IJMS
Vol 51, No 5, May 2026

Review Article

Revolutionizing Stem Cell Sorting with Machine Learning: A Review of Trends, Tools, and Future Directions

Marziyeh Mousazadeh[1], PhD; Atieh Jahangiri-Manesh[1], PhD; Hossein Soltaninejad[2], PhD; Farzaneh Yazdi[3], PhD; Karim Rahimian[4], MSc; Kathleen M. Curran[5], PhD; Patricia Khashayar[6], PhD

[1]Department of Nanobiotechnology, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran;
[2]Department of Stem Cells Technology and Tissue Regeneration, Faculty of Interdisciplinary Science and Technologies, Tarbiat Modares University, Tehran, Iran;
[3]Endocrinology and Metabolism Research Center, Kerman University of Medical Sciences, Kerman, Iran;
[4]Bioinformatics and Computational Omics Lab (BioCOOL), Department of Biophysics, Faculty of Biological Sciences, Tarbiat Modares University, Tehran, Iran;
[5]UCD School of Medicine, Dublin, Ireland;
[6]International Institute for Biosensing, University of Minnesota, Minnesota, USA

Correspondence:
Hossein Soltaninejad, PhD;
Faculty of Interdisciplinary Science and Technology, Tarbiat Modares University, Jalal-Al Ahmad Hwy., Nasr Bridge, Postal code: 14115-336, Tehran, Iran
Tel: +98 21 82884041
Email: hossein.soltani@modares.ac.ir
Received: 10 June 2025
Revised: 30 July 2025
Accepted: 15 August 2025

What's Known

- Stem cells are critical tools in regenerative medicine, large-scale cell production, drug discovery, and cell-based therapies.
- Accurate stem cell sorting enables improved therapeutic outcomes, efficient production pipelines, and more reliable biological studies.
- Recent advances in machine learning techniques based on image and video processing, have revolutionized stem cell sorting by enabling rapid, automated, and highly accurate classification.

What's New

- In this review, various machine learning-driven strategies for stem cell sorting, with a particular focus on visual and non-visual data processing methodologies and their applications in different stem cell types have been comprehensively explored.
- Furthermore, an overview of the historical development of stem cell sorting technologies and machine learning applications was introduced, and emerging automated systems, software solutions, start-ups, and future directions for this type of stem cell sorters were discussed.

Abstract

Stem cells are critical tools in regenerative medicine, large-scale cell production, drug discovery, and cell-based therapies, making their precise identification and sorting essential for advancing both research and clinical applications. Accurate stem cell sorting enables improved therapeutic outcomes, efficient production pipelines, and more reliable biological studies. Traditional sorting methods, while effective, face challenges related to speed, scalability, cost, and human error. Recent advances in machine learning (ML) techniques based on image and video processing have revolutionized stem cell sorting by enabling rapid, automated, and highly accurate classification. In addition to visual data approaches, non-visual processing methods using ML have also emerged as powerful tools for stem cell analysis and separation. In this review, various ML-driven strategies for stem cell sorting, with a particular focus on visual and non-visual data processing methodologies and their applications in different stem cell types, have been comprehensively explored and categorized based on the input data types, ML techniques, stem cell types, study objectives, and performance metrics. Furthermore, an overview of the historical development of stem cell sorting technologies and ML applications was introduced, and emerging automated systems, software solutions, start-ups, and future directions for this type of stem cell sorters were discussed.

Please cite this article as: Mousazadeh M, Jahangiri-Manesh A, Soltaninejad H, Yazdi F, Rahimian K, Curran KM, Khashayar P. Revolutionizing Stem Cell Sorting with Machine Learning: A Review of Trends, Tools, and Future Directions. Iran J Med Sci. 2026;51(5):285-304. doi: 10.30476/ijms.2025.107395.4197.

Keywords ● Stem cells ● Cell biology ● Machine learning ● Artificial intelligence

Introduction

The ability to identify, purify, or enrich cell populations plays a central role in life sciences and medicine. This is because the overall success of the cell- or cell product-based therapies is highly dependent on the qualitative or quantitative characterization of the cells.[1] Cell enrichment is usually achieved by eliminating unwanted cells or isolating the target cells from a heterogeneous population.[2] Drug screening, the automation of cell sorting, and high-scale production are the main applications of stem cell image processing.[3]

Today, the use of artificial intelligence (AI) and one of its subsets, ML, has become popular in all aspects of human life. AI is the development of computer systems capable of thinking, deciding, and acting in a logical way and similar to the decision-making

process of human brains. A rapidly growing medical AI application has been sparked by the recent advancements in AI software and hardware, including the development of deep learning (DL) algorithms and graphics processing units (GPUs).[4] Clinical diagnostics based on AI computer vision and isolated cell sorting devices manufactured by Sony Inc. are two important examples of such AI-based machines.[5, 6] Since stem cells are important undifferentiated cells and their enrichment or classification during the process of differentiation, trans-differentiation, or reprogramming in a heterogeneous context is crucial, AI- and ML-based algorithms can help to accelerate this process in a precise manner. Stem cell therapy is central to regenerative medicine. Traditional reactive medicine is shifting toward the novel regenerative approaches that replace damaged cells with new ones that perform the same function more effectively.[7] This review discusses stem cells, AI, and their convergence from both scientific and historical perspectives. Moreover, stem cell sorting methods are categorized and elaborated by examples and applications. Ultimately, different software and start-ups in this area are introduced.

Stem Cell Types

Stem cells are characterized as undifferentiated cells capable of repairing and regenerating damaged tissues through their ability to differentiate into specific cell types. They are categorized based on their origin into five subcategories: Embryonic stem cells (ESCs), Adult stem cells (ASCs), Fetal stem cells (FSCs), Mesenchymal stem cells (MSCs), and induced pluripotent stem cells (iPSCs). ESCs are derived from the blastocyst inner mass. ASCs are tissue-specific stem cells and are present in most tissues. They are responsible for replacing dead cells. FSCs are obtained from the fetus or fetal tissues, such as umbilical cord blood. MSCs can be obtained from different tissues, especially bone marrow, adipose tissue, and skin. They mostly generate the connective tissue. iPSCs are stem cells created using Yamanaka factors to reprogram the differentiated cells into undifferentiated stem cells.[8] Cancer stem cells are a type of cancer cells inside the tumor that are characterized by normal stem cell potencies. They are usually resistant to chemotherapy drugs, which makes them an important target in cancer treatments.[9]

Stem Cell Potency

Stem cells are classified into five categories—totipotent, pluripotent, multipotent, oligopotent, and unipotent—according to their differentiation potential, which describes the range and diversity of cell types they are capable of producing. Totipotent stem cells can differentiate into all cell types of the body, including extra-embryonic tissues such as the placenta. Pluripotent stem cells can give rise to cells derived from all three germ layers: ectoderm, mesoderm, and endoderm. Multipotent stem cells are restricted to producing multiple cell types within a particular lineage or tissue. Oligopotent stem cells can generate only a few closely related cell types. Unipotent stem cells are limited to differentiating into a single cell type, while retaining self-renewal capacity. Understanding this hierarchy is fundamental to advancing stem cell research and therapeutic applications.[8]

Waddington's Landscape

Regarding the cells' potency and their energy levels, a diagram called Waddington's landscape has been defined. Based on this landscape, the most potent stem cells (totipotent) have the highest energy level. During embryogenesis, other stem cells with different potencies (pluripotent, multipotent, and oligopotent) and different energy levels (less than the totipotent cells) are generated. In the reprogramming process, in a redirection way, cells will go through fewer energy levels to the higher energy levels (low to high entropy) thanks to the Yamanaka factors' discovery. At the lowest part of this landscape, unipotent stem cells with the lowest energy level among other stem cells exist. Thus, the more potent stem cells are, the higher entropy or energy levels they have (figure 1-A).[10] This understanding is crucial for identifying key factors that influence stem cell fate, enabling researchers to better control their differentiation and behavior.[11]

The high diversity of stem cells and their differentiation outcomes creates a need for precise, fast, and high-throughput data monitoring and analysis. Manual calculations cannot meet this demand, but ML and DL algorithms help address it.[12]

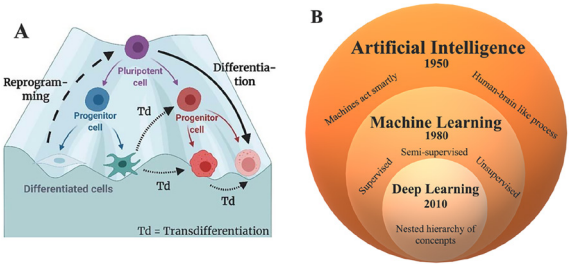

Figure 1: General overview over the stem cells and AI: A) Waddington's landscape shows that more differentiated cells have a lower energy status than stem cells (Reproduced with permission from Huanhuan Chen). B) AI-based systems are categorized into different parts, which were first discovered in the 1950s. Td: Trans differentiation

Artificial Intelligence

AI, as a fast-propagating definition, is defined as any system that acts smartly, similar to human brain intelligence. John McCarthy defined AI as "the science and engineering of making intelligent machines".[13] AI development is highly owed to the availability of powerful GPUs and infinite data storage spaces such as clouds.[14]

Machine Learning

ML, as a subset of AI, is the process of learning from raw data and generating clustered classifiers or regression output algorithms. Basically, ML learns from data to perform a specific task, instead of following some pre-programmed rules and codes. ML and DL both have some internal adjustable parameters to reduce the error of classifiers, which are called weights. Increasing and decreasing these weights through the output scores will result in more accurate data prediction and categorization. ML algorithms work based on three different learning procedures, which are supervised, unsupervised, and semi-supervised learning.[14]

Supervised ML

In a supervised ML, input data are tagged with the desired label. It means that the model training procedure is done by a certain training set of data.[14]

Unsupervised ML

In unsupervised ML, input data are not tagged with a suitable label, and ML algorithms try to classify different inputs into different output clusters based on their similarities. Unsupervised ML helps identify hidden patterns embedded in the input data, which are not easily recognized by human intuition.[14]

Semi-supervised ML

In semi-supervised ML, a small amount of labeled and a large amount of unlabeled input data are used to train the model and leverage the desired output.[15]

Among different ML methods, supervised learning has the highest predictive accuracy due to its training on labeled datasets, and it is well-suited for classification and regression tasks with clearly defined labels. On the other hand, this high accuracy requires large, high-quality labeled data, which is costly and time-consuming to generate. In supervised ML, there will be the risk of overfitting due to the model training on limited and non-generalized labeled data. In this case, performance will decrease based on the biased input data. In another view, unsupervised ML is highly suitable for discovering hidden structures, clusters, or relationships in large datasets, and experiencing broader exploratory analysis and reducing dataset bias effects compared to supervised learning. Output evaluation is basically harder in unsupervised than supervised learning, so choosing the suitable method in unsupervised ML is more crucial than others.[16] In semi-supervised learning, improving performance while reducing labeling efforts is observed compared to other ML methods. Their output is also more generalized rather than purely supervised learning. Similar to the two other methods, semi-supervised learning depends on both labeled and non-labeled data and usually requires more careful design of algorithms.[14]

Deep Learning

DL is a subset of ML. DL is actually a kind of ML algorithm that works with big data and breaks the input into different layers of features and then learn from them in a purpose-oriented manner. These layers of features are usually not designed by human code. This hierarchy-layered learning procedure is the key point in DL algorithms. Artificial Neural Network (ANN) is one of the DL algorithms that is inspired by the human brain and how its neurons work in a multilayered and interconnected network. While DL is suitable for a set of large data, ML is more professional in solving problems with a smaller amount of input data. Hence, ML needs less time to be trained than the DL algorithms.[14] Figure 1-B summarizes different categories of AI systems.

Furthermore, AI and ML-based algorithms can be used for different biological and cell-based approaches such as cell classification, cell culture qualification, cell function monitoring, disease modeling, drug screening, and genetic analysis,[17] predictive modelling, personalized medicine, tissue engineering, clinical trial design, patient monitoring,[18] synthetic biology,[19] genome sequencing, predicting protein structure, drug discovery,[20] advanced hypothesis generation in life sciences, crop improvement, advanced environmental sensors, personalized training planning,[21] electronic noses, biomarker prediction,[22] point-of-care biosensors,[23] gene editing, and precision medicine.[24]

Phase-contrast and fluorescent microscopes are the main image-based methods for AI-based cell sorting. Using AI algorithms can help reduce human intervention and thus related errors, providing more accurate, controlled, and quantitative results than human eyes.[25]

The current sorting procedures for cell detection are mainly based on antibody-antigen binding, where fluorescent markers are physically added to the cells which pass through the fluorescence-activated cell sorting (FACS) device to help identify specific cell populations.[26] Additionally, the microfluidics, micropipette aspiration, osmotic selection, laser capture dissection, optical traps, and antibiotic selection techniques are used as other separation methods.[27-34]

Meanwhile, biochemical approaches in stem cell sorting, such as molecular staining, have two opposing sides. These methods provide a clear outlook of the molecular characteristics and are capable of isolating the target cells with an acceptable accuracy. Since they are not cost- or time-efficient, their use is limited to markers with fluorescence tags,[35] and they may cause unwanted side effects, such as chemical toxicity.[36-38]

In such stem cell sorting techniques, cellular biomechanical or structural characteristics such as elastic or equilibrium modulus, apparent viscosity, and cell size help distinguish or differentiate between cell types or even adult stem cells.[39] Differentiating cells using such features, however, could be challenging due to the large number of parameters that must be measured. In this respect, AI and ML approaches, such as ANN, provide a potential means for classifying, filtering, and sorting large collections of properties by exploiting their ability to discern patterns in complex data.[40, 41]

The main advantage of utilizing neural networks in this regard is their ability to easily analyze large, high-dimensional datasets for distinct groupings of similar cases. There is no limitation on the number of inputs, eliminating the need to determine *a priori* or predefine a parameter list for the analysis. Moreover, the relative weight of each feature is determined in the neural network, providing an alternative approach to identifying the most influential features for a given population. Recent studies on unsupervised ML neural network models, such as Kohonen's self-organizing feature maps, have shed light on this process and how neighboring groups or nodes are related to each other.[41-43]

In the current paper, different ML-based methods for stem cell sorting are reviewed. Based on the input data, two main groups were considered: visual and non-visual datasets. The visual datasets contain image- or video-based data and flow cytometry results. The non-visual data, on the other hand, include microarray, expression patterns, and electrical and biomechanical data (figure 2-A). Table 1 also lists various ML-based algorithms, types of input datasets, stem cell types, performance, and dataset sizes. Additionally, the historical evolution of stem cell and AI/ML technologies has been discussed in this article. Different automated devices, software, start-ups, and the future probabilities of AI-based stem cell sorters have also been investigated. Figure 2-B represents a general view of the ML-based methods for stem cell sorting.

The Timelines and Convergence of Stem Cell and AI Discovery

The integration of AI and ML into stem cell research and sorting has evolved dramatically over the past few decades (figure 3).

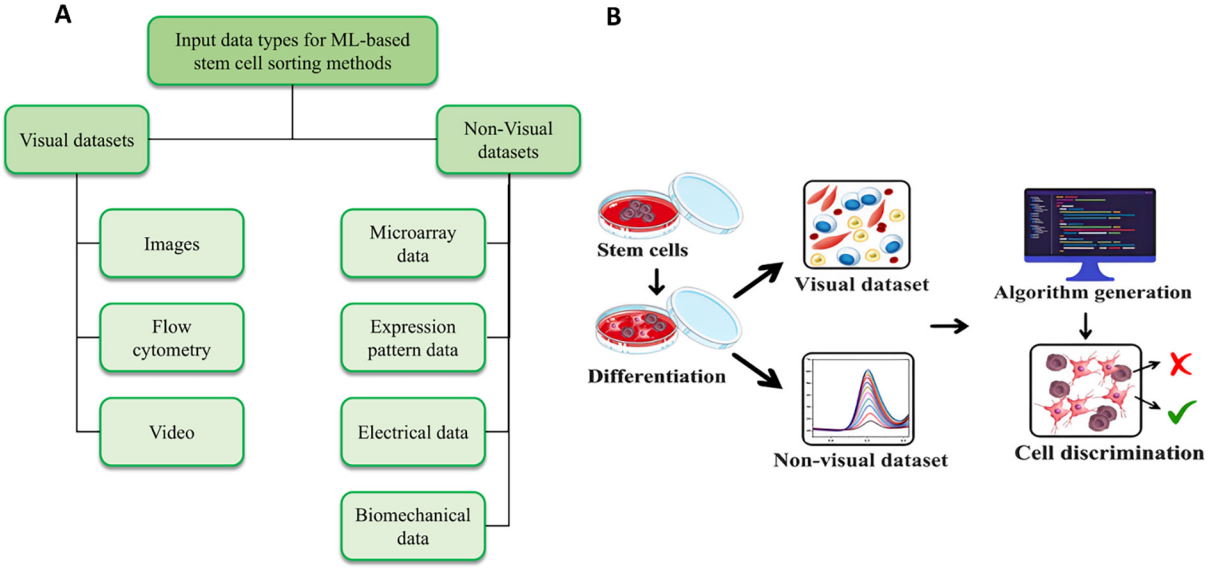

Figure 2: Categorization of different stem cell sorting methods: A) Different ML-based input datasets for stem cell sorting are divided into visual and non-visual datasets. B) The main concept of ML-based data processing for stem cell discrimination shows that visual and non-visual datasets are mostly used in this regard. ML: Machine learning

Table 1: ML approaches in stem cell sorting applications

Reference	Input data	Stem cell type	Methods	Dataset size	Purpose	Performance
Fan, 2017[66]	Time-lapse-based bright-field imaging	iPSC	DCNN, AlexNet, and HMM	2×10^6 standard patches	Determining the morphological changes during iPSC colony formation	Pearson coefficient=0.877
Mota, 2021[67]	Phase-contrast micrographing	MSC	Conventional U-Net, LSVM, RSVM, LDA, KNN, and LR	107 images, 658 cells	Classifying the low and mid-dense MSCs	Sensitivity=88% Precision=86% AUC=0.816/0.787 (early and mid-logarithmic expansion, respectively) Sorensen-Dice coefficient=0.849±0.106
Waisman, 2019[68]	Transmitted light microscopy	PSC	CNN, DenseNet, ResNet50	1000 images per group	Detecting PSCs from early differentiated cells	Accuracy>0.9
Zhu, 2021[70]	Bright field and dark field single cell imaging	NSC	CNN, multi-layer perceptron, ResNet, VGGNet, and InceptionV3	59287 datasets	Identifying the NSC differentiation to neurons	Accuracy=0.923 (BF)/0.998 (FluMM) AUC (average)=0.95
Joutsijoki, 2016[71]	Inverted microscope imaging	iPSC	LS-SVM and KNN classifiers	173 images	Controlling the quality of iPSCs	Accuracy=62.4%
Kusumoto, 2018[72]	Phase-contrast images	Endothelial cells derived from iPSCs	LeNet & AlexNet	800 images	Identification of differentiated cells by morphology	Accuracy>0.95
Orita, 2020[73]	Bright-field microscopic time-lapse video images	hiPSC-CMs	UMAP, SVM	556 videos	Discrimination of functionally normal and abnormal	Accuracy=0.89±0.02
Guo, 2021[74]	HCS image series (bright-field, DAPI, GFP, and RFP)	iPCS-derived ITS embryos	PhenoLOGIC™ software package	NA	Investigation of stem cell-based embryogenesis	Accuracy=0.92
Sardiu, 2021[75]	Flow cytometry dataset	hHSC	TDA, X-shift clustering, t-SNE	A publicly available data set (Nilsson rare data)	Identification of rare cell populations	NA
Guan, 2021[79]	Video obtained from phase contrast microscopy	ESC	RandNet and autoencoded feature extractor	3559 images	Assorting the ESC into six different groups	Accuracy=97.23% ±0.94
Shibata, 2020[81]	Lectin microarray	PSC	PCA, LC, and NN	1577 cells	Classification of PSC from four other cell lines	Accuracy=89%/97% (NN)
Stumpf, 2019[82]	Expression patterns	PSC	PCA, GMM	46 measured features from two separate stem cell lines	Identification of pluripotency patterns	NA
Cunha, 2019[83]	EIS	NPC	LSTM –ARNN	NA	Differentiation between proliferation and differentiation states	Accuracy=91% (4603c27)/76% (IMR90c01)/91% (IMR90-4)

Heylman, 2015[84]	Electrophysiological data	hiPSC-derived Cardiomyocytes	MATLAB TreeBagger function	457 waveforms	Assessment, classification, and prediction of depolarization of the hiPSC membrane	Accuracy>70% AUC=1 for classification, AUC=0.92 for control, 0.95 for propranolol, and 0.95 for isoproterenol
Darling, 2008[41]	Biomechanical properties obtained from AFM	ASCs and MSCs	NN	49 Superficial cells, 6 factors	Isolating ASCs and MSCs from primary mesenchymal-derived cells	Mean purity=60% Efficiency=75% Enrichment=150%
Juhola, 2021[85]	Calcium transients' data	human iPSC-derived cardiomyocytes	Random forests	1173 signals originating from seven classes	Classification of different genetic cardiac diseases	Accuracy=73.7%
Ugawa, 2021[86]	Images of diffractive and scattering imaging	hiPSC	SVM	NA	Development of a stain-free A flow cytometry method called silico-labeled ghost cytometry	AUC=0.998±0.002 and 0.877±0.007 For discrimination of live cells from dead and apoptotic cells, respectively

iPSC: Induced pluripotent stem cell; MSC: Mesenchymal stem cell; PSC: Pluripotent stem cell; NSC: Neural stem cell; ESC: Embryonic stem cell; ASC: Adipose-derived stem cells; DCNN: Deep convolution neural network; HMM: Hidden Markov model; LSVM: Linear kernel support vector machine; RSVM: Radial-basis kernel support vector machine; LDA: Linear discriminant analysis; KNN: K-nearest neighbor; LR: Logistic regression; LS-SVM: Least-squares support vector machine; CNN: Convolution neural network; TDA: Topological data analysis; UMAP: Uniform manifold approximation and projection; SVM: Support vector machine; PCA: Principal component analysis; GMM: Gaussian mixture model; LSTM-ARNN: Long short-term memory artificial recurrent neural networks; AUC: Area under the curve; NA: Not available; AFM: Atomic force microscopy; ARNN: Autoregressive neural network; NPC: Neural progenitor cell; EIS: Electrochemical impedance spectroscopy; ITS: iPSCs and trophectoderm stem cells

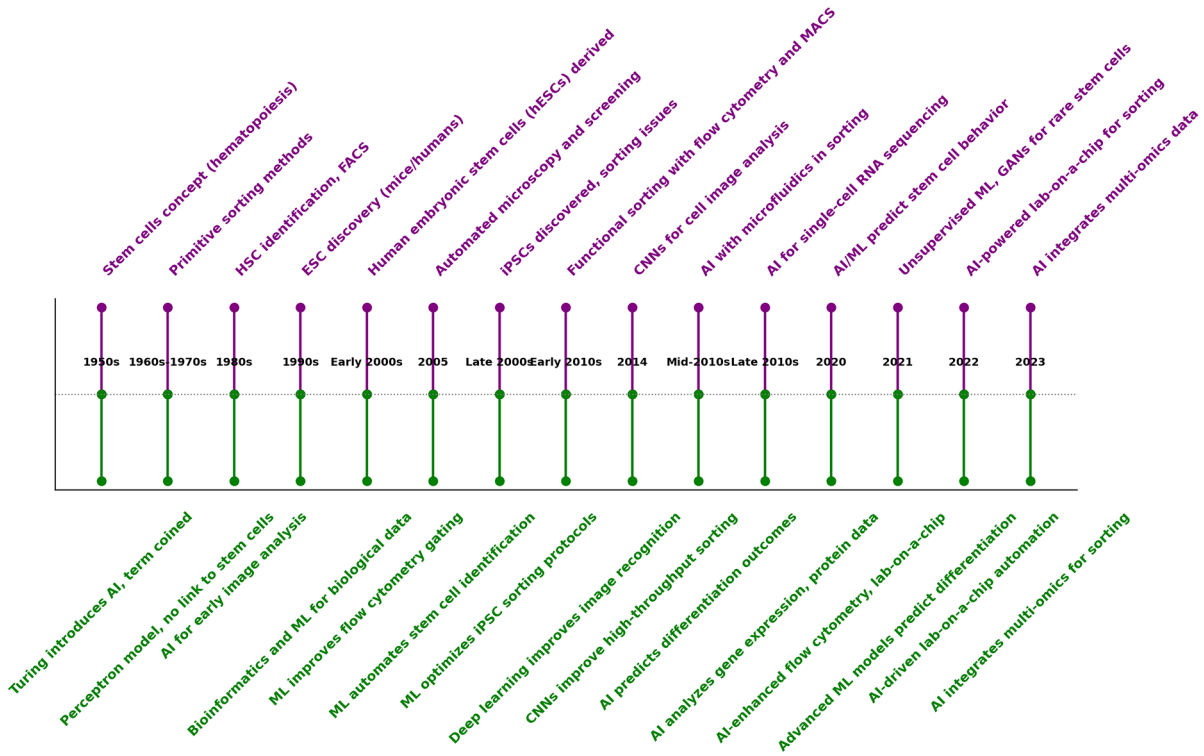

Figure 3: Stem cell and AI/ML evolution started in the 1950s and continued till now; FACS: Fluorescence-activated cell sorting; HSC: Hematopoietic stem cell; ESC: Embryonic stem cell; hESCs: Human embryonic stem cells; iPSCs: Induced pluripotent stem cells; MACS: Magnetic-activated cell sorting; CNN: Convolutional neural network; AI: Artificial intelligence; ML: Machine learning; GANs: Generative adversarial networks

Starting from basic computational tools, the field has progressed to sophisticated AI-driven systems that enable high-throughput, precise sorting of stem cells based on a wide range of characteristics. These advancements have not only improved our understanding of stem cells but also enhanced their application in regenerative medicine, making AI and ML indispensable tools in the field.[44]

1. Early Foundations of AI, ML, and Stem Cell Research (1950s-1980s)

1950s

Stem Cell Research: The foundational concept of stem cells emerges, focusing initially on understanding their role in hematopoiesis, or blood cell formation.[45]

AI Foundations: Alan Turing introduced the idea of AI, and John McCarthy coined the term "AI" in 1956. Early research explores AI's potential, primarily revolving around logic and reasoning.[46]

1960s-1970s

Stem Cell Sorting: Early stem cell sorting techniques rely on physical characteristics such as size and density, utilizing methods such as gradient centrifugation.[47]

ML Development: While rudimentary ML models, such as the Perceptron, are developed, there is little connection to stem cell sorting at this stage.[47]

1980s

Stem Cell Breakthroughs: The discovery of hematopoietic stem cells (HSCs) significantly advances the field. Techniques such as FACS enhance cell sorting precision.[48]

AI and Image Analysis: Emerging AI-driven tools begin to be applied in image analysis, laying the groundwork for future integration into cell sorting.[49]

2. Growth of Computational Biology and Advances in Stem Cell Sorting (1990s)

1990s

Stem Cell Research: The discovery of ESCs in both mice and humans catapults stem cell research into the spotlight, with improvements in isolation and cultivation techniques.[50]

Cell Sorting: Flow cytometry evolves, enabling more precise sorting based on surface markers.[51]

AI and ML Integration: As bioinformatics gains traction, ML starts being explored for biological data analysis, though its application to stem cell sorting remains limited.[52]

3. Application of ML in Stem Cell Sorting (2000s)

Early 2000s

Stem Cell Milestones: The first derivation of hESCs opens up new possibilities in regenerative medicine, with more targeted sorting based on specific cell markers.[53]

AI and ML Adoption: ML begins enhancing gating strategies in flow cytometry, improving the ability to identify and sort stem cells.[54]

2005

Stem Cell Sorting: The introduction of automated microscopy and high-content screening enables better sorting based on cellular morphology and behavior.[55]

AI in Sorting: ML helps automate stem cell identification in complex mixtures, improving accuracy and efficiency.[55]

Late 2000s

iPSCs: Shinya Yamanaka's discovery of iPSCs leads to advancements in personalized medicine, with new challenges in ensuring iPSC purity during sorting.[56]

ML Integration: ML is increasingly employed to optimize protocols for sorting iPSCs.[17]

4. AI-driven Stem Cell Sorting Technologies (2010s)

Early 2010s

Stem Cell Sorting Evolution: Sorting techniques evolve to focus on functional properties such as differentiation potential. Flow cytometry and magnetic-activated cell sorting (MACS) are widely used.[57]

AI in Sorting: The advent of DL vastly improves image recognition, which is used for analyzing stem cell images and refining sorting processes.[58]

2014

AI and Stem Cells: Convolutional neural networks (CNNs) are used to classify cellular images, offering more accurate identification and sorting in high-throughput environments.[59]

Mid-2010s

Stem Cell Sorting Innovations: AI is combined with microfluidic technologies to develop advanced platforms capable of sorting based on molecular, morphological, and functional characteristics.[60]

AI and ML Advances: Predictive models help foresee stem cell differentiation, guiding the sorting process to isolate cells with high therapeutic value.[58]

Late 2010s

Single-Cell Analysis: Single-cell RNA sequencing, powered by AI, allows for unprecedented resolution in sorting stem cells and understanding their heterogeneity.[61]

AI-Enhanced Sorting: AI-driven systems improve sorting accuracy by analyzing multi-dimensional data such as gene expression and protein levels.[62]

5. Modern AI, ML, and Stem Cell Sorting (2020s)

2020

AI in Stem Cell Research: AI and ML models predict stem cell behavior and fate, enhancing sorting efficiency and outcomes in regenerative medicine and personalized therapies.[60]

Advanced Sorting Techniques: AI-enhanced flow cytometry and lab-on-a-chip systems are introduced, enabling precise and automated stem cell sorting based on numerous biomarkers.[60]

2021

ML Applications: Advanced ML models, including unsupervised learning and GANs, are used to identify rare stem cell populations and predict differentiation outcomes.[63]

AI-Optimized Protocols: AI is applied to optimize sorting protocols, improving stem cell purity and yield, especially in clinical settings.[63]

2022

Lab-on-a-Chip Innovations: AI-powered lab-on-a-chip devices that merge microfluidics and ML provide fully automated, high-throughput stem cell sorting for tissue engineering and regenerative medicine.[64]

2023

Multi-Omics Integration: AI is employed to integrate multi-omics data in sorting, allowing for comprehensive analysis of stem cells based on genomic, transcriptomic, proteomic, and epigenetic profiles.[65]

Visual Data

Image-Based Methods: An AI method to differentiate the iPSCs formed during colony formation was developed by Fan and colleagues. CNN was used for image processing, whereas the system training was performed by the Hidden Markov Model (HMM). Such AI algorithms help both pharmaceuticals and research institutes to better choose the differentiated iPSCs. They converted the bright field images into binary images and then characterized the boundaries, location, and intensity of the spots in each image using a re-binary method. The OCT4-GFP expression in the iPSCs generated green fluorescent images, which were used for positive feedback training. The combination of CNN and HMM can perform a quantitative iPSC image analysis and differentiate between the four stages of iPSC expansion.[66] Traditionally, highly trained individuals were needed to differentiate the iPSC colonies, a task that may lead to misjudgment and batch-to-batch variation. ML approaches, on the other hand, can create big data to positively and negatively treat the system, helping to differentiate the shape, location, and polarity of cells. This can not only help reduce human errors drastically but also save lots of time.

MSCs are one of the most commonly used stem cells in tissue engineering and regenerative medicine. This points out the importance of image processing from microscopic pictures for cell analysis. Recently, Mota and colleagues introduced an AI-based algorithm for MSC detection through phase-contrast micrographs. Their algorithm consisted of three steps: finding cell clusters, determining specific cell features, and finally differentiating specific cells in the cell cluster region. This AI-based image processing helps reduce the time and cost of the frustrating traditional stem cell sorting. Different ML methods, such as Linear kernel support vector machine (LSVM), Radial-basis kernel support vector machine (RSVM), Linear discriminant analysis (LDA), K-nearest neighbor (KNN), Logistic regression (LR) models, and their combinations were trained to classify MSCs. The best result belonged to the combination of RSVM+LR and LDA+KNN based on the AUC results.[67]

Waisman and colleagues introduced a CNN-based algorithm for PSC differentiation. ESCs and iPSCs are two main PSCs that originated from the blastocyst's inner cell mass and engineered adult differentiated cells, respectively. These fast, easy, and intelligent methods will replace the traditional, expensive, hard, and time-consuming biomolecular cell analysis. Based on hundreds of transmitted light microscopy images, the CNN was capable of distinguishing between the PSCs and early-stage differentiated cells. For system training, the DenseNet and ResNet50 algorithms were used. Based on this method, each image was considered as a multilayer moiety, considered in a layer-by-layer manner. Each layer was linked to the size and depth of the image to increase the image contrast in all layers. During the early-stage differentiation process of PSCs, the expression pattern of the cells changes, leading to the morphological transformation in the differentiated cells. CNN was used to learn the changes and predict the formation of the new differentiated colonies. It was also possible to train the algorithms to predict the cell differentiation status in different culture conditions, as well as their types.[68]

Neural stem cells (NSCs) are tripotent stem cells that can differentiate into neurons, oligodendrocytes, and astrocytes. These cells can be used for the treatment of central nervous system (CNS) diseases. As a result, fate identification of NSCs is essential since each differentiated cell has its own clinical application.[69] In a study by Zhu and colleagues, a

CNN algorithm was used to distinguish between the differentiated NSCs. Both bright and dark field single images were used to train and test the algorithm. Immunofluorescent imaging was used to monitor biomarkers during the cell differentiation process. The obtained accuracy was 0.923 and 0.998 for the bright field and fluorescent images, respectively. The CNN was designed based on an Xception algorithm to successfully identify the differentiated NSCs in both single and mixed culture media. The designed algorithm was trained by a multi-layer perceptron, ResNet, VGGNet, and InceptionV3. The final algorithm could identify the small morphological changes in almost one hour.[70]

Quality control (QC) of large-scale iPSCs production is crucial due to its applications in tissue engineering, disease modelling, and drug screening. Most QC tasks are performed manually; they are, therefore, laborious and tedious. Joutsijoki and colleagues introduced a classification AI method based on three classes: good, semi-good, and bad. Least-squares support vector machine (LS-SVM) and κ-nearest neighbor (κ-NN) classifiers were the main algorithms utilized in this study. One-versus-all (OVA), one-versus-one (OVO), directed acyclic graph support vector machines (DAGSVMs), and binary tree support vector machines were also incorporated in the algorithms as LS-SVM extensions. Ultimately, the κ-NN classifier with 0.62 accuracy values was concluded as the best iPSCs categorizer.[71]

The variations observed in stem cell differentiation highlight the necessity for an automated AI-assisted system for cell identification. While a human expert is involved in feature extraction in conventional ML, DL has the advantage of automated pattern extraction using multiple-layered deep neural networks. Hence, Kusumoto and others applied DL in CNN for the identification of iPSC-derived endothelial cells. The technique solely benefited from the morphological information of cells based on phase-contrast images, with no need for any molecular technique (immunostaining or lineage tracing).[72] They used two forms of differentiated iPSCs images: phase-contrast and binarized immunofluorescence-stained for CD31 (endothelial cell marker). The target blocks extracted from the phase-contrast images were classified as unstained or stained. Using LeNet and AlexNet networks, the target blocks extracted from the binarized immunofluorescence-stained images were predicted to be unstained or stained. Their investigation revealed the accuracy of the method to be in correlation with the depth of the network and the size of pixels in the analyzed images.[72]

Substitution of animal models with hiPSCs-derived cardiomyocytes (hiPSC-CMs) is an ongoing process to help overcome the issues in predicting cardiotoxicity of drug candidates due to the differences in cardiomyocytes of various species. This is because variations in the morphology and physiology of hiPSC-CMs, which depend on the experimental conditions or the product lots, hinder their usage.[73] Hence, Orita and colleagues used bright-field microscopic time-lapse video images for optical detection of contractility of cultured hiPSC-CMs and discriminating functionally normal from abnormal cardiomyocyte contractions.[73] They applied a sliding window method, fast Fourier transform, and uniform manifold approximation and projection (UMAP), for data augmentation, preprocessing, and nonlinear dimensional reduction, respectively. Finally, a support vector machine (SVM) was trained to assess the quality of individual contraction waves.

ML algorithms can also be beneficial in replacing manual morphology analysis of stem cell-based embryo models with less biased and more efficient automated workflows. The embryo-like structures can be formed from the self-assembly of embryonic and extra-embryonic stem cell lines. They are useful for *in vitro* investigation of embryo development. Manual analysis of the morphology of stem cell-based embryo models is researcher-dependent and associated with insufficient efficiency due to common variations in PSC cell lines secondary to differences in genetic background, culture, and methods of reprogramming.[74] To tackle this, Guo and others applied ML for high-content analysis of the embryo-like structures using 3D images.[74] Additional biological parameters (such as area, length, width, fluorescence intensity, and so on) were also quantified along with the machine-extracted features of the algorithm. The ML algorithm was trained by introducing good-morphology structures (high-grade ITS) as well as low-grade examples not representing ITS (iPSCs and trophectoderm stem cells) embryos (figure 4-A, B). The ML algorithm extracted multivariate features, capable of discriminating high- and low-grade ITS examples. Figure 4-C illustrates a summary of the method.

Flow Cytometry-Based Methods

The large-scale datasets obtained from the flow cytometry techniques are another input data with the potential for coupling with ML approaches for cell sorting purposes. Accordingly, Sardiu and colleagues used Nilsson's rare hHSCs flow cytometry dataset,

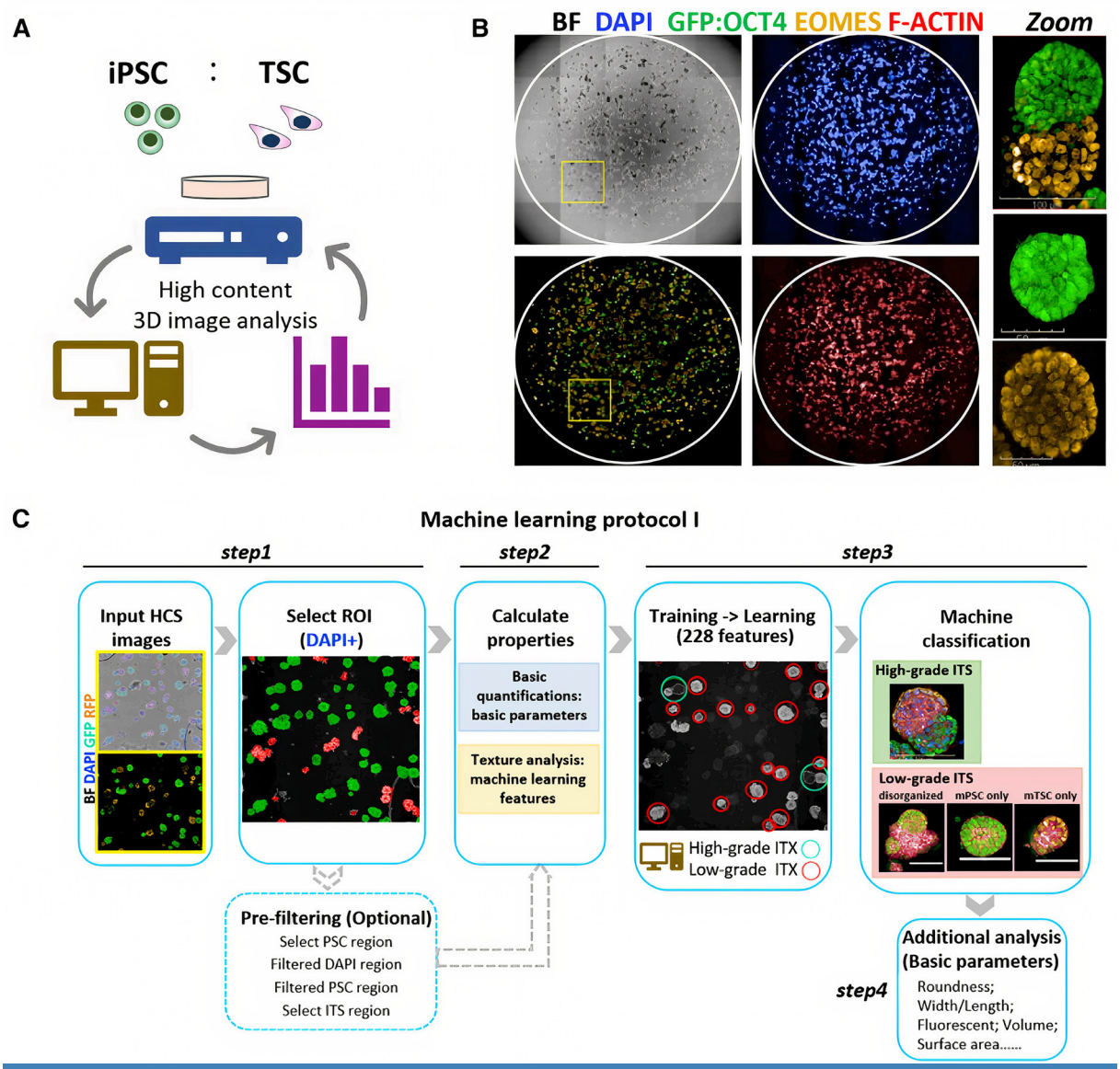

Figure 4: ML algorithm based on the visual dataset for identifying iPSCs and trophectoderm stem cells: A) ML–based high–content ITS embryos analysis method is capable of discriminating high- and low-grade ITS examples. B) Scanning the images by PerkinElmer Opera Phenix high–content confocal helps discriminate different stem cells. C) The strategy behind the ML method is based on applying different biological parameters (Reproduced from Elsevier under 5372900848943 permission number). iPSC: Induced pluripotent stem cell; TSC: Trophoblast stem cell; BF: Bright field; GFP: Green fluorescent protein; HSC: Hematopoietic stem cell; ROI: Region of interest; ITS: iPSCs and trophectoderm stem cells; ITX: iPSCs, trophectoderm stem cells, and extraembryonic endoderm stem cells; PSC: Pluripotent stem cell

an immunological publicly available dataset, to provide an effective approach to analyze large-scale flow cytometry data and identify rare cell populations.[75] Their method consisted of coupling the Topological Score or TopS (a previously developed proteomics data analysis method)[76, 77] with ML approaches such as topological data analysis (TDA), X-shift clustering, and t-Distributed Stochastic Neighbor Embedding (t-SNE). The original data was first processed by Weber's arc-sinh transformation[78] to generate a so-called "original/transformed data" matrix. Following the generation of topological values, a Pearson correlation and a hierarchical clustering were performed to obtain an illustrated sample classification. The ML approaches were then utilized to show the capability of TopS in rare cell discrimination.[75]

Video-Based Methods

As previously described, ESCs are PSCs with self-renewal abilities. Guan and colleagues developed a DL AI-based random network (RandNet) with an auto-encoder feature extractor to divide human ESCs (hESCs) into six groups (namely, cell clusters, attached cells, unattached cells, apoptotically blebbing cells, dynamically blebbing cells, and debris) (figure 5-A). In this research, time-lapse videos were captured from hESCs using phase contrast microscopes.

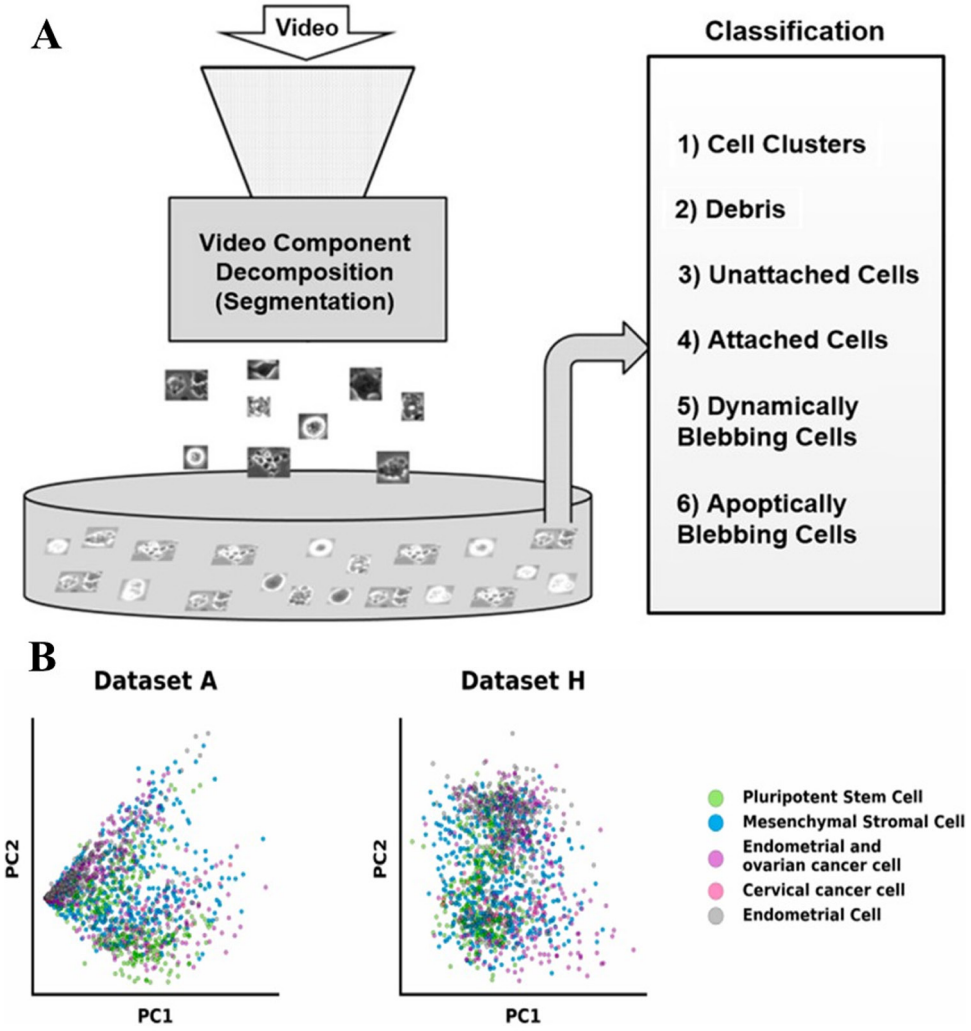

Figure 5: Using RandNet algorithm for visual and NN algorithm for non-visual datasets to cluster and identify stem cells: A) hESC can be sorted using RandNet AI–based algorithm into six different groups based on the microscopic videos (Adapted from Guan and others, 2021 under CC BY license). B) PCA was applied on five different cell lines to identify PSCs using LC and NN algorithms with the help of lectin microarray (Reproduced from Elsevier under 5372901428561 permission number). PC: Principal component

Then, different video parts were assembled into various groups, on which AI algorithms were applied. This helped to categorize the six described elements in the ESCs culture media by an accuracy value of 0.97±0.0094. It also helped with improving the ESCs' sorting for regenerative medicine and clinical applications.[79]

Non-Visual Data

Microarray-Based Methods

Shibata and colleagues studied the lectin microarray using AI models for human pluripotent stem cell (hPSC) identification (figure 5-B). Lectin microarray is the study of glycan patterns and heterogeneity in different cell types. Since most cellular proteins are glycosylated, the lectin microarray can be considered as a biomarker for recognizing normal cells and disease patterns.[80] In this study, principal component analysis (PCA) was first applied to the glycan profile of five different cell lines, including hPSCs, mesenchymal stromal cells, ovarian and cervical cancer cells, and endometrial cells. Subsequently, the linear classification[66] and neural network (NN) were applied on the lectin microarray to discriminate these cells from each other. The target cell (PSC) was successfully discriminated by 0.89 and 0.97 accuracy values using LC and NN, respectively.[81]

Expression Pattern-based Methods

Complex biological dynamics of the molecular regulatory networks are involved in the formation of a functional pluripotency state. The research of Stumpf and colleagues is another example of using ML to study complex dynamics in cell communities.[82] They used a PCA-similar approach to reconstruct the pluripotent cell identities from single-cell expression data. The method consisted of three main steps:

1) Snapshot collection from the regulatory activity of individual cells using their expression data, 2) PCA application on snapshots to extract the regulatory network patterns and weighting, and 3) Fitting a Gaussian Mixture Model to weight based on the low-dimensional representation of the activity of the regulatory network. Their method helped identify three different stem cell network configurations and their association with specific combinations of regulatory network activity patterns involved in controlling various aspects of the cell response to environmental circumstances.[82]

Electrical Data-Based Methods

Although ML has extensive applications in the analysis of cellular images as the primary data, there are examples of using other types of input data with AI for biological study purposes. Electrical impedance spectroscopy (EIS) of live cells is a non-invasive, label-free, and quantitative method for the assessment of the stem cells in terms of proliferation and differentiation.[83] This method, however, requires broad post-processing analysis. Cunha and others developed three ML models to classify the stem cell differentiation or proliferation status based on the electrical impedance data.[83] Their method comprised of recording the impedance data every 5 days for a 35-day duration, and subsequent use of long short-term memory (LSTM) artificial recurrent neural networks for classification purposes. To overcome the overfitting problem, a regularizer (a connected hidden layer using Ridge regularization) and a random Dropout (dropping selected classifiers randomly during the training process for a second recovery, resulting in enhanced accuracy and less overfitting) were applied. Finally, a Softmax activation function was used to translate the probability of the model into differentiation or proliferation classes.[83]

Cell electrophysiological data is another input data that can be coupled with ML algorithms. For instance, Heylman and colleagues developed an algorithm by applying a supervised ML (TreeBagger function of MATLAB, which uses random forests with bootstrap aggregation for training) for drug development purposes.[84] They used the algorithm on cell electrophysiological data, based on the assessment of voltage-sensitive dyes (VSDs) located in the cell membrane of hiPSC-derived cardiomyocytes by 2-photon microscopy. The algorithm was successful in the accurate assessment, prediction, and classification of membrane depolarization of the target cells upon exposure to the chronotropic drugs.[84]

Biomechanical Data-Based Methods

One of the most commonly used techniques for cell enrichment is FACS, which analyzes the cell surface markers. Darling and Guilak introduced a biomechanical-based method to train the AI algorithms using the data acquired from an atomic force microscopy (AFM) and its biomechanical properties, such as equilibrium modulus, elastic modulus, structural features, and viscosity, instead of the routine FACS and cell surface property methods. Cell enrichment is divided into two main categories: removing the unrelated cells and isolating the target cells. In the current study, the isolation was assessed based on AFM data, whereas the neural network was used to categorize similar cells. In one of their experiments, two adult stem cells, adipose–(ASCs) and bone marrow–derived (MSCs) stem cells, were successfully isolated from primary mesenchymal-derived cells after NN training based on the AFM data. Such approaches are important due to the growing demand for tissue engineering techniques requiring cell sorting and isolation methods.[41]

Recent research in the ML applications for stem cell sorting is summarized in table 1, and compared in terms of their input data, stem cell, ML methods, study purpose, and performance values.

The analysis of stem cell sorting approaches revealed that image-based data dominate among input types, comprising 53% of all data modalities used (figure 6-A). This substantial reliance on image data can be attributed to the high information content and resolution provided by imaging techniques, which allow for detailed assessment of stem cell morphology and phenotypic characteristics. Imaging is particularly well-suited for ML algorithms, especially DL models, due to its structured pixel-based format and large datasets that can be annotated or augmented for training.[87] In contrast, other data types, such as electrical, biochemical, or flow cytometry data, represent a smaller fraction of use, likely due to limitations in scalability, standardization, or compatibility with automated image analysis pipelines.

In terms of stem cell types studied (figure 6-B), iPSCs represent the largest category (44%), followed by MSCs (11%) and PSCs (17%). The dominance of iPSCs is likely due to their clinical relevance, broad differentiation potential, and growing interest in personalized regenerative therapies. Moreover, iPSCs often exhibit distinct morphological features that are readily captured and analyzed via imaging, reinforcing the predominance of image-based data in ML-driven stem cell research.[88] It also may

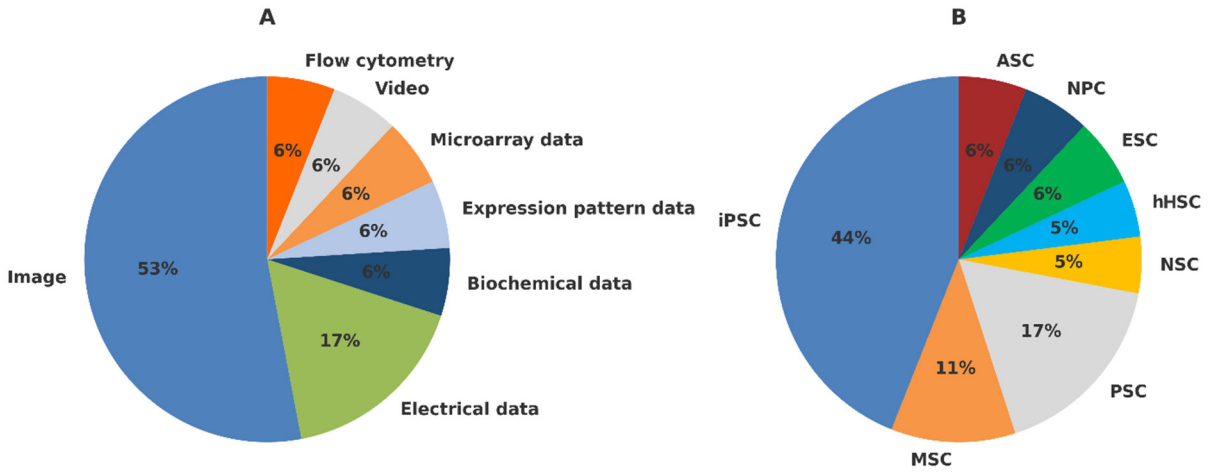

Figure 6: AI-assisted stem cell sorting researches listed in table 1 are summarized in this figure. A) Input data distribution shows that image data consists about 53% of all input data, and B) Distribution of stem cell types classified by AI-based methods represents that iPSC is responsible for 44% of cell types; iPSC: Induced pluripotent stem cell; ASC: Adipose-derived stem cell; NPC: Neural progenitor cell; ESC: Embryonic stem cell; hHSC: Human hematopoietic stem cell; NSC: Neural stem cell; PSC: Pluripotent stem cell; MSC: Mesenchymal stem cell

be attributed to their biological reprogramming ability and the provision of reduced reliance on ESCs and adult stem cells, which are more limited in availability.[89] These trends highlight the critical role of computer vision in advancing stem cell sorting technologies and the need for robust image analysis tools tailored to stem cell-specific applications.

Future Perspective

The landscape of stem cell research and therapy is being profoundly reshaped by advancements in AI and ML. For a more comprehensive look at the future perspectives of the AI-driven stem cell sorting algorithms, it is better to first discuss the current marketed AI-based stem cell sorter technologies. This can be divided into two main categories, including automated devices with advanced ML algorithms to sort and separate special stem cells, and the online or offline software, which can be used for stem cell sorting in combination with other devices.

Automated Devices with Advanced ML Algorithms

Automated cell sorting devices incorporate ML algorithms to enhance the precision and efficiency of identifying and isolating specific stem cell types. These devices are designed to handle complex sorting tasks that require high-throughput capabilities and precise control.

1. *ThinkCyte's VisionSort* utilizes a technology called Ghost Cytometry, which combines image-based sorting with real-time AI analysis to sort cells without the need for fluorescent markers. This system is particularly adept at distinguishing subtle morphological differences between cells, which is crucial for applications such as regenerative medicine, where cell purity can significantly impact therapeutic outcomes.[90]

2. *NanoCellect's WOLF G2 Benchtop Microfluidic Cell Sorter* is another example, celebrated for its gentle handling of cells, thereby preserving their viability and functionality. It uses microfluidic technology to sort cells under low pressure, reducing the stress inflicted on delicate stem cells during sorting.[91]

3. *Deepcell* offers the REM-I platform, an AI-powered system that integrates single-cell imaging, sorting, and high-dimensional analysis. This platform is highly effective in areas such as cancer research, stem cell studies, and gene therapy, providing deep insights into cell biology through advanced AI-based morphology analysis.[92]

Software Solutions for Stem Cell Sorting

Complementing physical sorting devices are software solutions designed to enhance the functionality and usability of cell sorters. These programs allow researchers to define sorting parameters more precisely, analyze results in real-time, and integrate with existing laboratory information systems for seamless data management.

1. *BD FACSChorus Software*, designed for use with BD's cell sorters, simplifies the cell sorting process with an intuitive interface and automated features that guide users through setup, execution, and analysis of sorting procedures.[93]

2. *MACSQuantify™ Tyto Software* by Miltenyi Biotec is designed to handle complex

cell sorting experiments. It offers advanced data protection, streamlined standardization, and easy integration into GMP environments, making it ideal for high-level cell sorting tasks.[94]

Startups Shaping the Future of AI-based Stem Cell Sorters

Several startups are pivotal in driving innovation in this sector. These companies are not only enhancing current technologies but are also paving the way for new applications and methodologies in stem cell sorting.

1. *Somite*, a rising leader in the field, is using AI to create innovative models that enhance the production of human tissues from stem cells for therapeutic purposes. Their efforts are essential in overcoming the current challenges of scalability and efficiency in stem cell therapies.[95]

2. *CytoRecovery* specializes in microfluidics technology to separate specific cell types, including cancer stem cells, from complex biological samples. Their advancements aim to boost the accuracy of cell therapies and improve cancer treatment methods.[96]

AI-based technologies are revolutionizing the field of stem cell research and therapy. Automated devices and sophisticated software solutions are making stem cell sorting more precise and efficient, while startups play a critical role in advancing these technologies and expanding their application in medicine. As AI continues to evolve, its integration into stem cell technologies holds the promise of significant breakthroughs in healthcare and therapy.

Conclusion

The successful application of stem cells in regenerative medicine necessitates the precise separation of undifferentiated from differentiated cell types. This is predominantly achieved through the integration of FACS or biochemical methods with traditional techniques, wherein the identification of molecular characteristics and accurate cell isolation are accomplished through costly and time-intensive processes. Recent advancements in AI technology have also proven advantageous in biomedical applications. This review presents a comprehensive analysis of recent research on ML-assisted stem cell sorting approaches. These approaches are categorized based on the type of input data derived directly from biological samples (specifically stem cell populations), which are subsequently processed by computational ML-based methods. The integration of AI in stem cell sorting methodologies offers several advantages, including enhanced accuracy, predictive capabilities, the classification of complex cell populations, and increased time and cost efficiency. In addition, automated devices and software solutions are making stem cell sorting more precise and efficient, and startups can play a critical role in advancing AI-based technologies. Nevertheless, the dependency on the quality of laboratory-obtained data remains a significant limiting factor. Future research should aim to minimize human involvement in the collection of primary data to achieve a more accurate and automated process.

Acknowledgment

The research council of Tarbiat Modares University is acknowledged for supporting this work.

Authors' Contribution

M.M: Material preparation, data collection, analysis, drafting and reviewing the manuscript; A.J: Material preparation, data collection, analysis, drafting and reviewing the manuscript; H.S: Material preparation, data collection and analysis, drafting and reviewing the manuscript; F.Y: Study concept, supervision and reviewing the manuscript; K.R: Material preparation, data collection and analysis, validatin, and reviewing the manuscript; K.C: Data analysis and validation, and reviewing the manuscript: P.K: Data analysis and validation, and reviewing the manuscript; All authors have read and approved the final manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

AI Declaration

Artificial intelligence (ChatGPT, GPT-5.2) was used solely for creating figure 3. The authors are fully responsible for the content and integrity of the manuscript.

Conflict of Interest: None declared.

References

1 Knoepfler PS. Deconstructing stem cell tumorigenicity: a roadmap to safe regenerative medicine. Stem Cells. 2009;27:1050-6. doi: 10.1002/stem.37. PubMed PMID: 19415771; PubMed Central PMCID: PMC2733374.

2 Park PC, Selvarajah S, Bayani J, Zielenska M,

Squire JA. Stem cell enrichment approaches. Semin Cancer Biol. 2007;17:257-64. doi: 10.1016/j.semcancer.2006.04.001. PubMed PMID: 16814562.
3 Xu Z, Chen Z, Yang S, Chen S, Guo T, Chen H. Passive Focusing of Single Cells Using Microwell Arrays for High-Accuracy Image-Activated Sorting. Anal Chem. 2024;96:347-54. doi: 10.1021/acs.analchem.3c04195. PubMed PMID: 38153415.
4 Caudai C, Galizia A, Geraci F, Le Pera L, Morea V, Salerno E, et al. AI applications in functional genomics. Comput Struct Biotechnol J. 2021;19:5762-90. doi: 10.1016/j.csbj.2021.10.009. PubMed PMID: 34765093; PubMed Central PMCID: PMC8566780.
5 Dias R, Torkamani A. Artificial intelligence in clinical and genomic diagnostics. Genome Med. 2019;11:70. doi: 10.1186/s13073-019-0689-8. PubMed PMID: 31744524; PubMed Central PMCID: PMC6865045.
6 Sony [Internet]. Sony introduces new closed cell isolation system CGX10 for use in production of advanced cell therapy products. Tokyo: Sony Corporation; [cited 2022 Jan]. Available from: https://pro.sony/de_AT/press/closed-cell-isolation-system-cgx10-use-production-advanced-cell-therapy-products
7 Srinivasan M, Thangaraj SR, Ramasubramanian K, Thangaraj PP, Ramasubramanian KV. Exploring the Current Trends of Artificial Intelligence in Stem Cell Therapy: A Systematic Review. Cureus. 2021;13:e20083. doi: 10.7759/cureus.20083. PubMed PMID: 34873560; PubMed Central PMCID: PMC8635466.
8 Thanaskody K, Jusop AS, Tye GJ, Wan Kamarul Zaman WS, Dass SA, Nordin F. MSCs vs. iPSCs: potential in therapeutic applications. Front Cell Dev Biol. 2022;10:1005926. doi: 10.3389/fcell.2022.1005926. PubMed PMID: 36407112; PubMed Central PMCID: PMC9666898.
9 Alamdar N, Farivar S, Baghaei K, Hamidieh AA, Soltaninejad H, Aghdaei HA, et al. Effect of pioglitazone and cetuximab on colon cancer stem-like cell (CCSLCs) properties. Curr Stem Cell Res Ther. 2025;20:83-90. doi: 10.2174/011574888X28331824011811822. PubMed PMID: 38303523.
10 Ke X, Thakur A, Chen HJ. Transdifferentiation meets next-generation biotechnologies. StemJournal. 2022;4:1-11. doi: 10.3233/STJ-200003.
11 Li H, Long C, Hong Y, Luo L, Zuo Y. Characterizing cellular differentiation potency and Waddington landscape via energy indicator. Research (Wash DC). 2023;6:0118. doi: 10.34133/research.0118. PubMed PMID: 37223479; PubMed Central PMCID: PMC10202187.
12 Coronnello C, Francipane MG. Moving Towards Induced Pluripotent Stem Cell-based Therapies with Artificial Intelligence and Machine Learning. Stem Cell Rev Rep. 2022;18:559-69. doi: 10.1007/s12015-021-10302-y. PubMed PMID: 34843066; PubMed Central PMCID: PMC8930923.
13 Holmes J, Sacchi L, Bellazzi R. Artificial intelligence in medicine. Ann R Coll Surg Engl. 2004;86:334-8. doi: 10.1308/147870804290. PubMed PMID: 15333167; PubMed Central PMCID: PMC1964229.
14 Tiwari T, Tiwari T, Tiwari S. How artificial intelligence, machine learning and deep learning are radically different? Int J Adv Res Comput Sci Softw Eng. 2018;8:1-6. doi: 10.23956/ijarcsse.v8i2.569.
15 Reddy Y, Viswanath P, Reddy BE. Semi-supervised learning: a brief review. Int J Eng Technol. 2018;7:81-4. doi: 10.14419/ijet.v7i1.8.9977.
16 Tishan Y. Understanding the difference between supervised and unsupervised learning techniques. Plymouth: University of Plymouth; 2023.
17 Vo QD, Saito Y, Ida T, Nakamura K, Yuasa S. The use of artificial intelligence in induced pluripotent stem cell-based technology over 10-year period: a systematic scoping review. PLoS One. 2024;19:e0302537. doi: 10.1371/journal.pone.0302537. PubMed PMID: 38771829; PubMed Central PMCID: PMC11108174.
18 Nosrati H, Nosrati M. Artificial intelligence in regenerative medicine: applications and implications. Biomimetics (Basel). 2023;8:442. doi: 10.3390/biomimetics8050442. PubMed PMID: 37754193; PubMed Central PMCID: PMC10526210.
19 García Martín H, Mazurenko S, Zhao H. Special issue on artificial intelligence for synthetic biology. ACS Synth Biol. 2024;13:408-10. doi: 10.1021/acssynbio.3c00760. PubMed PMID: 38214972.
20 Praveen Chakravarthi G, Rambabu V, Ramamurty DSVNM, Prasad SVGVA. AI and machine learning in biotechnology: a paradigm shift in biochemical innovation. Int J Plant Anim Environ Sci. 2024;14:70-80. doi: 10.26502/ijpaes.4490166.
21 Gomes A, Gonçalves B, Inglês B, Silvério S, Pinto CA, Saraiva JA. Potential impacts of artificial intelligence (AI) in biotechnology. Appl Sci (Basel). 2024;14:11801. doi: 10.3390/app142411801.

22 Azimzadeh M, Khashayar P, Mousazadeh M, Daneshpour M, Rostami M, Goodlett DR, et al. Volatile organic compounds (VOCs) detection for the identification of bacterial infections in clinical wound samples. Talanta. 2025;287:127991. doi: 10.1016/j.talanta.2025.127991. PubMed PMID: 40132411.
23 Mousazadeh M, Nikkhah M. Advanced bladder cancer detection: innovations in biomarkers and nanobiosensors. Sens Bio-Sens Res. 2024;45:100667. doi: 10.1016/j.sbsr.2024.100667.
24 Khoshandam M, Soltaninejad H, Hosseinkhani S, Saltanatpour Z, Goudarzi MTH, Hamidieh AA. CRISPR/Cas and artificial intelligence to improve precision medicine: future perspectives and potential limitations. Hum Gene. 2024;42:201356. doi: 10.1016/j.humgen.2024.201356.
25 Lasarow RM, Isseroff RR, Gomez EC. Quantitative in vitro assessment of phototoxicity by a fibroblast-neutral red assay. J Invest Dermatol. 1992;98:725-9. doi: 10.1111/1523-1747.ep12499927. PubMed PMID: 1569321.
26 Brown M, Wittwer C. Flow cytometry: principles and clinical applications in hematology. Clin Chem. 2000;46:1221-9. doi: 10.1093/clinchem/46.8.1221. PubMed PMID: 10926916.
27 Bhagavati S, Xu W. Isolation and enrichment of skeletal muscle progenitor cells from mouse bone marrow. Biochem Biophys Res Commun. 2004;318:119-24. doi: 10.1016/j.bbrc.2004.03.192. PubMed PMID: 15110761.
28 Buican TN, Smyth MJ, Crissman HA, Salzman GC, Stewart CC, Martin JC. Automated single-cell manipulation and sorting by light trapping. Appl Opt. 1987;26:5311-6. doi: 10.1364/AO.26.005311. PubMed PMID: 20523522.
29 Eisenstein M. Divide and conquer. Nature. 2006;441:1179. doi: 10.1038/4411179a. PubMed PMID: 16810261.
30 Emmert-Buck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang Z, Goldstein SR, et al. Laser capture microdissection. Science. 1996;274:998-1001. doi: 10.1126/science.274.5289.998. PubMed PMID: 8875945.
31 Kovac J, Voldman J. Intuitive, image-based cell sorting using optofluidic cell sorting. Anal Chem. 2007;79:9321-30. doi: 10.1021/ac071366y. PubMed PMID: 18004819; PubMed Central PMCID: PMC2536524.
32 Kovanen K, Lukkari M, Purmonen S, Ylikomi T, Viitanen J, Kallio PJ, editors. Electric impedance assisted micropipette aspiration. In: 2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Piscataway (NJ): IEEE; 2007. p. 4489-92. doi: 10.1109/IEMBS.2007.4353671. PubMed PMID: 18003337.
33 Liu YJ, Guo SS, Zhang ZL, Huang WH, Baigl D, Xie M, et al. A micropillar-integrated smart microfluidic device for specific capture and sorting of cells. Electrophoresis. 2007;28:4713-22. doi: 10.1002/elps.200700212. PubMed PMID: 18008303.
34 Parekkadan B, Sethu P, Van Poll D, Yarmush ML, Toner M. Osmotic selection of human mesenchymal stem/progenitor cells from umbilical cord blood. Tissue Eng. 2007;13:2465-73. doi: 10.1089/ten.2007.0054. PubMed PMID: 17665999.
35 Perfetto SP, Chattopadhyay PK, Roederer M. Seventeen-colour flow cytometry: unravelling the immune system. Nat Rev Immunol. 2004;4:648-55. doi: 10.1038/nri1416. PubMed PMID: 15286731.
36 Fried J, Doblin J, Takamoto S, Perez A, Hansen H, Clarkson B. Effects of Hoechst 33342 on survival and growth of two tumor cell lines and on hematopoietically normal bone marrow cells. Cytometry. 1982;3:42-7. doi: 10.1002/cyto.990030110. PubMed PMID: 6180874.
37 Progatzky F, Dallman MJ, Lo Celso C. From seeing to believing: labelling strategies for in vivo cell-tracking experiments. Interface Focus. 2013;3:20130001. doi: 10.1098/rsfs.2013.0001. PubMed PMID: 23853708; PubMed Central PMCID: PMC3638420.
38 Patil RM, Thorat ND, Shete PB, Bedge PA, Gavde S, Joshi MG, et al. Comprehensive cytotoxicity studies of superparamagnetic iron oxide nanoparticles. Biochem Biophys Rep. 2018;13:63-72. doi: 10.1016/j.bbrep.2017.12.002. PubMed PMID: 29349357; PubMed Central PMCID: PMC5766481.
39 Darling EM, Topel M, Zauscher S, Vail TP, Guilak F. Viscoelastic properties of human mesenchymally-derived stem cells and primary osteoblasts, chondrocytes, and adipocytes. J Biomech. 2008;41:454-64. doi: 10.1016/j.jbiomech.2007.06.019. PubMed PMID: 17825308; PubMed Central PMCID: PMC2897251.
40 Frankel DS, Frankel SL, Binder BJ, Vogt RF. Application of neural networks to flow cytometry data analysis and real-time cell classification. Cytometry. 1996;23:290-302. doi: 10.1002/

(sici)1097-0320(19960401)23:4<290::aid-cyto5>3.0.co;2-l. PubMed PMID: 8900472.
41 Darling EM, Guilak F. A neural network model for cell classification based on single-cell biomechanical properties. Tissue Eng Part A. 2008;14:1507-15. doi: 10.1089/ten.tea.2008.0180. PubMed PMID: 18620486.
42 Kohonen T. Self-organizing maps. 3rd ed. Berlin: Springer; 2001. doi: 10.1007/978-3-642-56927-2.
43 Piraino P, Ricciardi A, Salzano G, Zotta T, Parente E. Use of unsupervised and supervised artificial neural networks for the identification of lactic acid bacteria on the basis of SDS-PAGE patterns of whole cell proteins. J Microbiol Methods. 2006;66:336-46. doi: 10.1016/j.mimet.2005.12.007. PubMed PMID: 16480784.
44 Mazalan M, Do TD, Zaman WSWK, Ramlan EI. Machine learning approaches for stem cells. Curr Stem Cell Rep. 2023;9:43-56. doi: 10.1007/s40778-023-00228-1.
45 Poliwoda S, Noor N, Downs E, Schaaf A, Cantwell A, Ganti L, et al. Stem cells: a comprehensive review of origins and emerging clinical roles in medical practice. Orthop Rev (Pavia). 2022;14:37498. doi: 10.52965/001c.37498. PubMed PMID: 36034728; PubMed Central PMCID: PMC9404248.
46 Chen JJ. From Turing's conception of machine intelligence to the evolution of AI in early childhood education: conceptual, empirical, and practical insights. AI Brain Child. 2025;1:1. doi: 10.1007/s44436-025-00002-6.
47 Choudhery MS, Arif T, Mahmood R. Applications of artificial intelligence in stem cell therapy. World J Stem Cells. 2025;17:106086. doi: 10.4252/wjsc.v17.i8.106086. PubMed PMID: 40951704; PubMed Central PMCID: PMC12427078.
48 Iyer JK. Applications of flow cytometry in scientific research and the clinic. In: Flow Cytometry: Principles, Applications, and Clinical Perspectives. Singapore: Springer; 2024. p. 197-212. doi: 10.1007/978-981-97-4553-1_12.
49 Yogo T, Iwamoto Y, Becker HJ, Kimura T, Ishida R, Sugiyama-Finnis A, et al. Quantitative phase imaging with temporal kinetics predicts hematopoietic stem cell diversity. Nat Commun. 2025;16:6496. doi: 10.1038/s41467-025-61846-3. PubMed PMID: 40659629; PubMed Central PMCID: PMC12260078.
50 Eguizabal C, Aran B, Chuva de Sousa Lopes SM, Geens M, Heindryckx B, Panula S, et al. Two decades of embryonic stem cells: a historical overview. Hum Reprod Open. 2019;2019:hoy024. doi: 10.1093/hropen/hoy024. PubMed PMID: 30895264; PubMed Central PMCID: PMC6396646.
51 Preffer F, Dombkowski D. Advances in complex multiparameter flow cytometry technology: applications in stem cell research. Cytometry B Clin Cytom. 2009;76:295-314. doi: 10.1002/cyto.b.20480. PubMed PMID: 19492350; PubMed Central PMCID: PMC2728153.
52 Cahan P, Cacchiarelli D, Dunn SJ, Hemberg M, de Sousa Lopes SMC, Morris SA, et al. Computational stem cell biology: open questions and guiding principles. Cell Stem Cell. 2021;28:20-32. doi: 10.1016/j.stem.2020.12.012. PubMed PMID: 33417869; PubMed Central PMCID: PMC7799393.
53 Amit M, Itskovitz-Eldor J. Methods for the derivation of human embryonic stem cell lines. In: Atlas of Human Pluripotent Stem Cells: Derivation and Culturing. New York: Springer; 2011. p. 1-14. doi: 10.1007/978-1-61779-548-0_1. PubMed PMID: 22387411.
54 Spies NC, Rangel A, English P, Morrison M, O'Fallon B, Ng DP. Machine learning methods in clinical flow cytometry. Cancers (Basel). 2025;17:483. doi: 10.3390/cancers17030483. PubMed PMID: 39941850; PubMed Central PMCID: PMC11816334.
55 Gu Y, Zhang AC, Han Y, Li J, Chen C, Lo YH. Machine learning based real-time image-guided cell sorting and classification. Cytometry A. 2019;95:499-509. doi: 10.1002/cyto.a.23764. PubMed PMID: 30958640; PubMed Central PMCID: PMC6520141.
56 Shi Y, Inoue H, Wu JC, Yamanaka S. Induced pluripotent stem cell technology: a decade of progress. Nat Rev Drug Discov. 2017;16:115-30. doi: 10.1038/nrd.2016.245. PubMed PMID: 27980341; PubMed Central PMCID: PMC6416143.
57 Schmitz B, Radbruch A, Kümmel T, Wickenhauser C, Korb H, Hansmann M, et al. Magnetic activated cell sorting (MACS)—a new immunomagnetic method for megakaryocytic cell isolation: comparison of different separation techniques. Eur J Haematol. 1994;52:267-75. doi: 10.1111/j.1600-0609.1994.tb00095.x. PubMed PMID: 8020626.
58 Xu Y, Liu X, Cao X, Huang C, Liu E, Qian S, et al. Artificial intelligence: a powerful paradigm for scientific research. Innovation (Camb). 2021;2:100179. doi: 10.1016/j.xinn.2021.100179. PubMed PMID: 34877560; PubMed Central PMCID:

PMC8633405.
59 Ramakrishna RR, Abd Hamid Z, Zaki WMDW, Huddin AB, Mathialagan R. Stem cell imaging through convolutional neural networks: current issues and future directions in artificial intelligence technology. PeerJ. 2020;8:e10346. doi: 10.7717/peerj.10346. PubMed PMID: 33240655; PubMed Central PMCID: PMC7680049.
60 Taha M, Pappa AM, Saleh H, Alazzam A. Enhancing cell characterization with microfluidics and AI: a comprehensive review of mechanical, electrical, and hybrid techniques. Biotechnol Rep (Amst). 2025;47:e00905. doi: 10.1016/j.btre.2025.e00905. PubMed PMID: 40735651; PubMed Central PMCID: PMC12305219.
61 Wen L, Li G, Huang T, Geng W, Pei H, Yang J, et al. Single-cell technologies: from research to application. Innovation (Camb). 2022;3:100342. doi: 10.1016/j.xinn.2022.100342. PubMed PMID: 36353677; PubMed Central PMCID: PMC9637996.
62 Ahmed Z, Wan S, Zhang F, Zhong W. Artificial intelligence for omics data analysis. BMC Methods. 2024;1:4. doi: 10.1186/s44330-024-00004-5.
63 Kim M, Hong S. Integrating artificial intelligence to biomedical science: new applications for innovative stem cell research and drug development. Technologies (Basel). 2024;12:95. doi: 10.3390/technologies12070095.
64 Zare Harofte S, Soltani M, Siavashy S, Raahemifar K. Recent advances of utilizing artificial intelligence in lab on a chip for diagnosis and treatment. Small. 2022;18:e2203169. doi: 10.1002/smll.202203169. PubMed PMID: 36026569.
65 Ren F, Wei J, Chen Q, Hu M, Yu L, Mi J, et al. Artificial intelligence-driven multi-omics approaches in Alzheimer's disease: progress, challenges, and future directions. Acta Pharm Sin B. 2025. doi: 10.1016/j.apsb.2025.07.030. PubMed PMID: 41049729; PubMed Central PMCID: PMC12491694.
66 Fan K, Zhang S, Zhang Y, Lu J, Holcombe M, Zhang X. A machine learning assisted, label-free, non-invasive approach for somatic reprogramming in induced pluripotent stem cell colony formation detection and prediction. Sci Rep. 2017;7:13680. doi: 10.1038/s41598-017-13680-x. PubMed PMID: 29044152; PubMed Central PMCID: PMC5647349.
67 Mota SM, Rogers RE, Haskell AW, McNeill EP, Kaunas R, Gregory CA, et al. Automated mesenchymal stem cell segmentation and machine learning-based phenotype classification using morphometric and textural analysis. J Med Imaging (Bellingham). 2021;8:014503. doi: 10.1117/1.JMI.8.1.014503. PubMed PMID: 33542945; PubMed Central PMCID: PMC7849042.
68 Waisman A, La Greca A, Möbbs AM, Scarafía MA, Velazque NLS, Neiman G, et al. Deep learning neural networks highly predict very early onset of pluripotent stem cell differentiation. Stem Cell Reports. 2019;12:845-59. doi: 10.1016/j.stemcr.2019.02.004. PubMed PMID: 30880077; PubMed Central PMCID: PMC6449871.
69 Nie L, Yao D, Chen S, Wang J, Pan C, Wu D, et al. Directional induction of neural stem cells, a new therapy for neurodegenerative diseases and ischemic stroke. Cell Death Discov. 2023;9:215. doi: 10.1038/s41420-023-01532-9. PubMed PMID: 37393356; PubMed Central PMCID: PMC10314944.
70 Zhu Y, Huang R, Wu Z, Song S, Cheng L, Zhu R. Deep learning-based predictive identification of neural stem cell differentiation. Nat Commun. 2021;12:2305. doi: 10.1038/s41467-021-22758-0. PubMed PMID: 33972525; PubMed Central PMCID: PMC8110743.
71 Joutsijoki H, Haponen M, Rasku J, Aalto-Setälä K, Juhola M. Machine learning approach to automated quality identification of human induced pluripotent stem cell colony images. Comput Math Methods Med. 2016;2016:3091039. doi: 10.1155/2016/3091039. PubMed PMID: 27493680; PubMed Central PMCID: PMC4963598.
72 Kusumoto D, Lachmann M, Kunihiro T, Yuasa S, Kishino Y, Kimura M, et al. Automated deep learning-based system to identify endothelial cells derived from induced pluripotent stem cells. Stem Cell Reports. 2018;10:1687-95. doi: 10.1016/j.stemcr.2018.04.007. PubMed PMID: 29754958; PubMed Central PMCID: PMC5989816.
73 Orita K, Sawada K, Matsumoto N, Ikegaya Y. Machine-learning-based quality control of contractility of cultured human-induced pluripotent stem-cell-derived cardiomyocytes. Biochem Biophys Res Commun. 2020;526:751-5. doi: 10.1016/j.bbrc.2020.03.141. PubMed PMID: 32265031.
74 Guo J, Wang P, Sozen B, Qiu H, Zhu Y, Zhang X, et al. Machine learning-assisted high-content analysis of pluripotent stem cell-derived embryos in vitro. Stem Cell

Reports. 2021;16:1331-46. doi: 10.1016/j.stemcr.2021.03.018. PubMed PMID: 33891867; PubMed Central PMCID: PMC8185434.
75 Sardiu ME, Box AC, Haug JS, Washburn MP. Identification of stem cells from large cell populations with topological scoring. Mol Omics. 2021;17:59-65. doi: 10.1039/D0MO00039F. PubMed PMID: 32924050; PubMed Central PMCID: PMC9112426.
76 Sardiu ME, Florens L, Washburn MP. Generating topological protein interaction scores and data visualization with TopS. Methods. 2020;184:13-8. doi: 10.1016/j.ymeth.2019.08.010. PubMed PMID: 31476375; PubMed Central PMCID: PMC7048650.
77 Sardiu ME, Gilmore JM, Groppe BD, Dutta A, Florens L, Washburn MP. Topological scoring of protein interaction networks. Nat Commun. 2019;10:1089. doi: 10.1038/s41467-019-09123-y. PubMed PMID: 30850613; PubMed Central PMCID: PMC6408525.
78 Weber LM, Robinson MD. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. Cytometry A. 2016;89:1084-96. doi: 10.1002/cyto.a.23030. PubMed PMID: 27992111; PubMed Central PMCID: PMC7912900.
79 Guan B, Bhanu B, Theagarajan R, Liu H, Talbot P, Weng N. Human embryonic stem cell classification: random network with autoencoded feature extractor. J Biomed Opt. 2021;26:052913. doi: 10.1117/1.JBO.26.5.052913. PubMed PMID: 33928769; PubMed Central PMCID: PMC8084167.
80 Dang K, Zhang W, Jiang S, Lin X, Qian A. Application of lectin microarrays for biomarker discovery. ChemistryOpen. 2020;9:285-300. doi: 10.1002/open.201900326. PubMed PMID: 32154049; PubMed Central PMCID: PMC7050261.
81 Shibata M, Okamura K, Yura K, Umezawa A. High-precision multiclass cell classification by supervised machine learning on lectin microarray data. Regen Ther. 2020;15:195-201. doi: 10.1016/j.reth.2020.09.005. PubMed PMID: 33426219; PubMed Central PMCID: PMC7770415.
82 Stumpf PS, MacArthur BD. Machine learning of stem cell identities from single-cell expression data via regulatory network archetypes. Front Genet. 2019;10:2. doi: 10.3389/fgene.2019.00002. PubMed PMID: 30723489; PubMed Central PMCID: PMC6349820.
83 Cunha AB, Hou J, Schuelke C. Machine learning for stem cell differentiation and proliferation classification on electrical impedance spectroscopy. J Electr Bioimpedance. 2019;10:124-32. doi: 10.2478/joeb-2019-0018. PubMed PMID: 33584893; PubMed Central PMCID: PMC7851974.
84 Heylman C, Datta R, Sobrino A, George S, Gratton E. Supervised machine learning for classification of the electrophysiological effects of chronotropic drugs on human induced pluripotent stem cell-derived cardiomyocytes. PLoS One. 2015;10:e0144572. doi: 10.1371/journal.pone.0144572. PubMed PMID: 26695765; PubMed Central PMCID: PMC4690607.
85 Juhola M, Joutsijoki H, Penttinen K, Shah D, Aalto-Setälä K. On computational classification of genetic cardiac diseases applying iPSC cardiomyocytes. Comput Methods Programs Biomed. 2021;210:106367. doi: 10.1016/j.cmpb.2021.106367. PubMed PMID: 34474196.
86 Ugawa M, Kawamura Y, Toda K, Teranishi K, Morita H, Adachi H, et al. In silico-labeled ghost cytometry. Elife. 2021;10:e67660. doi: 10.7554/eLife.67660. PubMed PMID: 34930522; PubMed Central PMCID: PMC8691837.
87 Islam T, Hafiz MS, Jim JR, Kabir MM, Mridha M. A systematic review of deep learning data augmentation in medical imaging: recent advances and future research directions. Healthc Anal (Amst). 2024;5:100340. doi: 10.1016/j.health.2024.100340.
88 Coronnello C, Francipane MG. Moving towards induced pluripotent stem cell-based therapies with artificial intelligence and machine learning. Stem Cell Rev Rep. 2022;18:559-69. doi: 10.1007/s12015-021-10302-y. PubMed PMID: 34843066; PubMed Central PMCID: PMC8930923.
89 Zaman WSWK, Karman SB, Ramlan EI, Tukimin SNB, Ahmad MYB. Machine learning in stem cells research: application for biosafety and bioefficacy assessment. IEEE Access. 2021;9:25926-45. doi: 10.1109/ACCESS.2021.3056553.
90 ThinkCyte [Internet]. VisionSort. San Francisco: ThinkCyte; 2024 [c2024]. Available from: https://thinkcyte.com/visionsort/
91 NanoCellect [Internet]. WOLF G2 benchtop microfluidic cell sorter. Woburn (MA): NanoCellect; 2024 [c2024]. Available from: https://nanocellect.com/
92 Deepcell [Internet]. Deepcell platform. San Francisco: Deepcell; 2024 [c2024]. Available from: https://deepcell.com/
93 BD Biosciences [Internet]. BD FACSChorus software. San Jose (CA): BD

Biosciences; 2024 [c2024]. Available from: https://www.bdbiosciences.com/en-eu/products/software/instrument-software/bd-facschorus-software

94 Miltenyi Biotec [Internet]. MACSQuantify Tyto software. Bergisch Gladbach: Miltenyi Biotec; 2024 [c2024]. Available from: https://www.miltenyibiotec.com/BE-en/products/macsquantify-tyto-software.html

95 Somite [Internet]. Somite AI platform. 2024 [c2024]. Available from: https://somite.ai/

96 CytoRecovery [Internet]. CytoRecovery platform. 2024 [c2024]. Available from: https://cytorecovery.com/