intellegens

WHITE PAPER

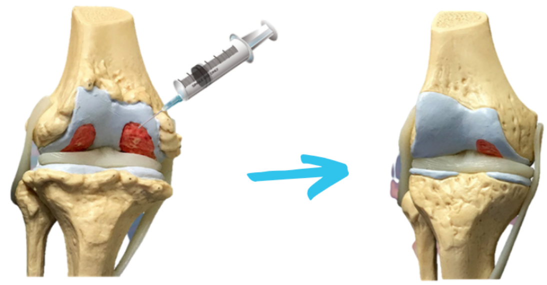
Machine Learning Applications to Precision Medicine - Predicting Stem Cell Therapeutic Efficacy

Executive summary

Advances in machine learning are opening a new chapter in precision medicine. Limitations in available data and computational shortcomings render the task of predicting therapeutic efficacy from clinical trials and animal studies information highly challenging. However, the machine learning approach presented in this white paper, based on the **Alchemite™** platform, provides significant insights into the optimisation and personalisation of treatment strategies. That is exemplified here with the use case of cartilage damage, which affects the life quality of hundreds of millions of people, causing chronic joint pain and disability. Stem cell therapy has emerged as a promising treatment to repair cartilage and reduce pain. Identifying the best treatment using computational methods will enable us to target therapeutic treatment and improve the patient's chances of successful recovery.

Intellegens Ltd., Eagle Labs, Chesterton Road, Cambridge, CB4 3AZ, UK

Introduction to cartilage injury



Articular cartilage is a critical tissue with multifaceted mechanical functions. It holds compression, absorbs shock, and enables smooth articulation at the joints. Cartilage injury is unfortunately common due to tears, accidents and arthritis, which often leads to joint pain, stiffness, and inflammation.

Cartilage disorders affect more than 10 million people of all ages, including children, in the UK

Globally, osteoarthritis alone affects over 200 million people. Adult cartilage has limited self-repair capacity due to its lack of blood vessels. Treatment is often necessary to accelerate repair and relieve pain during joint motion. Besides the conservative treatments and conventional surgical options, such as microfracture and autologous chondrocyte implantation that substitutes cartilage grown from the patient's own cells, the use of mesenchymal stem cells (MSC) is a therapy with high potential that has been widely investigated in recent decades.

Challenge: Inconsistency between MSC therapy studies

Although success has been achieved for MSC therapy in cartilage repair, therapy efficacy has been inconsistent. This is attributed to complex cellular mechanisms and dynamic interplay between different populations of cells in the stem-cell-assisted tissue repair processes. MSC therapy is complicated by cell heterogeneity, culture conditions, delivery methods, and recipients' conditions—all highly variable in clinical trials and laboratory studies. Thus, **disconnectedness between the in vitro, pre-clinical, and clinical performances of MSCs have been broadly observed**, rendering analysis of MSCs' therapeutic efficacy largely retrospective, rather than predictive.

Guidelines for MSC therapy require identification of critical properties that affect MSCs' therapeutic efficacy, so quantitative assessment of the significance of each property is needed. However, conventional controlled biomedical experiments can interrogate one or at most a few properties simultaneously. To overcome this challenge, **the machine learning tool Alchemite™, can capture multi-property correlations and leverage all of the historical experimental data**.

There is a lack of guidelines on MSC therapy strategy to promote optimal therapeutic effect

Data generated by different researchers often result from different experimental designs; therefore the properties considered in one study may not always be addressed in another, leading to the database containing "missing entries". Standard machine learning methods fail to perform analysis with incomplete information. This white paper presents a novel machine learning method that has a unique capability to "fill" the missing data by exploiting the correlations across multiple properties to generate the most complete model possible. The method also identifies the uncertainty of predictions raised from extrapolation and experimental noise, which allows clinicians to focus on the most promising treatments.

Standard machine learning methods fail to perform analysis with incomplete information

Machine learning: A novel approach for predictive therapeutic efficacy of stem cell therapy

We present a study on cartilage repair to demonstrate the benefits that machine learning can offer. This work was conducted by researchers from the University of Cambridge and Bioprocessing Institute (ATI), A*STAR, and was published in PLOS Computational Biology [1]. The method presented in this white paper is available through the Alchemite™ deep learning software.

The first challenge in the meta-analysis is to create a database comprising different research data that may have missing information. **To handle incomplete data, the Alchemite™ algorithm identifies links between existing properties and uses information from other completed entries to guide the extrapolation of the model.** **Alchemite™** applies an expectation- maximization algorithm, where an initial estimate for the missing data is provided and then the neural network iteratively improves that initial estimate. The trained model achieved the coefficient of determination (R^2) of 0.637—a statistical measure indicating that approximately 63.7% of the observed variation can be explained by the model's inputs, in the leave-one-out cross validation framework.

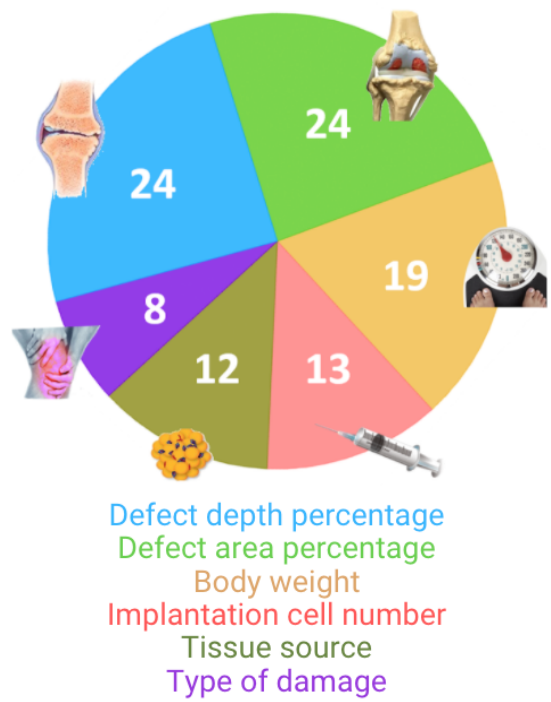


Figure 1. *Illustration of the relative importance of properties used in our model.*

Using the trained model, the tool can then select the most appropriate input treatment properties, as the available properties vary across different animal and human studies. The patients' pre-treatment conditions and therapeutic strategies were encoded within the input properties for the model to make predictions. Six critical input parameters (defect depth percentage, defect area percentage, body weight, implantation cell number, tissue source, and type of damage) have been identified by our tool. The relative importance of the properties on predicting the cartilage repair score is plotted in Fig 1. The pre-treatment conditions such as defect area percentage, defect depth percentage, and body weight play important roles in the treatment outcome. However, the treatment strategy properties, such as the implantation cell number and the tissue source, impact the outcome to a lesser extent.

The study showed that critical thresholds of damage exist for effective cartilage repair to happen, where a "critical size" osteochondral defect means it can not effectively repair by itself. Some studies attempted to experimentally determine the critical size of the defect in terms of depth and diameter. **In our machine learning model, however, it was straightforward to predict the critical size defect** as there was a rapid decrease in the normalized cartilage repair score when the defect area percentage was larger than 64% (Fig 2). An additional insight was that another sharp rise in the repair score was observed when the defect area percentage varied from 35% to 6%. These sharp improvements indicate the presence of multiple "critical sizes" that enable cartilage repair to different levels post MSC therapy.

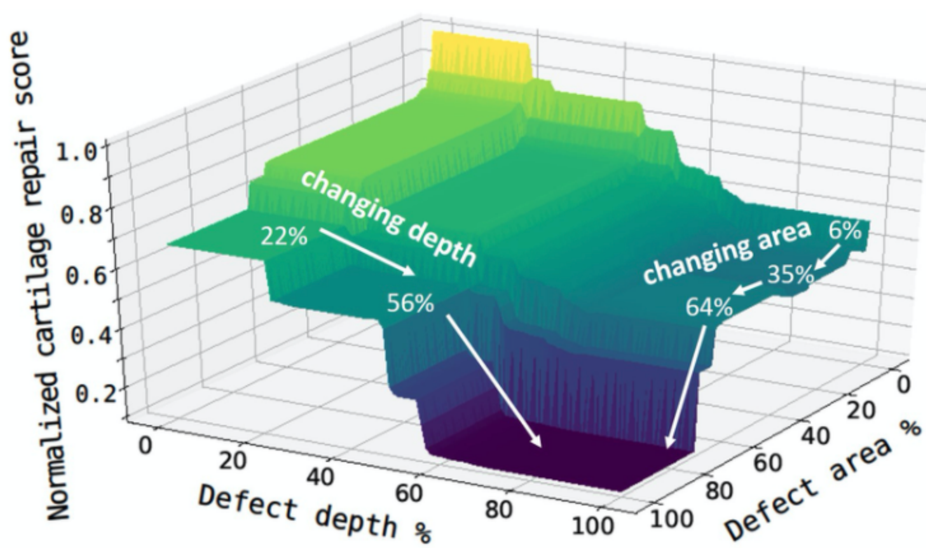


Figure 2. *Surface plot of the normalised cartilage repair score based on defect area percentage and defect depth percentage. The trajectory of changing area or depth is shown in white arrows.*

The determination of MSC dose for therapy remains intuitive in current practice. A wide range of implantation cell numbers has been found in the literature, ranging from a few thousand to 10 billion with the majority falling between 1 to 100 million. Besides the implantation cell number, these cells were also transplanted at a vast range of concentrations in different animal studies and clinical trials, between a thousand to a billion cells per millilitre of the delivery agents.

Controversial results on the cell dose-dependent influence on cartilage repair have been reported

Applying machine learning, we obtained a crucial insight: there is a near linear increase in the cartilage repair score with implantation cell numbers fewer than 17 million. The normalized cartilage repair score is above 0.9 when the implantation cell number is between 17 to 25 million, and remains around 0.8 in the 25 to 75 million range. The high cell density likely recapitulated the mesenchymal condensation process that occurred during embryonic development of cartilage, and promoted MSC differentiation towards chondrogenic lineage.

The optimal dose suggested by Alchemite™ of 17-25 million MSC for human therapy agrees with the recent dose-dependent MSC Phase II clinical trial

The optimal dose suggested by **Alchemite™** of 17-25 million MSC for human therapy agrees with the recent dose-dependent MSC Phase II clinical trial [2] to treat osteoarthritis patients, which was unseen to the machine learning model. Here, a trend towards improvement was seen in the 25 million MSC cell dose group compared to the higher dose groups (50, 75, and 150 million cells). This overturns the long-standing protocol of using fewer than 2 million cells for implantation.

In summary, our tool exploits the inter-property correlations to predict the therapeutic efficacy of MSC therapy for cartilage repair based on animal results and human clinical trials.

The predictive power of our model enables personalized therapy. We predicted the optimal therapeutic outcome based on individual patient's disease conditions, including defect area percentage, defect depth percentage, and body weight. For patients with severe cartilage damages beyond the threshold for effective repair or patients with unique conditions reflected in higher model uncertainties, other treatment strategies should be considered. The technology can also be adapted for MSC therapies to other medical indications, and to address other biomedical questions.

Alchemite™ predicts mesenchymal stem cell efficacy for cartilage repair

The **Alchemite™** platform yields the following benefits:

- **Design personalised treatment strategy** — based on the patient's condition.
- **Predict the therapeutic outcome** — understand the expected outcome with uncertainty before conducting the treatment to have a valuable reference for both clinicians and patients.
- **Explainable process design** — achieve the best possible understanding of how therapy parameters correlate to the final outcome and perform virtual experiments of changing parameters that cannot be tried on clinical or animal studies.
- **Utilise all available research data** — leverage cutting-edge science, across all studies from animals and humans, in order to improve the prediction accuracy and find optimized treatment.
- **Reduction of experimental costs** — use confidence levels to guide where to test new therapies and thus shorten time for new production to market and save money by eliminating research and development costs.

Demo

We provide an interactive demo based on the database for the published article at: https://stem-cell.intellegens.ai

About Intellegens

Intellegens has developed a unique deep learning engine, **Alchemite™** for training neural networks from the sparse and noisy data typical of real-world science and business challenges. The technique was first developed at the University of Cambridge where it has been used to develop aerospace alloys, guide the design of new drugs, and design next-generation battery technology. The tool is now being used to solve a wide range of industrial customer problems, optimising products and processes, saving time and cost in discovery and development, and enabling breakthrough insights.

www.intellegens.ai | info@intellegens.ai | @intellegensai