

Exploring Multi-Agent Reinforcement Learning for Cell Mechanics

Muhammad Waris
Department of Electronics,
Quaid-e-Azam University
Islamabad, Pakistan
mwaris.22411012@ele.qau.edu.pk

Mustafa Shah
Department of Electronics,
Quaid-e-Azam University
Islamabad, Pakistan
mustafamohmand59@gmail.com

Arsenio Cutolo
Department of Structures for
Engineering and Architecture
University of Napoli Federico II, Italy
arsenio.cutolo@unina.it

Musarat Abbas
Department of Electronics,
Quaid-e-Azam University
Islamabad, Pakistan
mabbas@qau.edu.pk

Abstract—Cell aggregation, where cells stick together, is a key process in many biological events like how embryos form, how tissues heal, and how microbes create communities. Studying this involves looking at different types of data, from detailed molecular information to images and patient data. With new technologies, we have access to large amounts of this data in public databases. Analyzing and combining this complex information requires advanced computer methods. While there are challenges in handling and integrating these diverse datasets, exploring them helps us understand basic biology, develop models for diseases, find new drugs, and advance regenerative medicine. This report reviews these data types, sources, and analysis methods to guide research in this important field.

Index Terms—Reinforcement Learning, MARL, Cell Mechanics, Cell aggregation

I. INTRODUCTION

CELL aggregation [3], the process by which individual cells adhere to one another to form multicellular structures [4], represents a fundamental biological phenomenon observed across the tree of life. This self-assembly is not merely a passive physical process but is frequently governed by intricate molecular mechanisms and dynamic cellular behaviors. Aggregation plays critical roles in diverse contexts, ranging from the formation of complex organisms during embryonic development to the establishment of resilient microbial communities known as biofilms [14]. It is also central to physiological processes such as hemostasis, where platelets aggregate to form blood clots, and immune responses, involving the clustering of lymphocytes and other immune cells at sites of infection or within specialized lymphoid tissues. Furthermore, in vitro cell aggregation is the foundational principle behind the generation of three-dimensional (3D) cell culture models, including spheroids and organoids [24], which serve as powerful tools for studying tissue development, disease modeling, and drug screening.

Understanding the intricacies of cell aggregation across these varied biological systems requires the collection and

analysis of diverse types of data. Modern high-throughput technologies, such as next-generation sequencing, advanced microscopy, and automated functional assays, are generating vast amounts of quantitative data related to cellular composition, molecular profiles, spatial organization, and dynamic behaviors within aggregating cell populations. Navigating and leveraging these extensive datasets, often stored in public repositories, presents both opportunities and challenges for researchers.

Artificial Intelligence (AI) and Machine Learning (ML) has significant uses in many areas including healthcare [9], vehicular communication [10], e-learning [2], rehabilitation [12] and risk management [7]. Reinforcement Learning (RL) is one the most promising type of ML [17] that has brought revolution in different areas and cell mechanics can also be benefited with this technology. Multi Agent RL (MARL) is the extension of RL where multiple agents are being used for multiple task within a bigger task.

This paper explores the landscape of data relevant to cell aggregation by examining key biological scenarios where it plays a critical role. The types of data generated by various experimental techniques are categorized, and prominent public data repositories where these data are stored and can be accessed are identified. The aim is to provide a structured overview for researchers seeking to utilize existing datasets to study cell aggregation phenomena.

II. BACKGROUND

Cell aggregation [18] is a fundamental process that underpins the formation, function, and maintenance of biological structures at multiple scales. Its significance spans numerous fields of biological and medical research [19].

In embryonic development, cell aggregation is a primary mechanism driving morphogenesis from the zygote. Following initial cell divisions [5], blastomeres aggregate to form the morula, a compact ball of cells. This compaction is critical

for establishing cell polarity and initiating the first lineage segregation, leading to the formation of the blastocyst with its distinct inner cell mass and trophoctoderm. Subsequent aggregation and rearrangement of cells within the developing embryo give rise to the three germ layers—ectoderm, mesoderm, and endoderm—which then differentiate and organize into the precursors of all tissues and organs. The precise timing and spatial control of these aggregation and differentiation events are governed by complex genetic programs and cell-cell communication mediated by signaling pathways.

Organoid formation in the laboratory directly leverages the inherent ability of cells, particularly stem cells, to aggregate and self-organize into 3D structures resembling native tissues. By providing specific biochemical cues, such as growth factors and signaling molecules, and appropriate physical environments, researchers can guide the aggregation and differentiation of pluripotent or adult stem cells to generate organoids mimicking various organs like the brain, intestine, kidney, or liver. These 3D models offer significant advantages over traditional two-dimensional cell cultures by better recapitulating the complex cell-cell interactions, tissue architecture, and physiological functions of their *in vivo* counterparts.

In the microbial world, biofilm formation is a widespread lifestyle characterized by the aggregation of bacteria and other microorganisms on surfaces, encased within a self-produced extracellular matrix (ECM). This aggregated lifestyle provides significant advantages, including enhanced resistance to environmental stresses, disinfectants, and antibiotics, as well as protection from host immune responses. Biofilms are implicated in numerous industrial issues and persistent infections, making the study of their formation and dispersal critical for developing effective control strategies.

Blood clotting, or platelet aggregation [20], plays a critical role in hemostasis [13], the body's natural process for stopping bleeding after a blood vessel is injured. Platelets quickly gather and stick together at the injury site, forming a plug that's strengthened by fibrin to seal the damaged vessel. While essential for survival, if platelet aggregation becomes uncontrolled, it can lead to dangerous thrombosis—the formation of clots within healthy blood vessels. This can result in serious conditions like deep vein thrombosis, pulmonary embolism, stroke, and heart attack. Understanding the mechanisms of platelet aggregation is therefore vital for diagnosing bleeding disorders and developing treatments to prevent clots.

Immune cell aggregation is a critical aspect of the adaptive immune response. Following recognition of foreign antigens, lymphocytes and other immune cells proliferate and aggregate in secondary lymphoid organs, forming structures like germinal centers within B cell follicles. These aggregates provide specialized microenvironments for processes such as B cell affinity maturation and the generation of memory cells and antibody-secreting plasma cells, which are essential for long-lived immunity and effective vaccination. Immune cell aggregation also occurs at sites of infection or inflammation, facilitating coordinated cellular interactions to clear pathogens or resolve tissue damage.

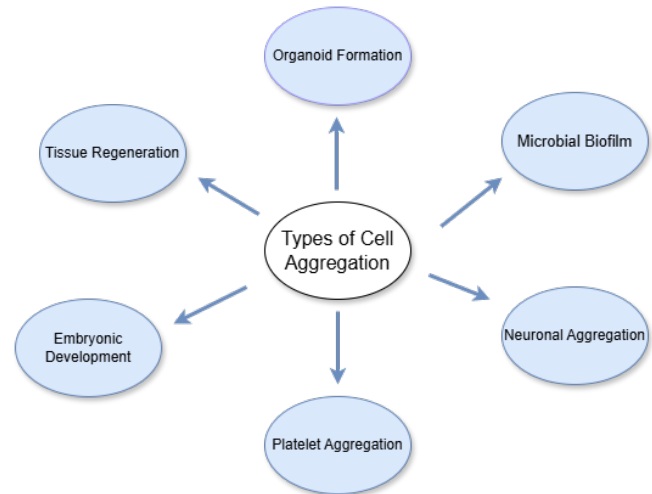


Fig. 1. Types of Cell Aggregation

In the context of tissue regeneration [15] and engineering, cell aggregation techniques are employed to create multicellular building blocks, such as spheroids or organoids, which can be used to repair or replace damaged tissues. Understanding how cells aggregate, maintain viability, and differentiate within these 3D structures is vital for developing effective regenerative therapies [23]. Mechanical forces and cell-cell interactions within these aggregates play a significant role in directing cell fate and tissue organization.

Finally, neuronal aggregation [21] is a key stage in the development of the nervous system. As newly generated neurons migrate to their final destinations in the brain, they aggregate with similar cell types to form distinct brain regions and layers. This process is guided by cell-cell recognition and adhesion molecules and is crucial for establishing the complex circuitry of the brain. Dysregulation [11] of neuronal aggregation and migration is implicated in various neurological disorders.

The study of these diverse cell aggregation phenomena is fundamentally important for unraveling basic biological principles, creating accurate models of human health and disease, and developing innovative therapeutic and biotechnological applications. Multi-Agent Reinforcement Learning (MARL) emerges as a particularly promising computational paradigm. MARL, a specialized subfield of artificial intelligence, is designed to model complex systems where multiple autonomous agents interact and learn in a shared environment. This report explores the application of MARL to the intricate domain of cell mechanics, aiming to address the inherent limitations of traditional computational approaches in fully capturing the multi-agent nature and emergent properties of cellular systems. The subsequent sections detail the necessary revisions to enhance the paper's technical depth and highlight the unique contributions of MARL to this vital field.

III. DATA TYPES RELEVANT TO CELL AGGREGATION STUDIES

Investigating the multifaceted nature of cell aggregation necessitates the acquisition and analysis of data across various scales, from the molecular interactions governing cell adhesion to the macroscopic morphology and dynamics of the resulting aggregates.

Molecular Data [16] provides insights into the genetic programs, protein machinery, and signaling networks that regulate cell aggregation and the subsequent behavior of aggregated cells. Transcriptomics, encompassing techniques like bulk RNA sequencing, single-cell RNA sequencing (scRNA-seq), and spatial transcriptomics, provides insight into gene expression patterns. These patterns, in turn, determine a cell's identity, its stage of differentiation, and how it responds to its surroundings within cellular groupings. Genomics data [1], including DNA sequence variations, copy number changes, and epigenetic modifications, provide the foundational genetic and regulatory landscape influencing aggregation potential and associated disease states. Proteomics data [8] identify the proteins present, their abundance, and post-translational modifications, detailing the molecular machinery of cell adhesion, ECM production, and signal transduction within aggregates. Data on signaling pathways, including the activity of receptors, kinases, and transcription factors, illuminate how cells perceive and respond to their environment and coordinate collective behaviors like aggregation, differentiation, and migration.

Cellular Data captures the physical characteristics and activities of individual cells and cell populations within aggregates. Imaging data, acquired through various microscopy techniques (light, confocal, electron, time-lapse, spatial), provides visual information on cell morphology, spatial arrangement, and the dynamic process of aggregation and structural development. Functional assay data quantifies cellular activities such as electrophysiological signaling in neuronal aggregates or organoids, transport function in epithelial structures, or responses to external stimuli like drugs or pathogens. Flow cytometry provides high-throughput, single-cell analysis of protein expression, enabling the identification and quantification of distinct cell types and their activation states within heterogeneous populations, particularly relevant for immune cells and platelets.

Clinical Data [6] provides essential context for studying cell aggregation in disease. This includes patient demographics, medical history, lifestyle factors, treatment regimens, disease severity, and clinical outcomes. Such data are critical for correlating in vitro findings with in vivo conditions and assessing the translational relevance of research, particularly in areas like thrombosis and immune disorders.

IV. EXPLORING CELL AGGREGATION DATA ACROSS BIOLOGICAL CONTEXTS

The application of these diverse data types varies depending on the specific biological context of cell aggregation being studied. Each scenario presents unique challenges and opportunities for data exploration .

In the study of organoid formation, a key aspect is understanding how these in vitro aggregates recapitulate the complexity of native organs. Single-cell RNA sequencing (scRNA-seq) is indispensable for dissecting the cellular heterogeneity within organoids, identifying the different cell types that emerge during differentiation, mapping their developmental trajectories, and comparing their molecular profiles to those of cells in primary tissues. Dedicated databases like OrganoidDB serve as valuable resources for exploring organoid transcriptomes, including extensive collections of scRNA-seq data. The inherent variability observed between individual organoids, even within the same culture, underscores the need for high-throughput quantitative data collection and analysis. This variability can be assessed through large-scale scRNA-seq studies of many organoids or through automated imaging analysis. Imaging data, particularly from brightfield, phase contrast, and confocal microscopy, provides crucial information on organoid morphology, size, growth kinetics, and the formation of complex structures like lumens. The large volume of images generated in high-throughput organoid screens necessitates automated image analysis tools, often employing machine learning, to segment, quantify, and track individual organoids. Datasets like MultiOrg specifically provide microscopy images of organoids with annotations for training such tools. Beyond structural and compositional analysis, functional assay data are critical for validating whether organoids truly mimic the physiological activities of their corresponding organs. This includes assessing barrier function, transport activity (e.g., in kidney or intestinal organoids), or electrophysiological signaling (e.g., in brain organoids). The combination of multi-omics (genomics, transcriptomics, proteomics, metabolomics) and functional data is essential for a thorough assessment of organoid authenticity, stability, and translational potential, particularly for applications in disease modeling and drug screening.

Investigating embryonic development requires unraveling precisely controlled spatiotemporal events, including cell aggregation, migration, and differentiation. Gene expression data, from bulk and single-cell transcriptomics, provides a molecular narrative of these processes, revealing which genes are active at different developmental stages and in different cell lineages. However, understanding development requires knowing where genes are expressed within the developing tissue. Spatial transcriptomics addresses this need by mapping gene expression profiles while preserving spatial information, providing molecular maps of embryonic structures and cellular organization. The four-dimensional nature of development (3D space over time) makes the integration of spatial and temporal data particularly crucial for linking molecular events to dynamic cellular behaviors and structural changes. Time-lapse microscopy captures the dynamic morphological aspects of embryonic development, including cell division timings, migration patterns, and the process of aggregation and morphogenesis in living embryos over extended periods. This generates massive datasets, particularly in applications like IVF, which necessitate advanced computational meth-

ods, such as machine learning, to automate analysis, extract morphokinetic parameters, and identify predictive patterns for embryo viability. Public repositories like GEO, TEDD, and the Allen Brain Atlas (Developing Mouse/Human) and BrainSpan provide access to vast amounts of gene expression and anatomical data from developing organisms.

The study of microbial biofilm formation relies heavily on understanding the transition from planktonic single cells to aggregated communities and the molecular mechanisms underlying this process. Genomic and transcriptomic data reveal the genes involved in surface attachment, cell-cell adhesion, ECM production, quorum sensing, and stress responses that are upregulated or downregulated during biofilm development. Multi-omics approaches, integrating genomics, transcriptomics, and proteomics, provide a more comprehensive view of the molecular changes and functional pathways involved in biofilm formation and resistance. Imaging data, particularly from confocal laser scanning microscopy (CLSM), is essential for visualizing the 3D structure of biofilms, including microcolonies, water channels, and the distribution of cells and ECM components. Time-lapse imaging allows tracking the dynamics of biofilm growth and dispersal. Public repositories like GEO and specialized biofilm databases (e.g., aBiofilm, BioOmics, Biofilms Structural Database, BRaID) serve as sources for genomic, transcriptomic, and sometimes image data related to biofilms. Understanding the molecular mechanisms driving phenotypic shifts during biofilm formation is significantly enhanced by integrating multi-omics data, while imaging captures the essential 3D structure and dynamic processes of aggregation.

Research on platelet aggregation and thrombus formation involves characterizing the rapid cellular and molecular events occurring at sites of vascular injury. Data from aggregometry, especially light transmission aggregometry (LTA) and impedance aggregometry, quantifies how platelets clump together and the degree to which they do so when exposed to different agonists. These assays provide quantitative parameters such as maximum aggregation, slope, and lag phase. Microscopy images, especially time-lapse fluorescence and DIC microscopy of thrombus formation under flow conditions, visualize the process of platelet adhesion, shape change, aggregation, and the incorporation of fibrin and other blood cells into the growing thrombus. These images allow for quantitative analysis of thrombus size, morphology, and dynamics. Flow cytometry is used to analyze platelet activation markers and identify distinct platelet subpopulations within blood samples. Clinical data from patients with thrombotic disorders [22] or bleeding tendencies are essential for identifying risk factors, correlating laboratory findings with clinical outcomes, and evaluating the effectiveness of antiplatelet and anticoagulant therapies. Public resources like clinical trial databases (e.g., ClinicalTrials.gov), disease-specific registries (e.g., ISTH registries), and genomic databases (e.g., NIH GTR) provide access to relevant clinical and genetic data.

The study of immune cell aggregation, such as in germinal centers or at infection sites, involves characterizing

the cellular composition, spatial organization, and functional interactions of immune cells. Flow cytometry, including high-dimensional techniques like CyTOF, is widely used to identify and quantify different immune cell subsets based on surface protein expression and analyze their activation states within heterogeneous populations. Repositories like ImmPort house extensive flow cytometry data from immunology studies and clinical trials. Imaging data, such as intravital microscopy, allows visualization and tracking of immune cell migration and interactions in real-time within tissues, providing spatial and dynamic context to flow cytometry findings. Databases like IDR and those linked through the Human Cell Atlas initiatives may contain relevant imaging data. Data on cytokines and chemokines are critical for understanding the molecular signals that mediate immune cell recruitment, activation, and communication within aggregates. Databases like ImmPort and specialized cytokine/chemokine resources (e.g., CYTOCON DB, Cell Interaction Knowledgebase) provide access to these data.

For tissue regeneration and engineering, data focuses on the behavior of cells within aggregates used as building blocks. Cellular data, including viability, proliferation, differentiation status (often assessed via markers), and the impact of mechanical forces or environmental cues, are critical. Imaging data captures the formation, growth, and structural organization of these cellular aggregates, as well as the integration of different cell types in co-cultures. Data on the composition and properties of the extracellular matrix within aggregates or surrounding them (e.g., hydrogels) is also important, as the ECM provides structural support and signaling cues influencing cell behavior.

Understanding neuronal aggregation during brain development involves characterizing the types of neurons, their migratory paths, and how they organize into specific brain structures. Imaging data, including light microscopy, electron microscopy, and various brain imaging modalities (MRI, fMRI), provides visual information on neuronal morphology, connectivity, and the large-scale structure of the brain formed by aggregated neurons. Specific data types include neuronal morphology reconstructions (neuronal tracing data), electrophysiological recordings of neuronal activity, and gene expression profiles (transcriptomics) related to neuronal development and cell type specification. Large public databases like NeuroMorpho.Org and the Allen Brain Atlas suite provide access to extensive datasets on neuronal morphology, gene expression, and connectivity.

V. PUBLIC DATA SOURCES

Access to publicly available data is crucial for advancing research in cell aggregation. Numerous repositories host relevant datasets, often specialized by data type or biological domain.

The Gene Expression Omnibus (GEO) serves as a prominent international public repository for a wide range of high-throughput functional genomics data, including genomic, transcriptomic, and epigenomic datasets, including microarray and next-generation sequencing data. GEO supports various

organisms and experimental conditions, making it a valuable resource for studying gene expression changes during aggregation processes in diverse contexts, including embryonic development, organoid formation, and biofilm development. Data can be searched and downloaded via the GEO DataSets and GEO Profiles interfaces, FTP, or programmatic access.

For organoid-specific transcriptomic data, OrganoidDB provides a comprehensive resource for bulk and single-cell RNA-seq profiles of human and mouse organoids, integrating data from GEO and ArrayExpress. It allows searching and browsing based on organoid type, source, protocol, and developmental stage.

For neuronal morphology and related data, NeuroMorpho.Org is a centrally curated inventory of digitally reconstructed neurons from various species, providing 3D morphological data and associated metadata. The Allen Brain Atlas suite provides extensive resources for neuronal data, including gene expression atlases for adult and developing mouse and human brains, connectivity maps, and single-cell characterization data (morphological, electrophysiological, transcriptomic). Data can be accessed via web portals, APIs, and SDKs. The BRAIN Initiative Cell Census Network (BICCN) also provides access to multimodal brain cell atlas data through various archives like NeMO, BIL, and DANDI.

For immune cell and cytokine/chemokine data, the Immunology Database and Analysis Portal (ImmPort) is a major repository for immunology research data, including clinical trial data, flow cytometry, and multiplex cytokine/chemokine data. ImmPort provides tools for searching, downloading, and analyzing shared data.

The Image Data Resource (IDR) serves as a public repository for imaging data, specifically microscopy images of cells and tissues. This resource archives image datasets from published scientific research, accommodating diverse imaging techniques and organisms. Users can search for and access high-quality biological image data through this platform.

Several resources are available for clinical data concerning thrombotic disorders. These include established clinical trial databases (such as ClinicalTrials.gov), registries specific to diseases (for example, those maintained by the ISTH for rare bleeding disorders or VTE), and certain extensive claims or electronic health record databases, though access to the latter might be limited. Additionally, the NIH Genetic Testing Registry (GTR) offers details on genetic tests relevant to thrombotic conditions.

For histological images, resources like the GTEx Tissue Image Library and specialized datasets like TissueNet or those linked through initiatives like TCGA or Human Protein Atlas provide access to tissue histology images, sometimes with annotations.

For biofilm genomic and transcriptomic data, in addition to GEO, specialized databases like BBSdb and the Biofilms Structural Database (BSD) exist, though access methods vary. Some data may also be available in generalist repositories like Dryad or institutional repositories.

VI. MULTI AGENT REINFORCEMENT LEARNING AND MACHINE LEARNING TO CELL AGGREGATION

A. Cancer: Histopathology and scRNA-seq Data Analysis

1) *Image Analysis (Histopathology)*: Deep learning models (e.g., Convolutional Neural Networks - CNNs) can be trained on histopathology images to identify cancerous aggregation patterns, tumor boundaries, and predict malignancy. Multi-RL can then be used to optimize image segmentation and classification by learning from different expert annotations or even guiding the sampling of new image regions for analysis.

2) *scRNA-seq for cell state and interaction*: ML algorithms such as clustering (e.g., t-SNE, UMAP, K-means) can identify distinct cell populations and their aggregation tendencies from scRNA-seq data. Multi-RL can be employed to model the dynamic interactions between different cell types (e.g., cancer cells, immune cells, stromal cells) within the tumor microenvironment. Each cell type could be considered an agent, learning optimal strategies for proliferation, migration, or interaction based on the transcriptional states of neighboring cells, allowing for prediction of tumor growth or response to therapy.

B. Wound Healing: Microscopy and scRNA-seq Data Analysis

1) *Time Lapse Microscopy for Cell Dynamics*: ML algorithms can track individual cell movements and aggregation dynamics from time-lapse microscopy images. Multi-RL can model the collective behavior of cells (e.g. fibroblasts, immune cells, keratinocytes) during wound closure. Each cell or a group of cells can act as an agent, learning policies for migration, proliferation, and extracellular matrix remodeling to optimize healing efficiency, potentially identifying bottlenecks or aberrant healing processes.

2) *Spatial Transcriptomics for Cellular Coordination*: Integrating scRNA-seq with spatial information allows us to understand how different cell types spatially interact during wound healing. ML can identify spatial gene expression patterns that indicate successful healing. Multi-RL can then simulate the "decision-making" of cells based on their local environment and gene expression, learning how to coordinate their actions (e.g., secreting growth factors, migrating towards specific cues) to achieve optimal tissue regeneration.

C. Embryogenesis: Live Imaging and Spatial RNA-seq Data Analysis

1) *Modeling Morphogenesis*: Live imaging data provides dynamic information on cell shape changes and movements. ML models can be trained to predict developmental outcomes based on initial cell configurations. Multi-RL is highly suitable for modeling complex, self-organizing processes of embryogenesis. Each cell or group of cells can be an agent, learning from its neighbors and environmental cues to make "decisions" regarding division, differentiation, migration, and adhesion, ultimately forming complex tissues and organs. The "reward" signal could be the successful formation of a specific tissue structure or stage of development.

2) *Spatial Transcriptomics for Developmental Programs:* Spatial RNA-seq data reveals gene expression patterns across developing tissues. ML can identify gene regulatory networks driving cell aggregation and differentiation. Multi-RL agents, representing different cell lineages, can learn optimal strategies for gene expression changes and physical interactions to achieve proper tissue patterning and organogenesis. This could involve simulating how cells interpret and respond to morphogen gradients and mechanical forces to reach their correct positions and fates.

D. Immune Swarming: Immune Imaging and scRNA-seq Data Analysis

1) *Tracking Immune Cell Dynamics:* Immune imaging data allows for tracking the movement and interactions of immune cells. ML can identify different immune cell subsets and their migration paths. Multi-RL can simulate immune swarming by treating individual immune cells or groups as agents. These agents can learn to chemotax (move along chemical gradients), interact with pathogens, and coordinate with other immune cells to effectively clear infections or respond to inflammation. The "reward" could be the successful containment of a pathogen or resolution of inflammation.

2) *Predicting Immune Response Outcomes:* scRNA-seq provides insights into the transcriptional states of immune cells during aggregation. ML can correlate these states with disease outcomes. Multi-RL can be used to model the adaptive strategies of immune cells in response to evolving threats, optimizing their aggregation and effector functions. For example, agents could learn to upregulate specific receptors, secrete cytokines, or initiate cell-to-cell contact based on the presence of pathogens or signals from other immune cells, leading to a more efficient and coordinated immune response.

E. Neural Aggregation: Brain Organoids and scRNA-seq Data Analysis

1) *Predicting Neuronal Migration and Circuit Formation:* ML models can analyze time-lapse imaging, gene expression data, and spatial transcriptomics data from brain organoids to predict the trajectories of migrating neurons and the formation of neural circuits. Multi-RL can simulate the intricate dance of neuronal migration and circuit assembly. Individual neurons or neuronal clusters can be agents that learn to navigate complex environments, form connections with appropriate partners (synaptogenesis), and integrate into functional networks. The "reward" signal could be the successful formation of a mature neural circuit with specific functional properties, as assessed by electrophysiological recordings or imaging data.

2) *Modeling Neuroplasticity and Disease Progression:* Multi-RL can be used to model neuroplasticity, where neurons learn to adapt their connections and firing patterns in response to stimuli. In the context of neurodevelopmental or neurodegenerative diseases, Multi-RL could simulate how aberrant aggregation or connectivity leads to dysfunction. Agents (neurons) could learn to compensate for damage or disease-related changes, or conversely, models could identify

tipping points where the system transitions to a diseased state. This could inform strategies for intervention or rehabilitation.

F. Cardiac Cell Repair: Heart Tissue Imaging Data Analysis

1) *Modeling Myocardial: Regeneration* Multi-RL can simulate the complex interplay of various cell types involved in cardiac repair, including cardiomyocytes, fibroblasts, and immune cells. Each cell type could be an agent, learning to respond to signals from the damaged microenvironment (e.g., inflammatory cues, growth factors) to contribute to tissue regeneration. This could involve learning optimal strategies for proliferation, differentiation, and secretion of extracellular matrix components to promote functional tissue repair and prevent maladaptive remodeling. This type of modeling could lead to the identification of novel therapeutic targets to enhance cardiac repair.

2) *Optimizing Cell Delivery and Engraftment:* ML algorithms can analyze heart tissue imaging (e.g., histology, gene expression profiles) to assess the survival, integration, and functional impact of transplanted cells (e.g., stem cells, cardiomyocytes) in damaged heart tissue. Multi-RL can then be employed to optimize cell delivery strategies. Agents (e.g., individual transplanted cells or surrounding host cells) could learn to interact optimally to promote engraftment, vascularization, and functional integration into the host myocardium. The "reward" could be measured by improvements in cardiac function, reduced scar tissue formation, or successful electrical coupling.

VII. FORMULATING A CELL MECHANICS PROBLEM INTO MARL

The process of formulating a cell mechanics problem into a Multi-Agent Reinforcement Learning (MARL) framework requires translating biological phenomena into computational elements while preserving the complex, emergent nature of multicellular systems.

In this formulation:

A. Agents

Agents correspond to autonomous biological cells (e.g., blastomeres, epithelial cells, immune responders), each acting based on local perceptions and internal states.

B. States

States encapsulate multidimensional cell features such as spatial coordinates, polarity vectors, cell cycle phase, gene expression profiles, mechanical tension, and adhesion strength. These may be derived from real-time imaging, transcriptomics, and biomechanical simulations.

C. Actions

Actions include discrete and continuous choices like migration, division, differentiation, polarity realignment, ECM remodeling, and intercellular signaling.

D. Reward Functions

Reward Functions are formulated to capture biologically meaningful objectives—such as optimizing tissue cohesion, minimizing energy expenditure, achieving correct positional fate, or synchronizing morphogenetic movements. These may include sparse or dense feedback and require multi-objective optimization.

E. Environment

Environment refers to the spatial-temporal tissue context, characterized by dynamic morphogen gradients, extracellular matrix properties, boundary conditions, and interactions with neighboring agents.

VIII. DISCUSSION

A. Advantages of MARL Over Single-Agent RL in Cell Mechanics

1) *Decentralized Coordination*: Biological cells function as autonomous entities, responding to local signals and engaging in self-organized behavior. MARL mirrors this natural decentralization, enabling accurate modeling of emergent developmental processes.

2) *Modeling Emergent Properties*: Complex multicellular phenomena such as morphogenesis and spatial patterning arise from local interactions. MARL is inherently suited to discover and simulate these emergent properties through distributed policy learning.

3) *Robustness to Perturbations*: In fluctuating and noisy biological environments, MARL provides resilience by allowing agents to adapt locally. This makes the system robust against disruptions, mimicking biological fault tolerance.

B. Challenges and Future Directions

Despite its transformative potential, applying Multi-Agent Reinforcement Learning (MARL) to cell mechanics is constrained by three core challenges. First, the vast spatial, temporal, and molecular complexity of multicellular systems creates high-dimensional environments that challenge MARL scalability. Second, designing biologically valid and multi-objective reward functions is non-trivial, requiring precise alignment with physiological outcomes. Third, integrating diverse data types like imaging, transcriptomics, and spatial omics into unified agent frameworks demands advanced modeling strategies. Addressing these challenges will require interdisciplinary advances in AI, systems biology, and data integration to fully leverage MARL for biological discovery.

Future research in MARL for cell mechanics should prioritize the development of biologically constrained multi-agent architectures, capable of encoding known intercellular signaling networks and mechanotransduction rules. Hybrid learning models that integrate reinforcement learning with supervised or self-supervised modules will be essential to leverage annotated biological datasets. Simultaneously, scalable data assimilation frameworks must be established to incorporate real-time spatial transcriptomics, live-cell

imaging, and dynamic tissue properties. Integrating these MARL systems with in vitro experimental platforms via co-simulation or closed-loop control could enable predictive modeling of morphogenesis and regeneration. Collectively, these efforts will transform MARL into a practical and predictive toolset for mechanobiology, synthetic development, and regenerative engineering.

IX. CONCLUSIONS

Cell aggregation is a fundamental biological process occurring across diverse scales and contexts, from the formation of multicellular organisms to the organization of microbial communities and the coordination of cellular responses in health and disease. Studying these phenomena requires integrating data from a wide array of experimental technologies, including genomics, transcriptomics, proteomics, advanced microscopy, functional assays, flow cytometry, and clinical data [12].

The exploration of cell aggregation data is significantly enhanced by the availability of public repositories. Databases like GEO, OrganoidDB, NeuroMorpho.Org, the Allen Brain Atlas suite, ImmPort, IDR, and specialized biofilm and clinical databases provide access to vast amounts of data, enabling researchers to investigate molecular mechanisms, cellular behaviors, and clinical correlations related to aggregation.

The inherent complexity and often high-throughput nature of data generated in cell aggregation studies, such as the large volumes of images from time-lapse microscopy of developing embryos or the high-dimensional data from single-cell transcriptomics and flow cytometry of organoids or immune cells, necessitate the use of advanced computational analysis methods, including machine learning and sophisticated visualization tools.

Future efforts in cell aggregation data exploration should focus on improving data integration across different modalities and repositories, developing standardized metadata and data formats to facilitate data sharing and reuse, and creating user-friendly computational tools that enable researchers from diverse backgrounds to effectively analyze and interpret these complex datasets. By leveraging the wealth of available data and developing innovative analytical approaches, the scientific community can gain deeper insights into the fundamental principles of cell aggregation and translate this knowledge into advancements in regenerative medicine, disease understanding, and therapeutic development.

ACKNOWLEDGMENT

This research has been partially supported by the European Union - Next Generation EU through the Project of National Relevance "Innovative mathematical modeling for cell mechanics: global approach from micro-scale models to experimental validation integrated by reinforcement learning", financed by European Union-Next-GenerationEU-National Recovery and Resilience Plan-NRRP-M4C1-I 1.1, CALL PRIN 2022 PNRR D.D. 1409 14-09-2022—(Project code P2022MXCJ2, CUP F53D23010080001) granted by the Italian MUR.

REFERENCES

- [1] Database resources of the national genomics data center, china national center for bioinformation in 2025. *Nucleic Acids Research*, 53(D1):D30–D44, 2025. <https://doi.org/10.1093/nar/gkae978>.
- [2] Fares Abomelha and Paul Newbury. A vark learning style-based recommendation system for adaptive e-learning. *Annals of Computer Science and Information Systems*, 41:1–8, 2024.
- [3] Anika Alexandrova-Watanabe, Emilia Abadjieva, Lidia Gartcheva, Ariana Langari, Miroslava Ivanova, Margarita Guenova, Tihomir Tiankov, Velichka Strijkova, Sashka Krumova, and Svetla Todinova. The impact of targeted therapies on red blood cell aggregation in patients with chronic lymphocytic leukemia evaluated using software image flow analysis. *Micromachines*, 16(1):95, 2025. <https://doi.org/10.3390/mi16010095>.
- [4] Mario Argenziano, Massimiliano Zingales, Arsenio Cutolo, Emanuela Bologna, and Massimiliano Fraldi. Competition between elasticity and adhesion in caterpillar locomotion. *Journal of the Royal Society Interface*, 22(225):20240703, 2025. <https://doi.org/10.1098/rsif.2024.0703>.
- [5] Hugo Cano-Fernández, Tazzio Tissot, Miguel Brun-Usan, and Isaac Salazar-Ciudad. A mathematical model of development shows that cell division, short-range signaling and self-activating gene networks increase developmental noise while long-range signaling and epithelial stiffness reduce it. *Developmental Biology*, 518:85–97, 2025. <https://doi.org/10.1016/j.ydbio.2024.11.014>.
- [6] Oluwadamilola M Fayanju, Elliott R Haut, and Kamal Itani. Practical guide to clinical big data sources. *JAMA surgery*, 2025. <https://jamanetwork.com/journals/jamasurgery/article-abstract/2828666>.
- [7] Mario Fiorino, Muddasar Naeem, Mario Ciampi, and Antonio Coronato. Defining a metric-driven approach for learning hazardous situations. *Technologies*, 12(7):103, 2024.
- [8] Tiannan Guo, Judith A Steen, and Matthias Mann. Mass-spectrometry-based proteomics: from single cells to clinical applications. *Nature*, 638(8052):901–911, 2025. <https://www.nature.com/articles/s41586-025-08584-0>.
- [9] Ahsan Ismail, Muddasar Naeem, Madiha Syed, Musarat Abbas, and Antonio Coronato. Advancing patient care with an intelligent and personalized medication engagement system. *Information*, 15:609, 10 2024.
- [10] Mansoor Jamal, Zaib Ullah, Muddasar Naeem, Musarat Abbas, and Antonio Coronato. A hybrid multi-agent reinforcement learning approach for spectrum sharing in vehicular networks. *Future Internet*, 16(5):152, 2024.
- [11] Kevin G Johnston, Bereket T Berackey, Kristine M Tran, Alon Gelber, Zhaoxia Yu, Grant R MacGregor, Eran A Mukamel, Zhiqun Tan, Kim N Green, and Xiangmin Xu. Single-cell spatial transcriptomics reveals distinct patterns of dysregulation in non-neuronal and neuronal cells induced by the trem2 r47h alzheimer’s risk gene mutation. *Molecular Psychiatry*, 30(2):461–477, 2025. <https://www.nature.com/articles/s41380-024-02651-0>.
- [12] Umamah Khalid, Muddasar Naeem, Fabrizio Stasolla, Madiha Syed, Musarat Abbas, and Antonio Coronato. Impact of ai-powered solutions in rehabilitation process: Recent improvements and future trends. *International Journal of General Medicine*, 17:943–969, 03 2024. <https://doi.org/10.2147/IJGM.S453903>.
- [13] Aziz Kubaev, Fadhil Faez Sead, Mohammad Pirouzbakht, Mobina Nazari, Hani Riyahi, Omolbanin Sargazi Aval, Alireza Hasanvand, Forough Mousavi, and Hamed Soleimani Samarkhazan. Platelet-derived extracellular vesicles: emerging players in hemostasis and thrombosis. *Journal of Liposome Research*, pages 1–11, 2025. <https://doi.org/10.1080/08982104.2025.2495261>.
- [14] Jiaming Lan, Jingyu Zou, He Xin, Jin Sun, Tao Han, Mengchi Sun, and Meng Niu. Nanomedicines as disruptors or inhibitors of biofilms: Opportunities in addressing antimicrobial resistance. *Journal of Controlled Release*, page 113589, 2025. <https://doi.org/10.1016/j.jconrel.2025.113589>.
- [15] Sang Jin Lee, Zhenzhen Wu, Mengyu Huang, Chao Liang, Ziqi Huang, Siyuan Chen, Vidhyashree Rajasekar, Mohamed Mahmoud Abdalla, Haram Nah, Dong Nyoung Heo, et al. Crosslinker-free in situ hydrogel induces self-aggregation of human dental pulp stem cells with enhanced antibacterial activity. *Materials Today Bio*, 31:101451, 2025. <https://doi.org/10.1016/j.mtbio.2025.101451>.
- [16] Tiqing Liu, Linda Hwang, Stephen K Burley, Carmen I Nitsche, Christopher Southan, W Patrick Walters, and Michael K Gilson. Bindingdb in 2024: a fair knowledgebase of protein-small molecule binding data. *Nucleic acids research*, 53(D1):D1633–D1644, 2025. <https://doi.org/10.1093/nar/gkae1075>.
- [17] Cu Kim Long, Vijender Kumar Solanki, Nguyen Viet Anh, Luu Hoang Bach, Cu Ngoc Son, et al. Machine learning-based prediction models for sentiment analysis on online customer reviews: A case study on airbnb. *Annals of Computer Science and Information Systems*, 42:103–116, 2024.
- [18] Muddasar Naeem, Mario Fiorino, Pia Addabbo, Antonio Coronato, et al. Integrating artificial intelligence techniques in cell mechanics. *ANNALS OF COMPUTER SCIENCE AND INFORMATION SYSTEMS*, 41:111–116, 2024.
- [19] Hafza Quayyum, Syed Rizvi, Muddasar Naeem, Umamah Khalid, Musarat Abbas, and Antonio Coronato. Enhancing diagnostic accuracy for skin cancer and covid-19 detection: A comparative study using a stacked ensemble method. *Technologies*, 12:142, 08 2024. <https://doi.org/10.3390/technologies12090142>.
- [20] Swathy Krishna Reghukumar and Iwona Inkielewiecz-Stepniak. Tumour cell-induced platelet aggregation in breast cancer: Scope of metal nanoparticles. *Biochimica et Biophysica Acta (BBA)-Reviews on Cancer*, page 189276, 2025. <https://doi.org/10.1016/j.bbcan.2025.189276>.
- [21] S Berlin Shaheema, Naresh Babu Muppalaneni, et al. An explainable liquid neural network combined with path aggregation residual network for an accurate brain tumor diagnosis. *Computers and Electrical Engineering*, 122:109999, 2025. <https://doi.org/10.1016/j.compeleceng.2024.109999>.
- [22] Christine Van Laer, Renaud Lavend’homme, Sarissa Baert, Koenraad De Wispelaere, Chantal Thys, Cyrielle Kint, Sam Noppen, Kathelijne Peerlinck, Chris Van Geet, Dominique Schols, et al. Functional assessment of genetic variants in thrombomodulin detected in patients with bleeding and thrombosis. *Blood*, 145(17):1929–1942, 2025. <https://doi.org/10.1182/blood.2024026454>.
- [23] Hilal Yilmaz, Israa F Abdulazez, Sevda Gursoy, Yagmur Kazancioglu, and Cem Bulent Ustundag. Cartilage tissue engineering in multilayer tissue regeneration. *Annals of Biomedical Engineering*, 53(2):284–317, 2025. <https://link.springer.com/article/10.1007/s10439-024-03626-6>.
- [24] Mengru Zhu, Hao Zhang, Qirong Zhou, Shihao Sheng, Qianmin Gao, Zhen Geng, Xiao Chen, Yuxiao Lai, Yingying Jing, Ke Xu, et al. Dynamic gelma/dna dual-network hydrogels promote woven bone organoid formation and enhance bone regeneration. *Advanced Materials*, page 2501254, 2025. <https://doi.org/10.1002/adma.202501254>.