
ABSTRACT FACE PAGE

1. **Presenting Author's name: Gary An**
2. **Presenting Author's affiliation: University of Vermont**
3. **Presenting Author's title: Professor of Surgery**
4. **Presenting Author's email: gan@med.uvm.edu**
5. **Presenting Author's gender (optional): Male**
6. **Presenting Author's race (optional): Asian**
7. **Presenting Author's ethnicity (optional):**
8. **Presenting Author's affiliation sector: (check one or more)**
    - Academia X
    - Industry
    - Federal Employee/Contractor
    - Private Foundation
    - Other: _____
9. **Presenting Author's Career stage: (check one)**
    - K-12 student
    - Undergraduate student
    - Graduate Student
    - Post-doctoral Trainee
    - Young employee (within first 3 year of post-training position)
    - Mid-level employee (3-10 years of post-training position)
    - Senior-level employee (10+ years of post-training position) X
    - Other: _____

Website / twitter handle / other public links (optional): https://www.med.uvm.edu/ancockrelllab/home

10.
11. **Is this the research presented in this abstract supported by IMAG MSM-related U01 funding? Yes or No**
12. **If the Presenting Author is a trainee, who is the trainee's primary research advisor? _____**

MACHINE LEARNING ON AGENT-BASED MODELS: TRAINING ARTIFICIAL NEURAL NETS TO AID IN CLINICALLY-RELEVANT SIMULATION EXPERIMENTS

Chase Cockrell[1], Zachary Silberman[1], Dale Larie[1], Andrew Becker[1] and Gary An*[1]
[1]University of Vermont, Burlington, VT, USA. Website: https://www.med.uvm.edu/ancockrelllab/home

BACKGROUND: Agent-based modeling is a widely used multi-scale modeling method that facilitates the translation and integration of basic science knowledge regarding cellular behavior into system-level phenomena seen at the tissue, organ and organism level. As such, agent-based models (ABMs) are particularly well suited to generating synthetic populations for *in silico* clinical trials. However, since there is currently no robust formalism able to capture the essential qualities of an ABM, they continue to be treated as empirical experimental objects, where their use essentially involves running numerous replicates under varying conditions. Our own insights into the role of parameter space in characterizing biological heterogeneity [1] further complicates the computational burden of performing simulation experiments with ABMs, recognizing that robust insights from the use of ABMs now requires an exponential increase in the number of simulations required. This, in addition to the generally greater compute time needed to execute an ABM (as opposed to equation-based models) has led to our interest in using machine learning to train artificial neural networks (ANNs) for a series of applications aimed at increasing the efficiency of clinically relevant simulation experiments.

METHODS: The reference model for this work is the well-established Innate Immune Response ABM (IIRABM)[2]. Machine learning tasks focused on training ANNs to: 1) improve the efficiency of sampling clinically-relevant parameter space of the IIRABM in order to constrain the size of a potential control discovery task, and 2) serve as surrogate model for trajectory forecasting. The first task used a Keras framework to train a ANN to identify clinical tipping points ("points of no return") that would guide the application of putative interventions based on the simulation run's position in parameter space, and then reducing the training set size to determine a performance inflection point. The second task involved training two ANNs to provide a forecast horizon for trajectories generated by the IIRABM: 1) a convolutional long short-term memory (LSTM) ANN that determines the future value for eleven cytokine parameters based on five consecutive previous values; and 2) a second ANN that accepts a list of the same eleven parameters as input and fits a single regression line of system health. Both networks were trained on *in silico* patient data until both converge to a near optimally fit model for near future prediction and regression, respectively.

RESULTS: Task 1: The ANN had 3 hidden layers of 256/128/128 nodes and was able to identify clinical tipping points across clinically relevant parameter space. Progressive reduction in the size of the training set demonstrated an inflection point at 30% reduction of the training set with a preserved 88% accuracy. Task 2: LSTM forecasting models accurately predicted *in silico* patient disease progress with sufficient data sampling, despite the inherent stochasticity in the cytokine and patient health trajectories present in a high-dimensional space. Iterative predictions were used to calculate the boundaries of clinically plausible behavior generated by the model. There is forecast horizon for useful prediction that arises from the stochasticity within the system, suggesting the need for a "rolling" forecast that uses iterated data feedback to update the position of the neural network over time.

CONCLUSIONS: These two examples of demonstrate how machine learning can be used in conjunction with ABMs to increase the efficiency of clinically-oriented simulation investigations. In the first example, the use of an ANN to define the sample size of a putative in silico clinical trial while retaining essential information about the behavioral dynamics of the system based on its position in parameter space (the clinical tipping point) can reduce the computational requirements in any control discovery task. In the second example, the ability to train ANNs as surrogates for the complete ABM offers the promise of developing "hybrid" simulation experiments, where the full mechanistic model is used to implement a potential control/therapeutic intervention and the exploration of subsequent consequences of that intervention can be greatly accelerated by the use of the ANN surrogates.

REFERENCES:
1. An, Bull Math Biol. 2018 Dec;80(12):3071-3080
2. Cockrell and An, J Theor Biol. 2017 Jul 18;430:157-168

ACKNOWLEDGEMENTS: This work is supported by NIH Grant #UO1EB025825