Live-Cell Imaging Intelligence and Organelle Dynamics Quantification: Real-Time AI Platforms for Monitoring and Analysis of Cellular Processes

Dr. Olga Volkova, Professor of Artificial Intelligence, National Research University – Information Technologies, Mechanics and Optics (ITMO)

1. Introduction to Real-Time AI Solutions in Cellular Biology

Developing real-time artificial intelligence (AI) solutions to investigate cellular biology in living conditions is gaining interest. As cellular biology evolves, and as we get to know more about mutually supporting networks of regulatory functions, including genetic and epigenetic networks, interest in tracing cellular operations increases. In this context, the real-time acquisition of single-cell data can show the biological systems in their whole complexity. In normal and in diseased conditions, events spanning over the whole range of temporal scales are going on at the same time—starting from slow and long-lasting to extremely fast and transient ones. The real-time monitoring of intracellular processes is claimed to be of high demand due to the pace of cellular processes and the continuum of disease. The real-time natural history of intracellular mechanisms was, until recently, unattainable. However, as recording the cells' internal communications remains important to biological research practice and clinical applications, cellular biology without real-time denotes incomplete acquisition conditions. Artificial intelligence (AI), the capability of computer systems to perform tasks that normally require human intelligence, is now regularly used to analyze data and make decisions. The current biological developments and the increasing data flood generate promising opportunities to exploit AI approaches and methods further. The advent of AI promises a complete review of fundamental science and an acceleration of cellular biology research. AI tackles the most exciting cellular endeavors in terms of scale and performance, including real-time intracellular tracking and long-term clustering, further exploring the potential behind the incredibly expanding biological process.

1.1. Overview of Cellular Processes and the Need for Real-Time Monitoring

A cell performs many functions at any given moment, including but not limited to dividing, metabolizing, and transmitting signals. These cellular processes are normally coordinated to ensure that the cell is functioning properly. The set of cell activities can be described as comprising a dynamic system of networks and pathways. This complex behavior can change under external stimuli, such as a drug. Cell division, in particular, is fundamental for life and understanding cancer. Pathological conditions, like cancer, neurodegenerative diseases, or disorders of the immune system, can occur through natural or acquired mutations in the macromolecular complexes that govern these pathways. More aggressive forms of cancer, for instance, result from malignant mutations of signaling pathways that suppress cell proliferation.

Standard techniques to observe these cellular processes provide at best a series of snapshots in time at low throughput. Stains for proteins show relative abundances, and antibody-based tests demonstrate the activity of a subset of proteins. RNA and DNA techniques generally require cell lysis and serve as static reporters of processes. In contrast, single-cell time-lapse microscopy captures many processes inside living cells in real time or near real time. It indicates that we can obtain dynamic insights into cell processes for exploration or intervention. In response, AI systems can be proposed for the end-to-end real-time analysis of cell behavior in combination with real-time microscopes or image processing techniques.

An additional reason for real-time monitoring is that cells are often transient in events of interest. Long-term assays usually require interventions to occur before starting to monitor. Technological advances have also allowed single-cell time-lapse microscopy to be performed with increasing throughput. A popular application is to quantify cellular responses to drug treatment. In addition, immunotherapy for cancer typically requires cells to be entered and manipulated before a patient's own cells can be re-administered to the patient. Measurements from these processes should be used in real time.

2. Fundamentals of Machine Learning in Cellular Biology

Machine learning and, in particular, deep learning are contemporary branches of artificial intelligence springing from the analysis of large datasets with a variety of potential applications. In the context of cellular biology, machine learning approaches are exploited to identify patterns in biological data, classify unhealthy from healthy samples, and predict cellular dynamic behaviors, among others. Principally, cells are complex entities that process signals and interact with their environment. All these functions unfold at different time scales through numerous signaling and intracellular pathways. Analyzing cellular dynamical behavior usually leads to huge amounts of complex high-dimensional datasets. Machine learning methods can help researchers interpret such complex cellular behaviors.

Machine learning is categorized by model-building paradigms such as supervised and unsupervised learning, among others. During supervised learning, models are trained using pairs of inputs and their corresponding outputs. These inputs are usually termed features, whereas the outputs are often referred to as the label or target variable. The training dataset provides the initial data so that the model can identify a prediction of the process. Once a model has been trained, it is then evaluated using a test dataset to explore its robustness and efficiency. One popular unsupervised technique is clustering, which aims to identify associations among the inputs in the absence of labeled data. Moreover, regression models are a family of supervised learning techniques that can model a response variable based on another variable or a combination of variables. To systematically analyze large datasets, researchers are seeking lightweight and non-parametric methods that can uncover hidden patterns in cellular dynamics or are capable of early detection of metabolic diseases in their early stages. In addition to using the appropriate classifier algorithm, building training datasets, selecting features, and preprocessing the data to remove any noisy or non-significant features can all affect the accuracy of the model. The high dimensionality, noise, missing values, and class imbalance, which are challenges associated with biological data, pose an extra source of variability and noise at the core of the dataset. The robustness of the machine learning techniques used, therefore, has to be taken into account.

2.1. Basic Concepts and Terminology of Machine Learning

To avoid errors in the following text, we provide a brief introduction to basic concepts and terminology of machine learning at this point. In general, machine learning uses algorithms that can automatically detect and reconstruct models from vast datasets, along with providing precise instructions for further analyses or new, unseen input data. The algorithm parameter allows for adaptation to increasing data, providing accurate results. Usually, learning is associated with biological terms, thus training to develop various types of models and serving as testing to ensure that the trained model is reliable and provides identical performance with respect to known labels. In this context, there are four learning strategies. • Supervised learning takes into account the training set of previously observed input data and output pairs, where input data denotes the input data and output label denotes the output label. • In contrast to supervised learning, unsupervised learning models are trained with input data, but without corresponding output labels. • Here, a given set of trials learns to discover the best possible decision at each state. • Finally, in evolutionary strategies, algorithms use combination, mutation, homeostasis, and genetic recombination to remodel selected parent models for a new generation of models by exploring new and efficient models. In biology, both supervised and unsupervised learning are intensively used. As an example, a sophisticated detection model recognizes cancer and normal samples of subcellular components and phenotypes from vast datasets that are used for developing the model. The optimal trained model is also used for testing accuracy, precision, recall, and F1 score. The performance metrics optimization depends on the respective task, such as predicting outcomes or classifying features. In prediction tasks, various label values are associated with an outcome, for which we try to learn the pattern and provide the most effective and accurate result. In contrast, classification models are used to decide and confirm whether the input falls into a certain category based on a restricted label.

3. Integration of Real-Time AI Solutions in Cellular Research

A variety of high-throughput laboratory methods have been developed to profile cellular behavior at an increasing pace and resolution, with corresponding leaps in the amount of data they generate. Integrating AI with these molecular methods opens a new world of more automated data analysis and hypothesis generation. Unlike conventional software in the laboratory, AI-driven tools can be taught to perform specific data analysis tasks without the need for explicit coding, and when made to work in close concert with biologists, the scientists know that data and biological context are compatible. Implementing AI can, for example, speed up drug discovery and development by quickly sifting through potential molecule candidates, identifying relevant cellular mechanisms in which previously undiscovered small molecules could interfere, and highlighting on- and off-targets of newly developed drug candidates. As these cell biological methods can now be performed robotically at high resolution and speed, AI becomes useful in integrating all the data and automatically making the best decision from it, as the biologist using the robotic microscope and sensor technologies does. Basically, this is just a robot making the data analysis part of its decision-making process more human-like. Having the AI available in real-time will make it easier to make decisions and adjust the experiment as needed in order to get the best possible results in experiments where the best biological insights are a direct result of continuous adjustments. By way of example, we cover case studies of how AI that facilitates the interpretation of cell tracings in parallel microfluidic systems can address some notable issues in cellular profiling and is successfully being used as a service in our laboratory. Challenges with moving data and decisions between a 'traditional' AI expert and a biologist in the context of these studies are discussed. Our findings rather point to the integration of AI experts and biologists in the same multidisciplinary research group or team, where understanding biological principles is good for training the subsequent models to perform on-the-fly data processing. To offset the fact that AI developed by non-biologists may not have the best biology package at the get-go and will need continuous checks and iterative training, setting the experts side by side is essential in leveraging and designing technological features that biophysicists in an AI tool will later come to rely on if they are to apply it to a biological question. Impressive microscopy imaging and cellular profiling methods are now flooding the biological research environment, producing high-dimensional data sets that can be used to exhaustively nurse hypotheses about cellular behavior, condition, and response.

3.1. Applications in Drug Discovery and Development

The ability to assess whether compounds are activating or inhibiting specific pathways and producing a phenotypic response allows the possibility of predicting off-target effects, one of the biggest safety challenges. These types of systems have been developed using algorithms to employ, where data on effector pathway modulation and resulting cellular responses is collected and stored, and new responses are compared to data to give a pathway hit. Machine learning and other algorithms are used to predict other pathways altered by a compound with the aim of identifying off-target effects earlier in the drug discovery process. Finally, using a cell-based system, real-time analysis can highlight key time points at which an active compound works, or time points when no effect is seen at any part of the pathway, which can also be helpful for prioritizing drug candidates. Real-time solutions that monitor cellular responses to compounds could, in the future, enable a rapid go/no-go decision on whether a series is reacting with a specific target or activating an effector pathway to give a phenotypic response, i.e., whether a hit is likely. This data could also be used in a variety of ways. For example, it can be used for individualizing predictions of drug-drug interactions, for identifying the syndrome of pathways activated by a compound, and evidence of synergy, or how a combination therapy is working.

Whether a hit can be associated with a particular safety hazard at an early part of the drug discovery/development process can be of benefit for lead optimization. Certainly, the ability to monitor responses to the hit through to use in the clinic offers potential for stratified or personalized medicine, enabling early assessment of patient response to drugs using cells such as kidney cells. In drug lead identification and optimization, toxicity and efficacy are not independent, particularly in cancer and some other indications. The concept of real-time hit identification and real-time data is gaining traction as these are key in determining the efficacy and toxicity of compounds at different stages of the drug development process. In particular, using a cellular phenotypic context is important for biological safety assessment, as it prevents discarding a potentially active drug due to suboptimal safety properties. What is considered an acceptable safety profile changes depending on the potential efficacy of a drug. The ability to track efficacy and off-target activation/toxicity non-invasively in real-time allows us to use data generated on the same subject to adjust the dose of the activating agent to the efficacy and the safety window. The potential to quantify and track efficacy in a dose-dependent way has a profound impact on how it is applied in the clinic. If we can monitor the progression of patients in real-time, they can potentially be maneuvered into more aggressive treatment pathways as their tolerance becomes known. Algorithms can also screen single-agent drugs for activity in new indications, based on the activation of pathways. Guided drug screening doubles activity hitting rates and significantly reduces the time to identify approved drugs with efficacy in DNA damage repair-deficient cancers by nearly half.

4. Challenges and Limitations of Real-Time AI Solutions in Cellular Biology

The prediction of cellular processes and their behavior at a system-wide level presents numerous research opportunities. Nevertheless, the translation of AI-based methodologies and algorithms to research environments is often hampered by a number of drawbacks and limitations.

Relevant issues in cellular biology comprise data quality, model bias, incomplete knowledge about cellular processes, and unicellular patients, among other factors. The acquisition of high-quality and well-annotated data is of fundamental importance since datasets whose data quality is compromised may limit the ability to achieve accurate predictions due to noisy data. Furthermore, real-time implementation of AI approaches capable of providing instant, partial, or biased results remains a challenge. Real-time outcomes provided by incomplete models may also mislead researchers by ignoring the overall biological system behavior. Such influences can occur in cases where an AI-based approach learns only from commonly accessible data patterns that may not fully represent the entire cell population under study.

Additionally, issues associated with the translation of AI applications into biomedicine will also be challenging to tackle. The ethical issues identified in AI-driven results for cellular biology mainly revolve around issues related to consent, data privacy, and obtaining and handling patient or donor data. Moreover, the potential unconscious model biases may influence AI-based predictions due to some subgroups' biased physiological behaviors. Additionally, the capability of translating AI approaches to the clinic is critical but not easy since produced individuals are currently being pseudo-mapped to genetic databases and have only a mild relevance for genetic surveillance or drug development. Therefore, this process will be inaccessible to widespread use for many years to come. For such valid concerns, reliable out-of-distribution detection will have to be implemented in AI paradigms to maximize their use for clinical cellular research. Evaluating the model's outputs using a thorough theoretical approach will also assist in interpreting the AI model's results. Nevertheless, the text emphasizes the need for active interaction between computer and biological scientists and provides a suggestion for delivering usable tools by combining microbiological and AI methodologies.

4.1. Ethical Considerations and Data Privacy Issues

The inclusion of real-time AI solutions for cellular biology raises general and discipline-specific questions concerning the ethics of using these solutions. Building a deep-learning-enabled live-cell platform unavoidably involves training algorithms, to which established guidelines for responsible AI should be applied. A trending anti-expert attitude makes it clear that relying purely on the intuition and empathy of the researchers, machine learning, or any other output of biological information on an individual is not enough to protect personal privacy or consent. Ethical usage and sharing of the data are necessary to protect from unintentional or malicious release of unanonymizable recordings. We believe that such a call for the adoption of AI ethics within our selection has a clear benefit because transparent data usage policies are a cornerstone for the establishment of trust, which is fundamental for obtaining participant understanding and engagement and ensuring high-quality or well-controlled experiments.

There is a risk that an AI solution can be inherently biased due to skewed data collection or designed-in algorithmic bias. In turn, such biases may have contributed to public harm because people perceived the AI as not capable of spreading misinformation. This process may, from a policy perspective, be too slow. A slow empirical error correction process can restrict the empirical information that individuals receive. Thus, the perception of errors in the emergent properties of AI can be constrained if data collection and data analysis preclude a true representation of the diversity of human beings. There are potential deep methodological and policy implications here, which raise concern over the kind of future that a given emergent property may anticipate. An ethics of caution stresses the importance of the incorporation of ethicists into technologies whose use across abolition, regulation, and intensification pathways may impact lives and livelihoods to prevent the innovation from being connected to a greater risk. It needs to be ethically neutral—AI innovation can and does involve humans in life-or-death decisions, can involve the manipulation of information to deceive, and can infringe personal freedom, expression, and privacy. All states are bound by data protection legal standards, and their derogations for research. System operators must, therefore, adhere to these standards when implementing an AI solution. There are technical challenges, but these are not insurmountable and must be considered early in the design because the integration of an AI solution late in the process incurs costs in scope and implementation feasibility. Generally, ethical research in cellular AI is the primary guiding principle. It is also the case that different countries may have different ethical requirements for using AI in real-time studies. For example, researchers working within the European Union are required to undertake a formal Data Protection Impact

Assessment, and researchers in the United Kingdom are required to assess compliance with the NHS Health Research Authority recommendations.

5. Future Directions and Emerging Trends in Real-Time AI Solutions for Cellular Biology

Real-Time AI Solutions for the Advancement in Real-Time Cellular Biology Future Directions and Emerging Trends: An important future direction for the development of the field of real-time AI for cellular biology is the enhancement of data analysis methodologies and predictive modeling capabilities. Improved AI algorithms can leverage the integration of multi-omics data and the dynamic cellular state to derive models with better biological relevance and predictability. Moreover, AI can uncover hidden data patterns that may be too complex and vast for traditional biological analysis techniques. On the technical aspects, we foresee improvements in both hardware and software that will aid the real-time processing of full multi-omic, real-time cellular analyses. One novel application for the use of AI hardware improvement is subject to real-time cellular behavior occurrences; we envision AI programs that are capable of dynamic learning and real-time responses to complex cellular stimuli. On the software side, an important focus is on the development of real-time AI solutions that are available as open source, allowing the scientific community to use, update, and deploy software for their specific research purposes. The second emerging trend concerning real-time AI applications in cellular biology is the importance of integrating with other components of drug discovery, in particular with artificial intelligence for drug repurposing and optimization. AI approaches have been proposed for the prediction of the efficiency of the drug and the rheostability of a gene in cells. Interdisciplinary research and training programs will also likely emerge to be implemented from an educational perspective. It is important to train students and professionals with a background knowledge of biology, data science, AI, and the ethical and legal implications of using AI for biological studies.

5.1. Advancements in Multi-Omics Data Integration

Improved advancements in multi-omics data integration are a prominent recent trend. Several R&D and implicit approaches result in the combined analysis of genomics, transcriptomics, proteomics, and metabolomics data, indeed reconstructing cell functioning. The cell is an intricate system that closely interacts with all living organisms' organs and systems. Disease pathologies are either congenital or acquired, and they often lead to disruptions in cellular operations in internal environments of space and time. In the past, some connections among the various cellular levels were already discovered. Still, even more unattainable vast amounts of data are now available for analysis, leading to redefining methodologies. A holistic view of cellular functions plays a pivotal role in considering multi-omics data holistically.

In the context of the integrative simplification trend, real-time AI results are the leading development focusing on predictive cellular function analysis from multi-omics data. Machine learning and deep learning algorithms show great ability to handle several functional data in a multi-omics context due to these advantages. In an effort to implement efficient strategies to collect, prepare, integrate, and learn from big data, AI algorithms have been proposed, based on either specifically designed multi-omics models or already available and customizable machine learning architectures. Not only are efficient AI algorithms needed, but also solid and frequently used visualization tools that can handle the complexity characterizing such a large amount of results will be necessary. Indeed, there is a complete need for collaborations among scientists, doctors, pathologists, biologists, and artificial intelligence researchers in order to profit collectively from such a modality of data analysis.

6. Conclusion

In light of the developments presented in the preceding sections, we conclude with various considerations on real-time AI solutions for cellular biology. Multidisciplinary research has highlighted that better biological outcomes may be achieved by monitoring biological systems in real time. This is especially pertinent for cellular processes and can lead to immediate feedback on success, for example, in microfluidic assays. Furthermore, cellular assays make insights gathered more relatable to in vivo human responses, such as those related to drug toxicity.

We have also considered that training the right CNN may provide important contextual information to determine cell responses that are not yet understood. We have also shown that control-like learning algorithms do exist for cell monitoring, and further testing may verify the importance of such control-based systems in cellular biology. The use of AI on imaging platforms is limited by available data and the cost of data generation. Researchers need to be cautious in using AI to identify outcomes; however, knowledge-based completion of these outcomes can provide important context for further cellular research. Although our capacity to integrate AI into our cellular research is in its infancy, the potential to rapidly test our hypotheses is very appealing for the future.

Automated image analysis of live cells is growing, facilitated also by the aggregation of multiple cellular modalities. For real-time monitoring, multidisciplinary research is also exploring the use of microfluidic organ-chips. Here we have described several new real-time imaging systems as part of immune assays to monitor immune cell responses. The next steps will also see a focus in industry to develop machine learning tools that will allow real-time monitoring and pre-warning during allograft rejection. Ethical development of AI is imperative, and data handling must strictly involve anonymization and conform with data protection initiatives. We are now seeing conferences also delve into so-called dry lab research, which helps to reiterate the importance of computational solutions in a wet lab setting. The natural synergy of these disciplines should help to drive forward the aim of invigorating biological research methods. Education and training in AI technology will thus be necessary for the future research generation to harness such powerful tools. This review has shown that research in AI in the cellular and molecular imaging field is growing and expanding into new frontiers. While AI and deep learning have made great leaps to provide identification of phenotypic differences, there are still opportunities to use these tools further to identify specific cell changes or discover new findings through algorithms that may work differently from the human brain. These AI tools are the future of biomedical discoveries, especially with the trend in moving more towards high-content imaging. With the right computational biologist using these tools and the right mix of talent, we may begin to unlock the future of AI in cellular research.