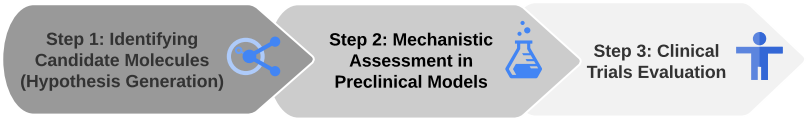
Adaptive Deep Learning of Drug Combination Mechanics for Accelerated Repurposing

Prashant Singh,* Ola Spjuth†

Motivation Drug discovery is a challenging endeavour hampered by high attrition rates in the pharmaceutical industry [1], problem complexity in certain therapeutic areas (e.g., several types of cancers), and changing regulatory requirements [2]. Drug repurposing (also known as repositioning, reprofiling or re-tasking) has emerged as a time and cost-efficient viable alternative towards identifying existing approved drugs for uses outside their original medical indication [2, 3]. The use of already approved drugs involves a lower risk (as the compounds have an acceptable safety profile), and lower research and development costs (US$300 million for a repurposed drug v/s US$2-3 billion for a new chemical entity on average [4]).



The steps involved in the drug repurposing pipeline can be seen in the figure above. The first step wherein candidate drug compounds of interest are identified, is of critical importance. There exist both computational and experimental approaches to the candidate identification problem [2]. Computational approaches typically involve leveraging different properties of the drug compounds and target genes/proteins from existing databases and electronic health records (EHRs) in order to identify synergistic drug combinations.

In contrast to single compounds, combinations of drugs are more effective since they are able to target multiple biological mechanisms. However, it is important to note that the search space is combinatorial, and efficient candidate combination identification is vital towards realising drug repurposing within practical timeframes. Existing computational approaches either require heavy manual intervention [2], or do not consider a comprehensive holistic view of information-driven combination generation [5].

Goal The project will address critical barriers in practical drug repurposing - automation, computational efficiency and effectiveness. The framework will manifest as a deep neural network driven adaptive sampling algorithm, iteratively generating promising candidate drug combinations, and interfacing with an automated cell profiling laboratory to experimentally evaluate the corresponding efficacy. The adaptive sampling process proceeds until a pseudo-optimal combination is found. The project is therefore highly interdisciplinary, and involves basic research in model-based sampling, design of experiments and deep learning.

Current State-of-the-Art A recent promising approach involves the use of machine learning (ML) regression models to learn the relationship between a pair of drug molecular signatures as input, and drug synergy as output [5]. Two drugs are said to be synergistic if the combined therapeutic treatment effect of the two drugs is greater than the sum of their individual effects. The regression model used in [5] is a combination multi-layer perceptron (MLP) network, which is composed of two individual MLPs - one for learning embeddings or features from single drugs, and the second for combining and mapping the features to a synergy score. There exist opportunities to improve the current state-of-the-art at multiple levels.

*Assistant Professor, Division of Scientific Computing, Department of Information Technology, Uppsala University
†Professor in Pharmaceutical Bioinformatics, Department of Pharmaceutical Biosciences, Uppsala University

- In addition to drug signatures, information about drug-target interactions, drug adverse effect profile, etc. can also be used as input to the regression model, and can potentially accelerate the search for pseudo-optimal drug candidates. Such information has not been utilised previously as part of an automated model-based sampling approach to the best our knowledge. It is also important to consider combinations of size >2 so that the treatment efficacy can be maximised.

- Given the large combinatorial search space, it is imperative to consider the (epistemic) uncertainty associated with unexplored/insufficiently explored candidates in a principled manner. Existing work uses deterministic regression models, and even when a probability distribution has been considered over the predicted synergy, it has been aleatoric in nature [5] rather than epistemic.

- Adaptive sampling combines space-filling *exploration* of the variable space (drug/dosage combinations) with *exploitation* of potentially promising combinations. There is scope to explore a wider interpretation of these two entities, including possibilities of defining exploration and exploitation with a view to minimise the number of preclinical experiments needed to arrive at pseudo-optimal drug/dosage combinations. Both model uncertainty and input-based (e.g., distances in drug combination feature spaces) approaches will be developed and evaluated for the exploration component, while exploitation will be based on sampling/acquisition strategies in classical surrogate-based optimisation literature (e.g., expected improvement, probability of improvement, etc.) [6, 7, 8].

- Given access to co-PI Spjuth's automated robotised cell profiling laboratory, the project entails a very unique opportunity to develop a highly integrative and symbiotic pipeline, harmonising the computational and experimental aspects of the drug repurposing.

Aims With a view of the motivation and opportunities outlined above, we define the following project aims.

- **Aim 1:** Systematically investigate and design a wider, more comprehensive and informative collection of inputs to the regressor (e.g., information about drug-target interactions, adverse effect profiles, other genotypic data) as the Bayesian prior.

- **Aim 2:** Design a bespoke deep Bayesian neural network (BNN) [9] architecture optimised to extract expressive features from the input collection outlined in Aim 1. The exact characteristics of the architecture will be dictated by the nature of inputs obtained from Aim 1 (as the prior distribution).

- **Aim 3:** Develop a balanced adaptive sampling strategy that can leverage the prediction uncertainty provided by the BNN. Different interpretations of exploration and exploitation will need to be evaluated and considered herein with the goal of minimising the number of experiments required to arrive at an effective drug combination.

Figure 1 presents an overview of the workflow corresponding to the proposed approach. Note that the deep learning driven adaptive sampling process will make use of existing data available in databases, in conjunction with live *in-vitro* experimental results. The feedback loop from the experiments will also iteratively enhance the accuracy of the neural network regressor.

Interdisciplinary Mathematics The project is highly interdisciplinary in nature, combining several mathematical disciplines including statistics (design of experiments, statistical sampling), applied probability (Bayesian methods, active learning), machine learning (deep learning, Bayesian optimisation), computational science (surrogate modelling), optimisation, and cross-disciplinary domains such as bioinformatics and pharmaceutical biosciences. The CIM environment will be an ideal setting for the development of the doctoral student, preparing a future interdisciplinary researcher towards tackling hard scientific problems.

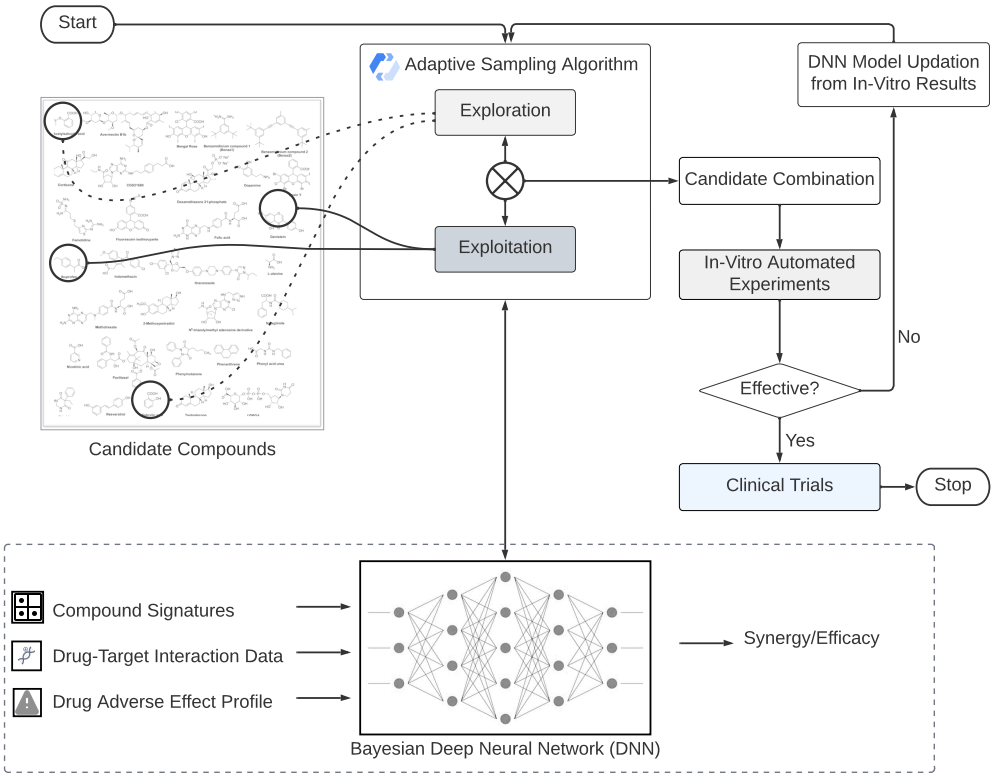


Figure 1: A high-level overview of the proposed approach.

Host Institution and Supervision Duties The PhD student will be hosted by the Division of Scientific Computing (TDB), Department of Information Technology, Uppsala University (UU). PI Singh (TDB, UU) will be the main supervisor, and co-PI Spjuth at the Department of Pharmaceutical Bioinformatics (Pharmbio, UU) will be the co-supervisor of the doctoral student. Both supervisors will be involved in all stages of the project. PI Singh brings extensive experience in the areas of design of experiments [8], adaptive sampling [10, 11, 6] and deep learning [9, 12], while co-PI Spjuth has made substantial contributions in AI-driven high-throughput and high-content cell profiling, particularly in drug development, and pharmacology in general.

Research Environment The student will be a part of the Scientific Machine Learning group (SciML) at TDB, with supportive colleagues from diverse backgrounds including machine learning, optimisation, mathematical modelling, distributed computing, and cybersecurity. The student will also be affiliated with co-PI Spjuth's Pharmbio lab with colleagues working in AI, cell biology and automation. PI Singh and co-PI Spjuth are part of the Science for Life Laboratory (SciLifeLab). The student will therefore have the opportunity to take part in career development activities, seminars and research visits organised by SciLifeLab, complementing activities at CIM. Project collaborators also include Wallenberg Advanced Bioinformatics Infrastructure (WABI), who will provide their expertise on the computational aspects of drug repurposing.

Financing CIM (50%), Division of Scientific Computing, Department of Information Technology (25%), Pharmaceutical Bioinformatics Research Group, Department of Pharmaceutical Biosciences (25%)

Data We will study Sarcoma cancer cell lines available in the Spjuth lab, and primary Soft Tissue Sarcoma (STS) cells from patients available via collaborators at the National Sarcoma Centre at the Karolinska University Hospital. The primary cells will be used in the later phase of the project when applying the developed methodologies to individual patient samples.

References

[1] Michael J Waring, John Arrowsmith, Andrew R Leach, Paul D Leeson, Sam Mandrell, Robert M Owen, Garry Pairaudeau, William D Pennie, Stephen D Pickett, Jibo Wang, et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nature reviews Drug discovery*, 14(7): 475–486, 2015.

[2] Sudeep Pushpakom, Francesco Iorio, Patrick A Eyers, K Jane Escott, Shirley Hopper, Andrew Wells, Andrew Doig, Tim Guilliams, Joanna Latimer, Christine McNamee, et al. Drug repurposing: progress, challenges and recommendations. *Nature reviews Drug discovery*, 18(1):41–58, 2019.

[3] Ted T Ashburn and Karl B Thor. Drug repositioning: identifying and developing new uses for existing drugs. *Nature reviews Drug discovery*, 3(8):673–683, 2004.

[4] Nicola Nosengo. New tricks for old drugs. *Nature*, 534(7607):314–6, 2016.

[5] Paul Bertin, Jarrid Rector-Brooks, Deepak Sharma, Thomas Gaudelet, Andrew Anighoro, Torsten Gross, Francisco Martinez-Pena, Eileen L Tang, Cristian Regep, Jeremy Hayter, et al. Recover: sequential model optimization platform for combination drug repurposing identifies novel synergistic compounds in vitro. *arXiv preprint arXiv:2202.04202*, 2022.

[6] Prashant Singh, Ivo Couckuyt, Francesco Ferranti, and Tom Dhaene. A constrained multi-objective surrogate-based optimization algorithm. In *2014 IEEE Congress on Evolutionary Computation (CEC)*, pages 3080–3087. IEEE, 2014.

[7] Prashant Singh, Dirk Deschrijver, Davy Pissoort, and Tom Dhaene. Accurate hotspot localization by sampling the near-field pattern of electronic devices. *IEEE Transactions on Electromagnetic Compatibility*, 55(6):1365–1368, 2013.

[8] Prashant Singh. *Design of experiments for model-based optimization*. PhD thesis, Ghent University, 2016.

[9] Fredrik Wrede, Robin Eriksson, Richard Jiang, Linda Petzold, Stefan Engblom, Andreas Hellander, and Prashant Singh. Robust and integrative bayesian neural networks for likelihood-free parameter inference. In *2022 International Joint Conference on Neural Networks (IJCNN)*, pages 1–10. IEEE, 2022.

[10] Prashant Singh, Joachim van der Herten, Dirk Deschrijver, Ivo Couckuyt, and Tom Dhaene. A sequential sampling strategy for adaptive classification of computationally expensive data. *Structural and Multidisciplinary Optimization*, 55(4):1425–1438, 2017.

[11] Prashant Singh, Dirk Deschrijver, and Tom Dhaene. A balanced sequential design strategy for global surrogate modeling. In *2013 Winter Simulations Conference (WSC)*, pages 2172–2179. IEEE, 2013.

[12] Mattias Åkesson, Prashant Singh, Fredrik Wrede, and Andreas Hellander. Convolutional neural networks as summary statistics for approximate bayesian computation. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021.