2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON) | 978-1-6654-0690-1/21/$31.00 ©2021 IEEE | DOI: 10.1109/UEMCON53757.2021.9666715

A Conceptual Deep Learning Framework for COVID-19 Drug Discovery

Mohammad (Behdad) Jamshidi*
Research and Innovation Centre for Electrical Engineering (RICE), University of West Bohemia,
Pilsen, Czech Republic
jamshidi@fel.zcu.cz
**Corresponding Author*

Jakub Talla
Research and Innovation Centre for Electrical Engineering (RICE), University of West Bohemia
Pilsen, Czech Republic
talic@fel.zcu.cz

Ali Lalbakhsh
School of Engineering, Macquarie University,
Sydney, NSW 2109, Australia
ali.lalbakhsh@mq.edu.au

Maryam S. Sharifi-Atashgah
Department of Education,
Wittenborg University of Applied Sciences,
Apeldoorn, Netherlands
maryam.sharifi@wittenborg.eu

Asal Sabet
Department of Pharmacy, Irma Lerma Rangel College of Pharmacy,
Texas A&M University,
Kingsville, TX 78363, USA
asal.sabet@tamu.edu

Zdeněk Peroutka
Research and Innovation Centre for Electrical Engineering (RICE), University of West Bohemia
Pilsen, Czech Republic
pero@fel.zcu.cz

Abstract- The analytical and experimental methods used for the development of drugs have some disadvantages in the aspect of the needed time for preparation of the desired parenthetical products and the efficiency of them, which not only can the risk for failure increase, particularly when pathogens are impossible to be cultivated under laboratory conditions, but these approaches can also lead to achieving arrays of antigens that are not able to provide sufficient immunity to combat the targeted disease. On the other hand, artificial intelligence (AI) and its new branches, including deep learning (DL) and machine learning (ML) techniques can be deployed for drug development purposes in order to alleviate the difficulties associated with conventional methods. Moreover, intelligent methods will provide researchers with the opportunity to use some user-friendly and efficient services to conquer such problems. In this respect, a conceptual DL framework has been studied in order to demonstrate the capability and applicability of these methods. Accordingly, a framework has been proposed to show how COVID-19 drug development can benefit from the potentials of AI and DL.

Keywords— Artificial intelligence, bioinformatics, covid-19, deep learning, drug discovery, RNA, machine learning.

I. INTRODUCTION

CVOID-19, as one of the most serious health problems that has occurred, significantly affects the world and it has become a big problem since December 2019. Although some vaccines have been approved, the new challenges are related to its mutations. One thing that makes this situation worst is difficulty associated with identify and designing the drug and vaccine for the virus and its mutations [1]. In this regard, many researches have been published to survey the different aspects of this pandemic. For instance, an analysis was presented fundamental information that could help researchers' human immune responses to this virus and achieve a better evaluation of candidates for performing diagnostic tests or developing vaccines [2].

Reverse vaccinology is known as a novel method in identifying the virus. In this method, genome sequencing [3]. plays a vital role as it makes accessible a catalog of all the protein antigens expressed by pathogen. In this approach, based on computer analysis some antigens are predicted which are assumed to be the best candidates to be a vaccine. However, this is an ignorant method as in the absence of information including the antigen abundance or its immunogenicity in the infection period or when expressed *in vitro* [3], it cannot regulate if the potential antigen candidate may conclude the provision of protective immunity [3]. As matter of fact, Artificial Intelligence (AI) techniques have shown good performance in modeling and estimating a number of complex systems and uncertain functions [4], not only engineering systems [5], but also in the fields of economics, psychology [6], healthcare [7], medicine[8]. Hence, utilizing these approaches to develop the drugs can be helpful. Although some efforts have been done to improve the drug process using AI methods, more researches and ideas should be considered to increase the speed of drug discovery. Particularly after experiencing this pandemic showing the importance of providing the drugs and vaccines at any cost. That is why, a large number of publications have been presented to explain the applications of AI and its related methods like Deep Learning and Machine Learning.

In this paper, we explain a methodology that can enable scientists and companies to work toward new vaccines and drugs using DL methods. The paper has been organized as follows; the next section gives information about identifying the virus with the use of the bioinformatic methods. The third section illustrates the proposed frame work. Section fourth has

discussed the results of this approach while the conclusion has been brought in the fifth section.

II. Identification of the Virus

In fact, one of the fundamental requirements to design and develop drugs and vaccines is to use bioinformatics tools. Such technique with focus on reverse vaccinology approaches provide facilities for designers, where identifying the models from allergens, cancer cells, parasites, viruses, or bacteria is sought to an immune response that could serve as a protection against disease [3]. Fig. 1 depicts the process of virus identification. This figure is the first layer in our methodology. As is observed, first the virus should be studied

Considering a priori epitope prediction by bioinformatics, potential targets for immune responses to the COVID-19 was indicated with the application of sequence homology with the use of similar SARS-CoV. Fundamental information could help researchers to observe human immune responses and achieve a better evaluation of candidates for performing diagnostic tests or developing vaccines. When two approaches are used, and the same regions are identified, it means that the identified regions are the right targets for immune recognition of SARS-CoV-2 [2]. Determining physicochemical and antigenic properties associated with suitable antigens is the next challenge in revers vaccinology. Various bioinformatics approaches should be employed to choose protein(s) with the most appropriate properties to test based on *in vivo* and *in vitro* and illustrate its immunogenicity and safety and examine. Indeed, every protein in the proteome under study should be analyzed to determine its characteristics [9]. While practicability of this technique is dramatically based on the existence of high-throughput system which is responsible for screening protective immunity. In fact, it makes the discovery of novel mechanisms of immune intervention possible [3].

The layer is dedicated to virus identification which is done through some experimental and bioinformatics methods to identify the crystal of virus. The sequence chain view and biological assembly are provided by bioinformatics operations where sequence analysis consists subjecting a peptide sequence, RNA, DNA, to each of analytical approaches with the purpose of comprehending desired features, function, structure, and evolutions. Searches against biological databases and sequence alignment are among the employed methodologies [10].

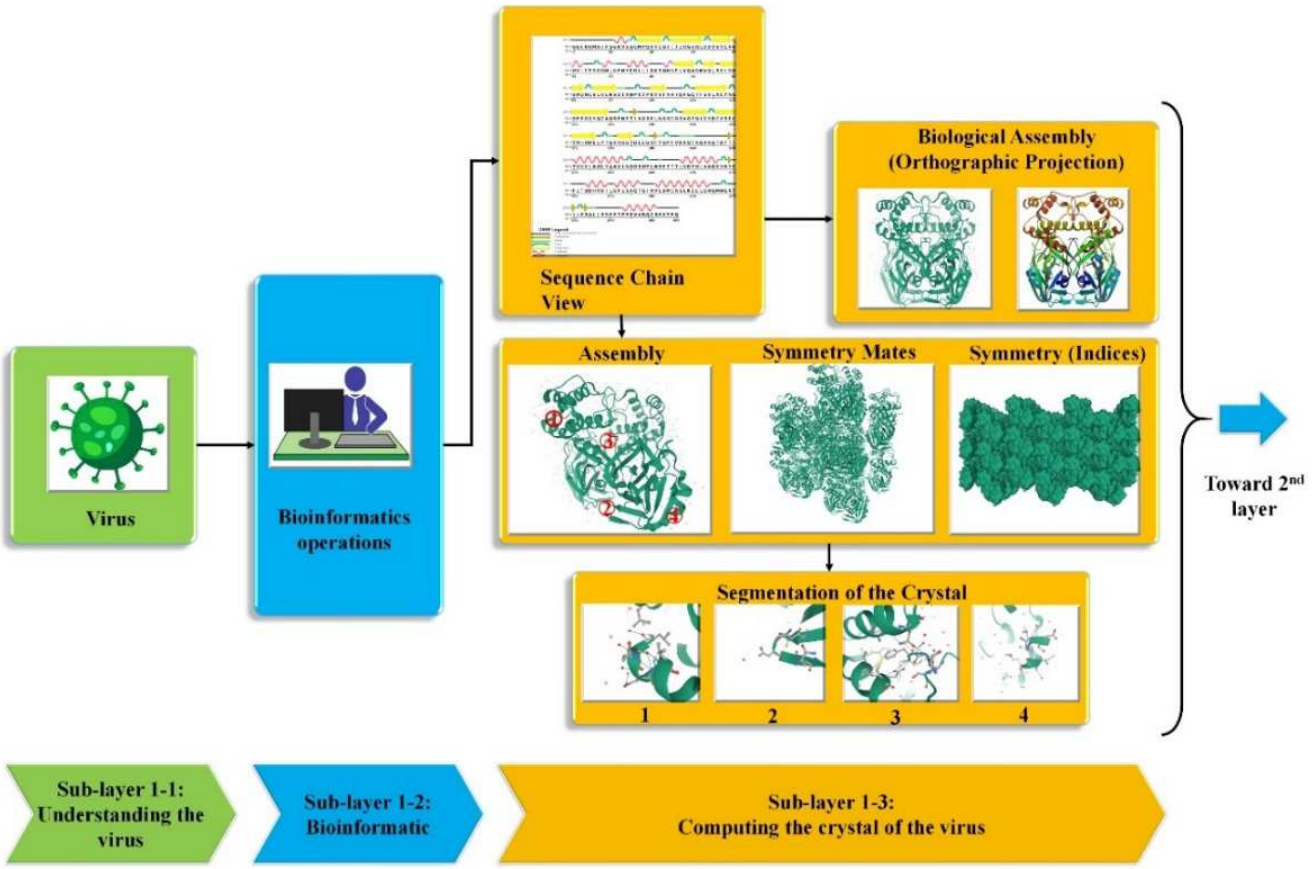

Fig. 1. Utilizing bioinformatic for identifing the crystal structure of COVID-19; The crystal structures shown in the schematic diagram is to explain the suggested method and they are extracted from [14].

As in the case of DNA the majority of biologically active RNAs, such as snRN, rRNA, tRNA, mRNA and other non-coding RNAs, consist of self-complementary sequences which let parts of the RNA fold [5] and be paired with itself to construct two-fold helices [11]. As a Betacoronaviruses, COVID-19, is almost the same as and the MERS HCoV and SARS HCoV [12, 13]. HCoVs are in general positive-sense and quite long (30,000 bp) single-stranded RNA viruses [10].

Ribonucleic acid (RNA) is a polymeric molecule known as a rudimentary element in a variety of biological parts such as expression of genes, decoding, regulation and coding [15]. It is assumed that genomes of all non-defective positive-strand RNA viruses encode RNA-dependent RNA polymerases. When amino acid sequences of the (putative) RNA polymerases of positive-strand RNA viruses are studied through a comparative analysis, the findings demonstrate that the only viral protein that contains motifs conserved throughout this class of viruses is polymerase, and this prepares the ground to tentatively identify polymerases among the proteins which are encoded by newly sequenced viral genomes [16, 17]. Along with other methodologies, sequence alignment, and searches against biological databases are employed [10]. Sequence analysis in chemistry is marked by techniques employed to determine how various monomers are sequenced and paired to form a polymer. This process of 'sequencing' in biology and genetics is very important because biological assembly is drawn according to sequence analysis. The biological assembly which is also called the biological unit is the macromolecular assembly which either presumably or demonstrably considered as the molecule's functional form [18]. The functional form of hemoglobin for instance consists of four chains [19]. To obtain the complete biological assembly, symmetry operations including rotations, translations or their combinations may be required, but it all depends on the crystal structure. As an alternative, it may be necessary to select a subset of the deposited coordinates to represent the biological assembly.

III. THE PROPOSED FRAMEWORK

In this section we explain how DL approaches, ANNs and their related inputs and targets can be used for achieving a practical framework for designing a vaccine or drugs against COVID-19 and its mutations. The method presented in Fig. 2. consists of 8 layers and illustrates the schematic diagram of the proposed strategy to develop a COVID-19 vaccine and drugs. The first layer has been demonstrated in section II.

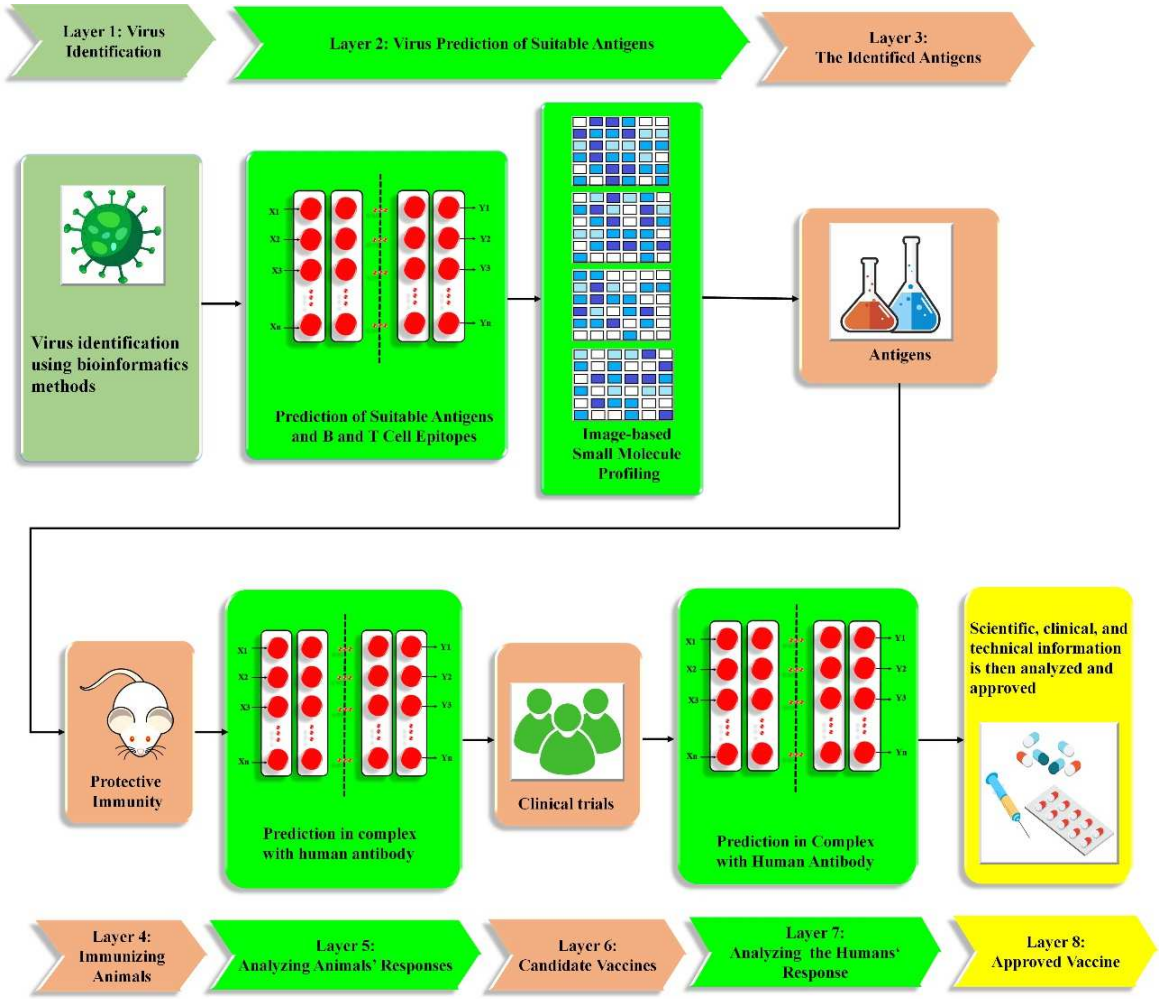

Fig. 2. description of the introduced AI-powered strategy for COVID-19 vaccine development

While the output of the first layer is small molecule profiling which is image-based, phenotypic screens of this sort can be successfully utilized to identify targets and compounds involved in modulation of phenotypes such as the subcellular localization of specific proteins. The first is that it provides direct protection, and the second is that also it works through herd immunity because it simply limits disease spread to individuals who are susceptible to it.

In second layer, the antigens are predicted and found using DL methods. In fact, that part of an antigen that is recognized by the immune system namely antibodies, B cells, or T cells is an epitope or antigenic determinant. Prediction of suitable antigen in this method takes place by a typical DL method for prediction and operation estimation. In other words, this layer is the space for predicting B cell epitopes. Identifying T- and B-cell epitopes on PmpD is essential to select various regions that could be used in designing subunits vaccines and inducing both humoral and cellular immune responses [20] Bioinformatics is able, therefore, to offer powerful tools that are capable of characterizing proteins that favor a rational design of vaccines [20].

The first type of vaccine to immunize animals are prepared in the third and fourth layers. Realizing animals' health and well-being is firmly dependent on vaccination of pets and farm animals and can reduce negative consequences that may harshly affect pets and livestock. While the goal of vaccination is generally inducing protective immunity against diseases, vaccination for infectious diseases functions in two different ways.

The animals' responses to the potential vaccine is evaluated and analyzed in the fifth layer using Restricted Boltzmann Machine (RBM). RBM that can be used in DL networks is a generative stochastic structure learning a distribution of probability upon its set of inputs. DL methods are desirable options to analyze animals' reactions to vaccines, whilst being able to discover complex relationships between data and potential faults, they outperform traditional ML methods as well [20].

Layer six is used for clinical trials that is an important stage in developing drugs. The general process of human clinical trials is to examine the performance of the designed vaccine on individuals. The purpose behind vaccine trials is to ensure that the vaccine is safe and efficient and could be licensed for application on human beings [21]. Vaccine candidates usually undergo preclinical testing, which includes the application of the vaccine on various animal models [22]. It is in this stage that a vaccine candidate drug is initially identified through preclinical evaluations to invoke an immune response. These preclinical evaluations, however, may include high throughput screening and selecting the proper antigen.

Layer 7 comprises an analysis of human's responses an is performed to determine vaccine trial regulation errors. Extreme measures and cautions need to be in place while performing vaccine trials and manufacturing because it is likely that recipients get infected or come across unwanted adverse effects. The case of Cutter Incident in which public health was seriously put at risk has made researchers and the public to be cautious when it comes to a matter of such gravity[23].

The last stage is marked by the application of the vaccine on people after it is approved for its safety and lack of risk. Food and Drug Administration (FDA) in the United States has the responsibility to determine the safety of the vaccine and its consequent approval [24]. However, despite the delineated strategy above, we present several developed approaches that while giving us a DL-attitude in problem-solving could be employed to speed up other relevant issues when designing a vaccine.

IV. Discussion

Combining medical information, AI approaches and ML methods with different fields of science and engineering, especially for medical and biological systems and chemical attributes and virus characteristics is an important advantage to reach an integrated approach to develop the drugs [5, 25-29]. Genetic information is employed by certain body cells to and produce elements of the virus can be considered as big data to utilize the AI methods. The main benefit of this technique is prevention of examination on humans before achieving a reliable drug. Vaccine-induced immunopathology may be considered as potential concern for SARS, MERS and even SARS-CoV-2 vaccine candidates; however, while concerns of this sort are vaccine-platform dependent, there have not been ant report with regard to existence of immune pathogenesis for SARS DNA vaccines in mice [30] and MERS DNA vaccines in non-human primate models or mice [31]. Therefore, performance and automation could be greatly enhanced and promoted.

Integration of attention mechanism is another important innovation to facilitate the interpretation of the model. DL and ML methods can affect the behavior of atom which explains how each reaction can contribute to the overall biological activity. Moreover, algorisms have the potential to analyze the structured and unstructured data more smoothly and with less limitations compared to conventional techniques. Additionally, since big data and cloudy space are used in the process, observing reactions in mouse clinical or human trails is performed with higher degrees of accuracy and measuring reactions of the drugs and their effects on immune systems can be done with the highest level of precision.

Another important feature of this study is its role in preventing or decreasing the risk of spoiling or wasting resources used by the pharmaceutical industry to realize vaccine mass production. After achieving official approval from the competent authorities, vaccines become available for the public. Furthermore, diseases can be treated either by themselves or by fresh and newly-launched treatments. at the same time, each time a medicine is produced, a part of the available resources for the manufacturer is lost which is precious and may be used for other purposes. Here, this study becomes more valuable as it proposes a framework that salivates reviewing a large number of drugs discovery methods and combines them with the effective DL-based techniques showing a suitable AI approach to both save the resources in the pharmaceutical industry and make a synergic approach that

helps to have better quality and a great balance in the performance of the distribution system. Furthermore, this approach has the potential to speed up the process of developing vaccines in the future.

V. Conclusion

Due to utilizing experimental tests and clinical processes, conventional methods for the development of the drugs are really time-consuming. Although some clinical stages of this process are necessary, deployment of machine intelligence approaches can be taken into consideration to solve the main part of such problems. In this paper, a framework based on DL methods has been presented in order to illustrate how AI can accelerate the process of drug development. This framework includes 8 layers, which are responsible for identifying, analyzing, and predicting the drug's performances in different stages. In addition, this methodology could be developed and generalized for other medical applications.

Acknowledgment

This research has been supported by the Ministry of Education, Youth and Sports of the Czech Republic under the project OP VVV Electrical Engineering Technologies with High-Level of Embedded Intelligence CZ.02.1.01/0.0/0.0/18_069/0009855.

References

[1] M. Jamshidi et al., "Artificial intelligence and COVID-19: deep learning approaches for diagnosis and treatment," vol. 8, pp. 109581-109595, 2020.

[2] A. Grifoni, J. Sidney, Y. Zhang, R. H. Scheuermann, B. Peters, and A. Sette, "A sequence homology and bioinformatic approach can predict candidate targets for immune responses to SARS-CoV-2," Cell host & microbe, 2020.

[3] R. Rappuoli, "Reverse vaccinology," Current opinion in microbiology, vol. 3, no. 5, pp. 445-450, 2000.

[4] M. B. Jamshidi, J. Talla, and Z. Peroutka, "Deep Learning Techniques for Model Reference Adaptive Control and Identification of Complex Systems," in 2020 19th International Conference on Mechatronics-Mechatronika (ME), 2020, pp. 1-7: IEEE.

[5] M. B. Jamshidi and S. Rostami, "A dynamic artificial neural network approach to estimate thermal behaviors of li-ion batteries," in 2017 IEEE 2nd International Conference on Automatic Control and Intelligent Systems (I2CACIS), 2017, pp. 116-121: IEEE.

[6] M. B. Jamshidi, et al. Artificial neural networks: a powerful tool for cognitive science," in 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2018, pp. 674-679: IEEE.

[7] F. Jiang et al., "Artificial intelligence in healthcare: past, present and future," vol. 2, no. 4, 2017.

[8] B. J. N. M. I. Mittelstadt, "Principles alone cannot guarantee ethical AI," vol. 1, no. 11, pp. 501-507, 2019.

[9] R. María, C. Arturo, J. A. Alicia, M. Paulina, and A. O. Gerardo, The impact of bioinformatics on vaccine design and development. InTech, Rijeka, Croatia, 2017.

[10] R. Durbin, S. R. Eddy, A. Krogh, and G. Mitchison, Biological sequence analysis: probabilistic models of proteins and nucleic acids. Cambridge university press, 1998.

[11] J. Tinoco, "I and C. Bustamante," How RNA folds,"" Journal of Molecular Biology, vol. 293, no. 2, pp. 271-281, 1999.

[12] J. F. Chan, S. K. Lau, K. K. To, V. C. Cheng, P. C. Woo, and K.-Y. Yuen, "Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease," Clinical microbiology reviews, vol. 28, no. 2, pp. 465-522, 2015.

[13] A. A. Elfiky, S. M. Mahdy, and W. M. Elshemey, "Quantitative structure‐activity relationship and molecular docking revealed a potency of anti‐hepatitis C virus drugs against human corona viruses," Journal of medical virology, vol. 89, no. 6, pp. 1040-1047, 2017.

[14] L. Zhang, Sun, X., Hilgenfeld, R., " Crystal structure of the free enzyme of the SARS-CoV-2 (2019-nCoV) main protease," World Wide Protein Data Bank PDB, 2020.

[15] Y.-C. Chang et al., "Potential therapeutic agents for COVID-19 based on the analysis of protease and RNA polymerase docking," 2020.

[16] E. V. Koonin, "The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses," Journal of General Virology, vol. 72, no. 9, pp. 2197-2206, 1991.

[17] E. V. Koonin, "Evolution of RNA-dependent RNA polymerases of positive-strand RNA viruses: a comparison of phylogenetic trees generated by different methods. Molekularnaya Genetika, Mikrobiolog (ja i Virusologija 3," 1988.

[18] Y. Wang, Z. Tang, S. Tan, and N. A. Kotov, "Biological assembly of nanocircuit prototypes from protein-modified CdTe nanowires," Nano letters, vol. 5, no. 2, pp. 243-248, 2005.

[19] E. Collazo et al., "Functional characterization of the purified holo form of hemoglobin I from Lucina pectinata overexpressed in Escherichia coli," The protein journal, vol. 23, no. 4, pp. 239-245, 2004.

[20] G. Hu, H. Li, Y. Xia, and L. Luo, "A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis," Computers in Industry, vol. 100, pp. 287-296, 2018.

[21] M. E. Halloran, I. M. Longini, C. J. Struchiner, and I. M. Longini, Design and analysis of vaccine studies. Springer, 2010.

[22] C. E. Mire, T. W. Geisbert, H. Feldmann, and A. Marzi, "Ebola virus vaccines–reality or fiction?," Expert review of vaccines, vol. 15, no. 11, pp. 1421-1430, 2016.

[23] S. S. Blume, "Lock in, the state and vaccine development: Lessons from the history of the polio vaccines," Research Policy, vol. 34, no. 2, pp. 159-173, 2005.

[24] C. f. D. Control and Prevention, FDA licensure of bivalent human papillomavirus vaccine (HPV2, Cervarix) for use in females and updated HPV vaccination recommendations from the Advisory Committee on Immunization Practices (ACIP), MMWR. Morbidity and mortality weekly report, vol. 59, no. 20, p. 626, 2010.

[25] A. Shafiei et al., "A Hybrid Technique Based on a Genetic Algorithm for Fuzzy Multiobjective Problems in 5G, Internet of Things, and Mobile Edge Computing," Mathematical Problems in Engineering, vol. 2021, 2021.

[26] M. B. Jamshidi et al., "Deep Learning Techniques and COVID-19 Drug Discovery: Fundamentals, State-of-the-Art and Future Directions," Emerging Technologies During the Era of COVID-19 Pandemic, vol. 348, p. 9, 2021.

[27] A. R. Belverdi, et al., "Novel donor-acceptor non-fullerene metal-organic solar cells: A first DFT and TD-DFT study," Physica B: Condensed Matter, vol. 542, pp. 37-43, 2018.

[28] S. Roshani, et al., Design and Modeling of a Compact Power Divider with Squared Resonators Using Artificial Intelligence, Wireless Personal Communications, vol. 117, no. 3, pp. 2085-2096, 2021.

[29] Jamshidi, M, et al. A neuro-based approach to designing a Wilkinson power divider. International Journal of RF and Microwave Computer‐Aided Engineering 30.3 (2020): e22091.

[30] P. Tebas et al., "Intradermal SynCon® Ebola GP DNA vaccine is temperature stable and safely demonstrates cellular and humoral immunogenicity advantages in healthy volunteers," vol. 220, no. 3, pp. 400-410, 2019.

[31] K. Muthumani et al., "A synthetic consensus anti–spike protein DNA vaccine induces protective immunity against Middle East respiratory syndrome coronavirus in nonhuman primates," vol. 7, no. 301, pp. 301ra132-301ra132, 2015.