# Automated Plate Analysis and Hit Confirmation Intelligence: Machine Learning Architectures for High-Throughput Screening Workflow Optimisation

*Dr. Sunita Singh, Associate Professor of Computer Science, Indian Institute of Technology Delhi (IIT Delhi)*

**1. Introduction to High-Throughput Screening (HTS)**

High-throughput screening (HTS) is a drug discovery process typically conducted at both the academic and industrial levels. It is employed during early discovery stages of drug development to validate bioactivity leading to the development of a clinical candidate as part of a disease treatment. Globalized drug discovery requiring the screening of vast chemical libraries demands such cost- and time-saving methods as HTS. Traditional HTS relies on the careful design of assays that facilitate the rapid screening of thousands of substances in a relatively short period of time, either to validate achievements post-synthesis or to identify potential hits contributing to the discovery of a lead structure.

High-throughput screening (HTS) for drug discovery was developed to facilitate researchers in the testing of large chemical collections of potential drug compounds. This is a way to carry out many biochemical, genetic, or pharmacological tests in the research and diagnostic field. HTS studies collect and generate a vast amount of data that needs to be validated and integrated into relevant biological contexts to provide insights for decision-making. HTS for the discovery of inhibitors of interactions associated with neurodegenerative diseases, HTS in cancer research to identify modulators of disease phenotypes, and HTS of natural products to identify drug-like molecules for therapeutic applications are a few examples of its use. However, current standard methods of HTS discovery are plagued by many limitations.

### 1.1. Definition and Importance of HTS

A considerable amount of research is based on data from high-throughput screening platforms. HTS is an automated process that allows large-scale screening of chemical, genetic, and pharmacological assays. The general idea behind imaging-based HTS is to observe and document changes between samples treated with chemical compounds or small interfering RNAs against RNAi targets and control samples that have not been treated. The readout of the experiments can be either cellular phenotypes or reporter assays in which proteins are tagged to a reporter protein detectable by fluorescent, luminescent, or colorimetric measurements. Observing changes in a reporter gene linked to a biological signaling event is very versatile and allows the assay of countless signaling pathways. Importantly, HTS allows hundreds to thousands of chemical compounds to be tested in one single experiment.

This simultaneous measurement of multiple samples drastically reduces the time requirement for screening libraries containing tens or hundreds of thousands of chemical compounds. The time and cost for developing a new drug are many-fold; a commonly used figure is $1,000 million, and between six and twelve years are required to get a new drug to market. HTS promises to tackle this problem. The Open Innovation initiative centers on revising the HTS process and already has several case studies with favorable results. Part of this process is integrating artificial intelligence, machine learning, and neural networks in analyzing biological data.

### 1.2. Challenges in Traditional HTS Methods

In spite of the technological advancements, HTS has multiple limitations when it comes to high-throughput treatment screening assays. A major limitation observed in the currently practiced HTS assays is a high hit rate, consisting of a large number of false positives and negatives. Such frequently unreliable assays are not suitable for drug or treatment screening because of their unacceptable assay quality. Both high false positive and high false negative rates require additional investigational costs in order to identify and analyze these compounds. Second, the traditional HTS methods are labor- and resource-intensive and do not provide additional metabolic and toxicity information. Third, manual activities in sample processing and assay result evaluation reduce the analysis speed, which creates a bottleneck in the processing of vast numbers of samples.

Fourth, the generation of a huge amount of data from discovery research has created a challenge in the management and standardization of this data.

Data collection methods and practices used in pharmaceutical companies are not standardized, which hinders the possibility of exchanging data between partners. Collecting a huge amount of data from compound evaluation becomes worthless if proper methods of data analysis are not available. Also, inhibition testing in microfluidic devices can offer similar solutions and advancements for high-throughput drug screening processes, with detailed drug evaluations in multiple organ-on-a-chip platforms, and can be expanded to assess drug-drug interaction studies. In terms of data, new challenges of data integration and analysis evolve and have evolved with recent data generation methodologies capable of monitoring, testing, and developing genome-scale metabolic assessments. Thus, there is an increasing need for automated screening methods such as biological assays for high-throughput drug and chemical testing, which are generally based on commercially available fluorogenic and bioluminescent substrates utilized in the horseradish peroxidase-based method. The challenge with these assays is that such products are often less sensitive and have extremely high background signals at a high substrate concentration in cell-based assays.

**2. Role of Artificial Intelligence in HTS**

The research field of artificial intelligence (AI) and machine learning (ML) has been rapidly developing. Machine learning is one of the subfields of artificial intelligence in the field of computer science. When studied independently, machine learning has been a part of AI research. Artificial intelligence and machine learning are implemented as powerful tools in a high-throughput screening (HTS) system. The implementation of AI and ML has revolutionized the drug discovery process by making it possible to improve assay design, results, and hit identification. The applications of AI/ML cover different aspects of HTS. They can be used to tackle the inadequacy of data in creating cumulative distributions or simulations for structure-activity relationships, improving the information harvested from dose-response curves, spending more efficient time on follow-up analysis through a refined hit calling process, and accurate molecular fingerprinting to flag and exclude compounds with assay artifacts.

One of the applications for artificial intelligence (AI) and machine learning (ML) is the innovation of HTS assay design that takes into account the current information and parameters required, with new ideas incorporated into the experimental design to improve the quality of the results. It also employs genomic technologies to analyze existing data and provide new insights, increasing the accuracy of computational predictive models. These applications have great potential in improving HTS workflow. AI and ML are valuable tools for the analysis of HTS data, which has provided an enhancement over present analysis approaches. Additionally, AI/ML can be used to predict hit compounds from HTS that can identify active and inactive compounds in a limited experiment. The applications of AI/ML in these aspects have shown that AI and ML can address the limitations of the current state-of-the-art HTS methods. AI and ML can be used in a variety of HTS methods. AI and ML, along with other screening technologies, should perform better than screening technologies alone if they refer to the same data.

**2.1. Overview of AI and Machine Learning**

Artificial intelligence (AI) has recently become an attractive technological approach for a broad range of applications, including scientific and biomedical inquiries. Within the field of AI, machine learning is a subfield that focuses on developing computational systems that automatically improve performance over time. This is achieved through the use of appropriate algorithms and architectures that can acquire applicable knowledge and utilize the learned information for the prediction of previously unseen instances. Thus, AI and machine learning are characterized by their ability to perform decision-making, generate insights based on the data, and automate the analysis of large and complex datasets. Among the various types of machine learning, supervised learning requires the model to be trained on an input data matrix that contains both the applied features and the appropriate output labels. The feature matrix and output variable(s) are continuously adjusted until the model is able to generalize the data without the need for further training. The aim is to achieve correct predictions when the input is altered. Unsupervised learning, however, is used to analyze datasets that do not have an output label; this includes the direct application of AI to high-content analysis images. Another machine learning approach is reinforcement learning, where the AI model learns to make decisions. In the context of scientific application, the supervised and unsupervised approaches can be linked to regression and clustering respectively, whereas

reinforcement learning can be employed in drug discovery strategies as it seeks to identify the optimal selection of lead compounds within the initial pipeline. The broader term 'data-driven AI' is given to approaches that involve the computer learning from and applying the knowledge of a large dataset to generate desired outcomes and predictions in previously unseen datasets. In the field of high-throughput screening (HTS), this effect has been leveraged to omit the need for rule-based targeting. Hence, the outcome is a more predictable and accurate set of predictions. In the following discussion, we will refer to 'AI' implicitly as a data-driven tool. The capability of AI to capture deeper insights from the data and detect subtle trends makes it a unique tool for analyzing datasets that are large and complex, dense and noisy, and contain an inherent degree of redundancy and multicollinearity. One of the main beneficiaries of AI today is high-throughput screening (HTS). Within the HTS industry, it is common to identify diverse sets of chemical compounds potent in a biological assay using very large libraries via various techniques. The approaches used could be target-specific, which are efficacious but lack rescue drugs or broad-target ligands. The search for new paradigms that can connect unidentified pharmacological mechanisms to diseases using novel techniques has turned to large-scale laboratory assays. Such techniques are made faster and safer by AI. To be effective, this AI approach will need to filter out a large variety of toxic and maladapted ligands. This aspect of AI not only makes the HTS experimentation efficient by omitting substantial amounts of ligand scrutiny but also reduces subjectivity and bias, increases decision-making and precision, and decreases time and resources simultaneously. Furthermore, AI and technology have a much lower error rate, particularly in large-scale, repetition-driven scans clearly influenced by human error. In developing drugs at a lesser expense, having a technique that is significantly less demanding on the budget is advantageous. Given the significant potential in total new drug growth, AI-driven HTS is recommended for subsequent therapeutic advances. Also, the field of biological high-content research has pushed the limits of AI. Furthermore, the inclusion of AI and technology has shown benefits of even more efficient analyses of large and more transcendent data flow. AI has come to be one of the current revolutions of science.

### 2.2. Applications of AI in Assay Design

2.2. Applications of AI in Assay Design. In drug discovery, AI has proved invaluable for developing the next generation’s most innovative biological assays. The assay design, or

the primary phase of high-throughput screening, is predicated on the experimental choice and modulation that is anticipated. At a system level, AI can assist in finding hits that down-modulate certain facets of cell-based assays. AI algorithms significantly reduce the time and effort involved in selecting relevant targets by predicting both a link to compound series and likely values, and can also reduce assay failure rates. While the linker choice during screening can differ depending on whether biochemical assays or cell-based assays are used primarily, it is important to keep this part of the assay selection routine scientific. Currently, AI is only used in the primary phase.

Assay development itself can include the creation of a systems biology model to predict the interactions seen in our multi-compartmental molecular assay and the standardization of primary data in order to build computational models for translation exercises to humans. AI-based design can reduce time at the bench by interpolating missing data points and choosing the nearest compound available to proceed. The automation of the design and assay writing process is also being handled by various organizations, reducing the need to manually describe commonly performed tasks, and thus making the life of the person more creative and investigative in the result analysis. Large and publicly available curated datasets can frequently be used for training. Further research has shown the rational application of a variety of machine learning approaches on data, which have been instrumental in building multi-gene molecular models and the discovery of novel, as yet undescribed activities. Values can all be systems-model flow predictions of compounds at their intracellular targets. While predicting systems efficacy is a useful validation of the human biology systems models, since you need to predict the pharmacokinetics-pharmacodynamics effect in the compound plan, the main applications are in the translation step.

### 2.3. Applications of AI in Data Analysis

HTS generates a large dataset, whereas this data can require complex experience design, multitasking, data format, data analysis, and management. In the last stage, unsuitable data analysis affects the performance of HTS, and traditional methods require much time and effort to accomplish the analysis. In general, traditional data analysis results in treating events based on the average of data points; these averages are sometimes obtained through data clustering or data segmentation. Consequently, the patient experiences of the formulation should be analyzed to respect what has just happened in

these experiences (both being linked in terms of statistical hypothesis as causal). In that circumstance, the choice of patients reproduces at the discovery stage, on one hand, and a lot of entangled effects should be controlled and examined completely. Moreover, the data are large enough that the error propagation can be managed in an industrial approach to drug development.

In this framework, artificial intelligence and machine learning in particular have their benefits. First of all, machine learning can handle big data and has the ability to learn from a large amount of data with various formats. Machine learning also has empowered the development of a high-throughput decision support system, making the analysis of HTS very easy and supporting the estimation of molecular activities. Then, if the machine finds an uncommon pattern, it can promptly alert the user. Furthermore, it can predict the expected result and set the robots to the sequence list for the next task in the laboratory. Today, machine learning is built indifferently for toddlers in neural networks, forming an artificial neural network. Mean squared error and checking the null hypothesis, i.e., that the error is normally distributed with a mean of zero, should be true. Consequently, multiple machine analysts pose some prediction errors to ensure whether their predictions are similar to the chemical supervisor of the clinical trial. Overall, machine learning increases accuracy and decreases the error of the resulting experience. The experience cost term is correlated with such a decrease in prediction errors. For instance, an artificial neural network can filter milligrams of samples quickly as set by using the predictions from machine learning, including the interest of a model called the random forest. The random forest quickly estimates the quantity of the same volume, but also reduces dimensions in various experiments.

### 2.4. Applications of AI in Hit Identification

Identifying hits has long been recognized as one of the most crucial steps in high-throughput screening (HTS) campaigns as it serves as a starting point for the whole drug discovery process. An identified hit is a chemical compound that shows promising in vitro activity against a given target and subsequently is considered to have potential as a starting point for further optimization. Traditional hit identification methods can generate an overwhelming number of hits that cannot all be progressed in the drug discovery process due to practical reasons, especially time and resources. Thus, it is important for researchers to identify the most promising compounds as early in the

discovery process as possible. It is a significant challenge for most rational drug discovery pipelines to have the capacity to fully explore the potential differences between compounds in a practically sized library and, in turn, typically need some experimental evidence before such research capacity can be invested to take place. In this context, there is a growing tide of opinions that high-quality screening hits are increasingly important. This has a carrying signal for HTS. Amidst the complexities and uncertainties of drug discovery, the holy grail of true qualitative hit minimization, not as a percentage of all the hits, is something that many drug hunters aspire to.

AI in the field of machine learning and, more specifically, in deep learning is promising as it can analyze datasets swiftly and get insights that are difficult to spot, for example, directly utilizing a predictive model instead of using experimental data from HTS alone. There are currently 53 companies and academic institutes that develop artificial intelligence techniques to discover and develop new drugs. Additionally, the number of articles discussing applications of artificial intelligence in drug discovery is expected to surpass the cumulative research published so far by the end of this year. Artificial intelligence can also be utilized with traditional read-write methods to harmonize HTS output. For instance, links may be made with existing pre-competitive knowledge and other data. Screening AI can therefore be a development, yet it will further develop the HTS workflow as a whole. A high-contrast observational overview of the HTS data and their relationship can be provided, rather than seeing data through the curtain of bargain modeling. All AI types are exclusively depicted to encourage meaningful comparisons to other AI modalities. To enable hit identification in a classical HTS scenario, AI instruments can be grouped into various sub-categories. That being stated, they produce a typical outcome. More candidates are identified, i.e., Pearson's correlation between experimental activity and model prediction, where the model prediction is made on the basis of the characteristic description of the drug, the basis to identify candidates in shape-based strategies of the inherent properties of the compounds and the drug target, following in silico methods that directly utilize generates of HTS runs that demonstrate an excessive rate of false positives. We hold that the existing methods can be used to address key drug-discovery demands. With the cautious explication of the contrast in the results for each traversed method, we illustrate our evaluation.

## 3. Case Studies and Examples of AI-Enhanced HTS

Achieving the quantitative and kinetic analysis of larger sample collections makes high-throughput screening a keystone to scientific discovery in pharmaceutical industries. Incorporating AI technologies is expected to truly pave the way for revolutionary developments in increasing the throughput and accuracy of high-throughput screening. In this manuscript, three case studies are reported and described in detail. The purpose of these case studies is to exhibit innovation in the high-throughput screening workflow as a result of AI solutions. The case studies are split categorically into two examples of high-throughput screening system integrations and one case study to illustrate the compatibility of AI systems with existing platforms.

Three major case studies illustrate the impact of employing AI technologies in high-throughput screening processes. The first is the application of AI models to a remote workflow, outputting a great enhancement of analysis speed and an increase in throughput as a result. Additionally, two internal collaborations instigated the designing and managing of complete robotic workflows using entirely integrated AI software stacks. Both instances delivered notable improvement and the expansion of new processing capabilities, with promising outlooks for their potential application to enhancing drug discovery rates. These investigations present novel applications of AI to impact high-throughput screening, with exciting potential for future advances. The introduction of AI models for high-throughput screening analysis has already accelerated drug candidate screening, with more data being processed each year than in the previous eight years combined.

### 3.1. Successful Implementations in Pharmaceutical Industry

The possibilities of artificial intelligence have been realized as early as the first day of the sequencing of the human genome. The work process—screening of a compound library against a specific drug target—lacked innovation for nearly a century. High-Throughput Screening (HTS) was the revolution that was needed and is considered as one of the most important advances in pharmaceutical research in the past 20 years. From the computational field perspective, HTS is mainly driven and defined by a data management process, so it was obvious to try to automate it through computer screens. Up to this day, the full scope of integration of artificial intelligence in laboratory automation is just a matter of time.

Almost two decades ago, a Brazilian research group used artificial intelligence to analyze HTS data. One of their studies used student-like decision trees, and the other employed a neural network to statistically classify dilution response patterns into six compound activity classes. These tools were robust enough to cluster multiple causal patterns in a dataset, as well as to reveal some patterns at the limit of significance. Predictions could serve as complementary information for chemists making decisions on which compounds to discard. In 2008, machine learning ensemble classifiers were used to remove false positives from hit lists. Each classifier was trained on the loci descriptor of compounds that had an activity spectrum resembling a known inhibitor. Collaborations and integrative community events are key to this successful phase. Isn't it an added value that distinct fields collaborated so that new HTS strategies could emerge? Data, previously used in artificial intelligence programs, has shown an association between the efficiency of the computer tool and the contextual relevance of the data when making a specific drug discovery decision. Therefore, it is interesting to review how the computational tools currently being developed are being designed and what types of data are being used as input so that we can have an overview of the opportunities and advantages of implementing a more sophisticated generation of screening computational tools. The environment for chromosomal-scale genome sequencing of the matocarpus and proud alga appeared without citation and reference to the article and its author(s). We apologize for the authors who should have been cited in order to support the discussion.

### 3.2. Impact of AI on Throughput and Accuracy

With a plethora of publications and white papers avowing the benefits of AI technologies in HTS, one would suspect that there is ample evidence supporting integrated AI leads to improved high-throughput capacity and saves time in drug discovery. Surprisingly, actual discussions of impact are relatively few. It might be imagined that AI will speed up drug discovery as it can foresee how to optimize complex biological systems. The ability of an integrated AI to generate data patterns resulted in hit rates of up to 25%, whereas the traditional process delivered only about a 10% hit rate. In another example, the AI process might deliver only half as many hits as the traditional process, but these hits might be twelve times more likely to show clinical utility. This leads to up to 24-fold better cost-per-treatment values in terms of the number of animals used or dollars spent on a hit-to-lead campaign, the not-so-easily

quantifiable cost of compounds discarded in these follow-through assays, as well as shortening the time to hit identification. AI results in a seven times improvement in the rate to find the right up-titrations, ten times faster at finding the most active ligands, and does not miss as many low-dose hits as traditional HTS does, even when analyzing only 20% of the dose-effect data. In tyramide amplification, the AI-enhanced process is six times faster at finding the active compounds that affect TNF alpha. A cell-based or whole-organism measurement is necessary to know that the quiescence is due to true antimalarial activity, and for speed, one cannot waste time determining causes of quiescence. Thus, a screen hit rate in a simple biochemical screen is expected to be 70% before sending for off-target analysis.

**4. Challenges and Limitations of AI in HTS**

Challenges

The implementation of AI in HTS workflows is linked to a few challenges and limitations. One of the main entry barriers is the lack of high-quality data. AI tools require a large and diverse set of data to perform robustly. Thus, data quantity and quality are decisive for AI performance, and low sample sizes can lead to biased results. Additionally, HTS is based on multiple complex biological and instrumentation-related parameters, leading to highly variable output data. The quality of HTS data can thus vary from being excellent to unsupportable. Thus, AI methods are only as accurate and robust as the input data. Further challenges are the ability of AI models to understand and learn from increasingly complex features in the training data. Some AI applications have a complex architecture which is not comprehensible to the user. This makes it difficult to understand and trust the AI-based HTS models.

Moreover, there is a risk of biased decision outcomes if the used data and the decision-making processes are influenced focally or unintentionally by human beliefs and behaviors, or if trained decision-making processes are neglecting the statutory regulations, for instance, on ethical standards or environmental policies. An important aspect to consider is the dynamic nature of AI technologies as research in the AI field is fast-paced and constant updates and releases are standard practice. Furthermore, new issues or challenges resulting from the use of AI in HTS may arise and emphasize the need for continuous monitoring, evaluation, and development of AI technologies. In conclusion, HTS and AI tools complement each other well when high-quality input data

and technologies to process this data are available. However, despite the high promise and expectations, a successful implementation comes with a number of open problems on data quality, model robustness and interpretability, legal and ethical concerns, user-friendly materials, and ecosystem challenges, which have to be discussed and solved in open dialogues and further developments.

### 4.1. Data Quality and Quantity

Key points Data quality and quantity are the major determinants in the effectiveness of using artificial intelligence to support high-throughput screening (HTS). "Garbage in – garbage out" is an old saying reflecting the characteristics of every statistical model. If the training data are of poor quality, then the developed model will also be of poor quality. Inadequate or poor experimental methodology and/or data processing also result in many misleading results from high-throughput screens. Applying these data as "confirmatory" by using sophisticated AI methods only supports the misinterpretation of results but does not overcome the basic flaw – insufficient and poor quality experimental results. The significance of the problem is that more and more data are now deposited in public repositories and made available through databases. This is propped up by the many calls for open access, sharing data, exchanging information, and the demand to reuse data that has been generated by public funding organizations. The quantity of the throughput and the perceived gold mine of information stored in these databases is seen as a gold standard. To many, the database is the end and not the means. One fact is, however, undeniable: all statistics in the world will not change; the conventional and traditional approach to produce the large swathes of raw data is by cultivating tissues in large amounts and then extracting using omics tools, where diverse reagents are used to move in a standardized manner through an analysis pipeline where we are pushed from common reagents and low content methods into high content methods. The significance of this basically flawed approach is left untouched in these reports.

### 4.2. Interpretability and Transparency

Given the increased complexity of AI systems, understanding the rationale behind the decisions they make and their degree of reliability can be challenging. This model complexity has spurred the development of interpretable AI, which suggests that specific information can be extracted from a model to improve the user's understanding

of its inner workings or the data that motivates its decision. Instead of mere interpretability, we see the need for 'interpretability of the user.' User interpretability as a concept aims to construct the description of a machine learning model using features that are related to the decision-making process of the user. This concept situates the interpretability of a machine learning algorithm within the context of the subjective decision-making of the user who must take in the recommendation of the model. In this context, the interpretability of AI not only requires that the AI produces recommendations that are interpretable but also to produce a model that the user can understand.

Openness and transparency in AI can also be seen as a mechanism to provide high interpretability to the user. Overall, model transparency involves not only codes that can be visually inspected but also methods such as model explanation or an algorithm that can be fed into AI to provide interpretability. This complementary framework brings together the two most important aspects that are necessary for open AI, i.e., transparency and interpretability from the user's perspective. Openness in AI can be seen as more than mere transparency in decision-making around the AI process, results, and raw data. However, any AI recommendation or decision must be easy to understand and be visible to users, data advocates, and relevant stakeholders. This 'AI for the user' strategy that focuses on openness and transparency in the AI decision-making process is not just about the necessity to give as many people access as possible to company AI or the models. However, openness in including relevant people or stakeholders from a wide variety of backgrounds, and input including human experts, is essential for creating a transparent and open AI model that can be understood and used by everyone at risk. Furthermore, involving stakeholders has helped in the incorporation of background knowledge on the topic being studied or discussed into the AI model, including mathematical modeling, to get an output or decision that is understandable to all involved.

### 4.3. Ethical Considerations

Ethical considerations are likely to increase in significance with the design of AI/ML-based models and their use in HTS, both in terms of the scientific results as well as the potential impact of in-house scientific practices. These issues are likely multi-impact due to both the nature of the utilized data as well as the potential results. Regarding the

handling of data, interest will most likely be the datasets, their composition, and the process of data curation or normalization. Specifically, the following two dimensions might increase ethical and social issues: lack of inclusiveness and unintended biases that can be carried over within human and computational judgment or be perpetuated in the decision-making process. A potential lack of inclusiveness can arise from the simple accessibility or collection of data.

Such a bias can occur due to both intrinsic difficulties in involving or the presented choice of representatives acting as stakeholders, law-making bodies, or socially and structurally significant communities. The use or misuse of HTS results can not only have an economic impact on industry for founders or investors in start-ups, but also, social and ethical issues are brought into focus when considering HTS results. Examples may include a model inadvertently identified from historical data showing biases or providing inequitable outcomes. People with fair complexions may backtrack because darker skin pigment may be less visible when using facial images for consumer-personalized drug development. These may have both legal and social ramifications as many countries condemn extreme examples of algorithmic discrimination. For example, if translating data is a step-by-step process, transparency becomes an issue if the underlying algorithms are confirmed or proprietary steps that require authority to be deciphered. Furthermore, using AI-generated predictions alters the responsibility of decision-making as a solution in itself. Thus, should HTS-driven algorithms be not only scientifically valid but also socially and behaviorally tested before use? We need to establish guidelines and strategies for a range of such issues that we, in particular the AI and ethics community, do not have. Therefore, a multidisciplinary and participatory approach to the use of AI is suggested. Have you seen something that you think is unethical? Talk to your staff! Talk to us! Let 2023 be the year we reimagine AI ethically and inclusively.

**5. Future Directions and Emerging Trends in AI-HTS Integration**

Emerging trends and future directions in AI-HTS integration are promising and under heavy investigation. As machine learning continues to evolve, future directions will likely see the introduction of adaptive algorithms that will not be based solely on initial data, but will in a way "learn" and "adapt" to changing conditions of the screened cell models during the experiment, thus continuously modifying data interpretation and

revealing drug properties not only in concentrated 2D cell culture but also in more complex and realistic 3D tumor microenvironments. In vitro robotics equipped with AI functionality will continue to develop, joining HT direct Next Generation DNA or long-read RNA sequencing technologies, fluorescence microscopy, and ELISA with a database system to propose an AI-based customer-friendly discovery, not only for pharmaceutical companies but also at an academic level.

An AI-integrated IoT technology will be where diseases and the latest drug information will be constantly updated and transmitted throughout the world. The meeting of pharmaceutical companies with academia is essential and must be deepened in terms of hardware, data communication, big data, and especially the latest trends in AI, for the purpose of consulting the consequences of research and development from the drug perspective. In addition, the funds invested by support systems or all other equivalent places will allow the development of AI technologies in 5, 10, up to 20 or even more years for AI to reach the state considered the most advanced and the most accurate and acceptable not only by the academic community or the world of drug production but also by the final recipient. In conclusion, methods like AI-photonic imaging offer hope for increased throughput and simplification in drug screening. The investment in new technologies will enhance the applicable capacities in the area of high-throughput drug screening.

**6. Conclusion**

In conclusion, we discussed the impact of artificial intelligence, comprising machine learning and big data analysis, in the field of high-throughput screening. As described, ML techniques are already widely applied in medicinal chemistry and support many steps in the HTS workflow. AI was shown to revolutionize compound screening workflows in a solid-state drug discovery start-up and enabled the reduction of assay deviation, optimization of inhibitors, and provision of new insights that were missed in the raw data. However, while countless intriguing AI-based screening approaches are currently in diffusion and undoubtedly more of them will appear in the near future, we emphasize that they still represent a subdomain of HTS. A priority area to address in the future will be to have integrated approaches spanning both the biochemical and cell-based HTS workflows to impact and influence biologists and cell assay scientists. Key issues and challenges encountered when employing AI in HTS are related to data

quality, transparency and interpretability, and ethical considerations. Multi-disciplinary bridging efforts will be important in the years to come in order to propose novel and transformative solutions, effectively coupling AI, automation, machine learning, and big data approaches to offer insight and intuition in complex biological assays, simultaneously providing improved public health outcomes. In concerted efforts, our vision may well be to establish an unambiguous way forward, establishing AI anchors that enable us to more definitively leverage insights learned when employing AI in HTS, validating AI concepts that reach the clinic, and helping to possibly replace poor or biased human decision-making. More broadly, there is a responsibility to ensure we unlock the potential of AI treasures in order to reveal the unexpected in our ambition to address outstanding clinical needs.