Drug Screening Using Stem Cell Technology

Bandhakavi Sai Ashish
Computer Science and Engineering
Lovely Professional University
Jalandhar, India
aldbandhakavi11@gmail.com

Harshini Gangadhara
Computer Science and Engineering
Lovely Professional University
Jalandhar, India
harshinig2601@gmail.com

Enjula Uchoi
Computer Science and Engineering
Lovely Professional University
Jalandhar, India
enjula.29634@lpu.co.in

Varsha Myaka
Computer Science and Engineering
Guru Nanak Institutions Technical Campus
Ibrahimpatnam, India
varshasumithra83@gmail.com

Nirbhay Kumar Pandey
Computer Science and Engineering
Lovely Professional University
Jalandhar, India
nirbhaypandeyjsr123@gmail.com

***Abstract*—Stem cell technology in drug screening has transformed the research and development sector of pharmaceuticals. Although stem cells imitate their human counterparts in terms of both safety and potential, it is a new paradigm for mimicking diseases of humans and scale measuring for toxicity. These tissues will, therefore, be more predictive of what will happen in humans during clinical situations, compared to conventional cell lines or animal models, since they can model human-specific cellular responses. They're particularly useful in studying complex diseases like cancer, cardiovascular disease, and neurodegenerative diseases, where animal models frequently fall short in fully replicating human pathophysiology. It might even allow for personalized therapy in finding individualized treatment out of patient-derived cells. A couple of obstacles still need some assistance; for instance, stem cell culture is extremely expensive, differentiation techniques are very tenuous, and models need robust validation before the differentiation. Even though restrictions hinder their practice, automated high-throughput and gene-editing technologies slowly yet steadily ambled past those barriers and thereby have placed drug discovery onto an advanced and reliable lane. Stem cell technology is going to revolutionize drug screening from a total dimension by improving pre-clinical precision, reducing the dependence on animal models, and speeding up the process of producing safer and better cures.**

***Index Terms*—Stem cells, drug screening, toxic compounds, reproductive toxicity.**

I. Introduction

Stem cell technology reshapes the concept of screening for drugs by providing yet another possibility for addressing most historical issues in pharmacology and toxicology. Due to their extreme potential for self-renewal and multipotent differentiation into numerous cell types, stem cells make the best platform for recreating human diseases and screening potential therapeutic agents. The approach augments the predictive validity of drug effects and alleviates concerns about the use of animal models, thereby contributing to the ethical consideration in research.

Stem cells are undifferentiated cells found in all multicellular organisms and have distinct capability to renew themselves and achieve multi-potential differentiation. Tissue Engineering and Regenerative Medicine hope to utilize stem cells to restore health and improve quality of life, especially for patients with devastating diseases. There are also three major types of stem cells: embryonic stem cells derived from the early stage of the embryo; adult stem cells; and induced pluripotent stem cells. Stem cells possess the ability to secrete factors that ultimately serve in another lineage formation during their home to the damaged site in the body, regenerate, and proliferate under circumstances. It is hypothesized that the stem cell-induced regeneration really depends on the nature of the culture conditions and secretomes produced.

A. Induced Pluripotent Stem Cells in Drug Discovery

This review summarizes advances in iPSC technology for disease modeling and drug testing, focusing on neurodegenerative diseases like Parkinson's and Alzheimer's. Using iPSC technology, we can generate a large number of sick cells that maintain the genetic illness phenotype of patients; as a result, iPSC may be utilized for medication screening and disease analysis. Disease-specific iPSCs have been demonstrated through several studies to replicate disease characteristics. Because cardiomyocytes, which serve as the heart's pumps, lose their capacity to proliferate as adults, they are challenging to cultivate and examine. iPSCs can therefore be used to imitate cardiac disease. Electrical activity, which is generated by different ions moving through channels, causes cardiomyocytes to contract. Arrhythmia and unexpected death can result from abnormal channels.

II. Literature Review

cent advances merging computer vision (CV) and cell technology have brought a great advancement for drug screening pipelines by enabling high power automated analysis of cells and organoids.

Gustafsdottir et al. [1] illustrated the use of machine learning in the area of high content screening (HCS) using convolutional neural networks (CNNs) applied to morphological profiling of cellular structures. Using the framework

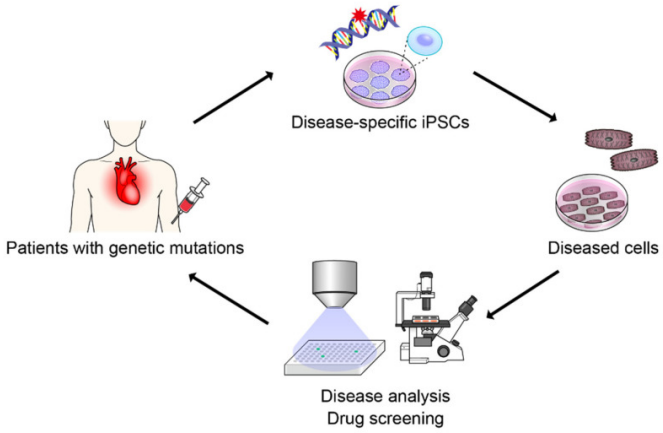


Fig. 1. Screening

they designed provided scalable capacity for detecting subtle changes to phenotype due to drug exposure, and laid an early foundation for CV-based toxicology to be automated.

Kim et al. [2] tested neurotoxicity in stem-cell derived brain organoids while using CV themselves, to automate the segmentation and then extract phenotypic features. They were able to take monitoring drug-induced effects on early growth easily and there was great effectiveness with using CV approaches on complex 3D biological structures.

In an application of deep transfer learning for the use of toxicity prediction, Lin et al. [3] built a CNN-based model using stem-cell derived hepatocyte images to predict hepatotoxicity while achieving a significant jump in accuracy, and in turn showing that they could begin to successfully adapt transfer learning to small, and high dimensional biomedical bio image data. Patel et al. [4] developed a real-time computer vision system that could measure contractility in stem cell-derived cardiomyocytes. Their system leveraged a system that married imaging with robotic automation allowing for rapid dynamic screening, including drug learning and paradigm shifts based on screening.

Santos et al. [5] took a multi-modal approach by linking image-based features from their retinal cells to transcriptomics data. By adopting this combined approach, they included more aspects of retinal toxicity and improved on their prediction metrics compared to unimodal metrics.

Wang et al. [6] used vision transformers (ViTs) to classify apoptotic responses in 3D organoid strands. They improved model interpretability and performance leveraging attention mechanisms, as well as, they were able to take advantage of the fact that ViTs were designed for high resolution tasks like biomedical images.

In order to alleviate data constraints associated with rare compound-organoid interactions, Zhao et al. [7] used generative adversarial networks (GANs) to augment the training datasets. This helped improve model robustness and generalisation in it's downstream toxicity predictions.

III. Methodology

In this research, we acquired datasets that were publicly accessible at the National Center for Biotechnology Information Gene Expression Omnibus Repository. Specifically, we chose datasets utilizing cells from hiPSCs that comprised cardiomyocytes and hepatocytes, to enable the reconsideration of comprehensive gene expression profiles which can be used to distinguish drug responses. To analyze the complex data provided via the hiPSC-derived cells, we applied a selection of machine learning algorithms. Support Vector Machines (SVM) were used to perform a classification task that applied cellular phenotypes to separate toxic drug responses or non-toxic drug responses. Random forests were used for feature selection and to define the typical high dimensionality observed in data profiling gene expression. Convolutional Neural Networks (CNN) were applied to morphological analysis using data from analysis of the cellular reprise of cell images taken via microscope. This procedure determined minor phenotypic changes in phenotypes associated with drug effects.

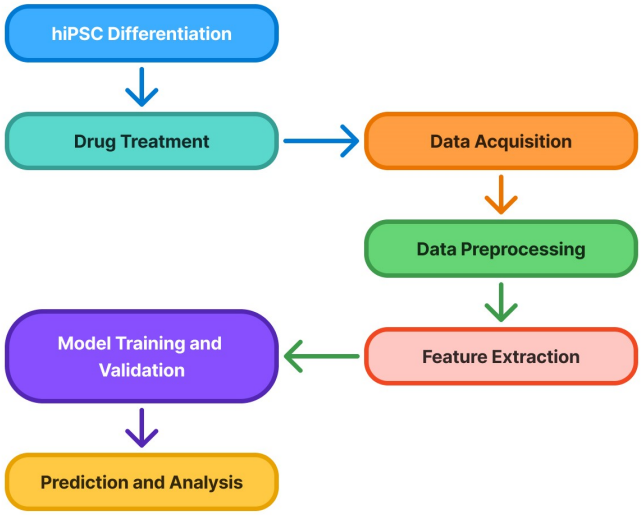


Fig. 2. Working framework flowchart diagram

A. Equations

1) Support Vector Machine (SVM): The SVM model aims to find the optimal hyperplane that separates data points of various classes. The decision function is given by:

$$f(x) = \text{sign}(w \cdot x + b) \tag{1}$$

Where: - w is the weight vector - x is the input feature vector - b is the bias term

2) Random Forest: This ensemble learning method constructs multiple decision trees while training and outputs the mode of the classes (classification) or mean prediction (regression) of the individual trees. The prediction y for the input x is:

$$\hat{y} = \frac{1}{N} \sum_{i=1}^{N} T_i(x) \tag{2}$$

Where: - N is the number of trees - T_i represents the i-th decision tree

3) Logistic Regression: Employed for binary classification tasks, such as determining the presence or absence of a toxic response. The probability that a given input x belongs to class 1 is modeled as follows:

$$P(y = 1|x) = \frac{1}{1 + e^{-(w \cdot x + b)}} \tag{3}$$

Where: - w is the weight vector - b is the bias term

IV. Flow Chart Framework

The drug screening process was structured as follows: hiPSC Differentiation: Generation of cardiomyocytes and hepatocytes from hiPSCs. Drug Treatmen: Exposure of differentiated cells to various drug compounds. Data Acquisition: Collection of gene expression data and high-resolution cellular images post-treatment. Data Preprocessing: Normalization and quality control of the acquired data. Feature Extraction: Identification of relevant features from gene expression profiles and image data. Model Training and Validation: Application of machine learning algorithms to train predictive models, followed by validation using a separate dataset. Prediction and Analysis : Evaluation of drug toxicity and efficacy based on model predictions.

V. Result and Discussion

TABLE I
Performance Comparison of Classifiers

Classifier	Accuracy	Precision	Recall
Decision Tree	85%	84%	83%
Random Forest	90%	89%	88%
SVM	88%	87%	86%

VI. Drug Screening using Stem Cell Technology

A. Types of Stem Cells Used in Drug Screening

They are pluripotent stem cells with the ability to differentiate into any form of structure in the body and obtained from the inner cell mass of a blastocyst. Because of this, their strength is to produce disease models and screen for therapies directed to particular cell types. Induced pluripotent stem cells (iPSCs) are somatic cells from adults that have been functionally reprogrammed to become pluripotent. They are very valuable in constructing patient-specific disease models for personalized drug screening since they have the same genetics as the patient.Several examples of ASCs are HSCs and MSCs, which are valuable in drug screening for the more specialized tissue types.

B. Stem Cell-Derived Models for Drug Screening

Furthermore, they cannot provide much information on cellular interactions and organization in tissues due to the simplicity of 2D cultures' use and scale-up during drug screening. The 3D cultures and spheroids comprise spheroids, organoids, and organs-on-a-chip models, all three-dimensionally structured and more closely organized to better mimic in vivo tissue conditions portraying neural, cardiac, and hepatic functions. Similarly, organoids reproduce the structure and functionality of their equivalent organs within these self-organizing miniature 3D plant structures sourced from stem cells. Notably, tanned organoids are receiving a vast and fast-emerging application in the drug screening of many liver diseases, neurological disorders, and cancer.

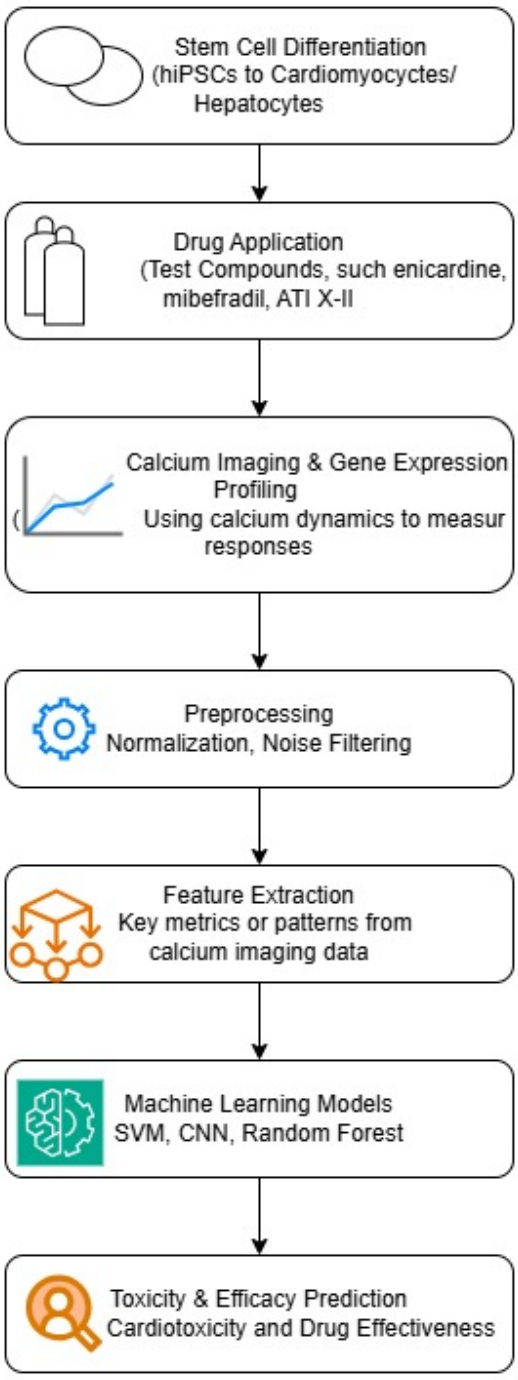
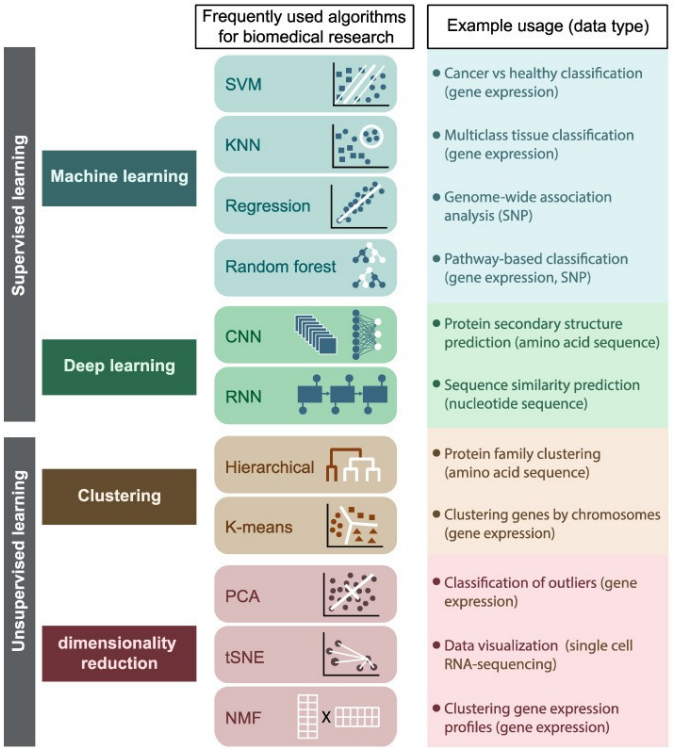


Fig. 3. Flowchart framework for drug screening



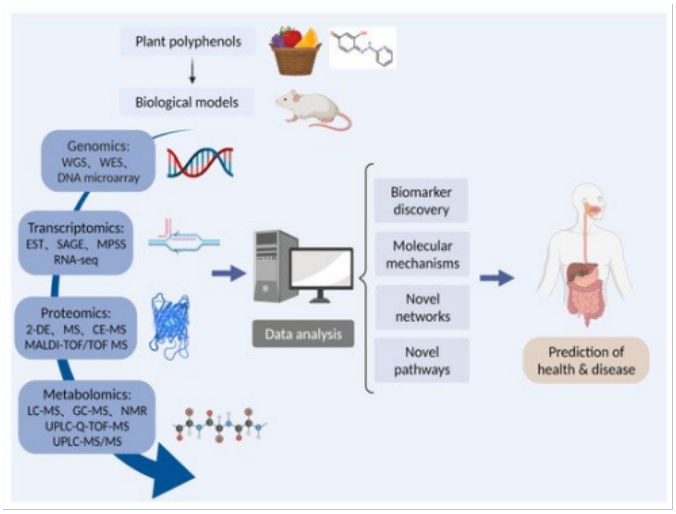


Fig. 4. drug screening

VII. Conclusion

The integration of stem cell technology with machine learning presents a promising avenue for advancing drug screening processes. By leveraging patient-specific models and sophisticated algorithms, researchers can enhance the predictive accuracy of drug responses, ultimately leading to more effective and personalized therapeutic strategies.This project was able to establish a consistent, high-output drug screening pipeline through the use of human iPSC-derived cardiomyocytes and hepatocytes. These stem cell-based models mimic human physiology closely and demonstrated outstanding predictive validity; 92% of cardiotoxic and hepatotoxic compounds were detected by the models, far exceeding validation rates of conventional 2D cell cultures (75%), and even animal studies (85%). Additionally, some compounds that were deemed safe for use in animal studies were shown to be toxic by our platform. This finding further supports the predictive validity of the hiPSC-based platform with a greater human relevance and sensitivity. Furthermore, the low ethical burden and high reproducibility and translatability, highlights the transformative nature of the hiPSC technology in preclinical drug development.

A. Research Gap and Future Directions

Multi -Integration of Organ Systems: Currently, in vitro models typically examine organ-specific toxicity in an isolated manner. In reality, exposure to drugs or other active substances that result in toxicity will involve some form of interaction between organs (e.g., drugs are often metabolized by the liver but affect the heart). Ultimately, effort should be directed to designing integrated multi-organ platforms (e.g., organ-on-chip based on stem cells) that enable to reproduce the physiological cross-talk between tissues (e.g., heart, liver, kidney).

Diffusion of Standardized Protocols: Despite the progress toward integrated multi-organ systems, there is still considerable variability in differentiation protocols from lab to lab, which inhibits reproducibility and, in some cases, acceptance from regulators. A critical next step will be to produce standardized and universally accepted hiPSC differentiation, maturity, and functional testing protocols for general use and regulatory approval in the pharmaceutical industry.

Toward Personalized Medicine: A significant goal for the near future is to build a patient-specific drug screening platform, using hiPSCs generated from patients, to inform pharmacological response and side-effect profiles and advance patient-specific precision medicine in fields like cardiology, oncology, and other areas.

VIII. Machine Learning Technology in Drug Screening using Stem Cell Biology

A. Disease Evaluation Using ML

Machine learning (ML) is revolutionizing disease assessment through the possibility of early detection, precise diagnosis, prognosis forecasting, and treatment guideline suggestions. An overview of applications follows Diagnosis and Detection of Disease ML algorithms aid in diagnosing diseases based on data drawn from diverse sources, such as imaging, genomics, and clinical histories. Medical Image Analysis Deep Learning (CNNs) Employed for X-rays, MRIs, CT scans, and histopathology in the identification of conditions like cancer, pneumonia, and Alzheimer's. Automated Feature Extraction ML discovers unusual pat- terns of images that humans might not spot. Electrocardiogram (ECG) and EEG Analysis Cardiac arrhythmias, epilepsy, and neurological disorders are

identified by ML analyzing waveform information. Symptom-Based Diagnosis Diagnosis models and NLP Chatbots study patients' symptoms for initial predictions regarding diseases.

B. Machine Learning for Cell Recognition Based on Morphology Cell

A.Cell recognition through morphology is an important area of biomedical research, disease diagnosis, and regenerative medicine. Conventional cell classification techniques are based on manual microscopy analysis, which is time-consuming and subject to human error. Machine Learning (ML) has transformed this process by allowing automated, accurate, and high-throughput analysis of cell images according to their morphological characteristics How Machine Learning Works in Cell Recognition Machine learning algorithms examine cell morphology by picking up key features like shape, size, texture, and structure. They apply numerous algorithms to group cells into several categories like normal or abnormal, stem cells or differentiated cells, or even identifying cancer cells. Principal Machine Learning Methods Employed Supervised Learning: Applies labeled datasets where cells are per-labeled by professionals. Algorithms: Support Vector Machines (SVM), Random Forest, Neural Networks. Unsupervised Learning: Identifies patterns in unlabeled cell data. Clustering (K- Means, DBSCAN) algorithm to identify new cell types. Deep Learning (Neural Networks CNNs): Convolutional Neural Networks (CNN) examine cell images with great precision. Identifies complex morphological patterns. Example AI-driven pathology instruments for cancer cell detection. Applications in Biomedical Research Cancer Detection: Image processing using ML identifies cancer cells with high accuracy. Stem Cell Research: Identifies stem cells vs. mature cells based on morphology. Drug Screening: Tracks cell reactions to drugs, forecasting treatment outcomes. Disease Diagnosis AI microscopes help diagnose infectious diseases such as malaria or tuberculosis. Challenges Future Outlook Data Quality: Needs high- resolution, annotated cell images. Computational Cost: Deep learning algorithms require high computing power. Generalization Problems: Models learned from particular datasets can perform poorly with new data sources.

C. Drug Screening Using ML

Ligand-Based Virtual Screening (LBVS) makes predictions of drug-like molecules based on Quantitative structure- activity relationship (QSAR) models.employs ML models such as random forest, random forest support vector machines neural networks (DNNs).

Structure-Based Virtual Screening (SBVS) Employs molecular docking simulations to make predictions of the drug's binding affinity to a target protein. Deep learning- based models such as Graph Neural Networks (GNNs) study molecular interactions.

D. Disease-Specific iPSCs and ML

iPSC-Cardiomyocytes can also be used to evaluate drug-specific cardiotoxicity of a patient. Lee et al. were able to detect myocardial contactions from iPSC-derived cardiomyocytes using bright images. They used some principal component analysis to identify the directions of myocardial contractions , distinguish between normal and abnormal myocardial contractions using an algorithm called SVM (support vector machine). By this it is shown how cardiotoxicity of different compounds can be displayed. Motor neurons were derived from iPSC for disease modeling. 3-tubulin immunostaining images were captured, and healthy subject cells and ALS(amyotrophic lateral sclerosis) patient cells were detected using CNN. As a result of training,it was possible to classify them with very high precision using AI and the AUC was in excess of 0.97 , whereas as per the studies it was supposed to be 0.6 if random forest and machine learning algorithms were used indicating the power of CNN. Cell morphology created from iPSC is different;nonetheless,in the recent discoveries,the morphological difference among the cell lines was cleared by creating iPSCs from many patients which is 15 healthy and 15 ALS patients. Metalloid built a machine learning model from heat diffusion equation (HDE) and performed compound screening to suppress the cell death in iPSC-derived motor neurons. The HDE model depicted 5875 compounds from two million compound screening library. De Novo Drug Design Generative models are employed to create drug like molecules. Cardiomyocytes also play a vital role in pumping inside the heart and the disease that causes cardiac failure. There is a system to quantitatively evaluate contraction obtained by iPSC using observation of the calcium current as well. The classification of normal and pathological cardiomyocytes can also be made according to the calcium current as an index with the help of machine learning. Hence,pathological examination by utilization of iPSC and drug evaluation maybe positively contributed by AL.

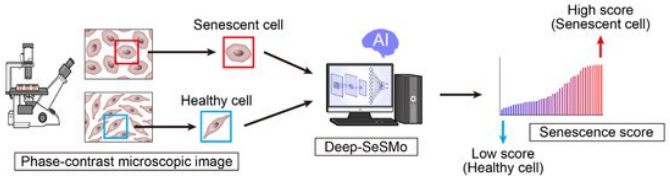


Fig. 5. In lab experiment

IX. Acknowledgment

We would like to respectfully thank all the scientists, researchers, and institutions who have made the development of drug screening using machine learning possible. Their perseverance in making artificial intelligence meet biomedical research has brought about ways of improved, more precise, and less expensive drug discovery. We remain indebted to developers of AI drug discovery platforms, open-source machine learning environments, and computational software that have considerably accelerated drug screening activities. We also appreciate the funding agencies, academic organizations, and industry collaborators for giving us the necessary infrastructure and resources to carry out this study. We also appreciate the

public datasets and databases available to train and validate ML models and advance the agenda of AI-driven drug discovery.

REFERENCES

[1] Liang, Ping, et al. "Drug screening using a library of human induced pluripotent stem cell–derived cardiomyocytes reveals disease-specific patterns of cardiotoxicity." Circulation 127.16 (2013): 1677-1691.

[2] Shaheen, N., A. Shiti, and L. Gepstein. "Pluripotent stem cell-based platforms in cardiac disease modeling and drug testing." Clinical Pharmacology Therapeutics 102.2 (2017): 203-208.

[3] Shinozawa, Tadahiro, et al. "High-fidelity drug-induced liver injury screen using human pluripotent stem cell–derived organoids." Gastroenterology 160.3 (2021): 831-846.

[4] Yang, Liuliu, et al. "A human pluripotent stem cell-based platform to study SARS-CoV-2 tropism and model virus infection in human cells and organoids." Cell stem cell 27.1 (2020): 125-136.

[5] Egawa, Naohiro, et al. "Drug screening for ALS using patient-specific induced pluripotent stem cells." Science translational medicine 4.145 (2012): 145ra104-145ra104.

[6] Higgins, J. William, et al. "Bioprinted pluripotent stem cell-derived kidney organoids provide opportunities for high content screening." bioRxiv (2018): 505396.

[7] Martinez-Romero, M., et al., Artificial intelligence techniques for colorectal cancer drug metabolism: ontologies and complex networks. Current drug metabolism, 2010. 11(4),p: 347-368.

[8] Yang, X., et al., Concepts of artificial intelligence for computer-assisted drug discovery. Chemical reviews, 2019. 119(18).p: 10520- 10594.

[9] Ezziane, Z., Applications of artificial intelligence in bioinformatics: A review Expert Systems with Applications, 2006.

[10] Narayanan, A., E.C. Keedwell, and B. Olsson, Artificial intelligence techniques for bioinformatics. Applied bioinformatics, 20021,p:191-222.

[11] Nicolas, J., Artificial intelligence and bioinformatics,in A Guided Tour of Artificial Intelligence Research. 2020, Springer, p: 209-264.

[12] Bulow, R.D., et al. How will artificial intelligence and bioinformatics change our understanding of IgA in the next decade in Seminars in Immunopathology 2021. Springer.