Biomedical
Communication

BBRC
Bioscience Biotechnology
Research Communications

Biosc.Biotech.Res.Comm. Vol 13 (2) April-May-June 2020 Pp-762-769

Machine Learning Model for Vaccine Development: A Perspective

Anubha Dubey
Independent Researcher and Analyst Bioinformatics Gayatri Nagar Katni, M.P. India

ABSTRACT

A vaccine is a hope to prevent disease while training the immune system to produce antibodies against pathogens. As it is safe, easy in use, and having no side effects, it is used for the cure of many diseases. Vaccines may be of different types, like subunit vaccines, attenuated vaccines, DNA vaccines, etc. The process of vaccine development is taking a long way, needs highly sophisticated labs, clinical trials, and much more. This whole process will take a long time to develop a vaccine. And then manufacturing in the specific environment will again time is taken to reach to the market. While clinical trials of vaccines sometimes failed to produce the desired results. So to improve these trial methods and making vaccine production successful. Here in this paper, it is tried to propose machine learning methods i.e. classification, clustering, association (mainly used algorithms) in different stages of vaccine development clinical trials. Because machine learning techniques are soft, time-consuming, and help to achieve a particular target with great sensitivity and accuracy. It is hoped that if machine learning methods are followed in a proper way time will be saved and vaccine designing will be done in less time with great accuracy, sensitivity, and specificity.

KEY WORDS: CLINICAL, DISEASE IMMUNE SYSTEM, MACHINE LEARNING, VACCINE.

INTRODUCTION

A biological preparation that provides active acquired immunity to a particular disease is called vaccine. Typically it contains an agent that resembles a disease-causing microorganism. It is seen that vaccines are made from weakened or killed forms of the microbe, its toxins, or one of its surface proteins. When a vaccine is administered to any person by injection is referred as vaccination. These vaccinations have some effects on our body, which is called Immunization. Some 80 percent of the world's infants are adequately immunized against six important diseases: measles, tetanus, pertussis diphtheria, tuberculosis, and polio. This is a remarkable achievement which protects children life at an early age (Tomic et al., 2019).Vaccines based on viral vectors (tools that deliver genetic material into cells) offer strong immune response, which are based on recombinant proteins for other diseases; these vaccine candidates have an advantage of large scale production capacity, (Tung et al., 2020).

Hence we can say that the vaccine stimulates the immune system so that it can recognize the disease and protect us from the future infections (i.e. provide immunity to the infection). Since vaccine is cost effective and it will reach many of the lives. So it is necessary that the vaccine delivery is also done in a proper way. Its tremendous effect

ARTICLE INFORMATION

*Corresponding Author: anubhadubey@rediffmail.com
Received 13th April 2020 Accepted after revision 27th May 2020
Print ISSN: 0974-6455 Online ISSN: 2321-4007 CODEN: BBRCBA

Thomson Reuters ISI Web of Science Clarivate Analytics USA and Crossref Indexed Journal

NAAS Journal Score 2020 (4.31) SJIF: 2020 (7.728)

*Online Contents Available at: http//www.bbrc.in/
DOI: 10.21786/bbrc/13.2/58*

and reach will soon eradicate global disease burden. The best example till date is poliomyelitis. Importance of Vaccines (Kallarup 2014) are noted as: a) Accelerate immunologic response, b) It has no side effects, c) Easy in use, d) Prevents from disease, e) Develops fast antibody against pathogens, f) Immunity is achieved with minimum doses, g) Mass production is available, h) Stable in storage conditions for long period of time. There are some vaccines which induces better immunity than any natural infection (Kallarup 2014, Chaudhary 2019):

Table 1. Modes of action of vaccines with infection caused are detailed below

Accine type	Infection Cured	Method of vaccine formation	Working of vaccine	References
Viral vector based (i.e. adenovirusas vector)	Alzheimer's disease, malaria,HIV	It uses chemically weakened virus to transport pieces of pathogen. The genes in such vaccines are usually antigen coding surface proteins from the pathogenic organism.		
Live, attenuated	Measles, mumps, rubella (MMR combined vaccine) Varicella (chickenpox) Influenza (nasal spray) Rotavirus Zoster (shingles) Yellow fever	A virus targeted for use in a vaccine may be grown through—"passaged" through—upwards of 200 different embryos or cell cultures.	It will unable to replicate and produce good immune response in future	Plotkin (2013); Plotkin (2018)
Inactivated/Killed	Polio (IPV) Hepatitis A	This can be made by inactivating a pathogen, typically using heat or chemicals such as formaldehyde or formalin. This destroys the pathogen's ability to replicate, but keeps it "intact" so that the immune system can still recognize it.	tend to provide a shorter length of protection than live vaccines are more likely to require boosters to create long-term immunity	Plotkin (2018) Plotkin (2018)
Toxoid (inactivated toxin) (Toxoids is considered killed or inactivated vaccines)	Diphtheria, tetanus (part of DTaP combined immunization) Rabies	It produces (tetanospasmin) neurotoxin. Immunizations for this type of pathogen can be made by inactivating the toxin that causes disease symptoms. This can be done via treatment with a chemical such as formalin, or by using heat or other methods.	Long term immunity, Used as a combination.	Plotkin (2018) Angsantikul et al (2018).
Subunit/ conjugate	Hepatitis B Influenza (injection) Haemophilus influenza type b (Hib)	One way of formation is both subunit and conjugate vaccines containing only pieces	Conjugate vaccines are sed to create a more u powerful, combined immune response	Carvelho 2010, Tomic et al 2019

	Pertussis (part of DTaP combined immunization) Pneumococcal Meningococcal Human papillomavirus (HPV)	of the pathogens they protect against. Another way is genetic engineering.		
Recombinant vaccines	Influenza, Rabies, Hepatitis B and other diseases	Researchers identify the region in virus DNA that is not necessary for replication. Hence this region is used for vaccine. Researchers put one or more genes code for immunogen of other pathogens.	Such modified virus is injected into a person's body; the immunogen is expressed and able to generate immune response.	Carvelho (2010), Tomic et al., 2019
DNA Vaccines	Parasitic diseases like malaria etc.	It consists of DNA coding for a particular antigen.	DNA itself insert into particular cells which then produce the antigen from the infectious agent and able to produce immune response.	Carvelho (2010)

- *Human papillomavirus* (HPV) vaccine – the specific protein is highly pure so the immunity will increase.
- Tetanus vaccine – it can prevent tetanus, an infection caused by Clostridium tetanii bacteria. This vaccine lowers the disease effect.
- Haemophilus influenzae type b (Hib) vaccine – Hib vaccines are of two types: The Hib vaccine protects children and adults from Hib disease. The DTaP-IPV/Hib vaccine protects babies' ages 2 through 18 months from Hib disease, tetanus, diphtheria, whooping cough and polio. The vaccine links sugar coated polysaccharide to a helper protein that creates a better immune response than would occur naturally.
- Pneumococcal vaccine – working of Pneumococcal vaccine is same as Haemophilus influenzae vaccine.
- Adjuvant: Adjuvant is derived from Latin word adjuvare meaning to help or aid. In immunology it is defined as any substance that acts to accelerate, prolong or enhance antigen-specific immune responses when used in combination with specific vaccine antigens". Adjutants enhance immunity to vaccines and immune response. Hence it is used in humans. For example, measles, mumps, rubella, vermicelli, rotavirus (Choudhry et al., 2019).

New vaccine types are currently developed by researchers and they also try to improve current approaches for vaccine delivery. DNA vaccines and Recombinant vector vaccines are in progressing stages because DNA vaccines are easy and inexpensive and provide long term immunity whereas recombinant vector vaccines making the immune system to fight germs. For example, inhaled vaccines are used usually in the form of nasal spray in some cases of influenza (Carvalho 2010).

Reverse vaccinology: It is improved by Bioinformatics and first used against Serogroup B meningococcus. Here Bioinformatics tools are used to screen entire genome of pathogens to determine whether the protein is good vaccine target (Bowman 2011). With the development of genomics in vaccine development it is needed to study whole genotype to phenotype and environment exposure. It is said that it is the future. Reverse vaccinology are improving the process of vaccine development by classification methods like Support vector machines (Bowman 2011).

Stages of Vaccine Development and Testing: Usually, vaccine development and testing follow a certain standard set of steps. The first stages are exploratory in nature. Regulation and oversight increase as the candidate vaccine makes its way through the process. First Steps: Study of Animal and laboratory Exploratory Stage: The basic need is to study laboratory where animal experiments are carried out. This will take time 2-4 years. The scientists of fully funded academic and government institutes are identifying natural or synthetic antigens that will help in disease prevention and treatment. These antigens can be virus-like particles, weakened viruses or bacteria, or it can be any other substances derived from pathogens.

Pre-Clinical Stage: In this stage it is very important to study tissue-culture and cell culture for assessing the candidate vaccine safety. The immunogenicity or ability to provoke an immune response is also very important criteria for finding vaccine response in animal of study. Basically animal subjects for study include mice and monkeys. These studies help the scientists to start a safe

dose for humans and also administering the vaccine in a safe way. The efficiency/ efficacy of vaccine are also very difficult task for scientists. Many candidate vaccines are not able to produce desired immune response. This often takes 1-2 years.

Any private company which wants to work on vaccines submits an application for an Investigational New Drug (IND) to the Food and Drug Administration of their country (The Central Drugs Standard Control Organization (CDSCO) is the national regulatory body for Indian pharmaceuticals and medical devices). It is the responsibility of sponsors of private company that they describes the manufacturing and testing processes, all the laboratory reports and proposal of study of vaccines in subjects. This should also include clinical trials. If the proposal of vaccine clinical trials are approved, the vaccine will have three phases of testing.

Next Steps: Clinical Studies with Human Subjects- this is one of the most important step in vaccine designing and administration. Phase I Vaccine Trials: Human intervention to vaccines involves small group of adults basically nearly 20-80 which can be subjects. Gradually come down to the age of each category of age, if vaccines are for children. This phase clearly indicates whether the particular vaccine will be used or not. In this phase, candidate vaccine safety is utmost important with its immune response in the subjects. All the participants of the study are carefully monitored and the laboratory conditions are fully controlled carefully. This phase success rate will decide the next stage.

Phase II Vaccine Trials: Phase 2 testing includes larger group of individual participates, sometimes may be hundred. There are some people which are at great risk of acquiring the disease. This may be randomized but in a well controlled way. This phase success will proves the safety, immunogenicity of candidate vaccine, proposed doses, schedule of immunizations, and method of delivery (Plotkin 2008).

Phase III Vaccine Trials: When phase II trials are successful, candidate vaccines move on to larger trials which involves thousands of people. Again these tests are randomized and double blind. The assessment of vaccine safety in a large group of people is most significant goal of phase III trials. In this, side effects of vaccines are also studied.

Vaccine efficacy includes: a) is candidate vaccine is able to prevent disease. b) Is candidate vaccine provides prevention against pathogen infection. c) Is candidate vaccine is able to produce immune response and antibody production against pathogen.

Approval and Licensure is another step: As phase III trial is completed successfully, the vaccine developer needs to have license. The license providing agency will inspect the manufacturer unit where the vaccine will be made and approval for vaccine label. The agency continuously monitors the production, potency, safety and purity of vaccines.Post-Licensure Monitoring of Vaccines & phase IV trial: This includes vaccine adverse event reporting system and the vaccine safety data link. Concern companies conduct studies after vaccine is released. They continuously monitor the vaccine safety, efficacy and other potential uses (Plotkin 2008).

Since it is very important for success of vaccine trials because it involves lot of efforts, finance and animal used under study could go through pains. To make these vaccine trials more effective and successful machine learning and artificial intelligence are come into existence. Because these techniques are cost effective, efficient, and time consuming. This will help scientist to make vaccine trials a success.

MATERIAL AND METHODS

Machine Learning: Machine learning is an emerging field which reaches everywhere. Not only in industries and solving big problems it will be one of the biggest challenging applications of artificial intelligence. It provides systems the ability to automatically learn and improve from experience without being explicitly programmed (Samuel 1959). Machine learning focuses on the development of computer programs that can access data and use it learn for themselves. There are algorithms that learn from the data as provided to train and this will create machine learning model that can perform a given task without any specific instructions. These models are used to make predictions or classifying images and emails. These techniques are nowadays widely adopted in every field i.e. identifying spam mails, diagnose diseases from X-rays, crop yield prediction, and in future it will help in driving cars (Tang 2019).

This technology has already shown its potential. Here in this paper it is tried to implement the machine learning algorithms to improve the approaches in vaccine development with great sensitivity and accuracy. Machine Learning is classified into three categories at a high level depending on the nature of the learning system (Han & Kamber2010):

1. Supervised learning: Machine gets labeled inputs and their desired outputs. The goal is to learn a general rule to map inputs to the output. That is classification algorithms.
2. Semi supervised learning: in this small amount of labeled data with a large amount unlabeled data during training.
3. Unsupervised learning: Machine gets inputs without desired outputs; the goal is to find structure in inputs. That is clustering and association rule.
4. Reinforcement learning: In this algorithm interacts with a dynamic environment, and it must perform a certain goal without any guidance.

Working of machine learning algorithm: In machine learning, algorithm works by learning

strategy to map input to output without being explicitly programmed.

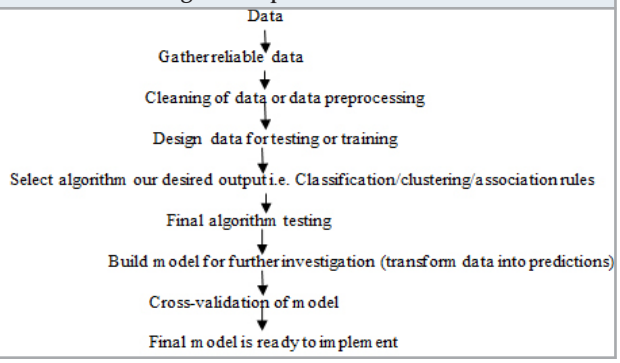
[A] Prediction and classification: Classification is a data mining method that assigns items in a collection to target categories or classes. The goal of classification is to accurately predict the target class for each case in the data. For example, a classification model could be used to identify loan applicants as low, medium, or high credit risks. Nowadays classification techniques are best way to use in health sector for predicting disease based on symptoms data and treatment (Dubey 2016; Dubey 2018). There are many classification algorithms i.e. J48 decision tree, random forest, Naïve bayes (Han and Kamber 2010) etc.

[B] Clustering: It is the assignment of a set of observations into subsets i.e. called clusters, so that observations in the same cluster are similar in some character (Parasian and Silitonga 2017). One of the best clustering algorithms is k-means clustering.

[C] Association rules: Association rules are generated which will be based on hidden patterns, correlations and other insights depends on data (http:link.springer.com//2february2020;Dubey 2014).

For implementing machine learning techniques, it is necessary to understand the biological data for vaccine development. Here is presented a schematic flow of working of machine learning technique.

Figure 1: A general flow diagram for implementing machine learning techniques.



RESULT AND DISCUSSION

If machine learning is applied to each of these above mentioned vaccine development stages it is suggested to save time and results obtained will be of great accuracy and efficiency.

A proposed way of implementation of machine learning in vaccine development: In this paper, author is tried to propose how biomedical scientists can implement different machine learning techniques: classification, clustering, association in different phases of vaccine development. The comparative chart of these techniques is given in table2.

To develop our model for vaccine, first requirement is– Data cleaning or preprocessing: There would be no noise or repetition in the data. There is need to remove all incorrect records.

Data testing and training: Training set is one on which we train and try to fit model according to parameters whereas test data is used for assessment of performance of model. Training data's output is ready to model but test data are unseen, only used for making predictions. There are many algorithms for data training and testing.

Cross validation of model: It is also called model assessment. In validation automatic computer check is done which ensure that the data entered for training is sensible and reasonable. All these methods involve high computation having mathematical and computer science background. For maximum efficiency of algorithms it is needed that it should minimize resource usage. Different machine learning methods need time, computing power, accuracy, space complexity etc. Hence which algorithm works better depends upon the type of data, goal of experiment and measure of efficiency of algorithm. In this paper author is taking the example of malaria and try to explain the role of machine learning algorithms in different stages of vaccine development (as proposed in table 2). Mostly decision tree, clustering and association rules are important to find the suitability in vaccine development.

Decision tree in vaccine development: The decision tree classifier is a simple and widely used classification technique. It poses a series of carefully crafted questions about the attributes of the test record. Each time it receives an answer according to follow-up asked questions, until a conclusion about the class label of the record is achieved. The following figure 1 shows an example decision tree for predicting whether the Antigen is specific and sensitive for disease i.e. malaria that can be use for vaccine development (as shown in table2). In the decision tree, the root and internal nodes contain attribute test conditions to separate records that have different characteristics. The entire terminal node is assigned a class label Yes or No. If the yes is found in testing the antigen then particular antigen is used for vaccine trial or development. Like this we can further move for experiments saving lot of time.

(b) Clustering (Han & Kamber 2010) in vaccine development: Here K-means clustering algorithm example is used to show how clustering is implemented (as given in table 2). All the antigens specific for malaria are grouped into small sub-groups /clusters according to their specificity i.e. more, medium, less. Now the k-means algorithm work as follows:

Specify number of clusters K.

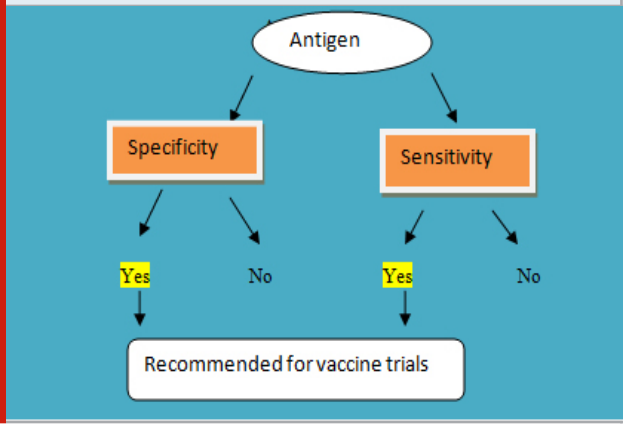
(i) The dataset is shuffled and centre is initialized. Now K data points are randomly selected.
(ii) Without changing centre, iterations are same.
(iii) Compute the sum of the squared distance between data points and all centroids.

(iv) Assign each data point to the closest cluster (centroid).
(v) Compute the centroids for the clusters by taking the average of the all data points that belong to each cluster. This will require mathematical and computation power.

Table 2. Proposed machine learning approaches in different pha ses of vaccine candidate development

Vaccine development PHASE I	Classification	Clustering	Association
1)Laboratory and animal studies			
[a]Exploratory stage	Specific antigens are classified on the basis of sensitivity and specificity.	Disease specific antigens are clustered in one group and used as per requirement	Host specific antigen associations are studied.
[b]Preclinical stage	Tissue culture or cell culture systems are classified as per their subject like monkeys or mice etc.	Animal vaccinate and target pathogen are clustered.	Candidate specific vaccines are in progress if association between candidate vaccine and target is studied perfectly.
PHASE II Clinical studies with human subjects			
Vaccine trial I	Several criteria of candidate vaccines are classified w.r.t. the type and extent of immune response.	Clustering of experimental group for particular injected participants with pathogens.	Association of participants i.e. mice etc with their pathogens is fully understandable for success of vaccine trials.
Vaccine trial II	Classify group that acquire the disease on certain parameters: i.e. (i)Vaccine safety, (ii)Immunogenicity, (iii)Proposed doses, (iv)Schedule of immunization and method of delivery	Cluster the group that work for immunization which produces immune response.	Association rule help to check particular vaccine safety to allergy and immunogenicity to disease. Schedule of immunization with age. Method of delivery of vaccine to humans.
Vaccine trial III	Classify vaccine efficacy on the basis of (i)disease prevention (ii)prevention from infected pathogen (iii)antibodies production	Cluster group of vaccines that proves better response while testing	Make associations of vaccines with disease prevention, vaccine inhibit pathogen prevention, vaccine able to produce antibody against infection.

Figure 2: Decision tree for particular antigen recommendation for vaccine trial



$$\text{Total distances} = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

Where k= numbers of clusters, n=number of points belonging to cluster j, cj=centroid of cluster j

(vi)Now the new centroid of each cluster is made by calculating the mean of all points assigned to that cluster.

(vii) Repeat from step 2 until we reach the required centroid which no longer moves. This antigen is more specific for disease (malaria the example taken) under study.

(c) Association rules (Han & Kamber 2010) in vaccine development: In case of malaria, it is tried to deduce the different pattern relations from the available data. Suppose a person having symptoms like malaria. So it is needed to understand two parameters: support and confidence for association rule study (as given in table 2).

Support: This indicates how frequently the if/then relationship appears in the database. This means the symptoms are supporting to having malaria or not based on if/then rule.

Confidence: It shows about the number of times these relationships have been found to be true. This means how the particular symptom is really associated with malaria.

For an association rule X--> Y, the support of the rule is denoted as sup(X-> Y) and is the number of transactions where XUY appears divided by the total number of transactions. The confidence is the number of transactions where XUY appears divided by the number of transactions where X appears. In our case, X- disease symptoms and Y is particular disease like malaria.

These are the different machine learning models that can be used according to the desired goal. These are achieved by implementing in a proper way in the path of vaccine development.

CONCLUSION

If classification and association rule mining techniques are implemented, it is possible to develop vaccine for infectious diseases where vaccine trials are not giving desired results. One most important factor to be studied for vaccine development is genotype to phenotype and their interactions with environment will surly give complete advantage to use machine learning methods in vaccine development. Machine learning and Artificial intelligence are the emerging technologies which paved the way for healthy way of living. The need of this modernized world is to develop more on these techniques and use the friendly nature of them. Robotic process automation is also emerging for making all things automatic and in a controlled manner. In future the method of immunization will be more simpler to administer, will provide long-lasting immune response and most importantly vaccines will survive in transport without any refrigeration. Lethal diseases like HIV/AIDS vaccines will also come into existence.

REFERENCES

Angsantikul P, Fang HR, Zang L (2018) Toxoid Vaccination against Bacterial Infection Using Cell Membrane-Coated Nanoparticles, Bioconjug Chem. 2018 March 21; 29(3): 604–612

Carvalho, J.A., Rodgers J., Atouguia J., Prazeres DM. F., Monteiro GA. (2010) DNA vaccines: a rational design against parasitic diseases. Expert Rev Vaccines. Feb; 9(2):175-91

Chaudhury S, Duncanb E H, Atre T (2019) Combining immunoprofiling with machine learning to assess the effects of adjuvant formulation on human vaccine-induced immunity, Human vaccine and Immunotherapeutics,1-12

Dubey A., (2016) Applications of Machine Learning: Cutting Edge Technology in HIV Diagnosis, Treatment and Further Research, Computational Molecular Biology, Vol 6 number 3: 1-6.

Dubey A., (2018) Potential Drug Target Sites of HIV Identified by Bioinformatics & Intelligent Machine Learning Techniques, Online Journal of Bioinformatics, Volume 19 number2:56-66.

Dubey A., (2014) Association Rules for diagnosis of HIV-AIDS", Computational Molecular Biology, vol4, no.3.

Eva K Lee (2018) Decision Analysis and Optimization in diseases prevention & treatment. WILEY online

Han N & K amber (2010) Data Mining: Concepts and

techniques, San Francisco: Morgan Kauffman,
Kallarp, R.S. (2014) Classification of vaccines, subunit vaccine delivery, pp15-29, Advances in delivery sciences and technology Book series.
Plotkin SA (2018) Vaccines, 5th ed. Philadelphia: Saunders.
Parasian D. and P.Silitonga, (2017) Clustering of patient disease data by using K-means clustering", International Journal of computer science and Information security, Vol5, no.7.
Rauch S, Jasny E, Schmidt KE, Petsch B (2018) New Vaccine Technologies to Combat Outbreak Situations. Front Immunol. 2018 Sep 19;9:1963. doi: 10.3389/fimmu.2018.01963. eCollection 2018.
Samuel ,Arthur L (1959) Some studies in machine learning using the game of checkers, IBM Journal of research and development.44,206-226.
Tang Z, Pan Z, Yin K., Khateeb A., (2019) Recent Advances of Deep learning in Bioinformatics and Computational Biology, Frontiers in Genetics, volume10
Tomic A, Tomic I, Dekker C L (2019) The FluPRINt dataset, a multidimensional analysis of the influenza vaccine imprint on the immune system, 6:214
Tung, Thanh Lee, Zacharias Andeadakis, Arun Kumar (2020) The COVID19 Vaccine development landscape, Nature review drug discovery