WHITE PAPER

Five ways machine learning can power life science data analysis

Drug discovery, translational medicine, study feasibility and product design, manufacturing, and design of experiments





intellegens.com | info@intellegens.com
Intellegens, The Studio, Chesterton Mill, Cambridge, CB4 3NP, UK

intellegens

Applied machine learning

Executive Summary

Life science researchers face an exponential expansion in data volumes, while wrestling with objectives including: minimising time-to-market, maximising efficacy at the same time as ensuring safety, and creating better-targeted therapies. Machine learning can turn this data from a time-sink into a resource that accelerates innovation. It can generate new ideas and provide insights that focus experimental programs and improve processes. In this paper, we discuss how machine learning addresses the data analysis challenge, drawing on examples of the use of our Alchemite™ machine learning software.

Drug discovery

Machine learning could help to identify small molecule or biological therapeutic entities, drug targets, or drugs that could be re-purposed. Case studies in this area include compound design and analysis of high-throughput screening data. We discuss work with **Genentech**, **Optibrium**, and **Takeda Pharmaceuticals**.

Translational medicine

Alchemite™ has been used, for example, to predict pharmacokinetic and ADME properties and to study personalised stem cell therapies. Examples come from projects with **AstraZeneca**, **Constellation Pharmaceuticals**, and the **A*Star Institute**.

Clinical study feasibility and product design

Machine learning can improve study design feasibility and provide insights into product design from population studies. Examples include work with **BAT** on studying physiological responses and projects on how patient characteristics impact disease progression.

Manufacturing and supply chain

Key focus areas are optimising process and reactor conditions and modifying formulations to improve yield or in response to regulatory requirements or procurement challenges for active pharmaceutical ingredients (APIs) and excipients. We cite examples from **CPI** and, since experience from other sectors is relevant here, from global food producer, **Yili**.

Design of experiments (DOE)

A common thread through all these application areas is improving the design of various types of experiment – whether in the laboratory or in studies of patient populations. We conclude with a discussion of some of the key strengths of the Alchemite™ machine learning method when compared with traditional approaches to DOE.

Introduction

The scope of this paper – data analysis in life science research

The term "AI" covers a large, growing, and much-hyped range of technologies. These are being applied across life science R&D. Some are still in proof-of-concept phase, whereas others are well-established. In this White Paper, we focus on a branch of AI, **machine learning (ML)** [1] for **data analysis**, which is relatively mature, but where opportunities remain for much more widespread adoption. ML is the class of methods that train themselves on a dataset, building an advanced statistical model that can then be applied to tasks such as classifying the data, understanding relationships within the data, or predicting performance for new data points.





The life sciences data analysis challenge and how ML helps

The data volumes available to life science researchers have grown exponentially in recent years with a range of advances in discovery technologies [2] such as high-throughput screening and genomics / proteomics, and the 10-fold growth in number of clinical trials [3]. At the same time, these researchers are set more demanding challenges: to find new drugs with unique efficacy in a crowded and competitive market; to build on new -omics knowledge and provide ever-more targeted (even personalised) medicines; and to compress time-to-market to maximise the benefits of patents and meet the demands of an aging population with increasing expectations.

> How can we ensure these huge data resources become an opportunity rather than a time-sink?

These huge data resources offer an enormous opportunity since hidden within them is information that can provide vital insights and guidance to improve productivity and evidence-based decision making. They also create a headache, since managing them and extracting value becomes a time-sink. Many traditional data analysis tools were designed for a much less data-rich time. Conventional statistical approaches struggle to scale for "big data". They get lost in the noise of complex multivariate or multimodal information. Where statistical tools can struggle, human intuition can fail entirely, creating a need for a new class of analysis tools.

> Machine learning can tease out hidden correlations and accelerate innovation

ML transfers the burden of performing statistical analysis and modelling into a computer-driven approach. In contrast to traditional statistical approaches and human intuition-driven research, it specialises in handling large, complex datasets, teasing out hidden correlations and turning data into valuable insights that can accelerate innovation.

Limitations of machine learning – and how to overcome them

ML has been around for decades, but it has yet to achieve its full potential. ML data analysis *should* be a standard tool on the desktop of every chemist or biologist. There are several reasons why this has not yet happened:

ML data analysis should be a standard desktop tool for every chemist or biologist

- ML methods often fail when trained using **real-world experimental data** due to dataset sizes, sparsity (the presence of gaps in the data), and noise.
- ML is **not well-integrated with the workflows** of chemists and biologists – methods can be hard to set up, requiring data science expertise and skills in coding or scripting.
- **Trust** – it is normal to be sceptical about the value of a new technology, particularly when the logic underlying ML models can be hard to inspect and understand.

We can address this final point through case studies of ML successes, as we do in this paper. These use our **Alchemite**™ software [4], which tackles the barriers to ML adoption as follows:

- **Real-world problems:** Alchemite™ uses a unique, self-consistent algorithm that *can* train models using sparse, noisy data, where other ML methods fail. It solves high-dimensional, multi-variate problems and is highly computationally efficient, making it practical for big datasets.
- **Integration with scientists' workflows:** a simple-to-use web browser interface [5] (Figure 1) enables scientists to ask questions of the method and analyse the results in familiar language and with no need for coding or advanced ML knowledge.
- **Trust:** a series of key capabilities build confidence for users, including accurate uncertainty quantification for predictions and 'Explainable AI' analytics [6] that add transparency, making it possible to drill into the logic of the ML model.

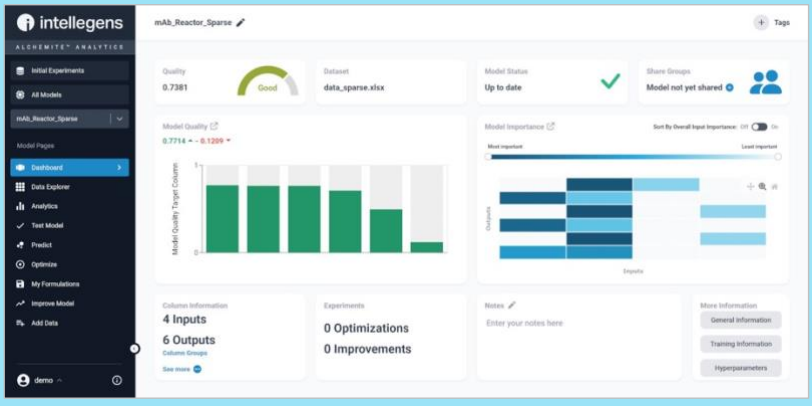

Figure 1. Alchemite™ Analytics web browser interface.

Example applications

Alchemite™ makes it possible to broaden adoption of ML methods for data analysis. In the rest of this paper, we explore how such adoption can help in five areas of life science research.

Drug Discovery

Background

Drug discovery is the process through which new therapeutic entities are identified, with the goal to develop them into safe and effective medicines. Most approved medicines are based on small molecules, but novel drug classes are increasingly important: biologics, gene therapy and RNA modulators, and cell-based therapeutics. The discovery process involves target identification and validation, lead optimisation, and, ultimately, candidate selection for non-clinical development and clinical trials. Despite advances, it is lengthy, costly, difficult, and has a high attrition rate. Drug discovery datasets are often large, but sparse and noisy, as early-stage compounds are rarely measured in all the assays of interest to an organisation.

Datasets are often large but sparse, as all assays of interest are rarely measured

The role of machine learning

A well-established application of machine learning is to train models from experimental or simulation data to **predict properties** such as solubility, bioavailability, and stability. This helps to identify potential drug candidates and optimise their properties, and to generate new ideas for molecules with increased potency or reduced toxicity. ML predictive models are excellent at identifying and exploiting **property-to-property** relationships and, where structure can be captured in molecular descriptors, **property-to-structure** relationships – providing vital clues to chemists in making evidence-based decisions with much greater confidence.

Machine learning excels at identifying property-to-property relationships

ML can also identify what missing data would best improve its model. This enables an **adaptive design of experiments** (DOE) approach in which ML suggests further experiments that will maximise knowledge about the system while minimising experimental time and cost (there is further discussion on adaptive DOE at the end of this paper).

ML could also be used to identify potentially valuable **drug-target interactions** based on known interactions and the molecular or biological properties of the drugs and targets, helping to inform target identification and candidate selection and to reduce the attrition rate of candidate molecules [7].

Finally, ML could identify existing drugs that might be repurposed [8] for new indications. **Drug repurposing** saves time and resources by bypassing the early stages of drug development and leveraging existing non-clinical safety and toxicity data.

Alchemite™ examples

Takeda wanted more value from a very sparse (<1% complete) dataset generated from high-throughput screens on over 700,000 compounds against 3,500 endpoints. Alchemite™ accurately modelled all of these endpoints simultaneously, significantly outperforming conventional QSAR models and other machine learning approaches (Figure 2), thereby increasing the information available for decision-making 100-fold [9].

Alchemite™ outperformed QSAR on Takeda's large, sparse dataset

Genentech optimised kinase profiling programmes

Working with our drug discovery partner, **Optibrium**, Alchemite™ was used by **Genentech** to optimise kinase profiling programmes [10]. The model accuracy was higher than traditional single target QSAR models and biological fingerprint similarity. Accuracy plus confidence in predictions enabled practical use of the machine learning in active learning, assay prioritisation, and virtual screening. Fabio Broccatelli, Principal Scientist at Genentech, commented: "Methodologically what is nice about this technology is that it uses the chemical information, but also uses the existing experimental information."

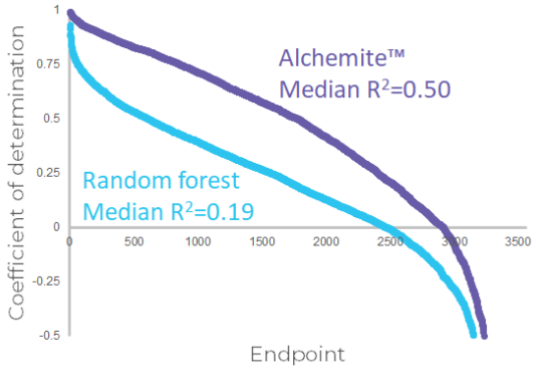

Figure 2. Comparing Alchemite™ predictions to alternative methods in the Takeda case study.

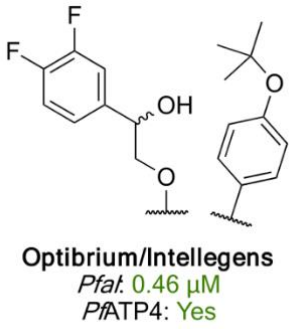

Figure 3. Proposed anti-malarial compound from the Open Source Malaria competition

As part of the **Open Source Malaria** competition [11] Alchemite™ proposed a novel antimalarial compound. The Director of competition, Prof Matthew Todd, described this as "something … none of the humans involved in the SAR over the years would have bothered making", opening a new development direction for the development of antimalarial compounds; the compound proposed was the only active compound found in the competition.

"This was something none of the humans involved would have bothered making"

Translational medicine

Background

Translational medicine is an emerging interdisciplinary field that seeks to bridge the gap between basic science research and clinical medicine. Many candidate drugs fail in human clinical trials because they are found to be unsafe or ineffective, despite promising nonclinical studies in animal and cell models. Translational medicine is driven in part by advances in genomics, proteomics, and other 'omics' technologies that have enabled the discovery of new biomarkers and therapeutic targets to improve the clinical success of therapeutics in development.

Precision medicine is an emerging approach to clinical research and patient care that takes this a step further, focusing on understanding and treating disease by integrating multi-modal or multi-omics data from an individual to make patient-tailored decisions.

The role of machine learning

Translational medicine studies how a drug will behave in the human body or in animal testing. ML algorithms can **predict the efficacy and toxicity** [12] of drugs by analysing datasets of chemical or biological structure, properties, and associated clinical information. This can focus efforts on the most promising candidates, improving drug safety, and reducing the number of experiments and scale of animal studies required to bring a drug to market.

Identify the most promising candidates, improve safety, reduce experiments and animal studies

In precision medicine, ML can **identify biomarkers that are predictive of disease outcomes** and expected treatment response. This can in turn identify patients most likely to respond to specific therapies, improving outcomes and potentially reducing risk of adverse effects. ML can also **analyse data derived from medical images,** such as MRI or CT scans, to help identify disease biomarkers and improve diagnostic accuracy. Analysis of multi-modal data extracts value from large datasets, enabling greater understanding of human health and disease [13].

'Explainable AI' is important for regulatory confidence

This is a new field, being addressed by Regulators to support developers. ML approaches have often been fairly criticised for their lack of transparency. It can be hard to inspect and understand the logic of the complex models that ML methods generate. '**Explainable AI**' – the ability to understand *why* a model is making its predictions – is particularly important in this context. Intellegens has a strong focus on analytics that support such understanding, such as the Importance Chart (Figure 4, overleaf).

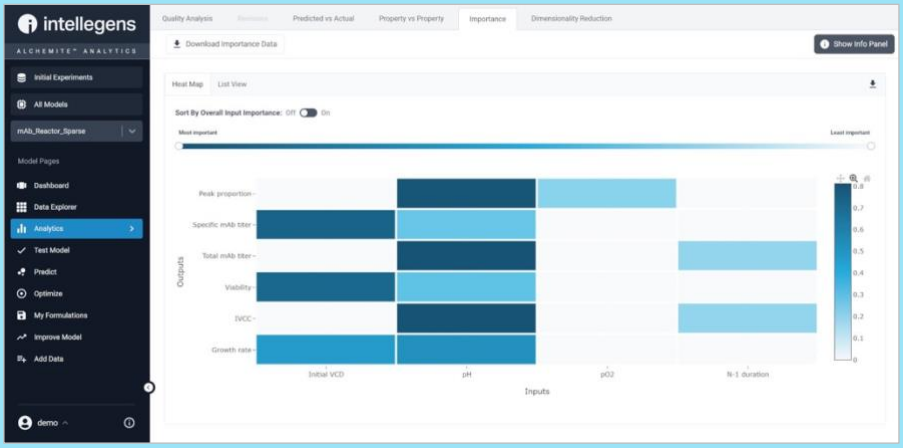

Figure 4. Importance chart in Alchemite™ Analytics enables understanding about how system inputs drive outputs.

Alchemite™ examples

Intellegens worked with **AstraZeneca** and our partner **Optibrium** to model rat PK data, enabling virtual screening of new compounds, which can reduce the number of animals required for testing [14]. Alchemite™ successfully combined descriptors, *in silico*, and *in vitro* data to predict PK parameters and curves, generating results comparable to the best in the literature.

Enabled virtual screening of compounds at AstraZeneca

With **Constellation Pharmaceuticals**, Intellegens predicted ADME properties for ongoing drug discovery projects [15]. Data from multiple projects was combined to create a holistic view of the chemical space of interest, enabling more accurate predictions of eight ADME and ten cell-based activity endpoints. Leveraging inter-endpoint relationships meant the models outperformed conventional QSAR models trained for each endpoint separately.

Working with the **A*Star Institute**, Alchemite™ identified the best personalised stem cell dosage for cartilage damage [16] based on information about the stem cell treatment and data on the patient's lesions. Figure 5 is part of the response surface learned by Alchemite™ showing how cartilage repair success changes based on the severity of the initial damage.

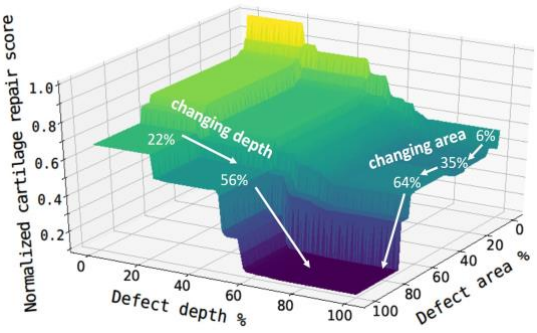

Figure 5. Modelling cartilage repair success.

Clinical study feasibility and product design

Background

The development process for therapeutics, medical devices, and consumer products that impact human health is lengthy and costly, with a high risk of failure [17]. Less than 10% of clinical therapeutics entering clinical development gain product approval and the process typically takes 10-15 years. Translation from non-clinical data and Phase II through pivotal studies for therapeutics are particularly high-risk. Recently, there has been an increase in Phase IV trials, suggesting a trend in post-approval commitments. An improvement in new therapeutic medicines success since 2015, which demonstrates focus on improved productivity, is most likely due to the increased use of biomarkers and novel therapeutic classes. Can machine learning be used to build on progress in targeting trials and product development more generally, further improving study design feasibility and informing product design?

The role of machine learning

Products often fail due to poor study design – machine learning can help

Products often fail due to poor study design. Studies can select individuals unlikely to benefit, dosing or regimen can be inadequate, participant compliance can be poor due to burden of invasive and/or complex clinical assessments, or there can be a failure to meet primary and secondary endpoints. Conducting better studies earlier on allows potential failures to be identified sooner, reducing the risk of more costly failures at a later stage, and saving time, energy, and resources. Machine learning can **optimise trial design**, for example, by identifying patient populations that are most likely to respond to specific treatments, determining optimal dosages, or helping to design adaptive trials. Machine learning can also **predict or identify potential safety issues during the trial**, enhancing safety monitoring and providing deep insights for market launch.

Knowledge gained can also be fed back into product design

Knowledge gained from trials or from other types of consumer or population studies can also be **fed back into product design**. In this process, machine learning can identify relationships between product attributes or performance and population characteristics or preferences. This can lead to optimised product performance, new product features, and, ultimately, increased sales.

Existing databases of clinical investigator sites and recruitment performance could be enhanced by integrating machine learning to **select optimal clinical trial sites and predict recruitment rates,** assisting trial planning for improved estimated duration and trial cost.

Alchemite™ examples

BAT conducted and analysed data from human studies. Predictive models enabled feedback into scientific functions and product design, and guided design of the next phase of studies. A published paper on this work is currently under review [18].

At BAT, insights informed design both of further studies and improved products

Another Intellegens project analysed the [OASIS] open access dataset [19], which covers cohorts of **Alzheimer's** patients. A model was built that was able to identify which patient characteristics contributed to disease progression. The aim was to make it easier to select patients through improved protocol design, likely to respond within a reasonable timeframe when conducting an interventional trial for a new therapeutic.

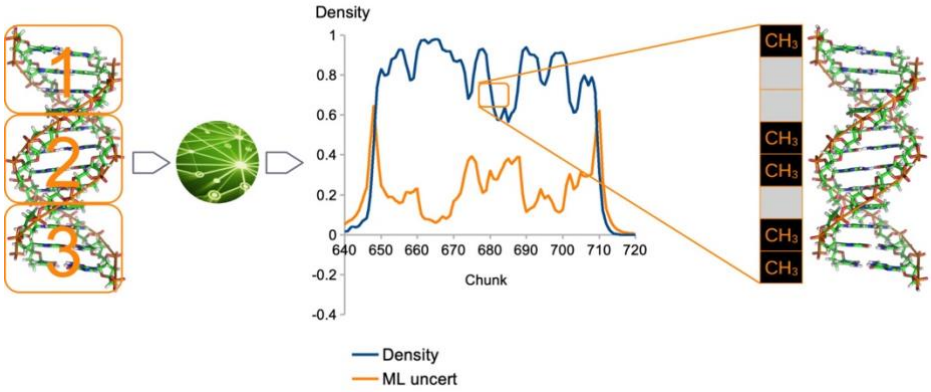

Figure 6. Studying methylation density and uncertainty using machine learning.

Alchemite™ has also been used in a **cancer** study [20], training a machine learning model for methylation density and uncertainty to extract insights into unexplained chaotic methylation due to rogue biology. The model predicted the probability of acute myeloid leukaemia within one year for a cohort of 40 patients with 90% accuracy.

Manufacturing and supply chain

Background

Discovery and development of therapeutics is only the start of the challenge. For commercial viability, optimising formulations and manufacturing processes is a vital step. Finding reactor conditions to enhance yields, reducing production costs, and lowering production times are all examples of manufacturing challenges with direct effects on product profitability. For biopharmaceuticals, it can be a challenge to manufacture reliably at scale at all.

It is also important to manage the impact of supply chain and other disruptions on these processes. How do you produce the same product at low costs if its contents are impacted by changing regulations or a key ingredient has become much more expensive?

The role of machine learning

Machine learning can be used to **optimise production processes** by identifying which process parameters impact product quality and yield, and how. These parameters can then be fine-tuned and controlled to reduce waste and improve efficiency. This modelling can be used throughout the development and production processes. For example, it might enable **quality by design (QbD)** [21] from design through development by understanding the impact of material attributes and process parameters for improved process controls and the manufacturing implications of pursuing particular candidates. During development, it can inform the **scale-up from laboratory** to factory. In production, ML models can identify **quality control** issues and anomalies by analysing data from in-service processes, reducing the likelihood of defects or recalls.

> Identify which process parameters impact product quality and yield, and how

Machine learning can be used to predict **supply chain disruptions** [22] and identify potential risks, allowing companies to take proactive measures to mitigate these risks. It can also be applied to rapidly identify **alternative ingredients or process steps** should some part of an existing process become unusable or too expensive.

Alchemite™ examples

An Intellegens webinar [23] included a presentation from **CPI** on a project to use the predictive capabilities of Alchemite™ to pre-screen potential monoclonal antibody (mAb) sequences, with the goal of finding which of a series of candidate mAbs would behave favourably within current

> CPI pre-screened monoclonal antibody sequences for early understanding of commercial viability

production systems. This modelling allows determination of mAb commercial viability very early in development, avoiding unnecessary expense due to further research into the production of unviable mAbs.

Yili was able to reduce the number of ingredients in a formulation

This is also an area where experience transferred from other sectors can aid life sciences. In the fast-moving consumer goods industry, global dairy products leader **Yili** used Alchemite™ to study UHT whipping cream formulations [24]. These products must remain stable over a nine-month shelf life, with the product reliability determined by the stabilisers and emulsifiers within the formulation, and by controlling processing conditions. Alchemite™ analysed years of formulation data (Figure 7) to understand the impact of each ingredient and process step on the eventual product performance. Matthias Eisner of Yili explained one of the key outcomes as the ability to "drop out a number of the ingredients we had been testing [...] because they did not affect the performance in the way we thought they would, or the supplier claimed. This wasn't obvious if you just looked at them one-by-one, because you always have some cross-interactions".

Due to successful work in the biologics processing area, Intellegens has recently secured a major project in the area of Oligonucleotide-therapeutic characterisation and manufacturing – details will be announced in June 2023 [25].

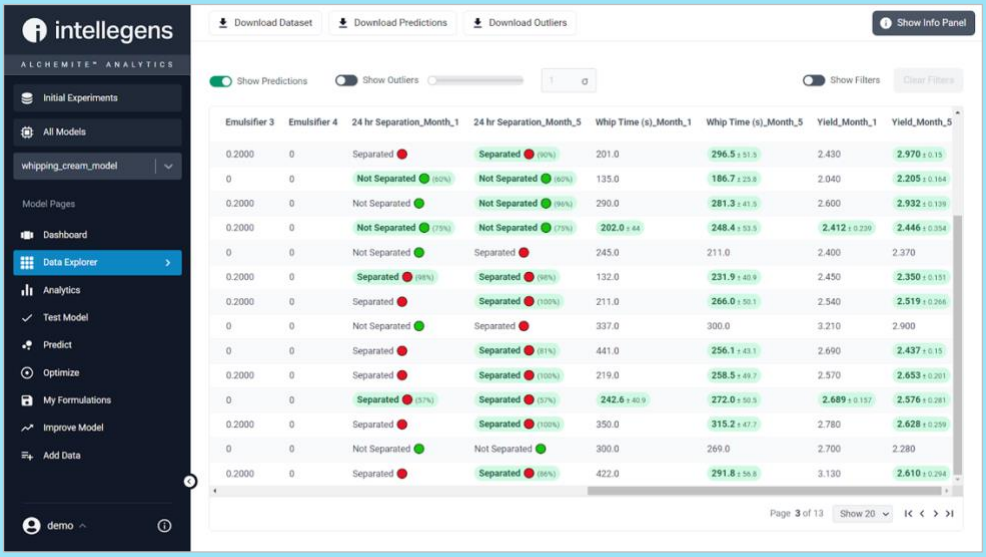

Figure 7. Yili used Alchemite™ to analyse sparse formulation data, imputing missing values and enabling the development team to remove unnecessary ingredients.

Design of experiments

Background – limitations of traditional DOE

A common thread throughout the application areas discussed above is improving the design of various types of experiment – whether in the laboratory or in studies of patient populations. Design of experiments (DOE) tools are already widely used in life science research, applying statistical methods to narrow down the search area for experimental programs, driving down costs and timescales. But these traditional DOE methods have limitations. Some of these limitations are itemised in table 1 along with the alternative offered by Alchemite™ machine learning, which is discussed in more detail below.

Table 1. Alchemite™ vs traditional DOE

Limitation of traditional DOE	Alchemite™ approach
They still result in a high experimental burden	Up to 90% reduction in experiments compared to traditional approaches
It's hard to address cross-correlations; they often model one parameter at a time	ML model captures complex, high dimensional, non-linear relationships
Can require statistical expertise	Method learns from the data provided to build a model – the user does not need statistical knowledge to set it up
Recommends experiments to explore the available design space, not the fastest route to your project objectives	Enables an adaptive, iterative approach that can optimise suggestions for target properties

Alchemite™ for Adaptive DOE

Machine learning models capture the relationships within the data used to train them, even when those relationships are high dimensional and non-linear. When data comes from partially completed experimental programs, the ability of the Alchemite™ method to train models from data that is sparse and noisy is particularly valuable. And, because the model is built by learning from the data, the user simply needs to provide that data rather than to apply advanced statistical knowledge to set up and interpret the tool.

Alchemite™ explores high-dimensional spaces using a guided Bayesian framework [26] and can suggest what experiments should be done next. It can recommend the experiments most likely to improve understanding of the overall design space, or those most likely to lead to solutions optimised for a specific set of target properties. Once these experiments have been performed, the new data can be imported to Alchemite™ and the machine learning model recalculated to recommend further experiments until a model of the desired quality is reached. This *Adaptive* DOE has been shown to achieve project objectives with up to 90% fewer experiments than conventional DOE [27].

Adaptive DOE can achieve objectives with up to 90% fewer experiments

Highly accurate uncertainty quantification is important

Integral to this process is the ability to quantify the uncertainty in the machine learning predictions. Alchemite™ performs this calculation based on nonparametric probability distributions for many properties simultaneously. Such methods are more accurate than other approaches that compute uncertainty based on assumptions about probability distributions [28]. Alchemite™ understands risk, and thus the experimental route most likely to succeed, more reliably.

Finally, Alchemite™ can even provide value at the start of an experimental program in the absence of any historic experimental data. It offers generic advice on where to perform those first few experiments to gain a broad sweep of the system, before training a model that can then guide where additional experiments should be performed.

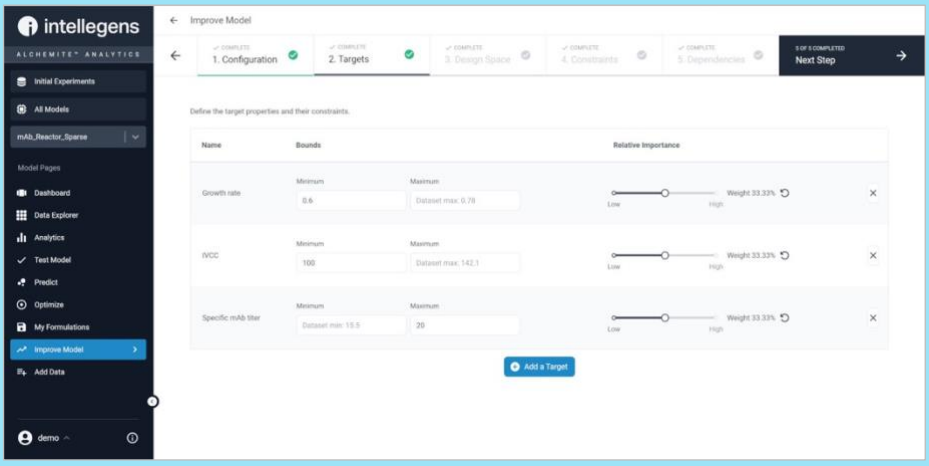

Figure 8. Setting up a Design of Experiments project in Alchemite™. It is possible to define target properties to enable more efficient exploration of design space.

Conclusion

In this white paper, we've studied the potential of machine learning to accelerate life sciences R&D by extracting more value from various types of data – for example, from experiments, simulation, processes, clinical studies, or patient cohorts. We have drawn on our experience of applying the Alchemite™ machine learning technology. We introduced this technology and discussed its application in four key parts of the life science value chain as well as for the widely occurring challenge of design of experiments. Benefits from the examples cited include valuable insight into new drugs and therapeutics, reduced costs and timescales for lab experiments and population studies, and optimisation and control of manufacturing processes.

We have also acknowledged some of the challenges that have limited use of machine learning thus far, including its ability to handle sparse, noisy data, difficulties in implementation and use, and the need for excellent 'Explainable AI' capabilities.

Our focus at Intellegens is on providing software to overcome these challenges to deliver the potential benefits of machine learning for life sciences R&D. Contact us if you would like to discuss collaboration or apply Alchemite™ to your R&D challenges.

intellegens.com/contact

info@intellegens.com

References

[1] K. Kersting, Machine Learning and Artificial Intelligence: Two Fellow Travelers on the Quest for Intelligent Behavior in Machines, *Front Big Data*, **1**: 6 (2018)

[2] G.K. Kiriiri, P.M. Njogu, A.N. Mwangi, Exploring different approaches to improve the success of drug discovery and development projects: a review, *Futur J Pharm Sci* **6** 27 (2020)

[3] US National Library of Medicine, *ClinicalTrials.gov – a database of privately and publicly-funded clinical studies conducted around the world*, Bethesda, MD, USA. Available at: https://clinicaltrials.gov

[4] B.D. Conduit, N.G. Jones, H.J. Stone & G.J. Conduit, Design of a nickel-base superalloy using a neural network, *Materials & Design* **131** 358 (2017)

[5] Alchemite™ Analytics, Intellegens Ltd., Cambridge, UK (2023). Information at: https://intellegens.com/products-services/alchemite-analytics/

[6] S.P. Warde, Blog: Think outside the black box – Explainable AI, Intellegens Ltd., Cambridge, UK (2022). Available at: https://intellegens.com/thinking-outside-the-black-box-explainable-ai/

[7] S. Priya et al, Machine learning approaches and their applications in drug discovery and design, *Chem Biol Drug Des.* **00**: 1–18 (2022)

[8] X. Zeng et al, Target identification among known drugs by deep learning from heterogeneous networks, *Chem. Sci.* **11** 1775-1797 (2020)

[9] B.W.J. Irwin, T.M. Whitehead, S. Rowland, S. Mahmoud, G.J. Conduit, M.D. Segall, Deep imputation on large-scale drug discovery data, *Applied AI Letters* **2** e21 (2021)

[10] Webinar: Optimising kinase profiling programmes with deep learning, Pistoia Alliance (2021) Details available at: https://intellegens.com/optimising-kinase-profiling-programs-with-optibrium-and-genentech/

[11] E.G. Tse et al, An Open Drug Discovery Competition: Experimental Validation of Predictive Models in a Series of Novel Antimalarials, *J. Med. Chem.* **64** 22, 16450-16463 (2021)

[12] L. Zhanga et al, Applications of Machine Learning Methods in Drug Toxicity Prediction, *Current Topics in Medicinal Chemistry* **18** 987-997 (2018)

[13] S.J. MacEachern, N.D. Forkert, Machine learning for precision medicine, *Genome* **64(4):** 416-425 (2020)

[14] O. Obrezanova et al, Prediction of In Vivo Pharmacokinetic Parameters and Time–Exposure Curves in Rats Using Machine Learning from the Chemical Structure, *Molecular Pharmaceutics* **19** (5) 1488-1504 (2022)

[15] B.W. J. Irwin, J. Levell, T. M. Whitehead, M. D. Segall, G. J. Conduit, Practical Applications of Deep Learning to Impute Heterogeneous Drug Discovery Data, *J. Chem. Inf. Model.* **60** (6) 2848–2857 (2020)

[16] Y.Y.F. Liu, Y. Lu, S. Oh, G.J. Conduit, Machine learning to predict mesenchymal stem cell efficacy for cartilage repair, *PLoS Comput Biol* **16**(10): e1008275 (2020)

[17] C.H. Wong et al., Clinical development success rates, *Biostatistics* **20:** (2) 273–286 (2019)

[18] K. Prasad et al., Modelling nicotine pharmacokinetic profile for e-cigarette using real time monitoring of consumer's physiological measurements and mouth level exposure, *BioData Mining*, under review (2023)

[19] Data were provided by OASIS-2: Longitudinal: Principal Investigators: D. Marcus, R, Buckner, J. Csernansky, J. Morris; P50 AG05681, P01 AG03991, P01 AG026276, R01 AG021910, P20 MH071616, U24 RR021382. More information at: https://www.oasis-brains.org/#data

[20] G.J. Conduit, Unveil the unseen: uncover hidden information with machine learning, Presentation at the A*STAR Institute of High Performance Computing, Singapore. Jan 2023. Available at: https://www.tcm.phy.cam.ac.uk/~gjc29/talks/ASTARTalk2023.pdf

[21] ICH guideline Q11 on development and manufacture of drug substances (chemical entities and biotechnological/biological entities), European Medicines Authority, London (2013). Available at https://www.ema.europa.eu/en/human-regulatory/research-development/quality-design

[22] E.B. Tirkolaee et al, Application of Machine Learning in Supply Chain Management: A Comprehensive Overview of the Main Areas, *Mathematical Problems in Engineering* Article ID 1476043 (2021)

[23] Webinar: Machine learning for life sciences R&D – a biopharmaceutical case study, Intellegens Limited, Cambridge, UK (2023) https://intellegens.com/webinar-23-may-machine-learning-for-life-science-rd-a-biopharmaceutical-case-study/

[24] Case study: Developing food formulations at Yili, Intellegens Limited, Cambridge, UK (2023) https://intellegens.com/developing-food-formulations-at-yili/

[25] Project announcement pending, May 2023. Search 'Oligonucleotides' on https://intellegens.com for details.

[26] J. Bergstra, D. Yamins, D. Cox, Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures, *Proceedings of the 30th International Conference on Machine Learning, PMLR* **28**(1):115-123 (2013)

[27] White paper: Machine Learning for Adaptive Experimental Design, Intellegens Limited, Cambridge, UK (2021). Available at: https://intellegens.com/machine-learning-for-guided-design-of-experiments-white-paper/

[28] M. Abdar et al., A review of uncertainty quantification in deep learning: Techniques, applications and challenges, *Information Fusion,* **76** 243-297 (2021)

About Intellegens

Our mission is to be the leading machine learning solution for real-world, sparse and noisy data problems in industrial R&D and manufacturing processes. Our focus is on making it easy to apply machine learning to accelerate innovation. Alchemite™ originated at the University of Cambridge and development is on-going at Intellegens, in close collaboration with our growing community of Alchemite™ customer organisations. These represent sectors including alloys, additive manufacturing, aerospace, batteries, ceramics, chemical processes, composites, consumer products, cosmetics, drug discovery and development, energy, food and beverage, formulated products, paints, plastics, printing technology, and translational medicine.

www.intellegens.com | info@intellegens.com | @intellegensai