclei.

Images were acquired on a Carl Zeiss Apotome Axiovert 200 M Fluorescence Microscope (Carl Zeiss, Jena, Germany) with a 40x objective (Plan-Apochromat 40x/1.3 Oil DIC (UV) VIS-IR M27) using an Axiocam HRm camera and the Zeiss Axion Vision 4.8 software. Multiple images were acquired along the Z axis (10 Z-stacks) and images were then analyzed. All images were acquired with the same acquisition settings and saved with a resolution of 16-bit and pixel width equal to 1 µm. In total, 12 RGB immunofluorescence images were gathered, examples of these are shown in Figure 4.1 (a).

4.3 Image Preprocessing

In this section, the image preprocessing methodology is described ((b-d) for nuclei and (e-i) for cytoskeletal structures). The aim was to segment the nuclei and cytoskeletal filaments

4.3.1 Deconvolution and MIP

The original RGB immunofluorescence images were preprocessed using deconvolution. Multiple images were acquired along the Z axis (10 Z-stacks) and images were deconvoluted for each channel using Deconvolution Express in Huygens Software (Scientific Volume Imaging). This step was crucial to diminish motion/blur effects and noise.

Subsequently, and since 10 Z-stacks were acquired, the Z dimension was collapsed to obtain single 2D grayscaled images for each channel. Therefore, Maximum Intensity Projection (MIP) was used to flatten the images on the (x, y) plane, where each pixel (i, j) of the resulting image is the maximum value of all the pixels along the z direction of (i, j). Figure 4.1 (b) and (e) represent images of nuclei and α-tubulin, respectively, after deconvolution and MIP projection.

4.3.2 Nuclei Segmentation and Feature Extraction

The deconvoluted blue channels containing nuclei were segmented using a pre-trained model from StarDist, a Python package for star-convex object detection in images [92]. The model was trained on 2D images of fluorescent-stained nuclei. A segmentation mask was obtained (Figure 4.1 (c)), from which individual nuclei could be analyzed separately. Their centroids and delimiting contours were computed (Figure 4.1 (d)). Centroids/contours associated with masks with an area below 600 pixels were discarded, corresponding to tiny nuclei and noise.

Actually, morphology-, intensity-, textural- and frequency-based features were also extracted from nuclei patches, as nuclei features can differ according to the cellular phenomenon and in cell migration/invasion in particular [7, 55]. However, since this work is focused on cytoskeletal organization and nuclear features are out of the scope of this project, they were not included in this Thesis.

Concluding, (weighted) centroids (C_N and C_N^W), area (A_N) and contours were extracted from post-segmented nuclei. Particularly, intensity values were used to calculate the weighted centroid, while centroid and area considered binary intensities (either 0, background, or 1, foreground (nucleus)). These features were helpful to better visualize the results and to study the cytoskeleton-nucleus interplay.

4.3.3 Cytoskeleton Enhancement

Deconvoluted cytoskeleton images (Figure 4.1 (e)) were Gaussian filtered, smoothing the fluorescence signal across the image (Figure 4.1 (f)). This operation is a convolution with the Gaussian kernel

$$G_\sigma(x,y) = \frac{1}{2\pi\sigma^2}\exp\Big(-\frac{x^2+y^2}{2\sigma^2}\Big)\,, \tag{4.1}$$

and structures below the standard deviation σ will be smoothed out and become indistinguishable [67]. As mentioned before, given the tiny lateral diameter of cytoskeletal fibers, the best possible resolution is desired so individual fibers can be distinguished. Counter-intuitively, a Gaussian filter has the characteristic of reducing the image resolution, but it is compensated by the ability to uniformize the signal and reduce noise. Additionally, some fibers often showed irregular borders, so smoothing was indispensable [67].

Then, a parameter-tuned Sato filter [70] was applied (Figure 4.1 (g)) to enhance continuous curvilinear structures (e.g., tubes, wrinkles, or rivers) and suppress other geometries (e.g., blobs or punctual noise). This filter calculates the eigenvectors of the Hessian matrix, and intensity values are consequently proportional to the degree of local curvilinearity [70].

For binarization, a parameter-tuned hybrid Hessian filter [71] was applied (Figure 4.1 (h)). Similar to the Sato filter, this filter is also based on the Hessian matrix and also aims to highlight vessels, wrinkles and other curvilinear objects in images. However, it outputs a binary image, only containing objects closer to a curvilinearity geometry, nulling other structures. The tuned parameters of the three used filters are displayed in Table 4.1.

The last step was a skeletonization using Zhang's method [94] (skeleton in Figure 4.1 (i)). Skeletonization reduces binary objects to 1-pixel-wide representations by identifying the line within the shape equidistant from each boundary.

Table 4.1: Parameters involved in the image preprocessing pipeline [19].

Method	Parameter	Description	Value/Mode
Gaussian Filter	`sigma` (σ)	Standard deviation for Gaussian kernel	1
Sato Filter	`sigmas`	Scales of the filter	1
Hybrid Hessian Filter	`mode`	Handling of values outside the image borders	reflect
	`sigmas`	Scales of the filter	0.03
	`gamma`	Correction constant that adjusts the filter's sensitivity to areas of high variance/texture/structure	500
	`beta`	Correction constant that adjusts the filter's sensitivity to deviation from a blob-like structure	0.5

4.3.4 Manual Cell Segmentation

After their automatic segmentation, individualized analysis of each cell describes an ideal/desirable, but unrealistic procedure. An attempt was made to segment the cells by contour detection automatically. This procedure proved to be effective in the segmentation of single, isolated cells. However, attached cells were observed due to the adherent phenotype that characterizes these cells and the intention to study the influence that different dysfunctions in an adhesion protein (E-cadherin) have on the cytoskeleton. This made the cells' automatic segmentation a difficult task with an acceptable success rate.

Thus, it was necessary to take a step back and manually segment cells. For this purpose, ROIPoly [95], a Python module for polygonal Region Of Interest (ROI) selection, was used to delineate the cell's outer contours (Figure 4.1 (j)). Individual binary masks were obtained by filling the polygon's interior (Figure 4.1 (k), where multiple ROIs were selected).

4.4 Image Processing

After manual segmentation, the binary mask obtained for a given cell can be multiplied by any preprocessing byproduct. Two are of particular interest: the cell's deconvoluted cytoskeleton version (as in Figure 4.1 (e)) was used to extract DCF (Figure 4.1 (l)), while its skeletonized version (as in Figure 4.1 (i)) was used to obtain LSF (Figure 4.1 (m)) and CNF (Figure 4.1 (n)) after processing.

4.4.1 Feature Extraction from the Cytoskeleton as a Grayscale Distribution

A feature selection pipeline was applied to single-cell microtubules in deconvoluted images (as in Figure 4.1 (e)). Morphological-, intensity-, textural- and frequency-related features were calculated from the image patch of the selected deconvoluted cytoskeleton (Figure 4.1 (l)). Table 4.2 summarizes the most

important features. The majority was left out because further interpretation is needed to correlate them with cytoskeleton organization if possible. More details can be found in the Appendix.

An image is a 2D function of space $f(x, y)$ where x and y are the spatial coordinates, and the value of f at the pixel coordinates (x, y) is called intensity [96]. Given an image patch of a cell with n cytoskeleton pixels, several features can be defined from its intensity distribution, $f(x, y)$. In particular, with the mean intensity, μ_f, defined as

$$\mu_f = \frac{\sum_{i,j=1}^{n} f(x_i, y_j)}{n} \ ,$$

and standard deviation, σ_f, given by

$$\sigma_f = \sqrt{\frac{\sum_{i,j=1}^{n} (f(x_i, y_j) - \mu_f)^2}{n}} \ ,$$

the skewness and CV of the intensity distribution were obtained to quantify *bundling* using

$$Skew = \frac{\sum_{i,j=1}^{n} (f(x_i, y_j) - \mu_f)^3}{(n-1) \cdot \sigma_f^3} \ , \tag{4.2}$$

and

$$CV = \frac{\sigma_f}{\mu_f} \ , \tag{4.3}$$

respectively.

To compute the area A_C occupied by the cytoskeleton of a given cell, Canny's edge detector [97] was applied to deconvoluted cytoskeleton image patches. Lower and upper thresholds, required parameters in Canny's edge detector, were set to 30% and 100% of Otsu's threshold, respectively. By visual inspection, these values were the most suitable to correctly separate the background from the cytoskeleton and segment it in all the manually selected cells in the dataset. The area A_C of the resulting image f', i.e., the number of pixels with non-zero intensity on the binary mask, was obtained according to

$$A_C = \#\{(x, y) : f'(x, y) > 0\} \ , \forall (x, y) \in \mathbb{N}^2 \ , \tag{4.4}$$

as an intermediate step to obtain N_l/A_C, used to quantify *compactness* and A_N/A_C to study the *cytoskeleton-nucleus interplay*.

The Average Microtubule Intensity (AMI) was calculated according to Liu, Y. et al. [13] to assess fiber *quantity*. Briefly, it is the ratio of total intensity from the first local minimal point (f_1) to the maximum (f_2) of the histogram divided by cell area (calculated in 4.4). Mathematically,

Table 4.2: 1D Deconvoluted Cell Features (DCF).

Deconvoluted Cell Features (DCF)		
Notation	Description	Equation
$Skew$	Skewness of intensity distribution	4.2
CV	Coefficient of variation of intensity distribution	4.3
A_N/A_C	Nucleus-cytoskeleton area ratio	4.4
AMI	Average Microtubule Intensity	4.5
D_{CN}	Nucleus-cytoskeleton centroid distance	4.6
D_{CN}^W	Weighted nucleus-cytoskeleton centroid distance	4.7

$$AMI = \frac{\sum_{f_1}^{f_2} \eta \cdot f(x,y)}{A_C} \ , \tag{4.5}$$

being η the count of pixels with brightness $f(x,y)$. Finally, to study the *cytoskeleton-nucleus interplay*, the (weighted) centroids of the cytoskeleton, (x_C, y_C) and (x_C^W, y_C^W), were computed and compared with the ones of the nucleus ((x_N, y_N) and (x_N^W, y_N^W), as described in Section 4.3.2) to compute the Euclidean distance between them, using

$$D_{CN} = \sqrt{(x_C - x_N)^2 + (y_C - y_N)^2} \tag{4.6}$$

and

$$D_{CN}^W = \sqrt{(x_C^W - x_N^W)^2 + (y_C^W - y_N^W)^2} \ . \tag{4.7}$$

4.4.2 Feature Extraction from the Cytoskeleton as a Set of Line Segments

From the skeletonized version of the cytoskeleton of a given cell (as in Figure 4.1 (i)), line segments were detected automatically using an algorithm called LSD (Figure 4.1 (m)). LSD is a linear-time tool to locally detect straight contours on images, i.e., line segments, providing accurate subpixel results [77]. Contours are regions of the image where pixel intensities change fast enough from dark to light, or the opposite [77]. According to user-defined parameters (Table 4.3), LSD approximates regions of the image as line segments.

For a given cell where N_l lines were detected in total, the output of LSD corresponds to a set $L = \{((x_i^1, y_i^1), (x_i^2, y_i^2)) : i \in \{1, ..., N_l\}\}$ where each element i corresponds to the two extremities (1 and 2) of the i'th detected line. With the information provided by the 2D spatial distribution of line segments, resembling the popular traditional game Mikado, features can be devised to quantify several properties,

Table 4.3: Parameters involved in automatic line detection with LSD [20].

Method	Parameter	Description	Value/Mode
Line Segment Detector (LSD)	`refine`	The way found lines will be refined (in advanced mode, false alarms are calculated and lines are refined according to size, precision, etc)	advanced
	`scale`	The scale of the image that will be used to find the lines	2.5
	`sigma_scale`	Sigma for Gaussian filter (sigma = sigma scale/scale)	0.01
	`quant`	Bound to the quantization error on the gradient norm	0
	`ang_th`	Gradient angle tolerance in degrees	90
	`log_eps`	Detection threshold	-50
	`density_th`	Minimal density of aligned region points in the enclosing rectangle	0.1
	`n_bins`	Number of bins in pseudo-ordering of gradient modulus	2048

such as the local and global distribution of orientations, lengths, and positions. Some features require the use of the nucleus' centroid of the cell (identified in 4.3.2) as an "anchor point", even though any point (x, y) can be used. This family of line segment-based features is called LSF, summarized in Figure 4.2 along with their notations.

With the extremity points (x_i^1, y_i^1) and (x_i^2, y_i^2), a given line segment i can be converted into a vector,

$$\vec{l_i} = (x_i^2 - x_i^1, y_i^2 - y_i^1) \ . \tag{4.8}$$

Even though line segments are seen as a dispersion of vectors in 2D space, vector direction will not be relevant for any LSF since no additional information (e.g. microtubule polarity or microtubule growth direction) is available to justify assuming a direction. The Euclidean norm of $\vec{l_i}$ gives its line segment length,

$$L_i = ||\vec{l_i}||_2 = \sqrt{(x_i^2 - x_i^1)^2 + (y_i^2 - y_i^1)^2} \ . \tag{4.9}$$

The angle θ_i (in radians) was measured between l_i and the horizontal axis to visualize *orientation*, i.e.,

$$\theta_i = \begin{cases} \arctan\left(\frac{y_i^2 - y_i^1}{x_i^2 - x_i^1}\right) & \text{if} \quad y_i^2 - y_i^1 > x_i^2 - x_i^1 \\ \pi - \arctan\left(\frac{y_i^2 - y_i^1}{x_i^2 - x_i^1}\right) & \text{if} \quad y_i^2 - y_i^1 < x_i^2 - x_i^1 \end{cases} \ . \tag{4.10}$$

Furthermore, the distance D_i between a given point inside the cell and line l_i is defined as the Euclidean norm of the vector that goes from that point to the midpoint of this line. Figure 4.2 illustrates this reasoning for the particular case of the point being the centroid of the nucleus, with coordinates (x_N, y_N), but it can be generalized for any point (x, y) inside the cell. Mathematically, it is defined as

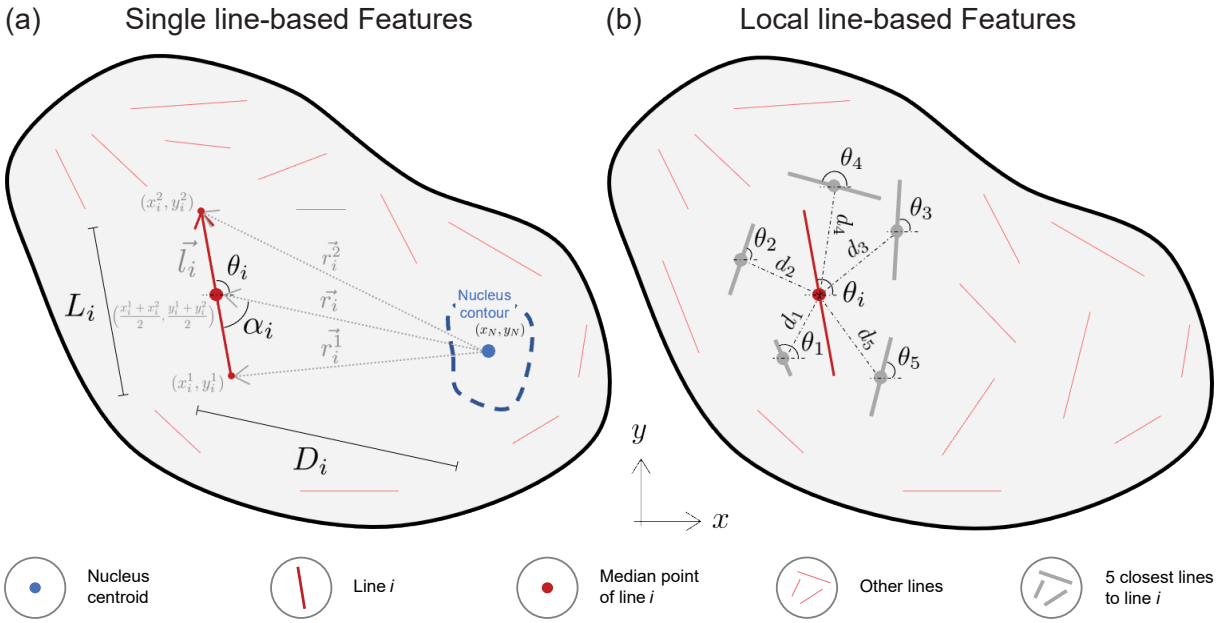


Figure 4.2: Schematic representation of single line-based features and local line-based features. For each and all line segments identified within a cell, (a) individual line-based features were calculated (with the centroid of the nucleus as the reference point) as well as (b) neighborhood-based features.

$$D_i = ||\vec{r_i}||_2 \quad \text{, where} \quad \vec{r_i} = \left(\frac{x_i^1 + x_i^2}{2} - x_N, \frac{y_i^1 + y_i^2}{2} - y_N\right) . \tag{4.11}$$

Still focusing on a given point (x_N, y_N), the smallest angle between $\vec{r_i}$ and $\vec{l_i}$, is defined as

$$\alpha_i = \min\{\hat{\alpha_i}, \pi - \hat{\alpha_i}\} \quad \text{, where} \quad \hat{\alpha_i} = \arccos\left(\frac{\vec{r_i}}{D_i} \cdot \frac{\vec{l_i}}{L_i}\right) , \tag{4.12}$$

and aims to quantify the *radiality* of line l_i from the perspective of (x_N, y_N) (or any point (x, y)). This angle is defined in $0 < \alpha_i < \pi/2$ and is $\approx 0°$ when l_i is oriented radially, and $\approx 90°$ when $\vec{r_i}$ and $\vec{l_i}$ are orthogonal. From any point (x, y) within the cell, all the angles can be considered, $\alpha = \{\alpha_i : i \in \{1, ..., N_l\}\}$, and radiality can be quantified through $RS_{x,y}$, defined as the probability of finding a line segment with $\alpha_i \in [0, 20]°$, i.e.,

$$RS_{x,y} = \frac{\#\{a_i : a_i \in [0, 20]°\}}{N_l} \, , \forall (x, y) \in \mathbb{N}^2 . \tag{4.13}$$

The radial score of a cell, RS, is found at the pixel (x^*, y^*) that maximizes $RS_{x,y}$,

$$RS = \max\{RS_{x,y} : \forall (x, y) \in \mathbb{N}^2\} . \tag{4.14}$$

Figure 4.3 illustrates how the radial score was computed. An evenly-spaced grid of pixels was overlayed on top of the cell's binary mask, similar to a subsampling procedure. This cell's binary mask was obtained

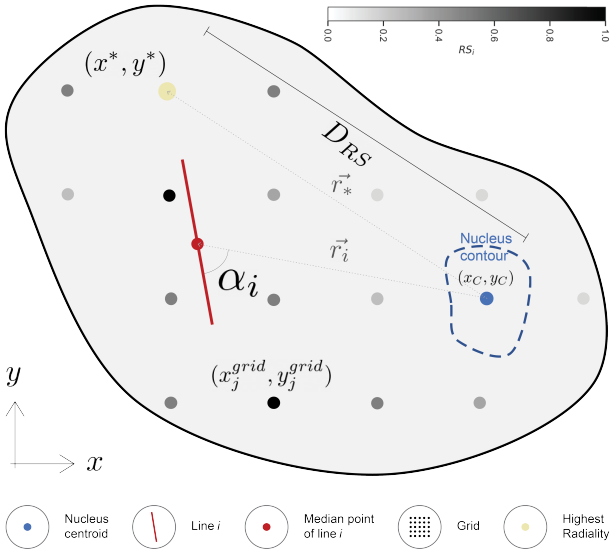


Figure 4.3: Schematic representation illustrating the measurement of the radial score. The point to be tested coincides with the nucleus centroid, but the radial score can be computed for any point in the grid.

after employing Canny's edge detector to the deconvoluted cell image patch as described in Section 4.4.1. Figure 4.3 shows that a value for $RS_{x,y}$ was computed for each coordinate in the grid. The pixel that yields the maximum $RS_{x,y}$ is the point where radiality is higher since lower α_i's were obtained.

Results were stored in a 2D array with 0's (background) and $RS_{x,y}$ evaluated for the pixel coordinates composing the grid. To obtain radiality maps and visualize radiality distribution within the cell mask, this 2D array was dilated with a square structuring element. The square width should be equal to the inverse of the subsampling factor; for instance, a grid with 5-pixel-spaced adjacent pixels should be dilated with a 5-pixel width square structuring element.

The Euclidean distance between the pixel with the highest radial score, (x^*, y^*), and the centroid of the nucleus located at (x_N, y_N) was measured using

$$D_{RS} = ||\vec{r_*}||_2 \quad \text{where} \quad \vec{r_*} = (x^* - x_N, y^* - y_N) \ . \tag{4.15}$$

to study *cytoskeleton-nucleus interaction* properties. The pixel (x^*, y^*) will be referred to as center of radiality.

Another subgroup of LSF adds local information, as suggested in Figure 4.2 (b). The pairwise Euclidean distances between line segment midpoints were calculated, and the c, $c \in \mathbb{N}$ closest line segments to a given line l_i were considered. The set of the c closest lines to line i was denoted as l_i^c.

To measure *parallelism*, the mean angle difference between neighbor lines was calculated by

$$\bar{\theta_i^c} = \frac{1}{c} \sum_{k=1}^{c} |\theta_i - \theta_k| \ , \tag{4.16}$$

with $\theta_k \in l_i^c,\ \forall k = 1, ..., c$. To measure *bundling*, the mean local line distance was used,

$$\bar{l_i^c} = \frac{1}{c} \sum_{k=1}^{c} d_k \ , \tag{4.17}$$

where d_k is the Euclidean distance between midpoints of line i and the k'th closest line midpoints, calculated with Equation 4.11.

Orientation was quantified with $CVar$ and OOP. To understand these metrics, line segment lengths will be neglected, focusing on the set of angles θ_i of a given cell, $\theta = \{\theta_i : i \in \{1, ..., N_l\}\}$. Each θ_i was converted to a unitary vector $\vec{v_i}$ defined as

$$\vec{v_i} = (\cos\theta_i, \sin\theta_i) \ . \tag{4.18}$$

These vectors can be averaged to obtain the mean resultant vector,

$$\bar{v} = \frac{1}{N_l} \sum_{i=1}^{N_l} \vec{v_i} \ .$$

The length of the mean resultant vector, $R = ||\bar{v}||_2$, is positively correlated with circular spread: the closer it is to one, the more concentrated the angles are around the mean direction [17]. Figure 4.4 illustrates the dispersion of two representative situations from [17].

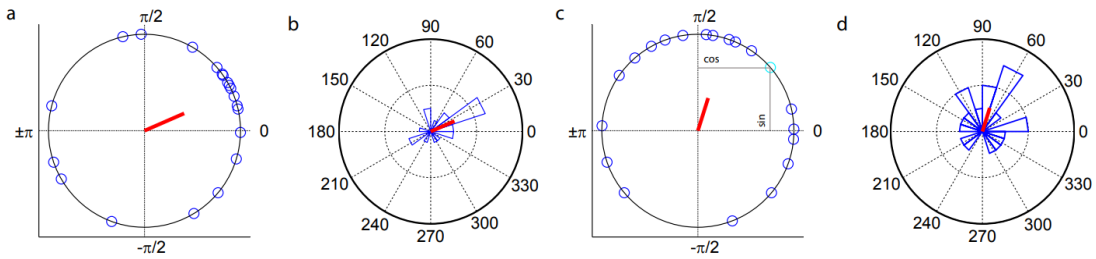


Figure 4.4: Data dispersion on the unit circle ((a) and (c)) and angular histograms ((b) and (d)) of two datasets A ((a) and (b)) and B ((c) and (d)) consisting of 20 samples. Red lines indicate the direction and magnitude of $\bar{r}$ [17].

Circular variance is, thus, defined as

$$CVar = 1 - R \ , \tag{4.19}$$

and is used in circular data statistical analysis and measures the spread of a set of angles. Therefore, it was used to quantify *orientation*

OOP was calculated according to Basu, A. et al. (2021) [12]. It was obtained from the order tensor,

$$\begin{bmatrix} \cos\theta_i \cos\theta_i & \cos\theta_i \sin\theta_i \\ \cos\theta_i \sin\theta_i & \sin\theta_i \sin\theta_i \end{bmatrix} \ .$$

Table 4.4: Line Segments Features.

Line Segment Features (LSF)					
2-Dimensional			1-Dimensional		
Notation	Description	Equation	Notation	Description	Equation
L_i^l	Line segment length	4.9	N_l	Number of lines	-
θ_i	Angle between line segment and horizontal axis	4.10	RS	Highest radial score	4.14
D_i	Distance between midpoint and nucleus' centroid	4.11	D_{RS}	Distance between $(x*, y*)$ and (x_N, y_N)	4.15
α_i	Smallest angle between anchor point and mid point	4.12	$CVar$	Circular variance	4.19
$RS_{x,y}$	Radiality from the pixel (x, y)	4.13	OOP	Orientational Order Parameter	4.20
θ_i^c	Mean angle difference between the c closest lines	4.16			
l_i^c	Distances of the c closest lines	4.17			

The mean order tensor (MOT) was calculated from the individual tensor of each line with angle θ_i, i.e.,

$$\text{MOT} = \langle 2 \cdot \begin{bmatrix} \cos\theta_i \cos\theta_i & \cos\theta_i \sin\theta_i \\ \cos\theta_i \sin\theta_i & \sin\theta_i \sin\theta_i \end{bmatrix} - \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \rangle \ .$$

Finally, OOP is defined as the maximum eigenvalue of the mean order tensor,

$$OOP = \max\{eigenvalue(\text{MOT})\} \ . \tag{4.20}$$

Figure 4.5 depicts the distribution of cytoskeleton fiber angles from a cell obtained by Basu, A. et al. (2021) [12], suggesting that OOP is lower in broad angular distributions and higher in narrow distributions.

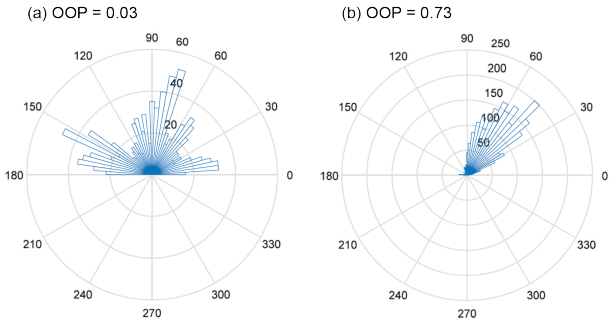


Figure 4.5: Angular distribution of fibers detected in two cells ((a) and (b)) and corresponding OOP values [12].

Table 4.4 summarizes the 2D and 1D line segment features extracted for a given cell. Notably, a 2D feature is referred to as a feature extracted for each line segment. Any 2D feature can be converted to a 1D one by applying statistics formulas. For instance, the mean, standard deviation, skewness, or kurtosis of line lengths, considering all identified line segments in a cell.

4.4.3 Feature Extraction from the Cytoskeleton as a Network

Before introducing the cytoskeleton as a network/graph representation, the fractal dimension, FD, was measured from the skeletons to quantify the degree of *complexity*. The nature of the fractal dimension does not perfectly fit the property profile of the features that compose each of the three categories (DCF, LSF, CNF). However, as the fractal dimension was measured directly on the skeletons of each cell and no line segments were extracted, by exclusion of parts and for the sake of document organization, the CNF category was considered the most suitable. Converting the skeletons into their graph representations in calculating the fractal dimension is unnecessary.

Minkovsky's box-counting method is widely used to estimate the fractal dimension of objects [18,88]. Briefly, an image is overlayed with a square grid (with squares of width q, for a total of $N(q)$ squares), and the number of squares with total non-zero intensity inside is counted. This calculation is repeated in the next iteration, but now with reduced width q. The method was repeated enough times until $q \to 0$, and the value of the fractal dimension was obtained from the limit

$$FD_{\phi} = \lim_{q \to 0} \frac{\log(N(q))}{\log(1/q)} \ . \tag{4.21}$$

When logarithms of box sizes are plotted against box counts logarithms in function of q, data points are spread almost linearly, and the value of FD_{ϕ} was estimated by the slope of the straight regression line [18].

The method implemented herein to estimate fractal dimension is similar to that of Revittser, A. et al. (2021) [18]. Before calculating FD_{ϕ}, an ellipse is fitted to the skeleton to find the major axis's orientation. Once found, the original image containing the cell was rotated so that the major axis became parallel to the horizontal axis. Then, the image was rotated by ϕ =15° clockwise six times, and the fractal dimension was estimated from each rotated image using 4.21. The final FD value is the average value from the seven measurements, i.e.,

$$FD = \bar{FD}_{\phi} \ . \tag{4.22}$$

Figure 4.6 illustrates this methodology.

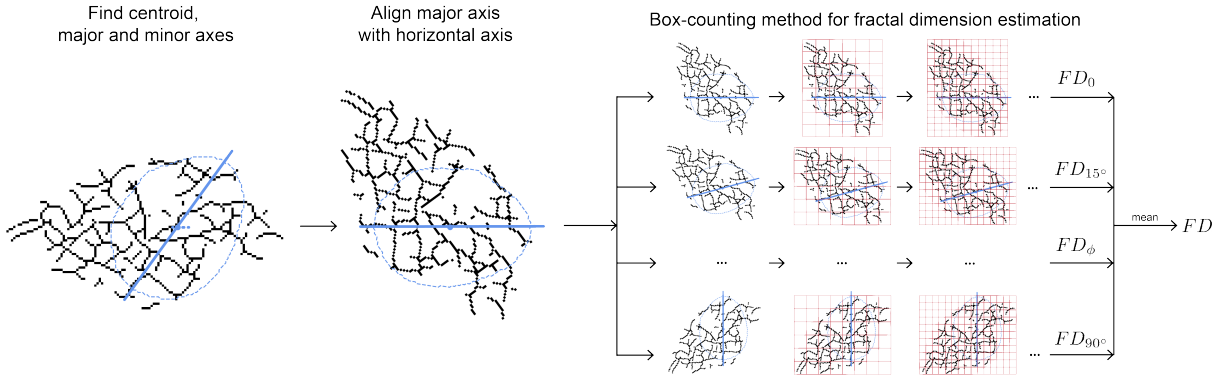


Figure 4.6: Method of Revittser, A. et al. (2021) [18] for cytoskeleton fractal dimension quantification to assess complexity.

The last class of features was calculated from the graph representation of the skeleton of a given cell. A graph is a structure made of vertices connected with edges. In this case, the concept of a graph was expanded, allowing the representation and quantification of the skeletons by calculating a new set of features, CNF, obtained in the light of graph theory concepts.

A skeleton can be seen as an undirected multigraph permitting loops, which is an ordered triple, $G = (V, E, I)$, where V and E are finite sets, called vertices and edges of G, and I maps each pixel to its intensity value. Nodes and edges were identified using the 4-connected Von Neumann neighborhood method to determine skeleton pixels that touch another pixel's edges, either vertically or horizontally. Such pixels were considered connected and, therefore, part of the same skeleton object [65]. This algorithm and other tools are included in Skan, a Python library to analyze skeleton images [16]. As Figure 4.7 (a) suggests, the output of Skan is a graph, a set of vertices connected by 1-pixel wide edges with an intensity distribution. Skan outputs each edge of G, e_i, which is a sequence of pixel coordinates that were encoded in a vector, given by

$$\vec{e_i} = ((x_i^1, y_i^1), ..., (x_i^{n_i}, y_i^{n_i})) \ \in V'^{n_i} \ .$$

The number of pixel coordinates composing an edge, n_i, varies for each i. Additionally, V' is a set of all pixel coordinates that compose G, with $V' \subset \mathbb{N}^2$. The set of edges is given by

$$E = \{\vec{e_i}, i = 1, ..., N_E\} \ .$$

For each i, the element (x_i^1, y_i^1) and $(x_i^{n_i}, y_i^{n_i})$ are called the extremities of the edge $\vec{e_i}$. The set of extremities of each edge is defined as

$$V = \{x_i^j \ , \ y_i^j \ \forall \, i \in 1, ..., N_E \ \ , \ \ j = \{1, n_i\}\} \ .$$

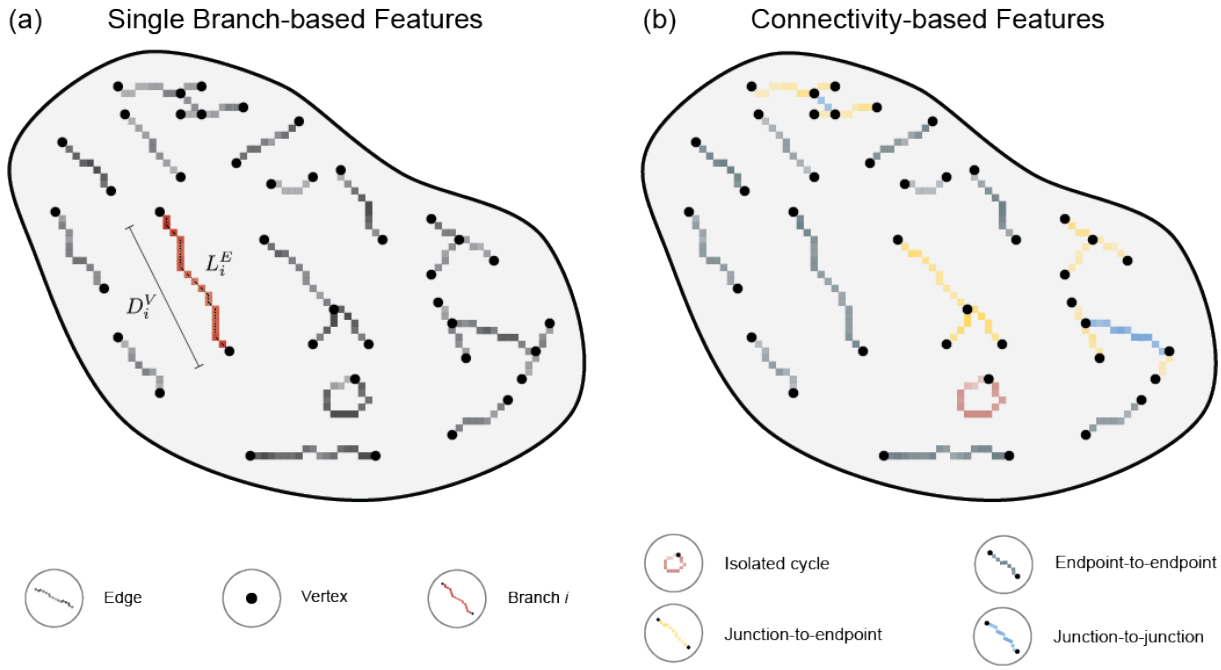


Figure 4.7: Schematic representation of connectivity features. (a) Features calculated for a given branch; (b) Branches classified based on connectivity.

Usually, a graph edge is represented as a straight line between its two vertices at the extremities. Here, an edge i of G represented by the vector $\vec{e_i}$ is actually defined as the 2D path starting at one extremity, located in (x_i^1, y_i^1), and following straight paths between adjacent pixels, (i.e., pixel (x_i^j, y_i^j) belongs to the 8-th neighborhood of pixel (x_i^{j+1}, y_i^{j+1}) $, \forall j \in 1, ..., n_i$) until the other extremity is reached, at $(x_i^{n_i}, y_i^{n_i})$. Therefore, it is ordered in the sense of adjacency, so "walking" in $\vec{e_i}$ from (x_i^1, y_i^1) to $(x_i^{n_i}, y_i^{n_i})$ establishes the same path as going in the opposite direction. Thus, edges link two vertices symmetrically, and G is an undirected graph. Additionally, an edge that joins a vertex to itself can also be outputted by Skan, i.e., $(x_i^1, y_i^1) = (x_i^{n_i}, y_i^{n_i})$, in structures called loops. Finally, G is a multigraph since two branches i and j, with $i \neq j$, can be found with the following properties:

$$\begin{cases} e_i^l = e_j^{l'} \ , \ l = 1, n_i \ , \ l' = 1, n_j \\ e_i^l \neq e_j^{l'} \ , \ c.c. \end{cases} ,$$

being e_i^l the l'th element of edge $\vec{e_i}$. This means it is possible that the same two vertices can be connected by multiple distinct edges (for example, $\vec{e_i}$ and $\vec{e_j}$). Pixel coordinates composing the edge $\vec{e_i}$ are unique compared to $\vec{e_j}$, except for the first and last pixels.

Finally, I is a function mapping a pixel belonging to the graph to its corresponding intensity value $w_{x,y}$,

$$I \colon V' \longrightarrow \mathbb{R}_0^+$$
$$(x, y) \longmapsto w_{x,y} \ .$$

The intensity variable was not addressed in detail as it was not possible to relate it to any specific aspect of cytoskeletal organization (Figure 3.4). However, the overall results were extracted for a separate analysis in the Appendix due to their relevance.

Figure 4.8 illustrates the above mathematical formalism. A cytoskeleton fiber will correspond to a branch (or path), this being an edge and its corresponding vertices. *Quantity* was measured by the total number of branches (or edges), i.e.,

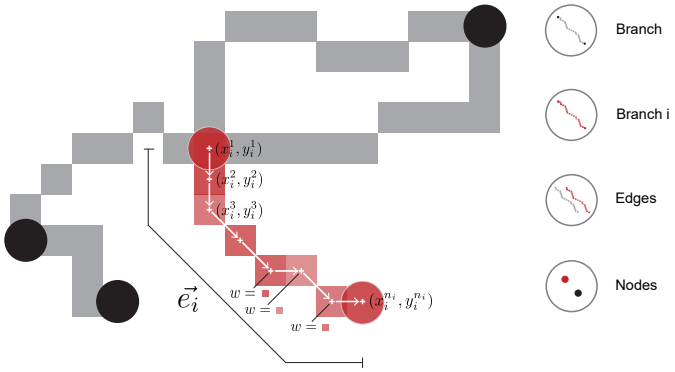


Figure 4.8: Skeleton represented as a graph. Each square represents a pixel. Red opacity is proportional to its intensity.

$$N_E = \#E \ . \tag{4.23}$$

Regarding fiber *morphology*, the length of branch e_i is given by

$$L_i^E = \sum_{j=1}^{n_i} \sqrt{(x_i^{j+1} - x_i^j)^2 + (y_i^{j+1} - y_i^j)^2} \ , \tag{4.24}$$

and the Euclidean distance between nodes is given by

$$D_i^V = \sqrt{(x_i^{n_i} - x_i^1)^2 + (y_i^{n_i} - y_i^1)^2} \ . \tag{4.25}$$

Additionally, contortion of a path can be evaluated with tortuosity, defined as

$$T_i = \frac{L_i^E}{D_i^V} \ . \tag{4.26}$$

The connectivity of a branch is related to its local neighborhood. Figure 4.7 (b) depicts the four

types of branches in terms of *connectivity*. Isolated cycles, or loops, are branches with only one vertex. Endpoint-to-endpoint branches are isolated. Junction-to-endpoint branches connect to only one branch, while junction-to-junction branches connect to other two or more branches. Connectivity is one of graph theory's basic concepts and may measure a network's resilience.

Concluding, CNF calculated in this work are shown in Table 4.5. Similarly to 2D LSF, 2D CNF can be converted to 1D features by applying statistics formulas; for instance, the mean, standard deviation, skewness or kurtosis of branch lengths.

Table 4.5: Cytoskeleton Network Features.

Cytoskeleton Network Features (CNF)					
2-Dimensional			1-Dimensional		
Notation	Description	Equation	Notation	Description	Equation
L_i^E	Branch length	4.24	FD	Fractal dimension	4.22
D_i^V	Euclidean distance between nodes	4.25	N_E	Number of branches	4.23
T_i	Tortuosity	4.26	$R_{(i.c)}, R_{(e,e)}$, $R_{(j,e)}, R_{(j,j)}$	Ratio of each connectivity types in a cell	-
$(i.c), (e,e)$, $(j,e), (j,j)$	Branch connectivity type	-			

4.4.4 Validation on Cells and Generalized Analysis

To further validate the applicability of the metrics developed to evaluate cytoskeletal architecture, each organization aspect was compared in small sets of cells with clearly different patterns (Figure 4.1 (o)). For instance, the organization in terms of fiber orientation was evaluated visually and quantitatively on three cell types: with aligned fibers, with disorganized fiber orientations, and with a combination of aligned and disorganized fibers

To investigate if microtubules of WT cells showed organizational differences compared to mutant cells, the masks of all non-excluded cells were manually segmented according to Section 4.3.4 (Figure 4.1 (k) and Figure 4.1 (o)). The exclusion criteria included cells touching borders, with a high degree of overlap, unclear cytoskeletal structures, and cells with missegmented nuclei. Using the ROIs of non-excluded cells, it was possible to process and compute the features (DCF, LSF and CNF) for all cells, saving their data values for analysis. Additionally, a cell label was assigned to each cell according to the experimental condition (either WT, NP, P1 or P2).

Statistical Analysis

A statistical analysis was applied to previously saved feature data. This approach aimed to identify statistically significant differences between WT cells and mutant cells for each of the extracted features

(Figure 4.1 (o)). A two-tailed unpaired t-test was performed to test for statistical significance, assuming that the values were normally distributed (Central Limit Theorem is valid since $n >> 30$) with unequal variances. The confidence intervals for these features are given by

$$CI = \bar{X} \pm Z\frac{\sigma}{\sqrt{n}} \ , \tag{4.27}$$

where $\bar{X}$ is the mean, Z is the Z-score for the desired confidence interval, σ is the standard deviation, and n is the number of observations. Using confidence intervals of 95% (Z =1.96), p-values were obtained and represented with asterisks ($*$) following the standard scale (ns (not significant): $p > 0.05$; $*$: $p \leq 0.05$; $**$: $p \leq 0.01$; $***$: $p \leq 0.001$; $****$: $p \leq 0.0001$).

5

Results and Discussion

Contents

In this Chapter, results obtained from the application of the described pipeline of (pre)processing methodologies will be presented and discussed. Cytoskeletal organization descriptors will be analy zed and quantified by the designed features (DCF, LSF and CNF), and compared in small groups of cells. After validating the strategy on these cells, cytoskeletal organization aspects differing between WT and mutant cancer cells will be interpreted in a global analysis performed on all non-excluded cells in the dataset.

5.1 Nuclei Segmentation and Centroid Identification

Nuclei segmentation was important to assign each nucleus to its respective cell and study the cytoskeleton-nucleus interaction. This task was performed with a high success rate confirmed by visual inspection. Still, exceptional cases of irregular nuclei morphology and outliers in terms of pixel intensity were present, whose segmentation was unsuccessful mostly due to over-segmentation. Examples of correctly and incorrectly segmented nuclei are shown in Figure 5.1.

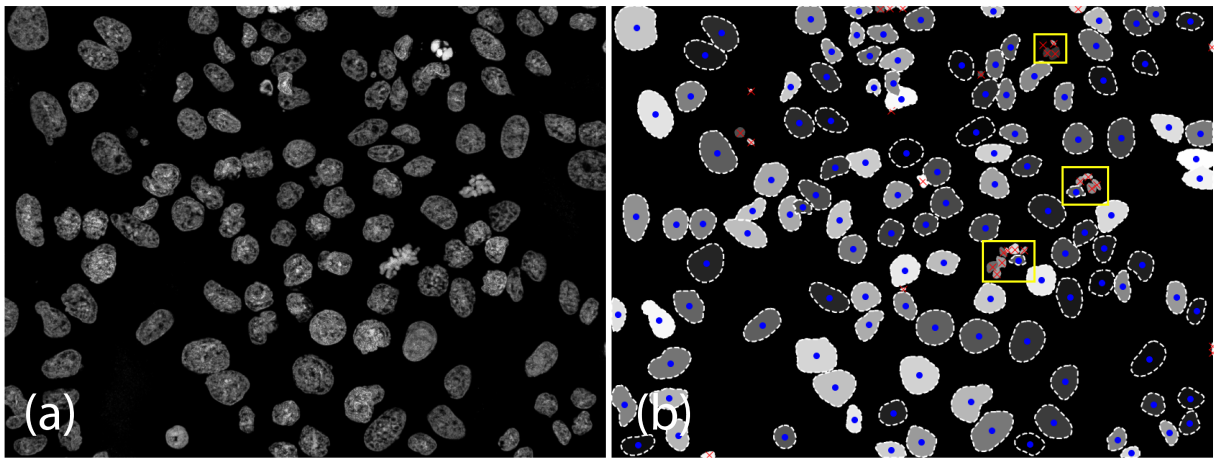


Figure 5.1: Nuclei segmentation and centroid identification using the Stardist pre-trained model with area threshold. (a) Original deconvoluted nuclei image; (b) Segmented nuclei image. The several monochromatic grayscaled blobs represent each individual nucleus mask. Blue dots depict their respective centroids, red crosses represent excluded nuclei, and white dashed lines delimit each nucleus contour. Yellow boxes correspond to examples of fragmented nuclei

Missegmentation cases, i.e., incorrectly segmented nuclei, could be explained by the limitations of StarDist's pre-trained model. This model was trained on grayscaled images of fluorescently labeled nuclei, similar to this work, but experimental settings like the magnification factor or cell line used in the acquisitions are unknown. Furthermore, available tuning parameters were insufficient to improve segmentation accuracy (verified by visual inspection). Another limitation is that StarDist's model could not correctly segment non-convex nuclei since it can only identify objects with such morphology, leading to over-segmentation.

To overcome these limitations and exclude incorrectly segmented nuclei, only those with an area above 600 pixels were considered. Nevertheless, in exceptional cases, this exclusion caused the only centroid identified for a cell to deviate from the ground-truth centroid (see fragmented nuclei in Figure 5.1).

In an attempt to obtain higher nuclei segmentation accuracies, an alternative strategy was developed. A thresholding step with Otsu threshold was included, followed by the detection of nuclei contours using a Canny edge detector, filling within them, and watershed segmentation to obtain binary masks of individual nuclei, as done in other works [32]. Unlike the StarDist model, this method segments nuclei with more irregular (non-convex) geometries. However, it was not sufficient to correctly segment overlapping nuclei, which existed in a considerable number in the dataset. Despite the advantages of this and other methods in the literature, nuclei segmentation was not the focus of this work. Thus, a comparative analysis with other state-of-the-art methods was not performed. If required, it would be necessary to implement these methods, make an appropriate choice of parameters to maximize performance and carry out a comparative analysis.

In summary, despite the effect of the above factors on nuclei missegmentation, the success rate of this method was above 90%, verified by visual inspection. Cells whose nuclei were missegmented or whose centroid was miscalculated were excluded from analysis.

After nuclei segmentation, it was essential to correspond each nucleus to its respective cytoskeleton. For a cell to be considered, it was not enough that its nucleus met the inclusion criteria. The cytoskeleton of the cell had to be mostly visible in the image frame, away from its margins. Cells that simultaneously met the inclusion criteria for both nucleus and cytoskeleton were considered for analysis, rendering a total of 652 cells, whose distribution is as follows: 176 (27.1%) WT cells, 155 (23.7%) NP cells, 197 (30.3%) P1 cells and 124 (19.0%) P2 cells.

5.2 Cytoskeleton Preprocessing

Fluorescence microscopy is used for many biological applications enabling the analysis of multiple cellular structures under distinct experimental conditions. For the particular case of cytoskeletal structures, this technique cannot reach a resolution that permits the precise localization of each cytoskeletal filament [56]. Indeed, it is known that microtubules have a lateral width of $\approx$24 nm, while actin microfilaments have $\approx$7 nm, which are both far beyond the resolution limit of fluorescence microscopy [56, 60]. Nevertheless, as performed in this work, fluorescence microscopy is a powerful technique and has been the strategy of choice to investigate cytoskeletal structures [10, 11, 83].

Despite the good temporal resolution of fluorescence microscopy, motion, blur effect and noise were still present in the original images (Figure 5.2, original $1Z$ image column). Thus, the preprocessing

pipeline aimed to nullify these effects without losing the resolution of cytoskeletal fibers. Deconvolution as the first step was crucial to remove motion, blur effects and noise from the original image and better separate the background from proper cell content (Figure 5.2, deconvolution column). In some cases, even though noise was present in the original RGB images, cytoskeletal fibers of some cells were clearer than in their deconvoluted versions since MIP was used, and resolution was lost. Nevertheless, the efficient noise removal justified the use of this technique even if loss of resolution was observed in some cells. Additional preprocessing methodologies were subsequently applied to the immunofluorescence images. Counter-intuitively, applying a Gaussian kernel further decreased the resolution of the cytoskeletal protein α-tubulin, but was beneficial for uniforming the signal within the fibers and further removing residual noise (Figure 5.2, Gaussian filter column). Next, the Sato filter was vital to enhance curvilinear structures, i.e., the cell's cytoskeletal filaments (Figure 5.2, Sato filter column). Of note, pixel intensity is proportional to the similarity of the filament to a tube [70]. Therefore, image binarization with a Hessian filter resulted in a mask of filaments that were closer to such geometry (Figure 5.2, Hessian filter column) by considering filaments with higher pixel intensities in the Sato image. Cytoskeletal structures are tri-dimensional fibrous networks, so with binarization, the tri-dimensional appearance of this structure is lost, causing frequent filament overlaps, which is a huge limitation as missegmentation was more prone to happen, as well as obtaining shorter filaments. Thus, skeletonization reduced binary objects to 1-pixel-wide representations and was crucial to obtain the representation of the skeleton as a graph and to identify line segments with LSD (Figure 5.2, skeletonization column). Finally, the overlayed skeleton column of Figure 5.2 gathers the input and output of the whole preprocessing methodology: nucleus contour and centroid are represented in blue, while the final skeleton (green channel) is multiplied by and overlayed above the original image (red channel).

5.3 Cytoskeleton Processing

The features devised in this work were designed to quantify several aspects of cytoskeletal organization. When processing the cytoskeleton, its structure was interpreted not only as a grayscale intensity distribution, but also as a 2D arrangement of line segments or a graph. Figure 5.3 depicts three examples of line segment and graph representations on skeletons from cells with distinct cytoskeletal structures.

Adjusting LSD parameters was a challenging trade-off problem, as higher refinement detection degrees could better capture the skeleton's morphology, but lower refinement degrees helped discard negligible skeleton paths. Figure 5.4 shows how different refinement degrees can impact line detection - adopting a more refined segmentation strategy results in detecting fragmented, smaller lines, with less sensitivity to longer ones. In contrast, smaller fibers can be ignored in a less refined analysis, and lines with longer lengths will be identified, which can better describe the structure of the cytoskeleton, but

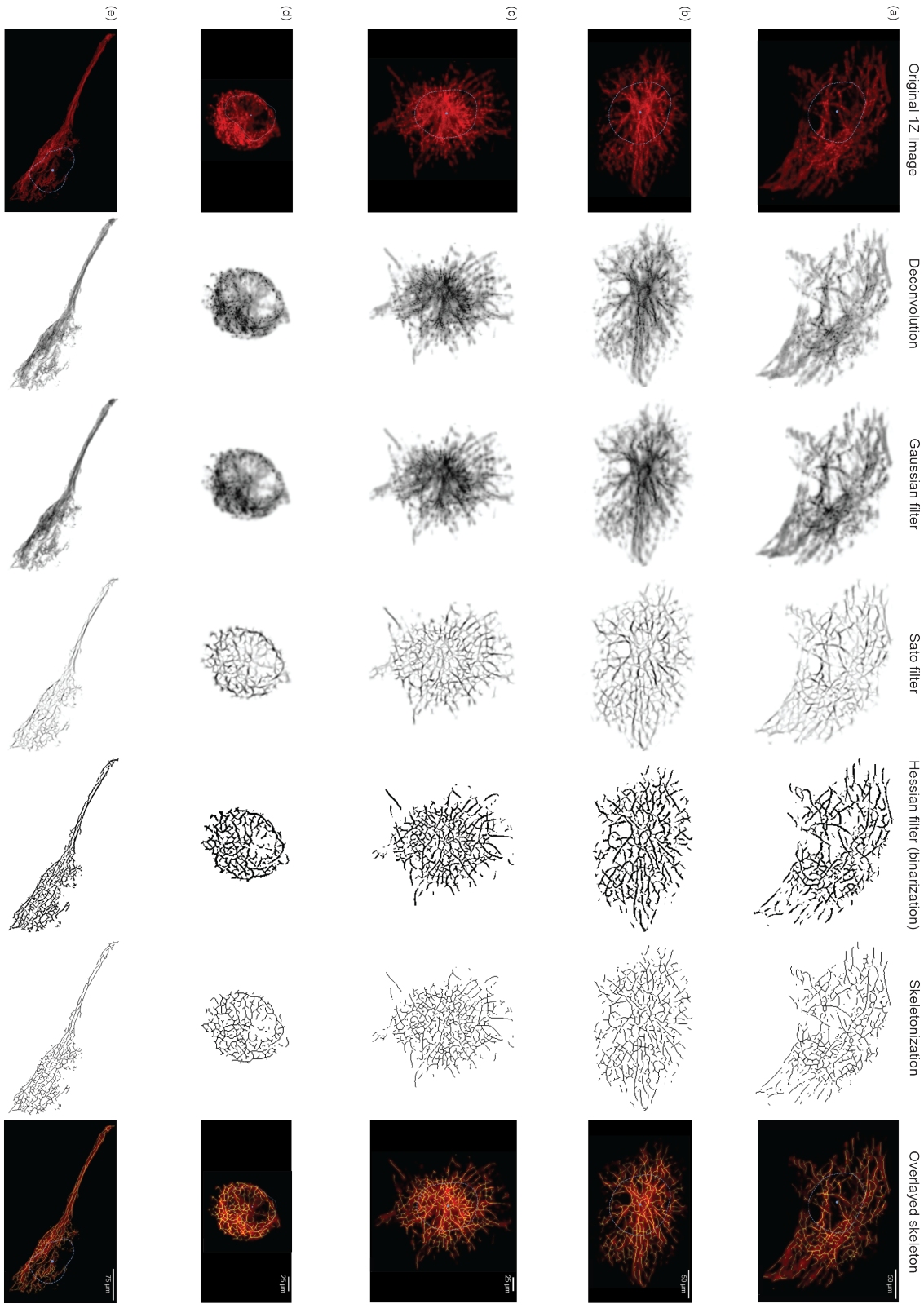


Figure 5.2: Cytoskeleton preprocessing byproducts in five cells with distinct cytoskeletal organization. Images were inverted in all columns except the first and the last for a better visualization.

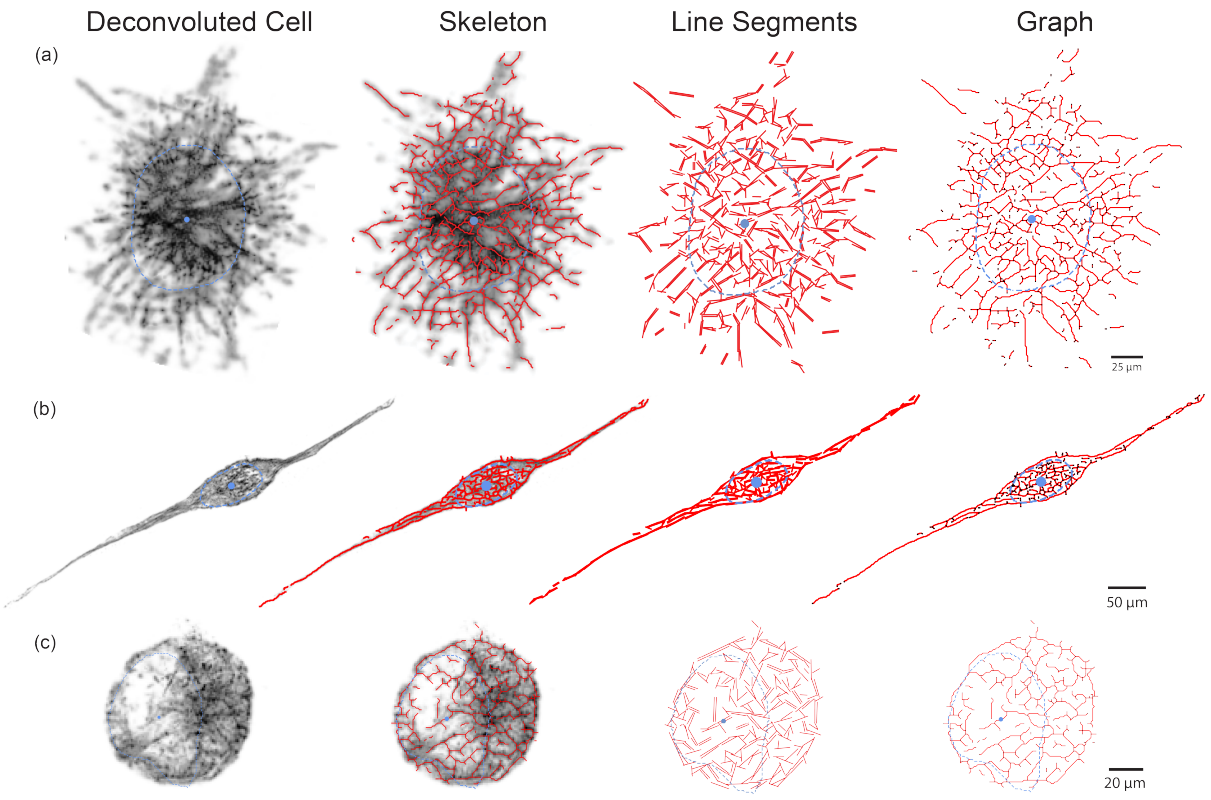


Figure 5.3: Cytoskeleton processing in cells with distinct cytoskeletal organizations. Three representative examples are shown in (a), (b), and (c). The first two columns represent the deconvoluted cells and the skeleton obtained after preprocessing overlayed above it, respectively. The third column shows the line segment detection with LSD. The fourth column depicts the graph representation of the skeleton, where black dots represent nodes, while the red pixelated paths indicate graph edges.

smaller fibers are lost.

An important feature of LSD is that this algorithm automatically detects the lines in a sub-pixelated contour image resulting from the input image. The algorithm will first draw the outer contour of this shape and identify the lines on this contour, and not directly within the fiber itself. For example, drawing the contour on a completely straight cytoskeletal fiber will cause two line segments to be identified and not just one. On the one hand, this factor causes over-identification of line segments. The advantage of using the resulting sub-pixelated contours from the input image is that it allows the identification of straight lines in more directions. If one were to hypothetically perform line detection directly at the $\top$ shape of Figure 5.4, there would be a very high chance of not being able to identify any lines at all, since a $\top$ shape does not resemble a line segment. Thus, lines were detected as shown in the middle case of Figure 5.4, corresponding to a middle-term refinement degree with parameters adjusted by visual inspection.

In contrast, the graph representation is straightforward once the skeleton is obtained. This representation allowed to make an individual analysis of the fibers, each one being a set of pixels defining a path

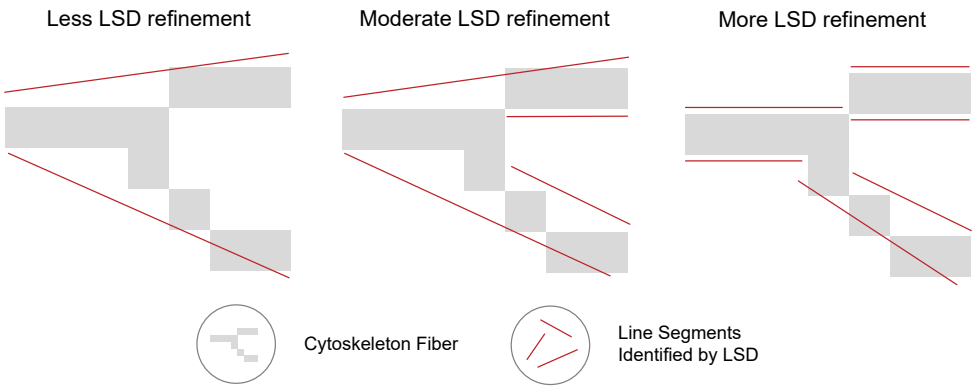


Figure 5.4: Skeletonized fibers can be segmented with different refinement degrees, leading to different segmentations.

between two nodes, where each pixel has a different intensity value (Figure 5.3).

5.4 Cytoskeleton Organization of Cancer Cells

The pipeline developed in this study aimed to investigate the cytoskeleton structural features associated with cancer cell invasion. For this purpose, E-cadherin mutant cells were used, leading to distinct adhesion and invasion phenotypes. Specifically, CHO cells were transiently transfected with WT or the mutant E-cadherin variants NP, P1, and P2. These cells were grown in laminin, a central ECM component [42], to closely mimic *in vivo* conditions.

Interestingly, from the analysis of the image dataset, a panoply of cytoskeletal structures was observed in cells from WT, NP, P1, and P2 groups. These cytoskeletal structures were shown to be dispersed in a radial, elongated, fusiform, and round profile, with and without projections, among other geometries and characteristics. However, before any computational strategy, it was not possible to identify evident trends toward a specific type of cytoskeletal architecture according to the experimental condition. The following sections describe the results obtained on quantifying cytoskeletal organization features of invasive and non-invasive cells based on the developed computational pipeline. Specifically, data on $\alpha-$tubulin expression patterns are presented concerning the following parameters: *orientation*, *bundling*, *parallelism*, *morphology*, *compactness*, *connectivity*, *quantity*, *complexity*, *radiality*, and *cytoskeleton-nucleus interaction* (Figure 3.4). For each structural feature, single and global analyses are presented.

Orientation

One descriptor of cytoskeletal organization is orientation. Figure 5.5 depicts three cells with different fiber orientation profiles, visualized using θ and quantified with OOP and $CVar$. The distribution of θ of each

line segment made it possible to observe the orientation profile of the fibers of a cell. The quantification of the disorganization of this angular profile was done through the one-dimensional metrics OOP and $CVar$, which presented lower and higher values in disorganized cells, respectively, (Figure 5.5 (a)) and the opposite in cells with aligned fibers (Figure 5.5 (c)).

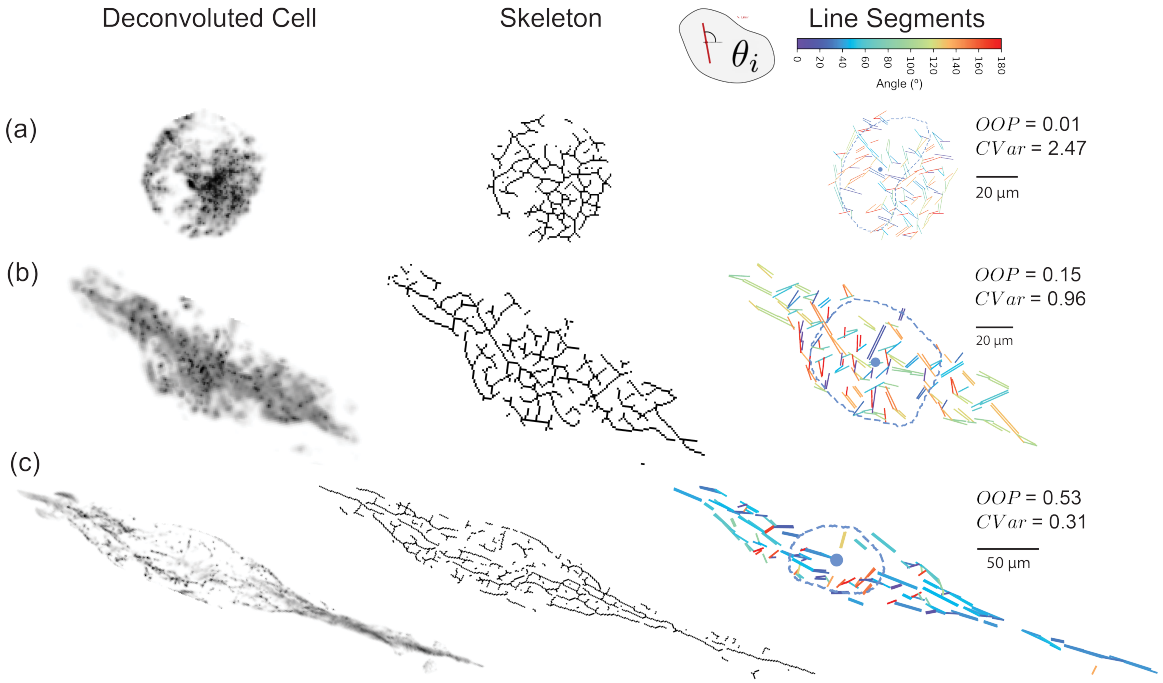


Figure 5.5: Fiber orientation of cytoskeletal structures. Three representative examples are shown in (a), (b), and (c). The angle θ of the line segments can reveal the distribution of fiber orientations across the cell; (a) cell with random orientation distribution; (b) cell with moderate orientation dispersion; (c) cell with aligned fibers, showing an organized orientation profile. Line segments are colored according to the angle θ_i. Cytoskeletal structures were obtained by α-tubulin staining.

The analysis of fiber orientation profiles and their quantification may allow the study of cellular phenomena involving fiber alignment. Indeed, Basu, A. et al. (2021) [12] used OOP to quantify the gradual increase of stress fiber alignment in the progression of mesenchymal phenotypes in lung cancer cells during EMT progression [12]. In those cells, this metric identified intermediate EMT states based on actin cytoskeletal orientations [12]. Thus, our data reinforce the idea that this feature could be an important tool to study cytoskeletal fiber orientations and help define its role in cancer progression.

Global Analysis

The results for the orientation feature using OOP and $CVar$, in which all cells of the dataset were separated by experimental condition, are shown in Figure 5.6.

OOP was obtained from the angular distributions of lines. Narrow angular distributions of well-aligned and broadly-distributed orientations correspond to a high and low OOPs values, respectively. Results

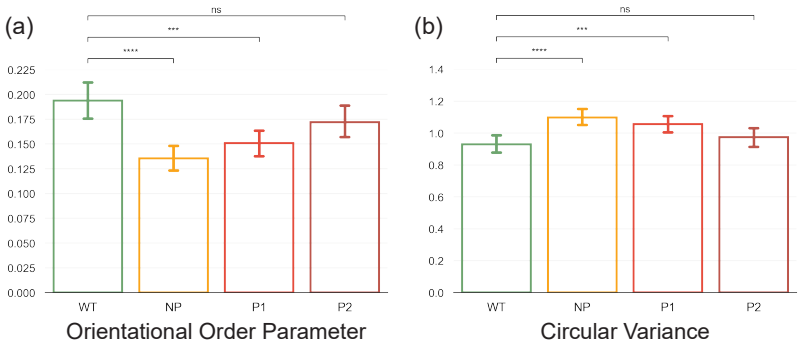


Figure 5.6: Quantification analysis of fiber orientation of cytoskeletal structures via OOP (a) and $CVar$ (b). Bars represent the mean $\pm$ 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

show that mutation groups NP and P1 presented broader angular distributions of microtubules since their OOP value is significantly lower than that of WT cells. However, mutation group P2 did not show statistically significant differences.

$CVar$ measures angular variations about the mean direction. Higher values indicate a higher spread of the cell's line orientation histogram, meaning that the orientations are highly dispersed around the mean. Similar to OOP, both NP and P1 groups have more dispersed patterns of microtubule orientations compared to WT cells, as their $CVar$s are increased. Again, mutation group P2 did not show statistically significant differences.

In conclusion, these results indicate that the cytoskeletal structures of NP and P1 cells are more disorganized in terms of orientation when compared to WT cells, with no significant differences observed in P2 cells.

Bundling

Fiber bundling was also studied, and its patterns in a cell were observed by calculating the average local line distance for a given line segment. Figure 5.7 depicts three cells with different bundling degrees, colored according to this feature. Bundling quantification was performed using two one-dimensional metrics: $Skew$ and CV of the cytoskeleton's intensity distribution. Both presented lower values in cells with compact but dispersed fibers (Figure 5.7 (a)) and higher values in cells forming bundles (Figure 5.7 (c)).

Global Analysis

These metrics did not reach a consensus when encompassing all cells in the dataset. The skewness of the pixel intensity distribution of NP, P1 and P2 cells were lower ($p = 6.15 \cdot 10^{-4}$), higher ($p = 2.43 \cdot 10^{-7}$) and without significant differences ($p = 1.29 \cdot 10^{-1}$), respectively, compared to WT cells. CV only

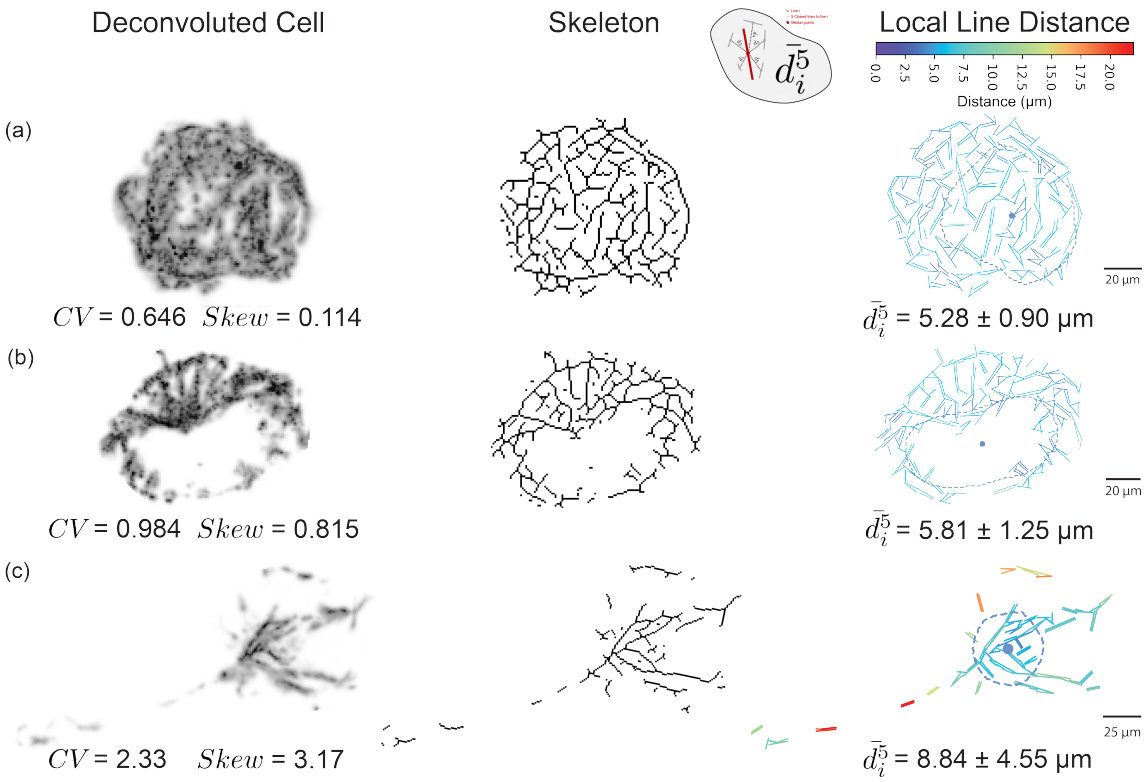


Figure 5.7: Fiber bundling of cytoskeletal structures. Three representative examples are shown in (a), (b), and (c). Feature $\bar{d}_i^5$ allows the visualization of bundles across the cell and reveals regions where lines were locally closer to each other; (a) cell with less bundling; (b) intermediate state; (c) cell with bundled fibers. Line segments are colored according to the angle $\bar{d}_i^5$. Cytoskeletal structures were obtained by α-tubulin staining.

remained lower in NP cells ($p = 3.34 \cdot 10^{-2}$), showing no significant differences in P1 and P2 cells. Due to page limit issues, these results were not further explored.

Parallelism

The local degree of parallelism was visualized in cells (Figure 5.8) by averaging the angle difference between line i and the nearest $c = 5$ lines, $\bar{\theta}_i^5$. The smaller the value of $\bar{\theta}_i^5$, the higher the degree of organization in terms of angle similarity, i.e., parallelism. To globally quantify the degree of parallelism, the average of $\bar{\theta}_i^5$, $\forall\ i \in \{1, ..., N_l\}$ was calculated. In Figure 5.8, cell (a) presented a high degree of global parallelism ($\bar{\theta}^5$ = 38.6°) since most of the fibers were horizontally oriented. Cell (b) also showed a high degree of parallelism (with $\bar{\theta}^5$ = 42.4°) due to the long protrusion with high bundling of fibers. The region around the nucleus was disorganized in terms of line orientations, contributing to an increase in the value of $\bar{\theta}^5$. Finally, cell (c) was round with a more random fiber distribution, corroborating this observation by the high value of $\bar{\theta}^5$ = 65.5°.

Notably, $\bar{\theta}_i^c$ and $\bar{\theta}^c$ are comparable to PAD and TAD in the work of Liu, Y. et al. (2018) [73]. However,

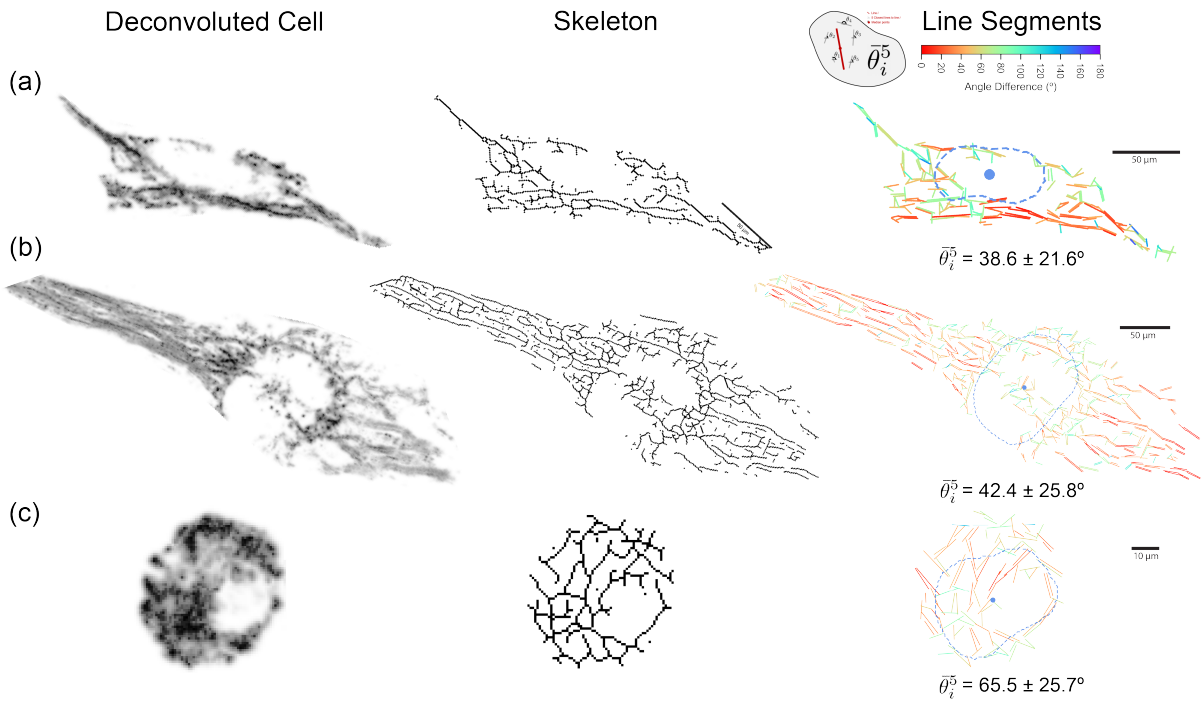


Figure 5.8: Fiber parallelism of cytoskeletal structures. Three representative examples are shown in (a), (b), and (c). The local degree of parallelism can be visualized across the cell with $\bar{\theta}_i^5$ in line segment rearrangements. (a) cell with high parallelism; (b) intermediate state; (c) cell with less parallelism. Line segments are colored according to the angle $\bar{\theta}_i^5$. Cytoskeletal structures were obtained by α-tubulin staining.

in this work, the c closest lines to a given line segment were used as sample data, instead of gathering them from sub-images to compute these metrics.

Global Analysis

Upon inclusion of all cells from the dataset, a global analysis was performed. The value of $\bar{\theta}^5$ was measured for all cells, obtaining the means and confidence intervals for this feature, separated by group as demonstrated in Figure 5.9. NP cells showed on average higher $\bar{\theta}^5$ values than the WT cells ($p = 8.51 \cdot 10^{-5}$), indicative of a lower degree of parallelism. This result corroborates the verification of these cells' greater disorganization of fiber orientations. The same was observed for the fiber orientation of P1 cells, which did not present significant differences in the degree of parallelism $\bar{\theta}^5$ in comparison to WT cells ($p = 2.49 \cdot 10^{-1}$). Finally, P2 cells had a slightly higher average value ($p = 4.61 \cdot 10^{-2}$), something that was not possible to verify in the orientation with OOP and $CVar$.

Morphology

Fiber morphology was analyzed in terms of length and level of contortion. Regarding fiber length, it was obtained by determining each line segment length L^l and edge distances L^E in graph representations.

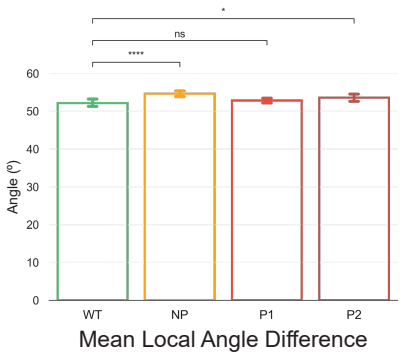


Figure 5.9: Quantification analysis of fiber parallelism of cytoskeletal structures. Bars represent the mean ± 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

Figure 5.10 shows three cells with distinct cytoskeleton length distributions. Results demonstrate that cell (a), which is smaller in size compared to (b) and (c), had the shortest filaments. In contrast, cell (c), with increased size, had longer filaments.

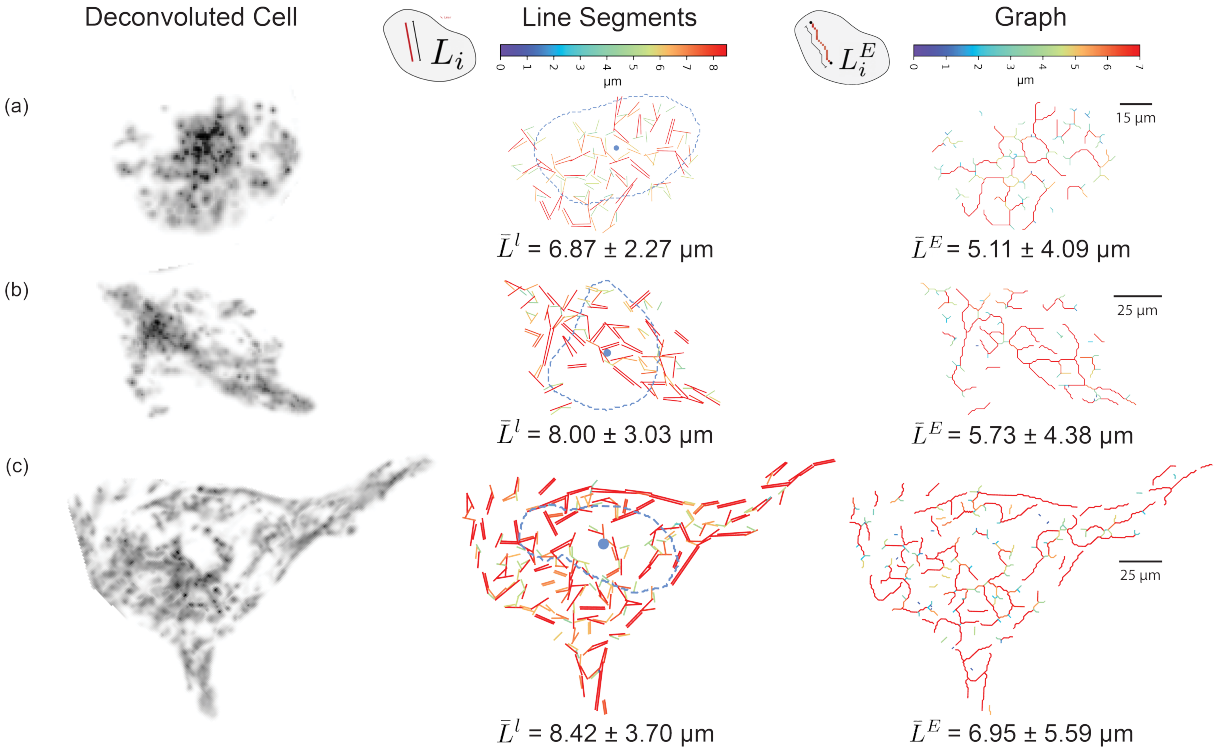


Figure 5.10: Fiber morphology of cytoskeletal structures. Three representative examples are shown in (a), (b), and (c). Fiber lengths can be visualized across the cell with L_i^l in line segment rearrangements and L_i^E in graph representations; (a) cell with shorter mean lengths; (b) cell with intermediate mean lengths; (c) cell with longer mean lengths. Line segments and graph edges are colored according to the value of L_i^l and L_i^E, respectively. Cytoskeletal structures were obtained by α-tubulin staining.

Notably, evaluation of fiber lengths can be helpful for performance assessment of drugs that inhibit

the polymerization of cytoskeletal proteins and prevent cancer progression [26].

Global Analysis

Microtubule morphology was one of the cytoskeletal features with statistically significant differences between WT and all groups of mutant cells, and for which methods of different natures (line segments vs. graphs) converged to the same biological conclusions. In particular, cells from NP, P1, and P2 groups had lower mean line segment lengths consistently compared with WT cells (Figure 5.11 (a)). Additionally, average graph edge lengths were also lower in all groups of mutant cells (Figure 5.12). In conclusion, cells with mutant E-cadherins had lower mean microtubule lengths compared to WT cells.

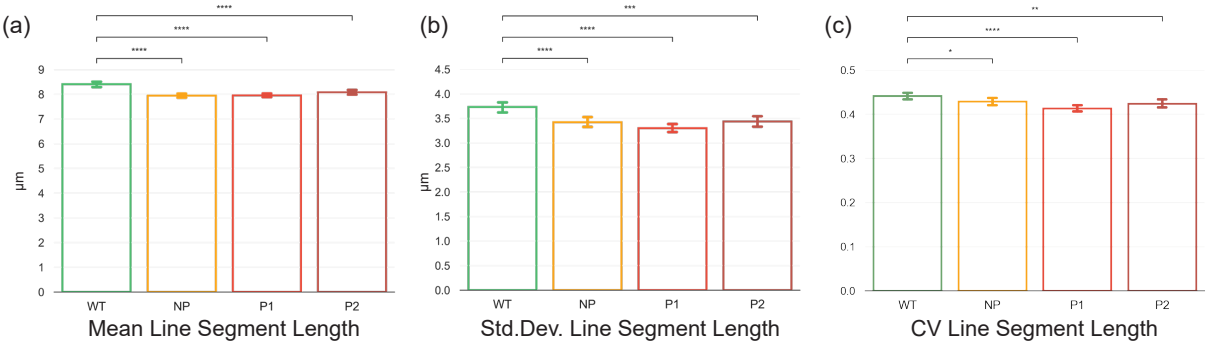


Figure 5.11: Quantification analysis of mean line lengths of cytoskeletal structures. (a) Mean line length; (b) Standard deviation of line lengths; (c) CV of line lengths. Bars represent the mean $\pm$ 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

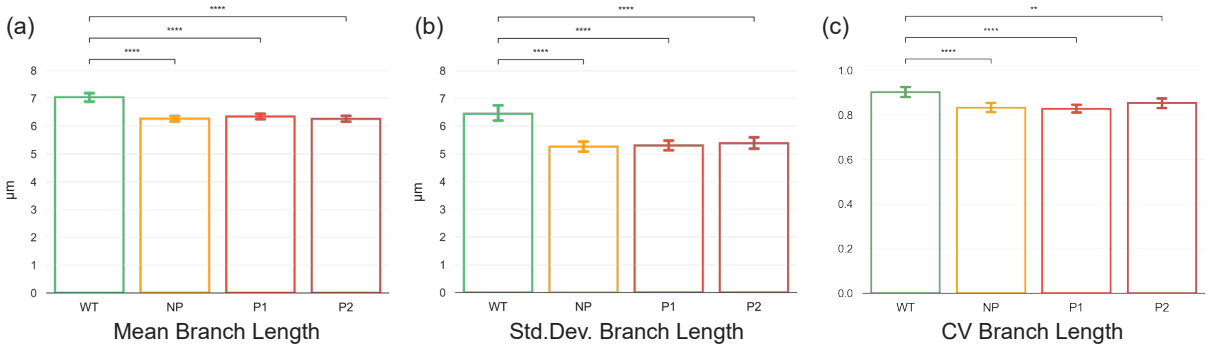


Figure 5.12: Quantification analysis of mean branch distance of cytoskeletal structures. (a) Mean branch distance; (b) Standard deviation of branch distances; (c) CV of branch distances. Bars represent the mean $\pm$ 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

Of note, the average length of the filaments of a cell estimated by approximating them by straight line segments was consistently higher than that estimated by graph representation. Over the whole dataset, the averages of both methods differ by an absolute error of 1.6 μm, which is still a considerable value.

This deviation is not only due to the different natures of the two methods but also to the parameter tuning process during refinement of LSD. These average length values are close to the lower bound of the individual microtubule length range, [5, 40] µm, found in CHO cells in other studies [98, 99].

The standard deviation of these values was also lower in all groups of mutant cells (Figure 5.11 (b) and Figure 5.12 (b)), indicative of the greater microtubule length uniformity in mutant cells. Since cell size is not taken into account in these calculations, and in order to avoid biased results, the coefficient of variations for the mean line segment lengths and mean graph edge length were also calculated (Figure 5.11 (c) and Figure 5.12 (c)). Herein, lower values were consistently obtained in mutant cells. Thus, mutant cells presented a profile of microtubule lengths with less dispersion around the mean, and the opposite for WT cells which presented a wider distribution of lengths around the mean.

Overall, statistically significant differences were found regarding cytoskeletal fiber morphology, as determined by α-tubulin structural pattern. In particular, on average, WT cells had longer microtubules, with a less uniform length profile, compared to mutant cells. Interestingly, Costigliola, N. et al. (2017) [15] found shorter vimentin filaments (intermediate filaments) in cytoskeletal structures of human foreskin cells during the mesenchymal stage of EMT. These results contrast the ones observed in this study for α-tubulin (microtubules), suggesting that different cytoskeletal components are modulated distinctly.

Contortion

Fiber contortion was determined by computing the tortuosity of each edge on the skeleton graph. Figure 5.13 illustrates another set of three cells with fibers with different contortion levels. The three cells indicate that the vast majority of graph branches had tortuosity levels approximately within the interval $\bar{T} \in$ [1.0,1.3]. However, more tortuous filaments were also observed. Cell (a) mainly had linear segments, while cell (c) had fewer branches but a higher fraction of more tortuous ones. The mean tortuosity of cell (b) is closer to the average $\bar{T}$ of all cells of the dataset, of 1.11$\pm$0.02. Similar values were obtained in other investigations, such as in microtubules of adult rat ventricular myocytes where average tortuosity was approximately 1.105 [100].

Overall, no significant differences were found between the four experimental conditions regarding tortuosity.

Compactness

Two metrics measured the compactness of a cell's cytoskeleton: the number of identified line segments per unit area, N_l/A_C, and the Euclidean distance between vertices in the graph representation. Figure 5.14 depicts three cells with different compactness levels. Cell (a) had a sparse cytoskeleton morphology, as the fibers were well dispersed and clearly visible. Consequently, lower N_l/A_C ratios were obtained, and wider Euclidean distances were observed. On the contrary, cell (c) had a denser

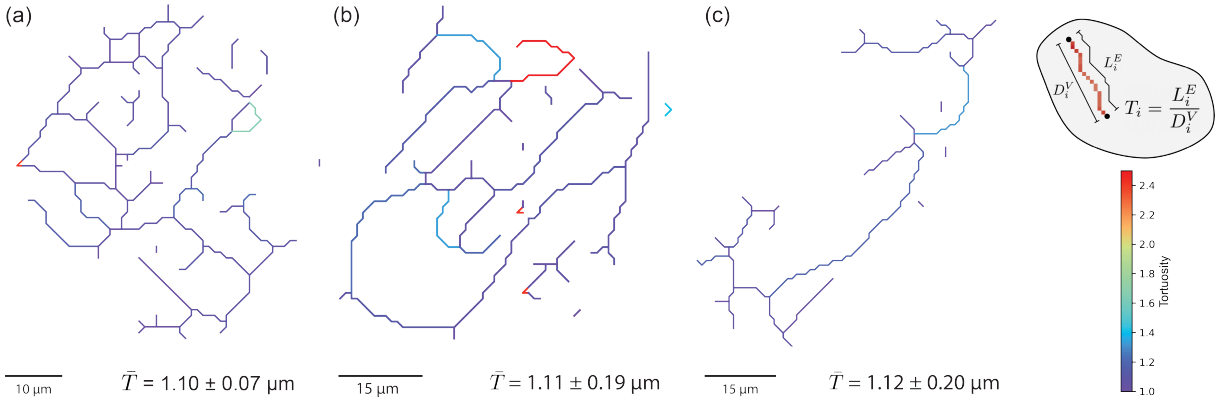


Figure 5.13: Branch tortuosity of cytoskeletal structures. Three representative examples are shown in (a), (b), and (c). Branch tortuosity is calculated by dividing its actual length by the Euclidean distance between nodes. Graph edges are colored according to the T_i value. Cytoskeletal structures were obtained by α-tubulin staining.

cytoskeleton morphology, as the fibers were closer to one another occupying less volume, thus obtaining an increased N_l/A_C ratio and decreased $\bar{D}_V$.

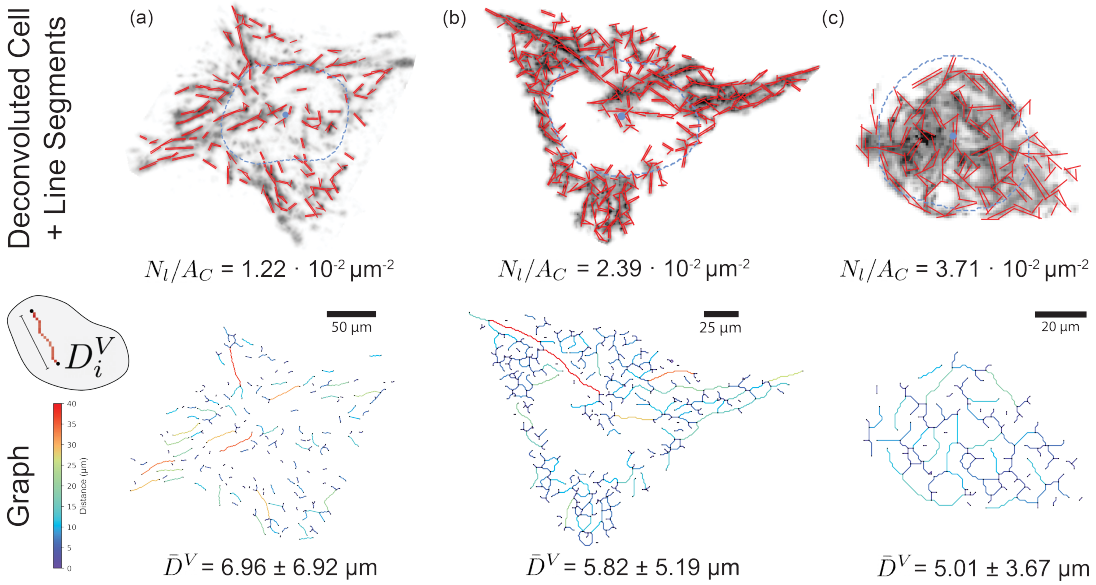


Figure 5.14: Fiber compactness of cytoskeletal structures. Three representative examples are shown in (a), (b), and (c). (a) Cell with sparse cytoskeleton; (b) Cell with moderate compactness; (c) Compact cytoskeleton. Graph edges are colored according to the D_i^V value. Cytoskeletal structures were obtained by α-tubulin staining.

The use of $\bar{D}^V$ as a metric to evaluate compactness deserves a more profound interpretation. This metric only considers the distance between the endpoints of a given filament, not providing information about the morphology of the filament (edge) itself. Between these endpoints, the edge may be slightly or very tortuous. For these reasons, this metric was not included in the morphology section. However,

given the relevance of these results in the present study, $\bar{D}^V$ was introduced as a compactness quantifier as it is closer to this aspect of organization comparing to the other nine parameters considered in the model.

Cells under mechanical stress conditions (like compressive stresses) may cause cytoskeleton compactness modifications [101] that can be visualized and quantified by the metrics developed.

Global analysis

Notably, statistically significant differences were found in cytoskeleton compactness across groups. Compactness quantification results by calculating the N_l/A_C ratio of cells are shown in Figure 5.15. All three groups of mutant cells presented significantly more compact cytoskeletons than WT cells, as more lines were detected per cell area in these cells.

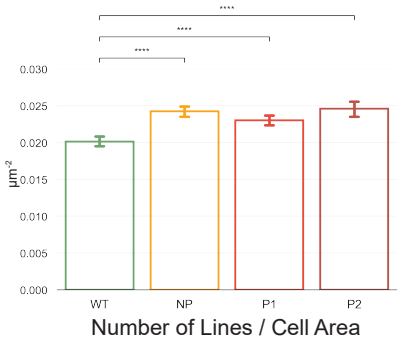


Figure 5.15: Quantification analysis of the number of lines of cytoskeletal structures per cell area. Bars represent the mean ± 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

Measuring cytoskeleton compactness using the Euclidean distance between adjacent nodes also yielded statistical differences between groups, as shown in Figure 5.16. Figure 5.16 (a) and (b) indicate that all three groups of mutant cells had lower mean and standard deviations of cellular Euclidean distance distributions, respectively. This means that nodes in graph representations of mutant cell's cytoskeletal structures are, on average, closer to one another compared to WT cells. Additionally, lower standard deviations indicate that these distances have a more uniform distribution in mutant cells. To account for the variety of cell sizes across groups, the coefficient of variation was also calculated (Figure 5.16 (c)), where smaller values were obtained for mutant cells compared to WT cells, thus indicating that mutant cells have narrower Euclidean distance distributions around the mean.

Both strategies indicate that microtubule networks of mutant cell groups are more compact than those of WT cells. As microtubules are more resistant to compression compared to other cytoskeletal proteins [42], possibly a greater cytoskeletal compactness may be advantageous for cells with disrupted E-cadherin, namely in the context of laminin, the ECM component used in this study.

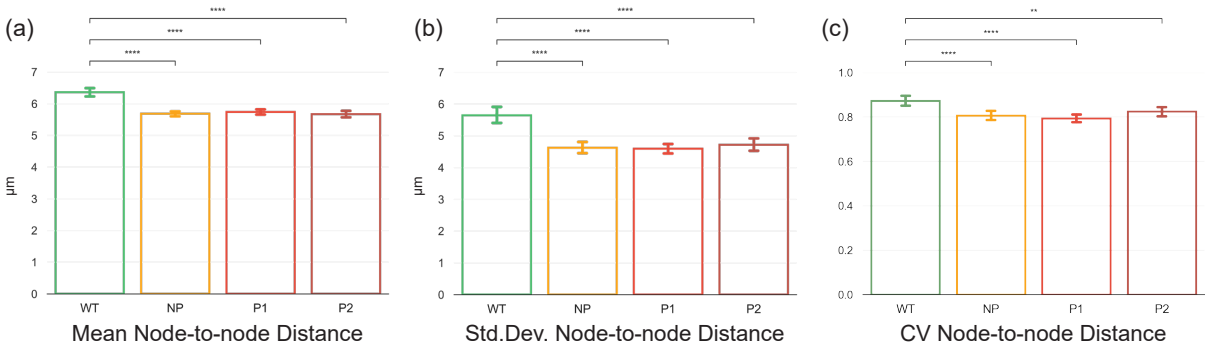


Figure 5.16: Quantification analysis of Euclidean distances between two connected nodes of cytoskeletal structures. Bars represent the mean $\pm$ 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

Connectivity

Figure 5.17 depicts three graph representations of cytoskeletons with different fiber-type compositions in terms of connectivity.

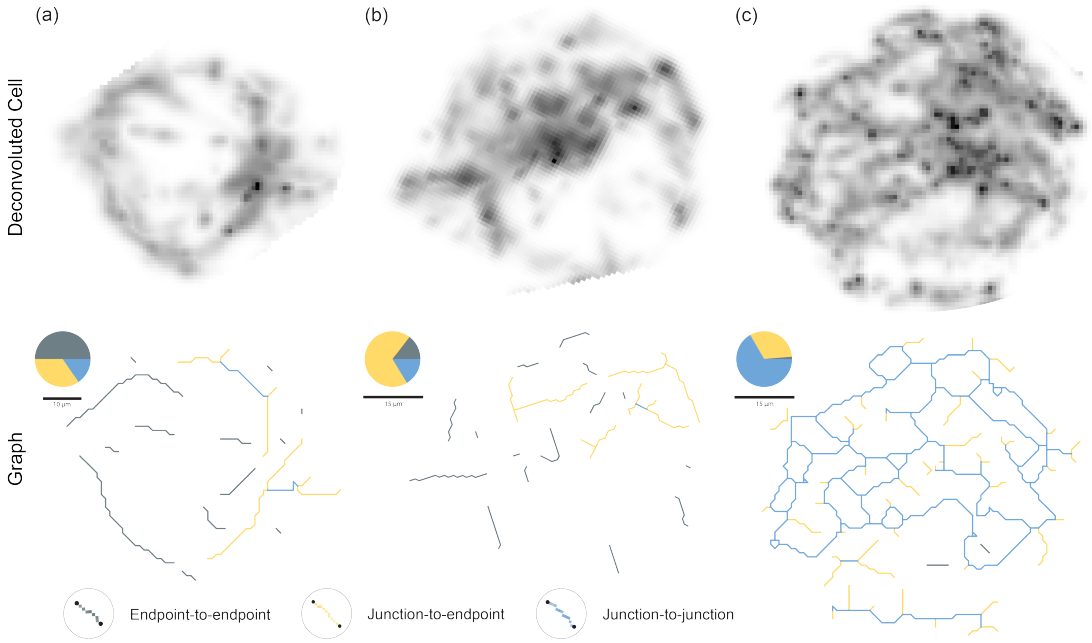


Figure 5.17: Fiber connectivity of cytoskeletal structures. Three representative examples are shown in (a), (b), and (c). Cytoskeletal structure (a), (b), and (c) present a high ratio of endpoint-to-endpoints (isolated branches), junction-to-endpoints, and junction-to-junctions, respectively. Graph edges are colored according to the type of junction. Pie charts represent the ratio of each type of junction. Cytoskeletal structures were obtained by α-tubulin staining.

Different proportions of fiber types in terms of connectivity may reveal cytoskeleton disruption phenomena. Cytoskeleton compactness may also be assessed with a connectivity analysis. A disrupted, sparse or disconnected cytoskeleton will present higher fractions of endpoint-to-endpoint branches,

while a more compact cytoskeleton will have an increased junction-to-junction ratio.

Global analysis

Differences were found between groups in the proportions of fiber connectivities. Figure 5.18 suggests that the most obvious differences are observed in the NP cells, which had a lower fraction of endpoints-to-endpoints (Figure 5.18 (a)) and junctions-to-endpoints (Figure 5.18 (b)), resulting in a higher fraction of junctions-to-junctions (Figure 5.18 (c)). As explained before, this may be another indicator for the higher degree of compactness acquired by NP cells. P1 cells, on the other hand, acquired lower junctions-to-endpoints ratios than WT cells, but had the highest junctions-to-endpoints ratio. Although no statistically significant differences were observed in the junctions-to-junctions ratio in these cells, these results also help to confirm the trend towards higher compactnesses in P1 cells. Surprisingly, no differences were found in P2 cells compared to WT in terms of connectivity, even though this group has the highest N_l/A_C ratio. Finally, since the number of isolated cycles was minimal (most cells in the dataset did not even have isolated cycles), this ratio was disregarded.

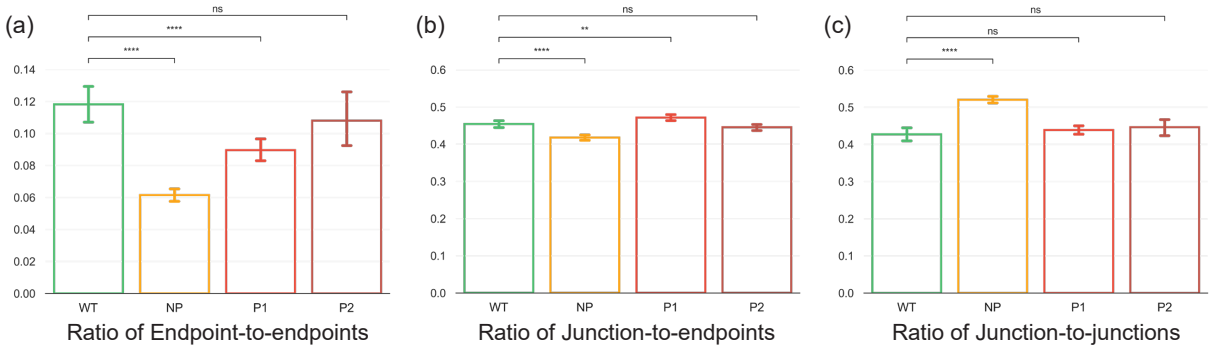


Figure 5.18: Quantification analysis of fiber connectivity ratios of cytoskeletal structures. Bars represent the mean $\pm$ 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

Quantity

Fiber quantity was measured by three metrics: through AMI, by the total number of detected line segments, N_l, and by the total number of branches in the graph representation, N_E. Figure 5.19 depicts three cells with different fiber quantities.

Notably, cytoskeletal structures may have the same number of fibers but spread over different areas. The cells in Figure 5.19 showed an ascending progression of AMI, as cell (a) shows reduced intensities compared to cell (b) or (c), which had filaments with very high intensities. However, this metric does not consider the spatial dispersion of these filaments, only the intensity distribution. Thus, the number of filaments was approximated by the total number of line segments and graph branches. Although cells (a)

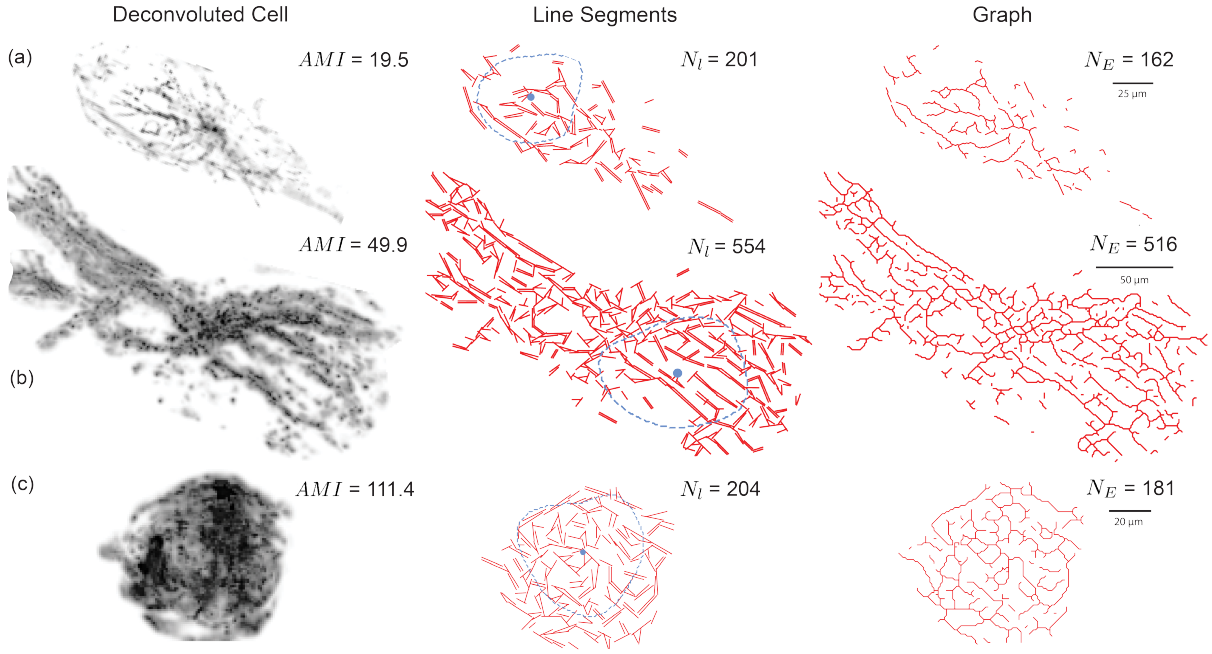


Figure 5.19: Fiber quantity of cytoskeletal structures. Three representative examples are shown in (a), (b), and (c). Cells (a), (b), and (c) have different microtubule intensities and quantities. Cytoskeletal structures were obtained by α-tubulin staining.

and (c) have quite distinct values for AMI, they present similar values for these features when compared to each other. Cell (b), with an intermediate AMI, was clearly a large cell and, for that reason, presented more line segments and branches in the graph.

Global analysis

AMI revealed statistically significant differences between groups (Figure 5.20 (a)). NP and P2 cell population tended to show higher values of AMI than WT cells, while in P1 cells the opposite pattern was observed.

Estimation of the amount of cytoskeleton through N_l (Figure 5.20 (b)) and N_E (Figure 5.20 (c)) only revealed a statistically significant decrease for both metrics in P1 cells. This finding is probably due to the fact that the P1 cell population mainly consisted of small and round cells, leading to the identification of simpler skeletons. Interestingly, the number of branches, as determined by N_E, was significantly different for P1 and P2 but not for NP cells, as compared to the WT.

Complexity

Complexity was estimated by calculating the fractal dimension. Figure 5.21 represents three cells with distinct fractal dimensions. The relationship between fractal dimension and the complexity of a structure

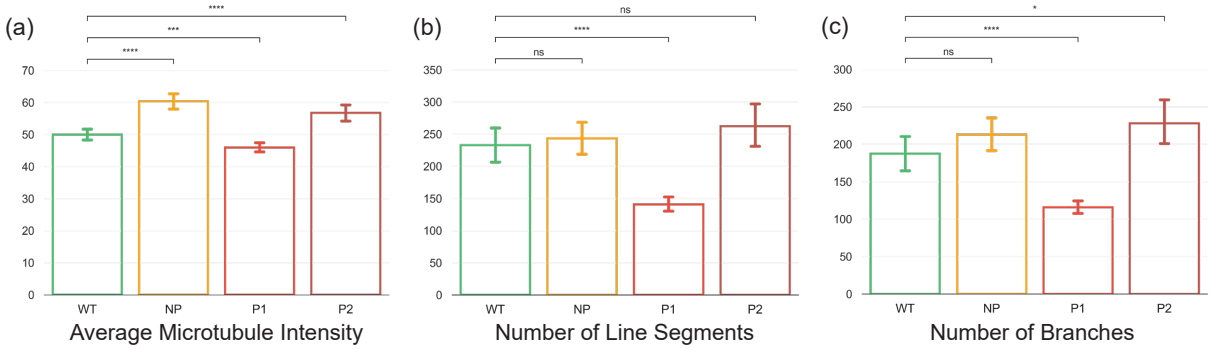


Figure 5.20: Quantification analysis of fiber quantity of cytoskeletal structures. Bars represent the mean ± 95% CI of (a) AMI; (b) N_l; (c) N_E ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

like a cytoskeleton is a challenging task to perform. For example, cells in Figure 5.21 present quite different fractal dimensions, but without information on the FD value, it would be difficult to reach this conclusion. However, quantifying cytoskeletal complexity through fractal dimension is a methodology widely adopted in the literature [18, 88, 89]. For the sake of coherence with previous studies, its analysis was performed in this work, though the global results did not show statistically significant differences between groups (data not shown).

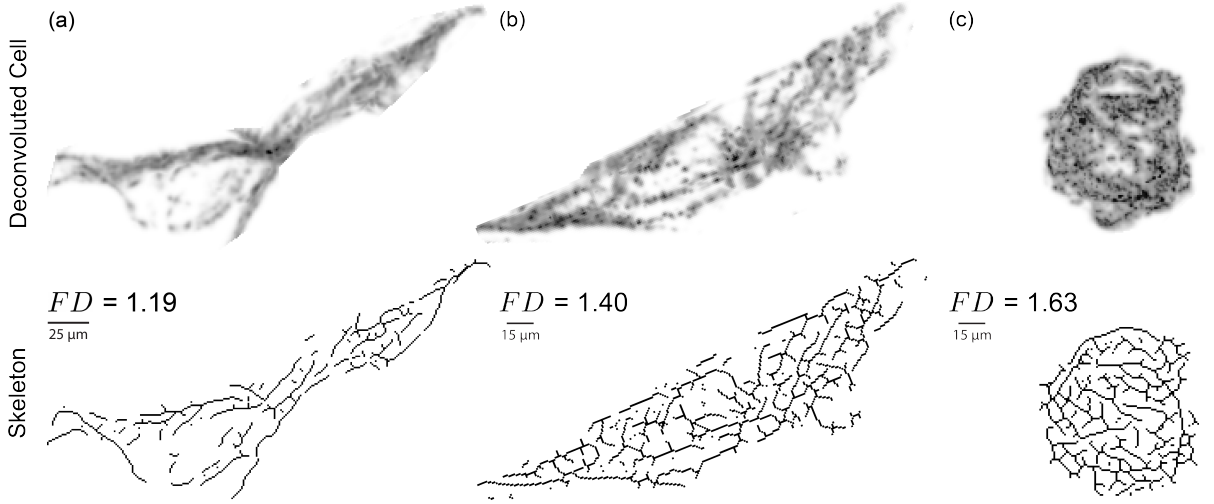


Figure 5.21: Estimated fractal dimension of cytoskeletal structures. Three representative examples are shown in (a), (b), and (c). Cytoskeletal structures were obtained by α-tubulin staining.

Radiality

Quantification of the degree of cytoskeletal radiality was performed using the radial score (equation 4.14). Briefly, the radial score $RS_{x,y}$ of pixel (x, y) is the density of line segments with angles $\alpha_i \in [0, 20]^\circ$ within the segmented cell. The density can be calculated at any point within the cytoplasm, but the pixel

(x^*, y^*) where this probability is maximized was chosen, thus obtaining the global radial score, RS. Figure 5.22 illustrates cells with different radial scores and the position where this score is maximized, represented with the yellow dot, to obtain RS.

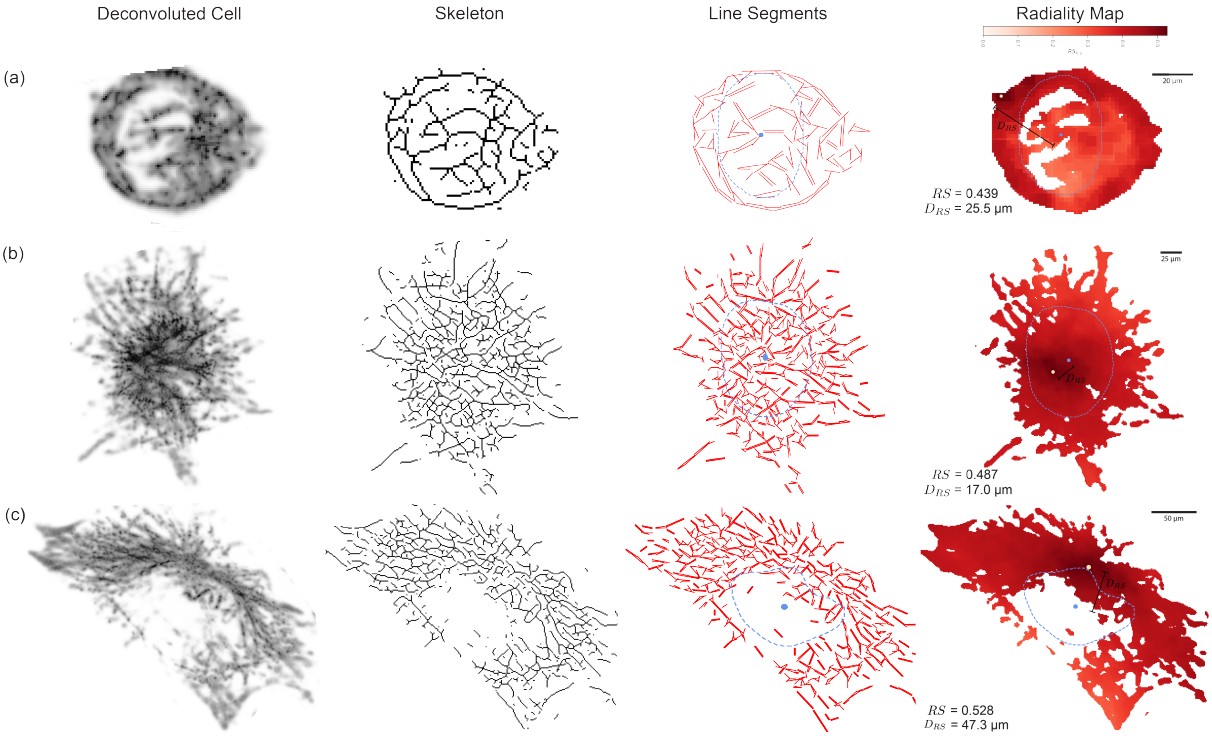


Figure 5.22: Illustration of the distribution of radiality ($RS_{x,y}$) of cytoskeletal structures. Three representative examples are shown in (a), (b), and (c). Yellow represents the position where the probability density (in percentage) of finding lower angled lines was maximized and equal to RS. Cytoskeletal structures were obtained by α-tubulin staining.

The point radial score $RS_{x,y}$ aimed to evaluate the radial organization of the filaments from a (x, y) coordinate reference. In Figure 5.22, cell (a) is small and round and did not appear to have an evident radial distribution, confirmed by the lower RS and the unpredictability of the corresponding pixel location. Cell (b), on the other hand, had a radial dispersion of fibers easily observed by the naked eye, nucleating from the center of the cell with a higher radial score than cell (a), of 48.7%. Finally, cell (c) had a more pronounced radial profile (52,8%), but nucleated further from the centroid of the nucleus compared to cell (b).

Global analysis

The global radial score RS was obtained for all cells, whose averages and confidence intervals by group are represented in Figure 5.23. The results in Figure 5.23 show that the microtubules of the NP ($p = 1.52 \cdot 10^{-6}$) and P1 ($p = 2.95 \cdot 10^{-3}$) groups were less radially arranged from the point where this parameter was maximized because they had lower RS values. No statistically significant differences

were found for the P2 mutation.

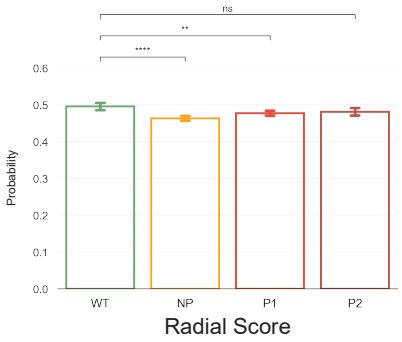


Figure 5.23: Quantification analysis of the radial score of cytoskeletal structures. Bars represent the mean ± 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

Cytoskeleton-Nucleus Interaction

As already discussed, the relationship between nuclear and cytoskeletal features in a cell may help understand specific cellular processes, including cell migration and invasion. For this purpose, the (weighted) distances between the centroids of both cytoskeleton and nucleus were measured, as shown in Figure 5.24 (Deconvoluted Cell) depicting cells with different inter-centroid distances. Additionally, line segment distances and radiality degrees were tested from the perspective of the nucleus using D and α, as shown in Figure 5.24 (distance to centroid D and angle α columns, respectively). Cell (a), being a small round cell, presents the shortest average distance from the line segments to the nucleus. In addition, the filaments are circumferentially oriented, so it has the highest average alpha value. On the contrary, cell (c), with an elongated profile, has a higher average distance of the line segments to the nucleus due to the formation of two pseudopodia. Pseudopodia are temporary arm-like projections of the cell membrane that emerge in the direction of movement, with a large alignment and bundling of microtubules. Consequently, lower α's were observed.

The distance between the centroid of the nucleus (blue dot) and the radiality center (x^*, y^*) (yellow dot) was also measured, being shown in three cells in Figure 5.22. Cell (a) presented an abnormal positioning of the radiality center. In cell (b), it is located within the nuclear contour, while in cell (c) touches the nuclear borders in which, interestingly, the centrosome is usually positioned [42].

Global Analysis

Across groups, statistically significant differences were found in the interplay between the cytoskeleton and the nucleus, whose results are displayed in Figure 5.25.

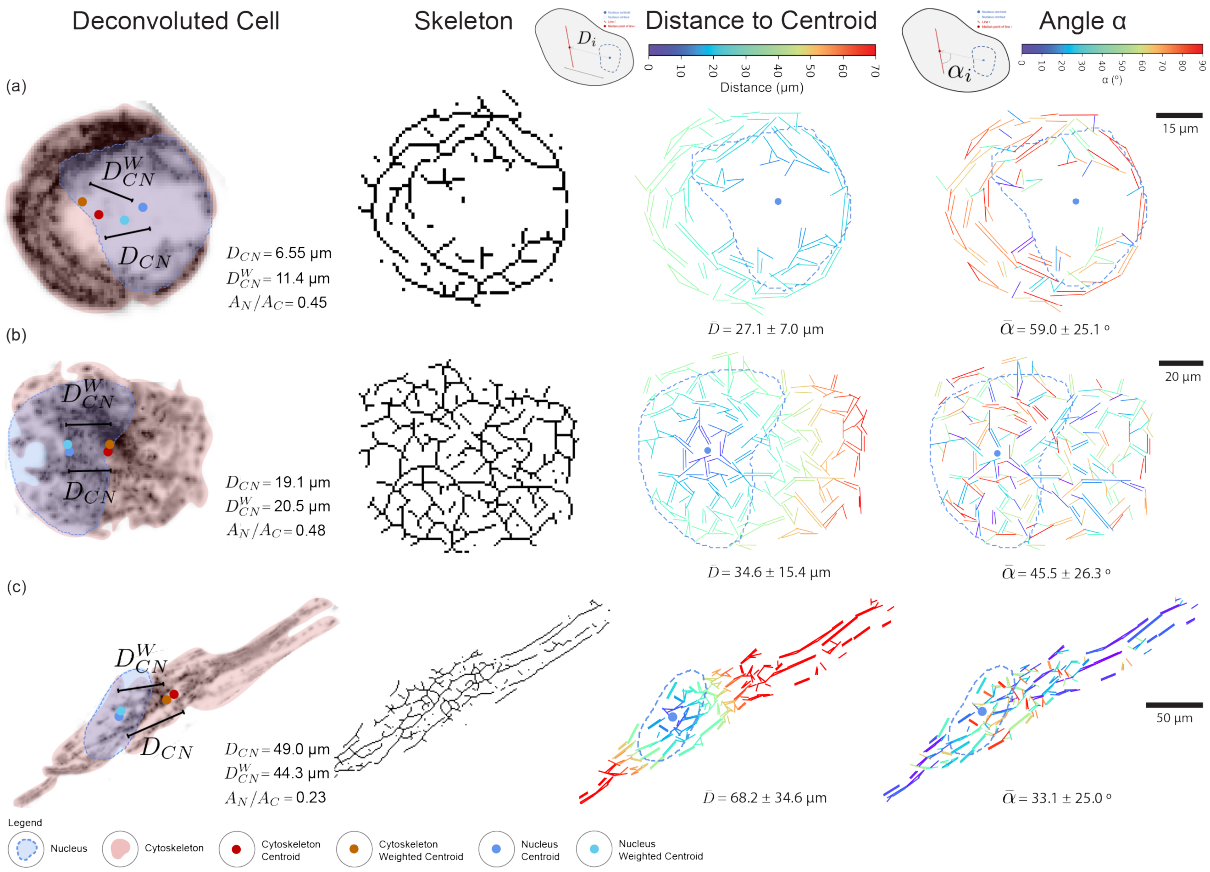


Figure 5.24: Cytoskeleton-nucleus interaction of cytoskeletal structures. Three representative examples are shown in (a), (b), and (c). Line segments are colored according to the value of D_i and α_i. Cytoskeletal structures were obtained by α-tubulin staining.

A prominent result is the longer (weighted) cytoskeleton-nucleus distance in P2 cells (Figure 5.25 (a) and (b)). This means that the nucleus is further away from the center of the cell, positioning itself at the periphery. A peripheral position of the nucleus within the cytoplasm is indicative of cell migration [7]. Conversely, NP cells presented questionable differences in both distance measurements. Finally, P1 cells showed a marked decrease in the weighted distance (Figure 5.25 (b)) but not for the normal distance (Figure 5.25 (a)). Higher intensity values could be located near the nucleus centroid in P1 cells, indicative of a significant concentration of α-tubulin near the nucleus. This could cause a significant decrease in weighted distances.

Unfortunately, no significant differences were found between groups for $\bar{D}$ and $\bar{\alpha}$ features, except for P1 cells. From this analysis, it is only concluded again that cells in this group tended to be smaller and rounder because, respectively, mean $\bar{D}$ distances were lower ($p = 2.61 \cdot 10^{-12}$) and $\bar{\alpha}$ angles were higher ($p = 6.31 \cdot 10^{-5}$) in P1 cells compared to WT cells.

The results for D_{RS} (Figure 5.25 (c)) show that while no statistically significant differences were observed for NP cells, P1 cells had a marked decrease in D_{RS} distance ($p = 3.78 \cdot 10^{-7}$). As previously

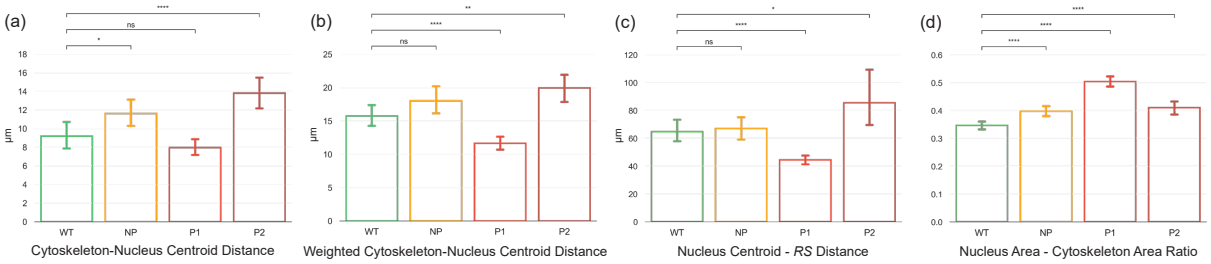


Figure 5.25: Quantification analysis of cytoskeleton-nucleus interaction. Bars represent the mean $\pm$ 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

mentioned, most P1 cells were small and round, whose morphology could explain these values. The results also showed a debatable increase in D_{RS} in P2 cells relative to WT cells ($p = 3.84 \cdot 10^{-2}$), as the former had a higher mean distance, but the confidence interval for D_{RS} in the latter cell group was very wide.

According to Figure 5.25 (d), mutant cells showed an increased nucleus-to-cytoskeleton area ratio. Global results regarding nuclear and cytoskeletal area indicated that only P1 cells had a marked decreased in both cases, while NP and P2 did not present statistically significant different nuclear/cytoskeletal areas compared to WT cells (data not shown). This finding could be another indicator of the increased cytoskeleton compactness in mutant cells considering the area of the nucleus.

Understanding which metrics can discriminate the cytoskeletal architecture of non-invasive cells and cells with invasive potential, coupled with features extracted from nuclei and their interplay with the cytoskeleton, remains an unmet goal.

5.5 Discussion

The cytoskeleton is a highly dynamic structure involved in many cellular processes. However, the evaluation of cytoskeletal remodeling remains challenging despite technological advances. In this study, a novel computational approach was developed to investigate the cytoskeletal architecture of cancer cells. A new pipeline is proposed involving preprocessing and processing strategies that ultimately can characterize the cytoskeletal structure of cells. The flower plot in Figure 5.26 depicts data obtained following the application of the proposed pipeline and illustrates the cytoskeleton organization metrics of five distinct cells, aiming to highlight cytoskeletal architecture similarities and differences between them.

For a more detailed analysis, two-dimensional features related to different aspects of the cytoskeleton organization may be assessed by color maps across the cell. These can also be converted to a histogram or other data visualization method. Furthermore, one-dimensional features allow organization features to be quantified numerically and compared between cells. Remarkably, the proposed

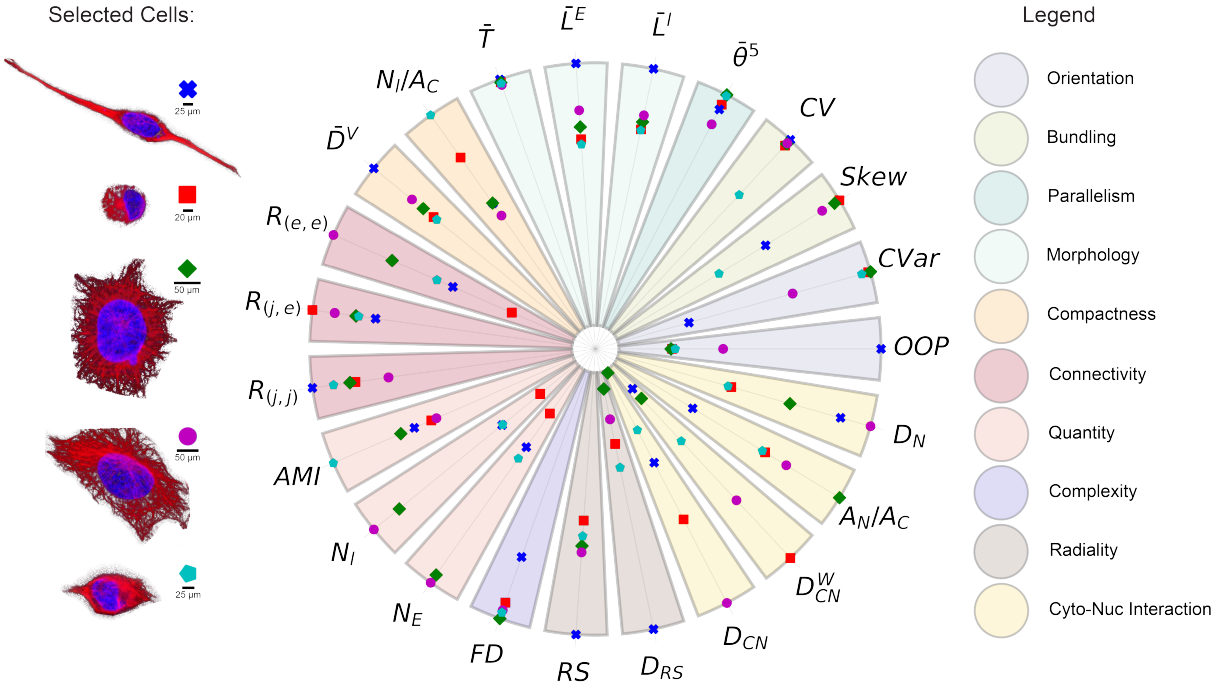


Figure 5.26: Full cytoskeleton description and comparative analysis between five cells. Immunofluorescent images were obtained following α-tubulin (red) and DAPI (blue) staining.

framework was used to investigate α-tubulin, but it can also be applied to evaluate other cytoskeleton components.

Despite the difficulty in selecting and tuning the right image processing operations in intensity-threshold methods, these encompass simple strategies and commonly output a binary mask for the cytoskeleton pixels. An accurate cytoskeleton segmentation using classical methods may be compromised when challenging images are used, since denoising, filament-enhancement filtering, and thresholding may not be enough to deal with images with low signal-to-noise ratios, blurred/discontinuous contours, and noise. The consequent lack of precision in the localization of the filament tracks is a problem for intensity-thresholding methods. It may result, for instance, in a pixel mask that is wider than the physical width of the filaments. Morphological thinning (using skeletonization or other algorithms) tries to overcome this problem but fails to guarantee their precise localization and might generate artifacts [56]. With the limitations of intensity-threshold methods in mind, other more complex solutions are available in the state-of-the-art. However, their by-hand implementation is difficult due to methodology complexity, and sometimes authors do not include open-access software.

Noteworthy, a drawback of the developed methodology is that it still requires manual cell segmentation and parameter tuning to overcome problems like blur, Poisson noise, filament enhancement properties and line segment refinements, which can compromise the results. In this study, automatic segmentation was not possible, given the high confluence observed in the images. User-to-user variability

and high confluence index can also influence cell segmentation and therefore further improvements in these computational models should be performed in the future. During preprocessing, the Gaussian kernel can be modulated to provide different levels of smoothing. Depending on the image properties, the resolution and thresholds of the Sato filter can be adjusted, just as the thresholds in the binarization may need changes. Regarding processing, different combinations of parameters in detecting line segments generate different rearrangements. Regarding parameter tuning, errors or biases will inevitably arise, but since the same methodology was used in all cells, any bias was propagated through the whole dataset.

Most investigations on cytoskeleton organization are aimed towards specific cellular processes that cause, for instance, the alignment/orientation [12, 15, 82] or disruption of fibers [11, 64]. Consequently, metrics are built to optimally quantify changes in these specific processes. However, they are often limited to the incorporation of easy-to-interpret data, like total cytoskeleton intensity, average straight-line length and orientations. In other studies, the evaluation of cytoskeleton dynamics is limited to the analysis of the fluorescence intensity distribution of stained cytoskeletal structures. State-of-the-art software that allows the segmentation and extraction of individual filaments (either by line segments or by graph representations) are already available. In particular, ImageJ [102], a state-of-the-art software for various applications in biological image analysis, provides tools to obtain skeletons and plugins like AnalyzeSkeleton [103] that agglomerate straightforward skeleton information. Line detection-based software for skeleton analysis is also possible with FilamentSensor [79]. However, methods for analyzing this information remain very limited, requiring the creation of new metrics and relating them to aspects of the cytoskeletal organization (that also need to be established), taking into account the 3D network morphology of this structure. The lack of state-of-the-art papers establishing models to describe cytoskeletal organization may be due to the specificity of their objectives. The focus is often turned to the study of one or two aspects of cytoskeletal architecture, such as orientation/alignment, amount of fluorescence, among other issues, and the methods of analysis usually do not need to be very developed to meet those objectives. Thus, information scarcity regarding cytoskeleton remodeling in cancer cells and the complexity of this structure prompted the generation of a thorough qualitative description model of cytoskeletal structures from a holistic perspective. Since it is not well understood at which level the cytoskeleton of a non-invasive cancer cell varies from that of an invasive cancer cell, a comprehensive approach was performed to assess at which points of view there are significant and clinically relevant differences.

In the proposed framework, the ”workbenches” on which feature extraction was based on were the pixel intensity distributions (DCF), line segments (LSF) and graph representation (CNF). Other features can be developed from the information available by each of these methodologies, or by a combination of existing features. From any two-dimensional feature distribution, one-dimensional features were ob-

tained like the mean, standard deviation, kurtosis, skewness, and interquartile distance, among other statistics. Overall, this study has shown that the features analyzed are distinct in the different groups of cells with WT or mutant E-cadherin, reflecting different cytoskeletal organizations. In particular, the microtubules of cells with disrupted E-cadherin were shown to be shorter, to have uniform length patterns and to be more compactly distributed as compared with WT E-cadherin cells. Thus, the presence of an E-cadherin mutation, either pathogenic or non-pathogenic, modulated the cytoskeletal organization. However, no pattern was obtained to distinguish invasive from non-invasive cancer cells. Indeed, cells with a non-pathogenic mutation and no invasive potential (NP) did not present distinct patterns for most features that could distinguish them from the pathogenic and invasive cells (P1 and P2). This result suggests that the non-pathogenic mutation in E-cadherin is sufficient to modulate α-tubulin organization, but that these modifications are not associated with invasive ability. Additional mechanisms and cytoskeletal components mediate cell invasion, which could explain the observed data. Interestingly, fiber quantity, as determined by the number of branches, was one of the features for which significantly differences were observed between WT and P1 and P2 but not NP. Of note, the mutation with the most striking effect in terms of E-cadherin loss of expression and loss of function (P1) [93] revealed significant different levels of fiber quantity as shown by the lower number of line segments and the lower number of branches. Notably, this pathogenic mutation (p.L13_L15del) is clinically relevant as it impacts the signal sequence of E-cadherin and constitutes a molecular mechanism underlying Hereditary Diffuse Gastric Cancer [93]. Thus, additional studies should be performed in the future to further characterize related features. In addition, to closely mimic *in vivo* conditions, these cells were grown in laminin. As previously demonstrated, distinct ECM substrates are able to modulate the cytoskeleton during *in vitro* tumor growth [32] and therefore experiments with additional ECM components with variable stiffness should also be tested. Indeed, it remains poorly understood how diverse levels of stiffness/tension impact cytoskeletal conformations and cell behavior [104, 105]. Thus, the use of additional systems to evaluate compression/stretching and mechanical properties should be investigated taking advantage of our pipeline to explore the effect on cytoskeletal architecture. Of note, along the different experimental conditions, morphological heterogeneity was observed. This could be due, at least in part, to the transient transfection procedure used in this study. Despite being a classical technique to evaluate the expression of proteins in cell lines, the transfection efficiency is not complete. Hence, stable transfections should be tested in the future.

This work distinguishes itself from others by building a qualitative model to comprehensively describe aspects of cytoskeletal architecture from fluorescence images. In addition, metrics were developed to quantify these aspects. This work includes existing analyses but also brings novelties such as radiality or the interaction between the cytoskeleton and the nucleus. By applying the developed methodology, it was possible to detect significant differences between cytoskeletal organization aspects in cell popula-

tions with different invasion phenotypes.

6

Concluding Remarks and Future Work

The application of the computational pipeline to the current dataset enabled to statistically assess how microtubules of cell populations reorganized upon E-cadherin deregulation. However, with the current knowledge, it is still not possible to relate the extracted features with a higher or lower invasive capacity. Nonetheless, this framework constitutes a powerful tool to further investigate the dynamic reorganization of the cytoskeleton during many cellular processes, including cell invasion. It is plausible to foresee that multiple features, rather than a single feature, will allow the characterization of cytoskeletal patterns, i.e. generate a cytoskeletal signature to be associated with specific cell mechanisms.

For a more complete evaluation of the cytoskeletal architecture, additional image datasets with an increased number of cells, different cell types, and interplay with diverse ECM components need to be explored. The refinement of the computational approach will further improve feature analysis. Future research will help to unravel how cytoskeletal organization cooperates to promote invasion. Moreover, with fine-tuned improvements, this pipeline has the potential to be applied to pathological tissue samples, with enormous implications in cancer research and clinical applications. The development of automatic detection methods to identify cancer cells with increased invasive potential could impact diagnosis and prognosis and the selection of therapeutic strategies, as anti-microtubular drugs are routinely used as cancer therapies.

7

Appendix

Appendix

Deconvoluted Cell Features

As mentioned in Section 4.4.1, features related to the distribution of grayscale intensities of a given cell were extracted. Some of these features were of more interest as they were indicators of specific aspects of cytoskeleton organization, for example, skewness and CV to quantify fiber bundling. The extracted features (see Table 7.1) were selected in order to survey a list of properties related to the representation of the microtubules of cells in images. Thus, morphology-, intensity-, texture- and frequency-based analyses were performed. Given the aims of this work, to create a model to characterize the cytoskeleton qualitatively and quantitatively, it remains as future work to make a thorough analysis of these features and, if possible, find those that describe the distribution of grayscale intensities in a way that characterizes specific aspects of cytoskeleton organization.

Table 7.1: Left out DCF.

Deconvoluted Cell Features (DCF)				
Morphology		Intensity	Texture	Frequency
(BB) Area	Eccentricity	Mean Intensity	(Invariant) Uniformity	Mean Gabor Power
Convex Hull Area	Circularity	Min & Max Intensity	(Invariant) Entropy	Gabor Variance
Perimeter	Roundness	Standard Deviation	(Invariant) Correlation	Gabor Entropy
Centroid Divergence	Orientation	Variance	(Invariant) Dissimilarity	Gabor Energy
Equivalent Diameter	Solidity	Kurtosis	(Invariant) Contrast	Mean Spectral Magnitude
Minor/Major Axis Length	Roughness	Entropy	(Invariant) Homogeneity	Mean Spectral Power
Fractal Dimension	Extent		(Invariant) Energy	

Even so, some features should be highlighted as they presented statistically significant differences between WT and mutant cells, represented in the barplots of Figures 7.1-7.3.

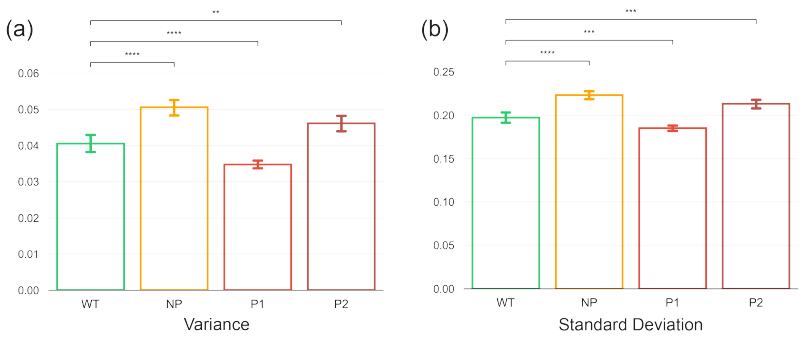


Figure 7.1: Quantification analysis of intensity-based features of cytoskeletal structures with statistical significant differences between WT and mutant cells. Bars represent the mean $\pm$ 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

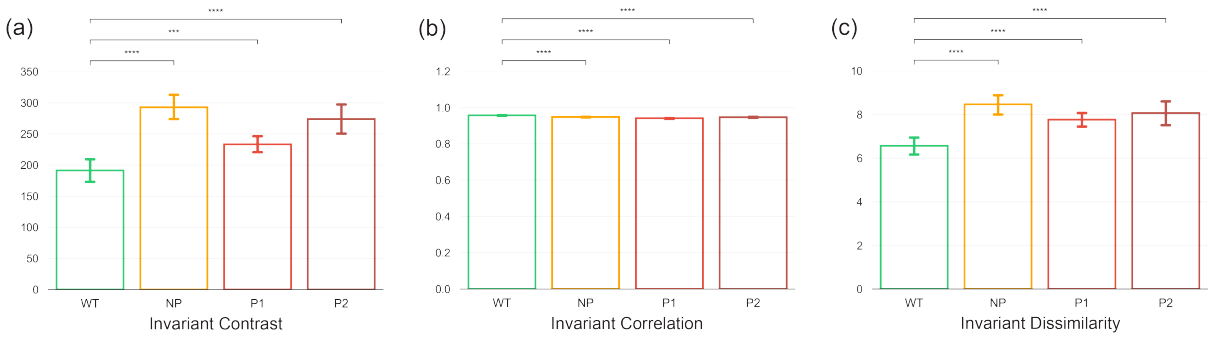


Figure 7.2: Quantification analysis of texture-based features of cytoskeletal structures with statistical significant differences between WT and mutant cells. Bars represent the mean ± 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

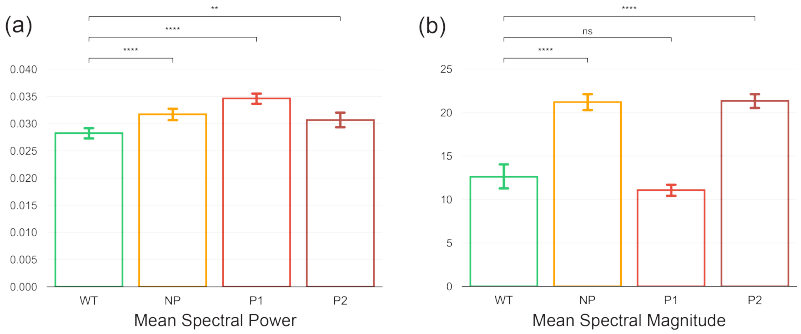


Figure 7.3: Quantification analysis of frequency-based features of cytoskeletal structures with statistical significant differences between WT and mutant cells. Bars represent the mean ± 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

Cytoskeleton Network Features

As mentioned in Section 4.4.3, graph-based features including the intensity variable were not addressed in detail as it was not possible to relate them to any specific aspect of cytoskeletal organization (Figure 3.4).

In graph representations, a cytoskeletal fiber is a set of adjacent pixel coordinates, each with a different intensity value. The mean and standard deviations of the distribution of pixel intensities of a fiber were obtained. Both results were then averaged across all graph edges. The resulting value was obtained for all cells, and the global results are shown in Figure 7.4.

The mean pixel intensity of NP cells was higher than in WT cells, while P1 had fibers with decreased mean intensities. For the P2 group, no significant differences were found.

The standard deviation of pixel intensities could be seen as a disorganization metric in terms of intensity values. Results showed that branches of mutant cells presented broader intensity distributions (Figure 7.4 (b)), while fibers of WT cells had more uniform distributions.

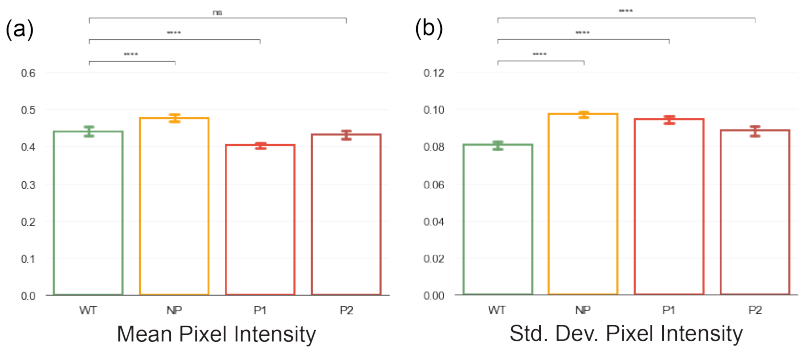


Figure 7.4: Quantification analysis of intensity-based features of graph representations of cytoskeletal structures, where statistical significant differences between WT and mutant cells were found. Bars represent the mean $\pm$ 95% CI. ns, not significant; $*$, $p \leq 0.05$; $**$, $p \leq 0.01$; $***$, $p \leq 0.001$; $****$, $p \leq 0.0001$. WT, Wild Type; NP, Non-Pathogenic; P1, Pathogenic 1; P2, Pathogenic 2.

Bibliography

[1] T. Hohmann and F. Dehghani, "The cytoskeleton—a complex interacting meshwork," *Cells*, vol. 8, no. 4, 2019.

[2] H. Lodish, A. Berk, C. A. Kaiser, M. Krieger, A. Bretscher, H. Ploegh, A. Amon, and K. C. Martin, *Molecular Cell Biology*, 8th ed. W. H. Freeman, 2016.

[3] J. Käs, H. Strey, J. X. Tang, D. Finger, R. Ezzell, E. Sackmann, and P. A. Janmey, "F-actin, a model polymer for semiflexible chains in dilute, semidilute, and liquid crystalline solutions," *Biophysical Journal*, vol. 70, no. 2, pp. 609–625, 1996.

[4] W. M. Becker, L. J. Kleinsmith, and J. Hardin, *The World of the Cell*, 9th ed. Pearson, 2017.

[5] A. L. Mescher, *Junqueira's Basic Histology: Text and Atlas*, 16th ed. McGraw Hill / Medical, 2021.

[6] P. Lu and Y. Lu, "Born to run? diverse modes of epithelial migration," *Frontiers in Cell and Developmental Biology*, vol. 9, no. 704939, 2021.

[7] F. J. Calero-Cuenca, C. S. Janota, and E. R. Gomes, "Dealing with the nucleus during cell migration," *Current Opinion in Cell Biology*, vol. 50, pp. 35–41, 2018.

[8] A. M. Moreira, J. Pereira, S. Melo, M. S. Fernandes, P. Carneiro, R. Seruca, and J. Figueiredo, "The extracellular matrix: An accomplice in gastric cancer development and progression," *Cells*, vol. 9, no. 2, 2021.

[9] J. Aseervatham, "Cytoskeletal remodeling in cancer," *Biology*, vol. 9, no. 11, 2020.

[10] M. Alioscha-Perez, C. Benadiba, K. Goossens, S. Kasas, G. Dietler, R. Willaert, and H. Sahli, "A robust actin filaments image analysis framework," *PLOS Computational Biology*, vol. 12, no. 8, 2016.

[11] D. Breuer, J. Nowak, A. Ivakov, M. Somssich, S. Persson, and Z. Nikoloski, "System-wide organization of actin cytoskeleton determines organelle transport in hypocotyl plant cells," *Proceedings of the National Academy of Sciences*, 2017.

[12] A. Basu, M. K. Paul, M. Alioscha-Perez, A. Grosberg, H. Sahli, S. M. Dubinett, and S. Weiss, "Statistical parametrization of cell cytoskeleton reveals lung cancer cytoskeletal phenotype with partial EMT signature," *Communications Biology*, vol. 5, no. 1, pp. 2399–3642, 2021.

[13] Y. Liu and J. Ren, "Modeling and control of dynamic cellular mechanotransduction (i): actin cytoskeleton quantification," *ASME 2018 Dynamic Systems and Control Conference*, 2018.

[14] H.-S. Wu, F.-Q. Wang, Y. Smicun, and D. Fishman, "Recognition of linear stress fibers based on hough transform," *Anal Quant Cytol Histol*, vol. 33, pp. 121–131, 2011.

[15] N. Costigliola, L. Dinga, C. J. Burckhardt, S. J. Han, E. Gutierrez, A. Mota, A. Groisman, T. J. Mitchison, and G. Danuser, "Vimentin fibers orient traction stress," *PNAS*, vol. 114, no. 20, p. 5195–5200, 2017.

[16] J. Nunez-Iglesias, A. J. Blanch, O. Looker, M. W. Dixon, and L. Tilley, "A new Python library to analyse skeleton images confirms malaria parasite remodelling of the red blood cell membrane skeleton," *PeerJ*, vol. 6, 2018.

[17] P. Berens, "Circstat: A matlab toolbox for circular statistics," *Journal of Statistical Software*, vol. 31, no. 10, 2009.

[18] A. Revittser, I. Selin, Y. Negulyaev, and V. Chubinskiy-Nadezhdin, "The analysis of f-actin structure of mesenchymal stem cells by quantification of fractal dimension," *PLoS ONE*, vol. 16(11), 2021.

[19] Scikit-Image. Module: filters. https://scikit-image.org/docs/stable/api/skimage.filters [accessed: 31.10.2022].

[20] OpenCV. Feature detection. https://docs.opencv.org/3.4/dd/d1a/group__imgproc__feature.html [accessed: 31.10.2022].

[21] F. H. C. Crick and A. F. W. Hughes, "The physical properties of cytoplasm. a study by means of the magnetic particle method," *Experimental Cell Research*, vol. 1, pp. 505–533, 1950.

[22] S. Chakraborty, M. Jasnin, and W. Baumeister, "Three-dimensional organization of the cytoskeleton: A cryo-electron tomography perspective," *Protein Science*, vol. 29, no. 6, p. 1302–1320, 2020.

[23] W. H. Organization. Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer [accessed: 31.10.2022].

[24] D. Lopes and H. Maiato, "The tubulin code in mitosis and cancer," *Cells*, vol. 9, no. 11, p. 2356, 2020.

[25] J. Yang, P. Antin, G. Ber, C. Blanpain, T. Brabletz, M. Bronner, K. Campbell, A. Cano, J. Casanova, G. Christofori, S. Dedhar, R. Derynck, H. L. Ford, J. Fuxe, A. G. de Herreros, G. J. Goodall, A.-K. Hadjantonakis, R. J. Y. Huang, C. Kalcheim, R. Kalluri, Y. Kang, Y. Khew-Goodall, H. Levine, J. Liu, G. D. Longmore, S. A. Mani, J. Massagué, R. Mayor, D. McClay, K. E. Mostov, D. F. Newgreen, M. A. Nieto, A. Puisieux, R. Runyan, P. Savagner, B. Stanger, M. P. Stemmler, Y. Takahashi, M. Takeichi, E. Theveneau, J. P. Thiery, E. W. Thompson, R. A. Weinberg, E. D. Williams, J. Xing, B. P. Zhou, G. Sheng, and E. I. A. (TEMTIA), "Guidelines and definitions for research on epithelial–mesenchymal transition," *Nature Reviews Molecular Cell Biology*, no. 21, p. 341–352, 2020.

[26] C. Ruggiero and E. Lalli, "Targeting the cytoskeleton against metastatic dissemination," *Cancer and Metastasis Reviews*, vol. 40, no. 1, p. 89–140, 2021.

[27] T. Mestre, J. Figueiredo, A. S. Ribeiro, J. Paredes, R. Seruca, and J. M. Sanches, "Quantification of topological features in cell meshes to explore E-cadherin dysfunction," *Scientific Reports*, vol. 6, no. 1, 2016.

[28] J. Paredes, J. Figueiredo, A. Albergaria, P. Oliveira, J. Carvalho, A. S. Ribeiro, J. Caldeira, Ângela Margarida Costa, J. Simões-Correia, M. J. Oliveira, H. Pinheiro, S. S. Pinho, R. Mateus, C. A. Reis, M. Leite, M. S. Fernandes, F. Schmitt, F. Carneiro, C. Figueiredo, C. Oliveira, and R. Seruca, "Epithelial e- and p-cadherins: Role and clinical significance in cancer," *Biochimica Et Biophysica Acta (BBA) - Reviews On Cancer*, vol. 1826(2), pp. 297–311, 2012.

[29] G. Corso, J. Figueiredo, S. P. D. Angelis, F. Corso, A. Girardi, J. Pereira, R. Seruca, B. Bonanni, P. Carneiro, G. Pravettoni, E. G. Rocco, P. Veronesi, G. Montagna, V. Sacchini, and S. Gandini, "E-cadherin deregulation in breast cancer," *Journal of Cellular and Molecular Medicine*, vol. 24, no. 11, pp. 5930–5936, 2020.

[30] worldometer. Deaths caused by cancer this year. https://www.worldometers.info/cancer/ [accessed: 31.10.2022].

[31] D. Hanahan, "Hallmarks of cancer: New dimensions," *Cancer Discovery*, vol. 12, no. 1, pp. 31–46, 2022.

[32] M. A. Al-Mamun, W. Srisukkham, D. M. Farid, L. Ravenhill, L. Zhang, A. Hossain, and R. Bass, "A quantitative image analysis for the cellular cytoskeleton during in vitro tumor growth," *Expert Systems With Applications*, vol. 92, pp. 39–51, 2017.

[33] M. S. Ong, S. Deng, C. E. Halim, W. Cai, T. Z. Tan, R. Y.-J. Huang, G. Sethi, S. C. Hooi, A. P. Kumar, and C. T. Yap, "Cytoskeletal proteins in cancer and intracellular stress: A therapeutic perspective," *Cancers*, vol. 12, no. 1, 2020.

[34] C. M. Fife, A. McCarroll, and M. Kavallaris, “Movers and shakers: cell cytoskeleton in cancer metastasis,” *British Journal of Pharmacology*, vol. 171, no. 24, pp. 5507–5523, 2014.

[35] D. A. Fletcher and R. D. Mullins, “Cell mechanics and the cytoskeleton,” *Nature*, vol. 463(7280), pp. 485–492, 2010.

[36] E. G. Stringham, N. Marcus-Gueret, L. Ramsay, and K. L.Schmidt, “Chapter eleven - live cell imaging of the cytoskeleton,” *Methods in Enzymology*, vol. 505, pp. 203–217, 2012.

[37] A. Hall, “The cytoskeleton and cancer,” *Cancer Metastasis Rev*, vol. 28, no. 1-2, p. 5–14, 2009.

[38] M. Nishita, T. Satake, Y. Minami, and A. Suzuki, “Regulatory mechanisms and cellular functions of non-centrosomal microtubules,” *Journal of Biochemistry*, 2017.

[39] A. Müsch, “Microtubule organization and function in epithelial cells,” *Traffic*, vol. 5(1), pp. 1–9, 2004.

[40] M. Toya and M. Takeichi, “Organization of non-centrosomal microtubules in epithelial cells,” *Cell Structure and Function*, vol. 41(2), p. 127–135, 2016.

[41] P. Binarová and J. Tuszynski, “Tubulin: Structure, functions and roles in disease,” *Cells*, vol. 8, no. 1294, 2019.

[42] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, and P. Walter, *Molecular Biology of the Cell*, 6th ed. Garland Science, 2015.

[43] T. M. H. Gall and A. E. Frampton, “Gene of the month: E-cadherin (CDH1),” *Journal of Clinical Pathology*, vol. 66, no. 11, pp. 928–932, 2013.

[44] I. Kaszak, O. Witkowska-Piłaszewicz, Z. Niewiadomska, B. Dworecka-Kaszak, F. N. T. 5, and P. Jurka, “Role of cadherins in cancer - a review,” *International Journal of Molecular Sciences*, vol. 21(20), p. 7624–, 2020.

[45] S. Melo, J. Figueiredo, M. S. Fernandes, M. Gonçalves, E. M. de Sá, J. M. Sanches, and R. Seruca, “Predicting the functional impact of CDH1 missense mutations in hereditary diffuse gastric cancer,” *International Journal of Molecular Sciences*, vol. 18, no. 12, 2017.

[46] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, “Adherens junction and E-Cadherin complex regulation by epithelial polarity,” *Cell Mol Life Sci.*, vol. 73(18), pp. 3535–53, 2016.

[47] T. Brabletz, R. Kalluri, M. A. Nieto, and R. A. Weinberg, “EMT in cancer,” *Nature Reviews Cancer*, vol. 18, no. 2, pp. 128–134, 2018.

[48] M. Sommariva and N. Gagliano, “E-cadherin in pancreatic ductal adenocarcinoma: A multifaceted actor during EMT,” *Cells*, vol. 9, no. 4, 2020.

[49] V. Braga, “Spatial integration of E-cadherin adhesion, signalling and the epithelial cytoskeleton,” *Current Opinion in Cell Biology*, vol. 42, pp. 138–145, 2016.

[50] E. Vasileva and S. Citi, “The role of microtubules in the regulation of epithelial junctions,” *Tissue Barriers*, vol. 6, no. 3, 2018.

[51] R. Robinson, “Link between cell junctions and microtubule cytoskeleton is critical for epithelial morphogenesis,” *PLOS Biology*, vol. 13, no. 3, 2015.

[52] J. Figueiredo, S. Melo, P. Carneiro, A. M. Moreira, M. S. Fernandes, A. S. Ribeiro, P. Guilford, J. Paredes, and R. Seruca, “Clinical spectrum and pleiotropic nature of CDH1 germline mutations,” *Journal of Hematology and Oncology*, vol. 56, no. 4, pp. 199–208, 2019.

[53] T. D. Pollard, J. Lippincott-Schwartz, W. C. Earnshaw, and G. T. Johnson, *Cell Biology*, 3rd ed. Elsevier, 2017.

[54] C. Uhler and G. Shivashankar, “Nuclear mechanopathology and cancer diagnosis,” *Trends in Cancer*, vol. 4, no. 4, 2018.

[55] M. T. Parreira, K. Lavrenyuk, J. M. Sanches, and K. N. Dahl, “A single stiffened nucleus alters cell dynamics and coherence in a monolayer,” *Cytoskeleton*, vol. 78, no. 6, pp. 277–283, 2021.

[56] B. Özdemir and R. Reski, “Automated and semi-automated enhancement, segmentation and tracing of cytoskeletal networks in microscopic images: A review.” *Computational and Structural Biotechnology Journal*, vol. 19, p. 2106–2120, 2021.

[57] J. Figueiredo, I. Rodrigues, J. Ribeiro, M. S. Fernandes, S. Melo, B. Sousa, J. Paredes, R. Seruca, and J. M. Sanches, “Geometric compensation applied to image analysis of cell populations with morphological variability: a new role for a classical concept,” *Scientific Reports*, vol. 8, no. 1, 2018.

[58] I. C. Rodrigues and J. M. R. Sanches, “Convex total variation denoising of Poisson fluorescence confocal images with anisotropic filtering,” *IEEE Transactions on Image Processing*, vol. 20, no. 1, pp. 146–160, 2011.

[59] M. Afonso, M. T. Parreira, and J. M. Sanches, “Noise decomposition using polynomial approximation,” *Pattern Recognition and Image Analysis*, pp. 157–164, 2015.

[60] M. Fernández-Suárez and A. Y. Ting, “Fluorescent probes for super-resolution imaging in living cells,” *Nature Reviews Molecular Cell Biology*, vol. 9(12), p. 929–943, 2008.

[61] S. Moccia, E. D. Momi, S. E. Hadji, and L. S. Mattos, “Blood vessel segmentation algorithms — review of methods, datasets and evaluation metrics,” *Computer Methods and Programs in Biomedicine*, vol. 158, p. 71–91, 2018.

[62] L. Acciai, P. Soda, and G. Iannello, “Automated neuron tracing methods: An updated account,” *Neuroinformatics*, vol. 14(4), p. 353–367, 2016.

[63] T. Higaki, “Quantitative evaluation of cytoskeletal organizations by microscopic image analysis,” *Plant Morphology*, vol. 29, no. 1, pp. 15–21, 2017.

[64] Y. Liu, W. Treible, A. Kolagunda, A. Nedo, P. Saponaro, J. Caplan, and C. Kambhamettu, “Densely connected stacked u-network for filament segmentation in microscopy images,” *Computer Vision*, vol. 11134, 2019.

[65] C. Faulkner, J. Zhou, A. Evrard, G. Bourdais, D. MacLean, H. Häweker, P. Eckes, and S. Robatzek, “An automated quantitative image analysis tool for the identification of microtubule patterns in plants,” *Traffic*, 2017.

[66] Z. Zhang, Y. Nishimura, and P. Kanchanawong, “Extracting microtubule networks from superresolution single-molecule localization microscopy data,” *Molecular Biology of the Cell*, vol. 28, no. 2, pp. 333–347, 2017.

[67] H. Rogge, N. Artelt, N. Endlich, and K. Endlich, “Automated segmentation and quantification of actin stress fibres undergoing experimentally induced changes,” *Journal of Microscopy*, vol. 268, no. 2, pp. 129–140, 2017.

[68] S. Xia, Y. B. Lim, Z. Zhang, Y. Wang, S. Zhang, C. T. Lim, E. K. Yim, and P. Kanchanawong, “Nanoscale architecture of the cortical actin cytoskeleton in embryonic stem cells,” *Cell Reports*, vol. 28, no. 5, p. 1251–1267, 2019.

[69] Y. Kimori, K. Hikino, M. Nishimura, and S. Mano, “Quantifying morphological features of actin cytoskeletal filaments in plant cells based on mathematical morphology,” *Journal of Theoretical Biology*, vol. 389, pp. 123–131, 2016.

[70] Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis, “Three-dimensional multi-scale line filter for segmentation and visualization of curvilinear structures in medical images,” *Medical Image Analysis*, vol. 2, no. 2, pp. 143–168, 1998.

[71] Skimage. Module: filters. https://scikit-image.org/docs/stable/api/skimage.filters.html#skimage.filters.hessian [accessed: 31.10.2022].

[72] K. Sandberg and M. Brega, “Segmentation of thin structures in electron micrographs using orientation fields,” *Journal of Structural Biology*, vol. 157, no. 2, pp. 403–415, 2007.

[73] Y. Liu, K. Mollaeian, and J. Ren, “An image recognition-based approach to actin cytoskeleton quantification,” *Electronics*, vol. 7, no. 443, 2018.

[74] X. Lan, L. Li, J. Hu, Q. Zhang, Y. Dang, and Y. Huang, “A quantitative method for microtubule analysis in fluorescence images,” *Microscopy and Microanalysis*, vol. 21, no. 6, p. 1582–1590, 2015.

[75] Y. Dang, X. Lan, Q. Zhang, L. Li, and Y. Huang, “Analysis of grayscale characteristics in images of labeled microtubules from cultured cardiac myocytes,” *Microscopy and Microanalysis*, vol. 21, no. 2, pp. 334–342, 2015.

[76] P.K.Bhagat, P. Choudhary, and K. M. Singh, “Chapter 13 - a comparative study for brain tumor detection in mri images using texture features,” *Sensors for Health Monitoring*, pp. 259–287, 2019.

[77] R. G. von Gioi, J. Jakubowicz, J.-M. Morel, and G. Randall, “LSD: a line segment detector,” *Image Processing On Line*, vol. 2, pp. 35–55, 2012.

[78] K. Sandberg and M. Brega, “Segmentation of thin structures in electron micrographs using orientation fields,” *Journal of Structural Biology*, vol. 157, no. 2, p. 403–415, 2007.

[79] B. Eltzner, C. Wollnik, C. Gottschlich, S. Huckemann, and F. Rehfeldt, “The filament sensor for near real-time detection of cytoskeletal fiber structures,” *PLoS ONE*, vol. 10, no. 5, 2015.

[80] D. Sturm, M. Jawad, A. Alonso, and C. Corbo, “A cytoskeleton linearity measure,” *2012 IEEE Southwest Symposium on Image Analysis and Interpretation*, pp. 45–48, 2012.

[81] G. Johnston, D. R. Burton, F. Lilley, A. Doyle, M. F. Murphy, G. Madden, M. A. Gdeisat, C. J. Moore, T. Marchant, and B. Matuszewski, “Analysis of microscopy and reconstructive images for applications in medicine and biology,” *2011 18th IEEE International Conference on Image Processing*, pp. 3069–3072, 2011.

[82] Y. Qu, Z. J. Smith, K. Tyler, S. Chang, S. Shen, M. Sun, and R. X. Xu, “Applying limiting entropy to quantify the alignment of collagen fibers by polarized light imaging,” *Mathematical Biosciences and Engineering*, vol. 18, no. 3, p. 2331–2356, 2020.

[83] P. Asgharzadeh, B. Özdemir, R. Reski, O. Röhrle, and A. I. Birkhold, “Computational 3D imaging to quantify structural components and assembly of protein networks,” *Acta Biomaterialia*, vol. 69, pp. 206–217, 2018.

[84] M. Louveaux, S. Rochette, L. Beauzamy, A. Boudaoud, and O. Hamant, “The impact of mechanical compression on cortical microtubules in *Arabidopsis*: a quantitative pipeline,” *Plant J.*, vol. 88, no. 2, pp. 328–342, 2016.

[85] Q. Ma, J. Sun, and T. Mao, “Microtubule bundling plays a role in ethylene-mediated cortical microtubule reorientation in etiolated hypocotyls,” *Journal of Cell Science*, 2016.

[86] T. Higaki, K. Akita, and K. Katoh, “Coefficient of variation as an image-intensity metric for cytoskeleton bundling,” *Scientific Reports*, vol. 10, no. 22187, 2020.

[87] S. Ciurică, M. Lopez-Sublet, B. L. Loeys, I. Radhouani, N. Natarajan, M. Vikkula, A. H. Maas, D. Adlam, and A. Persu, “Arterial tortuosity,” *Hypertension*, vol. 73, no. 5, pp. 951–960, 2019.

[88] G. Alhussein, A. Shanti, I. A. H. Farhat, S. B. H. Timraz, N. S. A. Alwahab, Y. E. Pearson, M. N. Martin, N. Christoforou, and J. C. M. Teo, “A spatiotemporal characterization method for the dynamic cytoskeleton,” *Cytoskeleton*, vol. 73, no. 5, pp. 221–232, 2016.

[89] S. O. Uppal, D. V. Voronine, E. Wendt, and C. A. Heckman, “Morphological fractal analysis of shape in cancer cells treated with combinations of microtubule-polymerizing and -depolymerizing agents,” *Microscopy and Microanalysis*, vol. 16, no. 4, pp. 472–477, 2010.

[90] Y. Cui, G. Zhang, Z. Liu, Z. Xiong, and J. Hu, “A deep learning algorithm for one-step contour aware nuclei segmentation of histopathological images,” *Med Biol Eng Comput*, vol. 57, pp. 2027–2043, 2018.

[91] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask R-CNN,” *Proceedings of the IEEE International Conference on Computer Vision*, p. 2961–2969, 2017.

[92] U. Schmidt, M. Weigert, C. Broaddus, and G. Myers, “Cell detection with star-convex polygons,” *Medical Image Computing and Computer Assisted Intervention*, pp. 265–273, 2018, gitHub: https://github.com/stardist/stardist.

[93] J. Figueiredo, S. Melo, K. Gamet, T. Godwin, S. Seixas, J. M. Sanches, P. Guilford, and R. Seruca, “E-cadherin signal sequence disruption: a novel mechanism underlying hereditary cancer,” *Molecular Cancer*, vol. 17, no. 112, 2018.

[94] Skimage. Skeletonize. https://scikit-image.org/docs/stable/auto_examples/edges/plot_skeleton.html [accessed: 31.10.2022].

[95] D. Kornhauser. Roipoly. https://github.com/jdoepfert/roipoly.py [accessed: 31.10.2022].

[96] R. M. Rangayyan, *Biomedical Image Analysis*, 1st ed. CRC Press, 2004.

[97] OpenCV. Feature detection - canny(). https://docs.opencv.org/3.4/dd/d1a/group__imgproc__feature.html#ga2a671611e104c093843d7b7fc46d24af [accessed: 31.10.2022].

[98] A. Ganguly, H. Yang, and F. Cabral, "Paclitaxel dependent cell lines reveal a novel drug activity," *Molecular Cancer Therapy*, vol. 9, no. 11, p. 2914–2923, 2010.

[99] H. Yang, A. Ganguly, and F. Cabral, "Inhibition of cell migration and cell division correlates with distinct effects of microtubule inhibiting drugs," *Cell Biology*, vol. 285, no. 42, pp. 32 242–32 250, 2010.

[100] R. R. Goldblum, M. McClellan, C. Hou, B. R. Thompson, K. White, H. X. Vang, H. Cohen, J. M. Metzger, and M. K. Gardner, "Oxidative stress pathogenically remodels the cardiac myocyte cytoskeleton via structural alterations to the microtubule lattice," *bioRxiv*, 2020.

[101] S. G. Lee, S.-N. Lee, J. Baek, J.-H. Yoon, and H. Lee, "Mechanical compression enhances ciliary beating through cytoskeleton remodeling in human nasal epithelial cells," *Acta Biomaterialia*, vol. 128, pp. 346–356, 2021.

[102] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: an open-source platform for biological-image analysis," *Nature Methods*, vol. 9, no. 7, p. 676–682, 2012.

[103] ——, "3D reconstruction of histological sections: Application to mammary gland tissue," *Microscopy Research and Technique*, vol. 73, no. 11, p. 1019–1029, 2010.

[104] J. K. Mouw, G. Ou, and V. M. Weaver, "Extracellular matrix assembly: a multiscale deconstruction," *Nature Reviews Molecular Cell Biology*, vol. 15, p. 771–785, 2014.

[105] Y. Jiang, H. Zhang, J. Wang, Y. Liu, T. Luo, and H. Hua, "Targeting extracellular matrix stiffness and mechanotransducers to improve cancer therapy," *Journal of Hematology and Oncology*, vol. 15, no. 34, 2022.