Analyzing the effect of mechanical properties of cellular processes on cellular packing patterns using unsupervised machine learning

Yasamin Modabber
251390444
Department of Physics and Astronomy
SciComp9502B GitHub Reposetory

1 Introduction

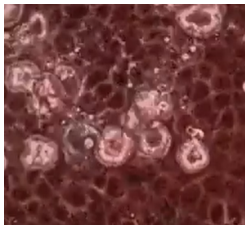
Understanding the spatial organization of cells within biological tissues is fundamental to unraveling the underlying principles governing tissue development and disease progression. In recent years, advances in computational modeling and simulation techniques have offered valuable insights into the complex interplay of cellular processes shaping tissues. This study follows the methodology proposed by [**Davies et al.(2023)**], which employs a radial angular distribution function to investigate the lipid conformation in the ripple phase. Here we apply that to epithelial tissues, Specifically, our focus lies in clarifying the mechanisms governing cell packing arrangements. In this report, we explore a range of parameters and conduct data analysis, aiming to uncover how mechanical factors influence the spatial organization of cells in ephitelial tissues.

2 Methods and methodology

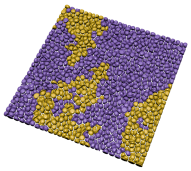
In this study, the approach of [**Davies et al.(2023)**] was used to utilize a radial angular distribution function to study the topology of the local structure of a system. Here we apply that to a biological system and study the and cell packing conformation in epithelial tissues. Mixed with an unsupervised machine-learning model, common emergent structures were observed which will be discussed further in the text. To get the raw data, the CellSim3D model and Software was used to study epithelial tissue growth in the presence of cancerous cells. CellSim3D allows the simulation of the mechanical aspects of growth, division, migration, and the interaction of cells with each other and their environment.Using the many parameters cellsim3D offers, Different systems were simulated in which Soft cell size, Soft cell stiffness, number of soft cells, Intermembrane friction, and medium friction were varied and the packed topology of the confluent state of each system was used in our analyses. The choice of parameter for the system was taken from the paper by [Madhikar et al.].

In order to find a system with a good sample population and a reasonable simulation time, systems with multiple sizes were simulated and observed, and a 2d Box size of 40 x 40 was chosen. This choice ensures the system reaches the confluent state in a reasonable time, and it provides enough space in the system for other patterns to emerge by reducing the dominance of the ordered arrangement of cells at the bounds. Through varying the different parameters, an ensemble of 200 different systems ($N_S = 200$) with different parameters was simulated. Data was extracted from the output files using custom-written scripts, and scikit-learn and scikit-image were used for the data analysis and model training.

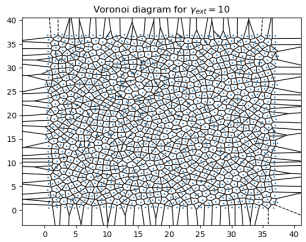
The metric choice for quantifying the local structure is the tree particle distribution function $g_3 = g_3(r, \theta)$, which here is proportional to the probability of finding a cell C at distance r from Cell B and its nearest neighbor A with vectors r_{AB} and r_{BC} forming angle θ. Here a cutoff distance of 2σ was chosen to prioritize neighborhood topologies, with σ being the unit length of the simulation. The bin sizes were chosen the same as the ones in the paper [**Davies et al.(2023)**].



(a) Snapshot of a real tissue



(b) A 3D render of a simulated cellular system simulated system



(c) The Voronoi diagram

Figure 1: CellSim3D allows us to effectively simulate cellular systems, in fig (a) you can see a snap-shot of a real tissue, and in fig (b) you can see a simulated system with two types of cells simulated using cellsim3D and rendered using blender. Fig (c) shows the Voronoi tessellation of fig (b)

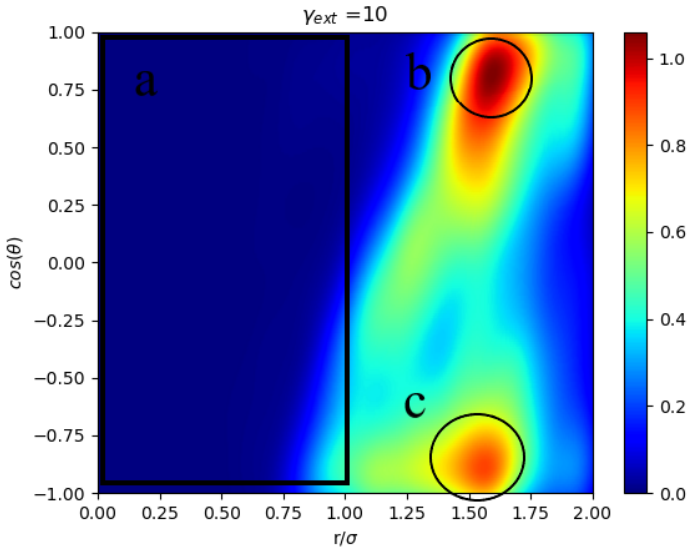
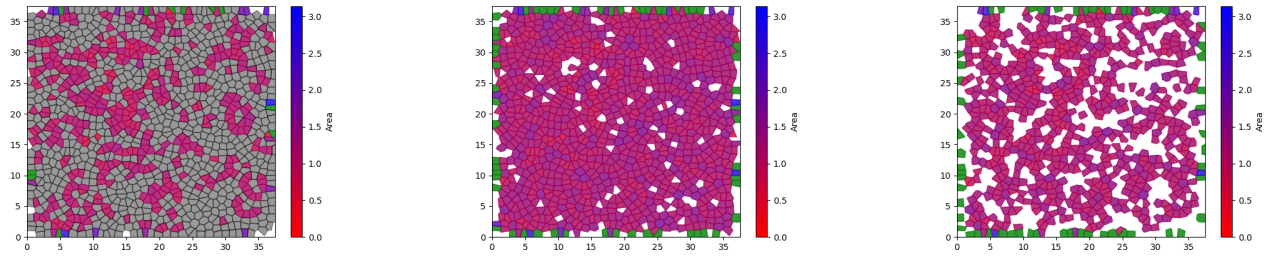


Figure 2: The g_3 density function for a system with $\gamma_m = 10$ and $\gamma_{int} = 10$, you can see there are two regions with High density positioned at $Cos\theta \in$ [0.75,1] and [-1,-0.75]. Each of the regions contributing to this density are shown in figure 3.

mean structural similarity index (SSIM)

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c1)(2\sigma_{xy} + c2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (1)$$

was chosen as a metric for comparing each two distribution functions. where x nd y refer to the window in each distribution matrix, μ is the window mean, σ^2 is the variance of the window, σ_{xy} is the covariance of the two windows,and c1 and c2 are small correction factors to avoid division by zero. The final SSIM is the average over the local windows reducing the two distributions to a scalar value in the range 0–1, with 1 being the result of comparing two identical distributions with the value decreasing the less similar the two distributions are. Here we use the Similarity-index function from scikit-image library that creates a similarity score between -1 to 1. values of 1 represent full similarity, 0 represents no similarrity and -1 is perfect dissimilarity. Using the same approach as [**Davies et al.(2023)**] window size was chosen as 7, and the data range was set by using the maximum and minimum values in all distributions.



(a) Packed cells belonging to the region with radius [0,1].

(b) Cells contributing to the population in section b

(c) Cells contributing to the population in section c

Figure 3: The tree Voronoi diagrams above show the cells contributing to each region shown in Fig 2. Section a was plotted to visualize the cells in close-knit packs.

In the next step, we create the Similarity index matrix by comparing each two distribution functions and calculating the SSIM index. once calculated, it was observed that the values range from 0.83 to 1. To increase the distance and emphasize the difference between these similarities, we scale the range [0.8-1] to [0-1]. In figure 8, indexes [0-100] represent variation in s_f, or soft cell fraction, which is the fraction of initial soft cells to the total initial cells in the system, with it ranging between 0 (no soft cells) to 1 (a system with only soft cells). We observe a noticeable shift across this range, and after observing the respective g_3 distribution functions in figure 5, we notice an increase in the population of observables with smaller r's, indicating more cells falling into the cut off radius meaning the system is packed. Indexes [100 - 200] represent a change in intermembrane viscosity, with γ_{ext} ranging between 0 to 10. indexes [200-300] represent a range in medium friction, with the value of γ_m ranging from [0-10].

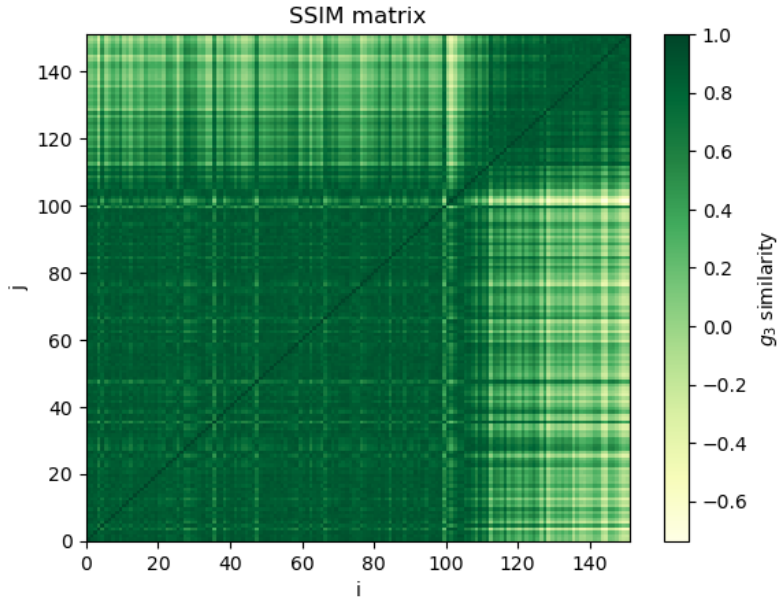
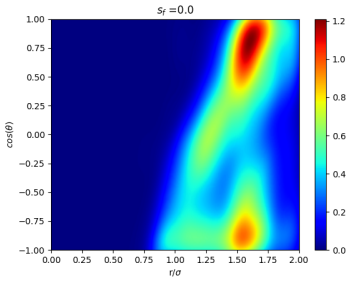
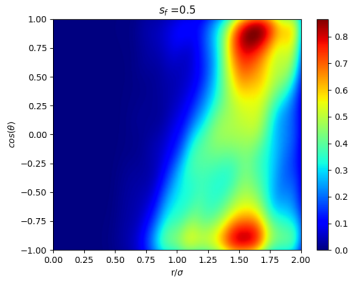


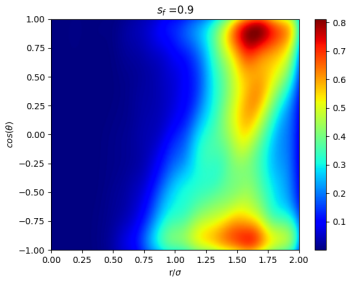
Figure 4: The SSIM matrix created for a sample of data. The first 100 frames represent a variation in medium friction, while the next 50 represent a change in s_f, soft fraction, the ratio of soft cells in the system to the rigid ones.



(a) g_3 density distribution of a system with only rigid cells



(b) g_3 density distribution of a system in which half of the initial cells are soft



(c) g_3 density distribution of a fully soft system

Figure 5: Difference in density function across different parameters for s_f, in fig 4 it seems as if there exists a critical s_f for which the soft cells fill the systems. Fig (a) shows the g_3 density function of a fully rigid system. As we increase s_f, the system starts to shift to a state (c) for which we have a higher density of neighboring points positioned at angles $cos\theta \in [-0.75 - 0.75]$

3 Results

After constructing the similarity matrix, we perform clustering using the DBScan method to find common patterns in the data. This unsupervised Machine learning method is chosen because it can distinguish the different patterns emerging in the systems and group the systems with similar behaviors. Analysis of the results hints the change in inter-membrane friction and medium friction do not bring about a notable change in local structure, while the most significant contributor to the distribution is the number of soft cells in the system.

In the next part of the analysis, we change the γ_{int} and γ_m from the previous value of 1 to 5, and simulate systems with s_f ranging from 0 to 1. After performing clustering on data, it was seen that the critical s_f for which the packing of the system changes shifts to a smaller value, hinting that soft cells were more easily able to fill the system.

Lastly, we perform simulations with a box size of 20 and compare the results with that of 40. It can be seen from the lack of correlation between the subsequent elements that the choice of a proper system size providing a good data population and reasonable simulation time is really important. Smaller box sizes tended to result in more localized cell packing arrangements, whereas larger boxes allowed for the emergence of global tissue structures.

Overall, our goal here was to provide insights into the effect of mechanical elements on cellular processes governing tissue organization. A more throughout investigation might contribute to our understanding of tissue morphogenesis and cancer progression, clarifying the underlying mechanisms driving tissue dynamics.

4 Discussion and conclusions

In summary, our study represents an initial exploration into the dynamics of cell packing within epithelial tissues using computational simulations. While our findings did not yield definitive results, our efforts in systematically varying parameters and conducting thorough data analysis is a valuable first step laying the foundation for future work in this area. An area worth further investigation is the effect of mechanical properties on the critical value for s_f in which we can see a transition to the state in which the soft cells take over the tissues.

Overall, to address the limitations identified in our study, future research should focus on increasing the number of simulations and extending the simulation time steps. We should also perform the simulations across a wider range of parameters. By doing so, we can overcome the inherent randomness within the system and obtain a more comprehensive understanding of tissue dynamics.

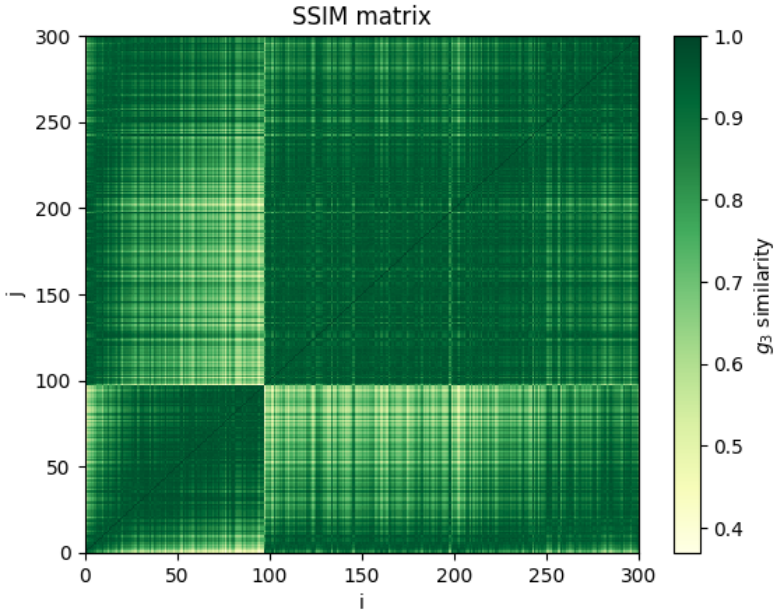


Figure 6: The structural similarity index matrix of all the data. indexes [0-100] represent variation in s_f, or soft cell fraction. Indexes [100 - 200] represent a change in intermembrane viscosity, with γ_{ext} ranging from [0 - 10]. Indexes [200-300] represent a range in medium friction, with the value of γ_m ranging from [0-10].

References

[1] Matthew Davies, A.D. Reyes-Figueroa, Andrey A. Gurtovenko, Daniel Frankel, Mikko Karttunen, Elucidating lipid conformations in the ripple phase: Machine learning reveals four lipid populations, Biophysical Journal, Volume 122, Issue 2,2023

[2] P. Madhikar, J. Åström, J. Westerholm, and M. Karttunen, CellSim3D: GPU accelerated software for simulations of cellular growth and division in three dimensions,

[3] https://github.com/SoftSimu/CellSim3D

[4] Madhikar, P., Åström, J., Westerholm, J., Baumeier, B., Karttunen, M. (2020). Coarse-grained modeling of cell division in 3D: influence of density, medium viscosity, and inter-membrane friction on cell growth and nearest neighbor distribution. Soft Materials, 18(2–3), 150–162.

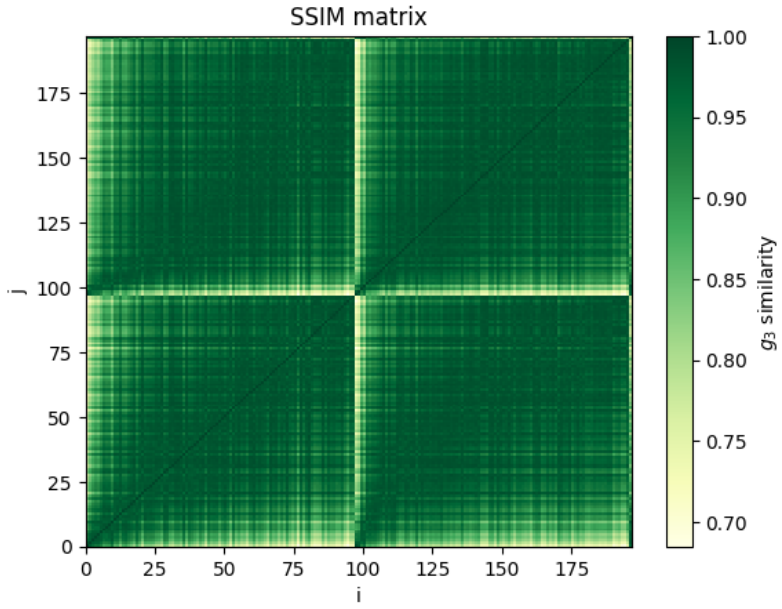


Figure 7: The structural similarity matrix with data showing a variation in s_f. The first 100 data are for a system with γ_{int} and γ_m of 1, and the next 100 data have γ_{int} and γ_m set to 5

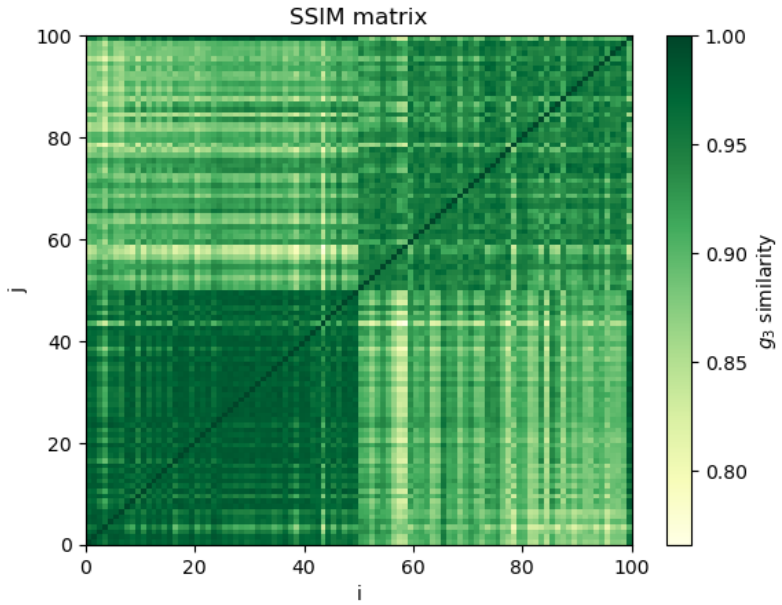


Figure 8: The structural similarity matrix for systems with similar mechanical parameters and different system sizes.