2021

TERM RESEARCH PAPER

Topic: Artificial Intelligence for Vaccine Development

Rohan Salvi

ID: VH10935

Email: rsalvi2@umbc.edu

Instructor: Dr. Carlton Crabtree

University Of Maryland Baltimore & County

Table of Contents

1. Introduction

Artificial intelligence (AI) is a broad field of computer science that focuses on creating intelligent machines that can accomplish activities that would normally need human intelligence. The goal of Artificial Intelligence (AI) from its inception in 1956 has been to solve problems that humans can't solve, but with higher speed and precision. Natural swarm intelligences, such as those seen in groups of animals and insects, have influenced AI greatly. Swarm intelligences may be found not only in our environment, but even within the human body. The vertebral immune system, for example, may be seen of as a swarm intelligence since it is made up of separate, self-organized individuals that combine to generate a higher intelligence, from (Jabbari & Rezaei, 2019). The immune system has benefited AI, and an artificial immune system has been useful in tackling issues such as intrusion detection, robot self-healing, optimization, and anomaly detection. As much as immunology has benefited AI, AI is now being applied in a variety of immunological domains. Apart from exceptional analytic abilities, antigen and phenotypic identification, prognosis and treatment outcome prediction, and so on. Antigen and phenotype identification, predicting prognosis and treatment results, and other applications of AI are examples of how it might help immunologists in the field.

Machine learning is a subset of ai technology that allows a machine to learn from prior data without having to design it explicitly. From (Jabbari & Rezaei, 2019), (ML) is an AI tool that can take large amounts of input data, process it, and produce models for it to fit into. As a result, AI can identify inputs and forecast consequences. At the molecular level, several algorithms have been created to anticipate the result of interactions and categorization. For example, machine learning can forecast genotypes linked to poor prognosis. Another example of AI's categorization skills is phenotype detection. The existence of a disease or its consequence can be determined by cellular phenotype detection and categorization. The relevance of phenotypic categorization at the molecular level is especially obvious in the case of specific diseases since it can directly impact prognosis and treatment strategy. For example, because some phenotypes are associated with medication resistance, HIV-1 response to therapy can be anticipated based on viral phenotype. In certain circumstances, determining the genotype that leads to resistant phenotypes is simple since these viruses are so common. Because these viruses are prone to point mutations, determining the genotype that leads to resistant phenotypes can be simple in some situations. Identifying the processes that lead to resistance in other viruses, on the other hand, is more difficult. Resistance is influenced by several factors, and evaluating the impact of each element is difficult. ML can identify and categorize viral strains based on resistance-associated phenotypic traits. It can also exactly calculate the weight of each element, which is a difficult and time-consuming operation for traditional statistical analysis. This paper will cover research regarding how artificial intelligence and its subset machine learning could be used to develop/study vaccines. How various methodologies or approaches can be done to analyse reverse vaccinology, ensuring the best possible way to overcome the mutations of a virus. Adding over understanding the Protegenicity score (Ong et al., 2020) of vaccines.

2. Literature Review

2.1 History

With COVID-19's spread showing no indications of stopping and its comparatively high fatality rate when compared to other viral-based infections like influenza, developing vaccines and antiviral treatments against SARS-CoV-2 is critical. The National Institute of Allergy and Infectious Diseases (NIAID) funded the first clinical trial of an AI-based flu vaccine in the United States in 2019. 2 Scientists at Flinders University originally built the vaccine with an AI tool called synthetic chemist, which synthesized billions of synthetic chemicals. 2 and 3 The researchers then employed another AI software called Search Algorithm for Ligands (SAM), which utilizes AI to sift through billions of chemicals and decide which ones might be suitable vaccine adjuvant candidates. 2 and 3 This method can cut the time it takes to create a vaccine by several years. Although a vaccination for SARS-CoV-2 has been tested in humans, its effectiveness is uncertain. Screening might be aided by an AI-based method. Screening compounds as possible adjuvants for the SARS-CoV-2 vaccine, as well as screening new compounds based on modelling of probable changes to the novel coronavirus, might be aided by an AI-based method. This will make it easier for us to design vaccinations as the virus evolves, from (Ahuja et al., 2020).

2.2 Associated Field of Technology

Reverse vaccinology (RV), which attempts to uncover viable vaccine candidates through bioinformatics study of the pathogen genome, has transformed vaccine research in recent years. RV has been used to develop vaccines for diseases like as Group B meningococcus, which resulted in the approval of the Bexsero vaccine, from (Ong et al., 2020)

Graph Convolutional Neural Networks (GCNN) have been the preferred technology. These networks can process graphs and extract features by embedding adjacency information in the features. In drug property prediction, successful representation learning from compounds using GCNNs has been demonstrated, from (Arora et al., 2021).

3. Technical Details

The sections talk on the algorithm that is used to process the massive amounts of data. As a result, various forms of machine learning and other approaches were utilized to analyse the data.

3.1 Types of Machine Learning Models

- **SVM**
- **K-Nearest**

- **RF**
- **KNN**
- **Bernoulli Nave Bayesian**
- **AdaBoosted Decision tree**

From (Gawriljuk et al., 2021), to see how different methods would perform in identifying new chemicals for Yellow Fever activity, several machine learning models (Rf, SVM, K-nearest neighbours (KNN), Bernoulli Nave Bayesian (BnB), AdaBoosted decision trees (ada), and deep neural network (DNN) of three layers) were created in their research. The same training data was used for both approaches. Metrics from both training and external test data validations were used to assess model performance. The performance of each machine learning approach was examined using several statistical measures in both training data and prediction accuracy for external test data, utilizing fivefold cross validation. Specificity, recall, precision, F1-score, accuracy, Cohen's Kappa (CK), Mathew's correlation coefficient (MCC), and area under the receiver operating characteristic curve were among the measurements utilized (AUC).

3.2 Other Approaches

3.2.1 Agent-based models

From (Russo et al., 2020), agent-based models (ABM) are one of the foundations of artificial intelligence's computational arm, as well as artificial life in general . ABM's key concern is how systems might achieve a higher-order state through processes that lead to complicated emergent order behaviour using computational techniques. The bottom-up method to modelling complex systems includes agent-based modelling. It specifically describes each individual entity, defining them based on their conceivable status and traits, as well as their behaviours based on the processes through which they interact.

3.2.2 Knowledge Discovery Approaches

From (Russo et al., 2020), Scientific articles, functional annotation of antigens, specialist databases, and disease clinical trials are all part of the knowledge-based technique. It also mines such data using machine-learning techniques and visualizes the findings for downstream modules to easily comprehend.

3.2.3 Antigen and Epitope Prediction Approaches

- Reverse Vaccinology (RV)
  Reverse Vaccinology (RV) is a method of identifying possible vaccine candidates within a pathogen's proteome using computational methods.

  - **Vaxign-ML (Supervised ML Classification)**
    Vaxign-ML is a supervised machine learning algorithm that predicts bacterial protective antigens (BPAgs), (Ong et al., 2020). Vaxign-ML

is offered as a standalone version and is hosted on a publicly accessible web server.

- **Nerve**
  From (Vivona et al., 2006), NERVE (New Enhanced Reverse Vaccinology Environment) is a user-friendly software environment for identifying the best vaccine candidates in silico using bacterial pathogen proteomes. For protein analysis and comparison, the program incorporates a number of reliable and well-known methods.

- **VaxiJen**
  VaxiJen is the first server to forecast protective antigens without using alignment. It was created to allow antigen categorization exclusively based on protein physicochemical characteristics, rather than sequence alignment. The server can be used independently or in tandem with alignment-based prediction algorithms, from (Doytchinova & Flower, 2007).

- **Jenner-Predict**
  Jenner-Predict is a program that predicts PVCs (Protein Vaccine Candidates) from the proteomes of bacterial pathogens. The web server considers known functional domains from protein classes to target host-pathogen interactions and disease, from (Jaiswal et al., 2013).

3.3 Murcko Scaffold and Clustering Analysis

Murcko Scaffold approach divides a molecule into four parts: ring systems, linkers, side chains, and the Murcko framework, which is the intersection of ring systems and linkers in a molecule. To assess the structural variety of the collected data, a scaffold analysis was done. The free source program DataWarrior was used to create Murcko scaffolds, from (Gawriljuk et al., 2021)

3.4 Virtual Screening

Virtual screening (VS) is a drug discovery process that searches libraries of small molecules for structures that are most likely to bind to a therapeutic target, usually a protein receptor or enzyme, from (Arora et al., 2021).

3.5 Graph Convolutional Neural Networks (GCNN)

From (Arora et al., 2021), It is the process of multiplying the input neurons by a series of weights known as filters or kernels. CNNs may learn information from surrounding cells thanks to the filters, which operate as a sliding window over the entire image.

4. Risks & Limitations

4.1 Risks

More and more immunological issues are becoming "big data" challenges as a result of the present data amount, heterogeneity, and continuous flow. The discovery of signals and trends in highly noisy data sources, the integration of multi-omics immunological data, the incorporation of past information held in public repositories, and the necessity for automated, interpretable choices are all examples of big data difficulties in immunology. A new generation of computational techniques must be created to overcome these problems.

4.2 Limitation

In preclinical and clinical trials, AI may also imitate nonhuman and human participants. With the correct data set to train on, AI may be able to take the place of human volunteers in clinical studies, at least in the early stages. However, an AI implementation may raise ethical concerns because there may be features of in vivo interactions that cannot be foreseen in silico. Stakeholders and researchers must come to an agreement in order to go forward with in silico clinical trials. However, as computational medicine advances, several problems arise, such as how far AI can advance in medical disciplines. Should the ultimate objective of AI in medicine be to replace humans in clinical trials? How close are we to achieving this ostensibly final goal? Until 2016, neither the US Food and Drug Administration nor the European Medicines Agency allowed any primary evidence obtained through computer simulation to be used in the regulatory process of evaluating a medicinal product or medical intervention. However, the European Parliament and the US Congress have made specific recommendations to regulators to allow broader use of modelling and simulation in the regulatory process, introducing the urge, from (Russo et al., 2020).

5. Implication on Society & Industry

5.1 Impact on human lifestyle

Artificial intelligence (AI) has a great impact on our daily living and makes our lives more efficient and productive. Especially during the coronavirus disease (COVID-19) pandemic, AI has played a key role in response to the global health crisis. There has been a boom in AI innovation and its use since the pandemic. However, despite its widespread adoption and great potential, most people have little knowledge of AI concepts and realization of its potential. Apart from reducing time and money, AI-based technologies have the ability to simulate wholly artificial investigations. One of AI's primary strengths is its capacity to anticipate interactions like peptide-MHC affinity, which may be exploited to produce novel prophylactic/therapeutic treatments like neoantigen vaccines, from (Keshavarzi Arshadi et al., 2020).

5.2 Impact on Industry

Machine learning is still in its infancy when it comes to vaccine creation. A machine learning model is only as good as the data it is trained on, and current immunology models are taught on far smaller datasets than models for voice recognition or face identification, which are areas

where artificial intelligence has thrived. Immunology models will be more reliable if more and more diverse datasets are generated, and their potential effect on the field will be immense.

6. Suggested Course of Action

Pharmaceutical businesses and fundamental scientific institutes must collaborate with industry heavyweights with substantial AI capabilities, such as Google and IBM, at this time of need. With COVID-19 spreading throughout our communities, it's time to put our AI skills to work in the race to produce a vaccine and antivirals that work, from (Ahuja et al., 2020).

7. Conclusion

From (Ahuja et al., 2020), COVID-19 cases are expected to skyrocket in the coming weeks, both in the United States and throughout the world. Because of the high virality of SARS-CoV-2, the pandemic is causing a slew of challenges on the ground, including a lack of beds, ventilators, masks, gowns, and medical personnel, to name a few. Though the situation appears dire, it is critical to highlight AI technologies that have the potential to alleviate some of the most acute and relevant issues that this epidemic has caused. Trillions of molecules may be tested in a very short length of time using AI, enabling for the fast identification of therapeutic and vaccine candidates.

The pre-computational age of medicine produced enough data on illness status, prognostic markers, and therapeutic techniques to allow AI to utilise these data to develop better solutions. Computational medicine has advanced dramatically during the last few decades. Machines can now search databases, obtain relevant data, and use that data to anticipate outcomes and devise therapeutic actions for a specific individual. In combination with systems biology, which offers system-level interrelated data, a variety of artificial intelligence methods have been widely employed for improving vaccination development. However, approaches for managing the vaccine design process effectively are still required. A promising notion that shows how certain subfields of both methodologies are linked might help with vaccine candidate rational design.

8. Annotated Bibliography

Ong, E., Wang, H., Wong, M. U., Seetharaman, M., Valdez, N., & He, Y. (2020). Vaxign-ml: Supervised machine learning reverse vaccinology model for improved prediction of bacterial protective antigens. *Bioinformatics*, *36*(10), 3185–3191.

Overall, Vaxign-ML outperformed all other BPAg prediction approaches when it came to BPAg prediction. The first RV approach that integrate both biological and physicochemical features is Vaxign-ML. The importance of biological and physicochemical parameters in ML-based RV prediction was also proven in our research. Finally, Vaxign-findings ML's underlined the importance of physicochemical features, which might have implications for structural

vaccinology. ML approaches were evaluated with biological and physiochemical variables taken from well-defined training data to find the best ML method with optimum settings. To provide fair performance evaluation and the capacity to forecast vaccine candidates against a new developing disease, nested 5-fold cross-validation and leave-one-pathogen-out validation were applied. Using a high-quality benchmark dataset, the highest performing model was compared against three publicly accessible programs (Vaxign, VaxiJen, and Antigenic), one SVM-based approach, and one epitope-based method. Vaxign-ML outperformed other methods in predicting BPAgs. Vaxign-ML was used to compute and rank the respective protegenicity scores of five clinical trial MTB vaccinations and one commercial DTP vaccine as a final validation (C.diphtheriae, C.tetani, and B.pertussis). Vaxign-ML trained on the complete original data provided a percentile rank score for the protegenicity score. These six vaccinations contained a total of 20 proteins, all of which had a projected protegenicity score of > 90%. In other words, the Vaxign-ML classified these 20 proteins in the top 10% of BPAg candidates.

This paper was important since it evaluated and compared benchmarking of protective antigen prediction programs (Vaxign, VaxiJen and Antigenic), the Heinson–Bowman SVM-based method and the epitope-based method giving an understanding of how ML could be used in antigen prediction.

Ong, E., Cooke, M. F., Huffman, A., Xiang, Z., Wong, M. U., Wang, H., Seetharaman, M., Valdez, N., & He, Y. (2021). Vaxign2: The second generation of the first web-based vaccine design program using Reverse Vaccinology and machine learning. *Nucleic Acids Research*, *49*(W1).

Vaccination is one of medicine's most important inventions. RV (reverse vaccinology) is a cutting-edge method for predicting vaccine candidates from a pathogen's DNA (s). They upgraded Vaxign2, the first web-based vaccine design application that used reverse vaccinology and machine learning, to help in vaccine development. Vaxign2 is a web server that includes predictive and computational workflow components for rational vaccine creation. The old Vaxign filtering-based approach plus a new machine learning-based method, Vaxign-ML, make up the predictive component. The results of the benchmarking using a validation dataset revealed that Vaxign-ML outperformed other RV tools in terms of prediction performance. Aside from the prediction component, Vaxign2 added a number of post-prediction analyses to improve users' capacity to modify prediction findings based on multiple vaccine design rationales and minimize user time spent analysing Vaxign/Vaxign-ML prediction results. Users input proteome sequences, choose candidates based on Vaxign outputs and Vaxign-ML scores, and do post-prediction. Perform post-prediction analysis using Vaxign-ML scores. Vaxign2 also offers pre-calculated findings from over a million proteins in 398 proteomes from 36 infections. Vaxign2 was used to analyze SARS-CoV-2, the coronavirus that causes COVID-19, as a demonstration. Vaxign2's comprehensive architecture can aid in the development of better and more logical vaccines.

This paper was important since the author made a improvised version of Vaxign as Vaxign2 which is a web based vaccine design program using Reverse Vaccinology.

Vivona, S., Bernante, F., & Filippini, F. (2006). Nerve: New enhanced reverse vaccinology environment. *BMC Biotechnology*, *6*(1).

By applying in silico prediction stages to replace various experimental activities, Reverse Vaccinology has greatly improved the discovery of vaccine candidates. These procedures have allowed scientists to choose antigens from viruses' projected proteomes, which is difficult or impossible to do in cell culture, saving time and money. However, by refining in silico procedures and incorporating biologist-friendly tools, this good example of bioinformatics-driven immunology might be further expanded. They provide NERVE (New Enhanced Reverse Vaccinology Environment), a user-friendly software environment for identifying the best vaccine candidates in silico using bacterial pathogen proteomes. For protein analysis and comparison, the program incorporates a number of reliable and well-known methods. Vaccine candidates are graded and shown in an HTML table with pertinent information and links to main material. During the selection process, information about all proteins in the studied proteome is not destroyed, but instead goes into a SQL database for subsequent mining and analysis. introduce NERVE (New Enhanced Reverse Vaccinology Environment), a user-friendly software environment for identifying the best vaccine candidates in silico from whole bacterial pathogen proteomes. NERVE has been applied and fine-tuned as the first technology capable of ranking a limited pool of vaccine candidates (8–9% of the whole proteome) and demonstrating strong recall (75–80%) of known protective antigens. These vaccination candidates must be "safe" (taking into consideration the possibility of autoimmunity) and "simple" to test in high-throughput experiments (avoiding possibly not soluble antigens).

This paper helped to understand one of the reverse vaccinology methods.

Doytchinova, I. A., & Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinformatics*, *8*(1).

Models for predicting entire protein antigenicity were developed using bacterial, viral, and tumor protein datasets. There were 100 known antigens and 100 non-antigens in each batch. Internal leave-one-out cross-validation and external validation utilizing test sets were used to validate the resulting models. To assess the durability of the discriminating between antigens and non-antigens, an extra five training sets for each class of antigens were employed. In both validations, the models performed well, with prediction accuracy ranging from 70% to 89 percent. The capacity of the models to predict whether a protein sequence will be a protective antigen or not is the most important conclusion of this research. Subunit vaccinations are based on antigens like these. VaxiJen, a server that allows users to assess a protein's capacity to induce protection, was created to make the usage of the generated models easier. Single proteins as well as complete proteomes supplied in fasta format are handled by the server. Models for parasite and fungal antigens will be created in the future and incorporated in the VaxiJen server, as the approach is generic. VaxiJen is the first server to forecast protective antigens without using alignment. It was created to allow antigen categorization exclusively based on protein physicochemical characteristics, rather than sequence alignment.

The server can be used independently or in tandem with alignment-based prediction algorithms.

This paper helped to understand alignment-independent prediction method to understand reverse vaccinology in a better way.

Jaiswal, V., Chanumolu, S. K., Gupta, A., Chauhan, R. S., & Rout, C. (2013). Jenner-predict server: Prediction of protein vaccine candidates (pvcs) in bacteria based on host-pathogen interactions. *BMC Bioinformatics*, *14*(1).

Subunit vaccines based on recombinant proteins have proven to be beneficial in the prevention of infectious illnesses and are likely to satisfy future vaccine development demands. The potential for identifying protein vaccine candidates (PVCs) from a proteome using a computational technique, particularly the reverse vaccinology (RV) method, is immense. Due to the low prediction accuracy of current protective antigen prediction tools and web servers, vaccine development applications are limited. Apart from machine learning approaches, the software and web servers used just the protein's adhesin-likeliness as a criteria for PVC identification. Protection against bacterial infections is known to be provided by several non-adhesin functional groups of proteins involved in host-pathogen interactions and pathogenesis. As a result, understanding bacterial pathogenesis has the potential to help detect PVCs. The Jenner-Predict web server's prediction accuracy indicates that domains associated in host-pathogen interactions and pathogenesis are superior criteria for PVC(Protein Vaccine Candidates) prediction. The web server correctly estimated the maximum number of known PVCs for each functional class. NERVE, Vaxign, and VaxiJen algorithms were all surpassed by the Jenner-Predict server. The Protegen and VaxiJen datasets have sensitivity of 0.774 and 0.711, respectively, whereas the latter dataset has a specificity of 0.940.

This paper helped to understand prediction methodology in reverse vaccinology.

Rizwan, M., Naz, A., Ahmad, J., Naz, K., Obaid, A., Parveen, T., Ahsan, M., & Ali, A. (2017). Vacsol: A high throughput in silico pipeline to predict potential therapeutic targets in prokaryotic pathogens using subtractive reverse vaccinology. *BMC Bioinformatics*, *18*(1).

With advancements in reverse vaccinology methodologies, the prediction of potential vaccine candidates has steadily improved. Reverse vaccinology has revolutionized the field of discovery, allowing researchers to propose target identification in less time and with less effort. In this context, high-throughput genome sequencing technology and related bioinformatics tools have substantially aided the rapid examination of pathogens, with numerous projected candidates shown to be successful against various infections and disorders. They aimed to use the reverse vaccinology technique to overcome the constraints of already existing pipelines because the earlier prediction tools had minimal software restrictions. They focused on an in silico reverse vaccinology strategy to solve the problems with earlier pipelines and to precisely screen putative vaccine candidates from the whole bacterial genome in silico. They

created VacSol, an automated system for rapidly screening therapeutic vaccine molecules from the bacterial pathogen proteome, saving time and resources. VacSol, a pipeline based on a similar methodology, is aimed to forecast putative vaccine candidates quickly and effectively. VacSol screens the whole bacterial pathogen proteome quickly and effectively to discover a few projected potential vaccine candidate proteins. By effectively decreasing false positive candidate hits, this pipeline has the potential to reduce computational expenses and time. The outcomes of VacSol are not predicated on any universal set of principles and may vary depending on the input. Vaccine candidates are screened utilizing integrated, well-known, and robust algorithms/tools for proteome analysis, and the VacSol software output is delivered in five distinct forms based on the proteome sequence input in FASTA file format.

This paper helped to understand prediction methodology in reverse vaccinology.

Jabbari, P., & Rezaei, N. (2019). Artificial Intelligence and immunotherapy. *Expert Review of Clinical Immunology*, *15*(7), 689–691.

As much as immunology has benefited AI, AI is now being applied in a variety of immunological domains. Methodologies for machine learning that are commonly utilized in immunology include: Data must first be pre-processed before being processed by machine learning algorithms. Depending on the kind of data, numerous pre-processing approaches like as denoising, feature extraction, and so on are available. The data is then analysed using various machine learning methods. Image-based phenotypic detection is another method of phenotype detection. At both the cellular and molecular levels, image-based approaches show exceptional precision. Traditional microscopic visual assessment of tissues takes time and is vulnerable to subjectivity. Furthermore, they are unable to keep up with the volume of data generated by high-throughput investigations. Machines are capable of not just keeping up with the rate at which data is created in contemporary medicine, but also of detecting phenotypic alterations that would otherwise go undetected by human screening. The pre-computational age of medicine produced enough data on illness state, prognostic markers, and therapeutic techniques to allow AI to utilise these data to develop better solutions. Computational medicine has advanced dramatically during the last few decades. Machines can now search databases, obtain relevant data, and use that data to anticipate outcomes and devise therapeutic actions for a specific individual. They believe that more studies have the potential to recruit AI. This is particularly true in immunology, where there is a close link between the molecular and clinical levels.

This paper helped to understand association of AI and ML with immunotherapy. Basic information that how can one take help of AI and ML to use it in the clinical and pre-clinical fields.

Ahuja, A. S., Reddy, V. P., & Marques, O. (2020). Artificial Intelligence and covid-19: A multidisciplinary approach. *Integrative Medicine Research*, *9*(3), 100434.

The COVID-19 pandemic caused enormous misery and loss of life. Due to the pandemic's rapid spread, a large amount of new scientific research and data sharing is occurring. The authors used artificial intelligence (AI) and a rising number of coronavirus-related datasets and published articles to combat the pandemic by pushing innovative approaches to drug discovery, vaccine development, and public awareness. By cross-referencing papers and searching for patterns, AI algorithms may be used to mine this avalanche of new data and publications to extract new insights. AI algorithms could help uncover new possible treatments or aid in vaccine development. Drug development is a difficult undertaking, and AI technologies such as deep learning can assist speed up the process by predicting which existing medications or drug-like compounds could be used to treat specific conditions of COVID-19.

This paper was important as it explained how AI could be beneficial for making a COVID-19 drug discovery or vaccine development.

Ong, E., Cooke, M. F., Huffman, A., Xiang, Z., Wong, M. U., Wang, H., Seetharaman, M., Valdez, N., & He, Y. (2021). Vaxign2: The second generation of the first web-based vaccine design program using Reverse Vaccinology and machine learning. *Nucleic Acids Research*, *49*(W1).

Reverse vaccinology (RV) is a cutting-edge technique for predicting vaccine candidates from a pathogen's DNA, according to the paper. The authors modified Vaxign2, the first web-based vaccine design application employing reverse vaccinology and machine learning, to encourage vaccine development. Vaxign2 is a web server that includes predictive and computational workflow components for rational vaccine creation. The old Vaxign filtering-based method plus a new machine learning-based method, Vaxign-ML, make up the predictive part. The results of the benchmarking using a validation dataset revealed that Vaxign-ML outperformed other RV tools in terms of prediction performance. Aside from the prediction component, Vaxign2 added a number of post-prediction analyses to improve users' capacity to refine prediction results based on multiple vaccine design rationales and reduce user time spent analyzing Vaxign/Vaxign-ML prediction results. Users input proteome sequences, choose candidates based on Vaxign outputs and Vaxign-ML scores, and do post-prediction analysis. Vaxign2 also offers pre-calculated findings from over a million proteins in 398 proteomes from 36 infections. Vaxign2 was used to analyze SARS-CoV-2, the coronavirus that causes COVID-19, as a demonstration. Vaxign2's comprehensive architecture can aid in the development of better and more logical vaccines.

This paper used machine learning techniques on the proteins of SARS-CoV-2 coronavirus , Since machine learning is a subset of artificial Intelligence(AI) that provides systems the ability to automatically learn and improve from

experience with being explicitly programmed. This is the second generation of author's 1st vaccine design program.

Ong, E., Wong, M. U., Huffman, A., & He, Y. (2020). Covid-19 coronavirus vaccine design using reverse vaccinology and machine learning.

The entire virus, as well as the spike (S) protein, nucleocapsid (N) protein, and membrane protein, have all been studied for vaccine development against SARS and MERS, according to the author's literature and clinical trial review. To anticipate COVID-19 vaccine candidates, the authors employed the Vaxign reverse vaccinology tool and the newly built Vaxign-ML machine learning technology. The authors found that the N protein was found to be conserved in the more pathogenic strains (SARS/MERS/COVID-19), but not in the milder human coronaviruses. Six proteins, including the S protein and five non-structural proteins (nsp3, 3CL-pro, and nsp8–10), were anticipated to be adhesins, which are important for viral adhesion and host invasion, according to the authors. Vaxign-ML projected that the S, nsp3, and nsp8 proteins would generate high protective antigenicity. Aside from the regularly utilized S protein, the nsp3 protein has never been tested in a coronavirus vaccination study and was chosen for future research. The nsp3 gene was discovered to be more conserved in SARS-CoV-2, SARS-CoV, and MERS-CoV than in 15 coronaviruses that infect humans and other animals. Promiscuous MHC-I and MHC-II T-cell epitopes, as well as linear B-cell epitopes, were expected to be found in certain locations and functional domains of the protein. Our projected vaccine targets offer fresh approaches to developing a COVID-19 vaccine that is both effective and safe.

This is the author's 1st generation vaccine design code using machine learning and reverse vaccinology . This paper was very helpful for my topic since machine learning is a subset of AI

Keshavarzi Arshadi, A., Webb, J., Salem, M., Cruz, E., Calad-Thomson, S., Ghadirian, N., Collins, J., Diez-Cecilia, E., Kelly, B., Goodarzi, H., & Yuan, J. S. (2020). Artificial Intelligence for covid-19 drug discovery and vaccine development. *Frontiers in Artificial Intelligence*, *3*.

There were no new antiviral medicines or licensed vaccinations available for deployment as a frontline defence at the time this research was produced by the authors. By clarifying hitherto undiscovered viral pathways, scientists may be able to contribute in the identification of effective antivirals. Using computational approaches to find novel candidate medications and vaccines in silico is one way to accomplish this. Machine learning-based models trained on specific proteins have provided low-cost and quick-to-implement ways for finding successful viral treatments in the previous decade. These models can forecast inhibitor candidates in a structural-based manner given a target biomolecule. If a model is given enough data, it can help with the search for a medication or vaccine candidate by identifying trends in the data. The authors of this paper focus on current breakthroughs in COVID-19 drug and vaccine

research employing artificial intelligence, as well as the possibility of intelligent training for COVID-19 therapeutic discovery.

The authors highlight various molecular targets of COVID-19, suppression of which may increase patient survival, to make deep learning applications for SARS-COV-2 easier. Furthermore, the researchers published CoronaDB-AI, a library of chemicals, peptides, and epitopes found in silico or in vitro that may be used to train models to extract COVID-19 treatment. This review's data and information can be utilized to train deep learning-based models and speed up the development of successful viral medicines. This was a important paper relating to my research .

Keshavarzi Arshadi, A., Webb, J., Salem, M., Cruz, E., Calad-Thomson, S., Ghadirian, N., Collins, J., Diez-Cecilia, E., Kelly, B., Goodarzi, H., & Yuan, J. S. (2020). Artificial Intelligence for covid-19 drug discovery and vaccine development. *Frontiers in Artificial Intelligence*, *3*.

The authors of this review study emphasized the importance of artificial intelligence (AI) and machine learning (ML) techniques in finding a possible therapy for COVID-19. They also looked at how viromics and AI interact, in the hopes of finding a solution to the epidemic. For recent studies on the usage of AI, a review of different articles was conducted using the following databases: MEDLINE/PubMed, SCOPUS, Web of Science, ScienceDirect, and Google Scholar, looking for the spread of different infectious diseases using relevant MeSH subheadings. 30 papers were chosen after a careful review of various articles, and key information was extracted from them. Finally, in order to gather more data, the authors widened their focus. Their findings suggested that AI/ML could be a promising drug development strategy.

The topic of artificial intelligence (AI) holds immense potential for forecasting environmental changes. Breakthroughs could pave the path for novel vaccinations and antiviral medications if this technology is employed in pandemic conditions like COVID-19.

McGowan, E., Rosenthal, R., Fiore-Gartland, A., Macharia, G., Balinda, S., Kapaata, A., Umviligihozo, G., Muok, E., Dalel, J., Streatfield, C., Coutinho, H., Monaco, D. C., Morrison, D., Yue, L., Hunter, E., Nielsen, M., Gilmour, J., & Hare, J. (2020). Utilizing computational machine learning tools to understand immunogenic breadth in the context of a CD8 T-cell mediated HIV response.

The authors of this paper used the concept of assessing the HIV proteome for defined regions of immunogenicity using two key parameters: a diversity metric of individuals' HLA profiles within a population and consideration of sequence diversity in the context of an individual's CD8 T-cell immune repertoire. The discovery of areas within the proteome that offer high conservation, HLA recognition within a population, low prevalence of HLA adaptation, and shown immunogenicity using this method was made possible

by analysing HLA adaptation and functional immunogenicity data. As a supplement to vitro functional assays, the authors felt that this unique and original approach to vaccine design offered a tailored pipeline for accelerated and rational CD8 T-cell vaccine creation for HIV and potentially other infections, with the potential for both global and local coverage.

This paper utilized the use of conceptual machine learning tools, this helped to understand how a subset of AI could be useful in understand the T-cell meditated HIV response

Gawriljuk, V. O., Foil, D. H., Puhl, A. C., Zorn, K. M., Lane, T. R., Riabova, O., Makarov, V., Godoy, A. S., Oliva, G., & Ekins, S. (2021). Development of machine learning models and the discovery of a new antiviral compound against yellow fever virus. *Journal of Chemical Information and Modeling*, *61*(8), 3804–3813.

This paper mentions about Yellow fever ,which is a haemorrhagic viral disease spread by infected mosquitoes. When the virus is brought into densely inhabited areas with high mosquito density and insufficient immunization coverage, large outbreaks of YF develop. The lack of an unique small molecule medicinal treatment for YF, as well as analogous illnesses like zika and dengue fever, emphasizes the flaviviruses importance as a public health threat. New techniques based on machine learning approaches have been introduced into drug discovery as a means to use the rising high throughput screening (HTS) data collected to cut costs and speed up drug development, thanks to advancements in computer hardware and bioactivity data availability. In this research, predictive machine learning models were utilized to enable the identification of compounds with optimal bioactivity and absorption, distribution, metabolism, and excretion profiles utilizing previously published data from HTS campaigns or data available in public sources. In this research, predictive machine learning models were utilized to enable the identification of compounds with optimal bioactivity and absorption, distribution, metabolism, and excretion profiles utilizing previously published data from HTS campaigns or data available in public sources.

The authors gathered data on yellow fever virus cell-based assays from the literature and public databases. The information was used to create predictive models using a variety of machine learning techniques to help prioritize drugs for in vitro testing.

Russo, G., Reche, P., Pennisi, M., & Pappalardo, F. (2020). The combination of artificial intelligence and Systems Biology for Intelligent Vaccine Design. *Expert Opinion on Drug Discovery*, *15*(11), 1267–1281.

According to the author, a new body of evidence demonstrates the use of artificial intelligence and systems biology in vaccine research and design. Both ideas will transform healthcare by speeding up clinical trial processes and decreasing the expenses and time spent on medication research and development. The principles of artificial intelligence and systems biology techniques in the vaccine development pipeline are explored in this paper. A full discussion of epitope prediction tools for creating epitope-based vaccinations and agent-based models for immune system response prediction, as well as their potential to speed up clinical trial phases, are among the subjects covered. Artificial intelligence in silico trials approaches fed by systems biology data in the vaccine development pipeline are currently being approached by regulatory bodies for qualification/approval.

The collection of specific data from a combined strategy like this could help one better understand and monitor viral propagation and immune responses in the wake of the 2019-Novel Coronavirus (2019-nCoVrecent)'s debut and rapid spread, as well as the sickness it causes (COVID-19). Furthermore, a combined methodology like this can allow health experts and biologists to focus their time and energy on how to respond to infectious disease risks and quickly identify potential targets for immune responses to 2019-nCoV, allowing for the development of new vaccine interventions to be sped up.

Kimaina, A., Dick, J., DeLong, A., & Hogan , J. W. (2021). Comparison of forecasting models in the HIV epidemiology using Machine Learning Methods. *Applied Machine Learning and Multi-Criteria Decision-Making in Healthcare*, 98–123.

Their goal is to characterize and compare the predictive accuracy of several statistical machine learning methods for predicting viral failure at the first and second measurements following initiation of antiretroviral therapy using electronic health record data from a large HIV care program in Kenya. Sensitivity, specificity, and the area under the receiver-operator characteristic curve are used to assess predictive accuracy. They used data from over 10,000 patients in the Academic Model Providing Access to Healthcare care program in western Kenya to train and cross-validate 10 statistical machine learning models and algorithms. Parametric, non-parametric, ensemble, and Bayesian approaches were among them. 50 elements from the clinical record were chosen by hand in conjunction with physician specialists as input variables. 10-fold cross validation was used to derive predictive accuracy values. At both the first and second assessments, the viral load failure rate in this patient population is around 20%. Other approaches were often surpassed by ensemble strategies. The specificity of these approaches for predicting viral failure at the first

follow-up measure was above 90%, while the sensitivity was frequently in the 50–60% range. The second follow-up measure had a higher predictive accuracy, with sensitivities above 80%. Other approaches were regularly outperformed by Super Learner, gradient boosting, and Bayesian additive regression trees. The positive predictive value for the top-performing approaches is between 75 and 85 percent for a viral failure rate of 20%, whereas the negative predictive value is over 95 percent.

According to the findings of this study, machine learning algorithms have the ability to detect patients who are at risk of viral failure before they have their planned assessments. Finally, predictive virologic evaluation can aid in the delivery of early targeted interventions such as increased medication resistance monitoring, adherence counselling, or suitable next-line therapy switching.