THE USE OF AI AND ADVANCED COMPUTER SYSTEMS TO DEVELOP DRUGS AGAINST NEW EMERGING THREATS

Development Of A Ml/Al Model For Oligonucleotide Design For Respiratory Viruses

Stephen Hummel U.S. Army - West Point **Brian Lunn** United States Military Academy **Colette Muteteke** United States Military Academy **Tess Huchun-Walker** United States Military Academy **Bryant Nguyen** United States Military Academy **Richelle Ntumy** United States Military Academy **Jonah Landis** United States Military Academy

Early identification of pathogens, such as MERS, SARS, H1N1, RSV, and RV, is critical for reducing transmission through early detectior and providing the correct therapeutic interventions. Critical to detection and therapeutics are probes / binding molecules that demonstrate both high affinity and specificity. Single-strand oligonucleotides, or aptamers, due to their high specificity and affinity, have emerged as biosensor probes for early virus detection and as potential therapeutic agents for inhibiting viral entry or replication.

The aim of this project is to harness machine learning and artificial intelligence algorithms for the rapid and accurate design of high-affinity and specific single-stranded oligonucleotides, which can serve as biosensor probes or therapeutics. Leveraging the digitalSELEX platform, we focused on designing oligonucleotides targeting five respiratory viruses: MERS, SARS, H1N1, RSV, and RV, along with the human ACE2 receptor. These oligonucleotides were designed at varying lengths of 20, 30, 40, 50, and 60 nucleotides. Simultaneously, we implemented a design, build, test, and iterate model using the in silico data generated by digitalSELEX to initialize the classical machine learning methods, including Gaussian Naïve Bayes, Multilayer Perceptron, Decision Tree, and K-nearest neighbor. The preliminary results are promising yet the input data and training characteristics for these methods are unweighted. These classical machine learning techniques serve as a foundation for the subsequent implementation of deep learning methods, such as convolutional neural networks and recurrent neural networks.

The generated in silico design data, combined with the in vitro testing results, form the basis for constructing the machine learning training dataset. This integrated dataset will refine the machine learning models, further reducing the time and cost associated with the development of oligonucleotide probes and therapeutics. This project represents a significant advancement in the field, aiming to streamline the design process and facilitate the rapid development of biosensor probes and therapeutics for emerging pathogens.