Emerging Role for Artificial Intelligence in Testing COVID RNA-Human AntiRNA Antigenicity and Binding Affinity

Aritra Ghosh, PhD(c), Maria M. Larrondo-Petrie, PhD, and Mirjana D. Pavlovic, PhD
Computer & Electrical Engineering and Computer Science
Florida Atlantic University
Boca Raton, Florida 33431, USA
{aghosh2016, petrie, mpavlovi}@fau.edu

Abstract- Computational proteomics uses algorithms and models to investigate active predictable proteins. The principal research in this area is often focused on the rational vaccine design (RVD) and computer-aided medicines. Vital goals for this automatic body counting method are the results of a broken immune system where "themselves" are unable to distinguish from "unwanted ones" and thus lead to attacks on their structures (proteins and DNA, in particular) [1]. While we are not aware of the primary etiology of autoimmune diseases, there is a range of factors suspected to be the triggers of autoimmune diseases, including genetic predisposition, natural causes like viral infections, including gut microbiota, fungal viruses, and gluttonous infections, along with body and environmental agents, hormonal characteristics, and immune dysregulation [2]. On the other hand, Immune hyperactivation and excessive cytokine release, leading to multiorgan failure, are associated with severe/fatal conditions of COVID-19. It has been proposed that SARS-CoV-2 may be a factor in the development of acute autoimmune disease due to shared pathogenetic mechanisms, clinical characteristics, and radiological radiation between inflammatory diseases and Covid-19. Hence, the understanding of protein-RNA binding can provide crucial information on the efficacy and modification of mRNA vaccine targets. Therefore, the binding affinity and antigenicity of viral RNA and human antiRNA should be determined, and viral RNA antibodies should be distinguished from human/human antiRNA antibodies. In this paper, we are proposing a python-based software tool using machine learning approach for testing viral RNA-human antiRNA binding affinity and antigenicity.

Keywords- COVID-19, SARS COV2, coronavirus, antibodies, antigenicity, RNA vaccines, autoimmunity, molecular mimicry.

I. INTRODUCTION

Coronaviruses correspond to a broad group of viruses that, via zoonotic transmission, primarily affect humans. Some patients with respiratory failure and often with hyper-ferritinemia and multiorgan involvement are affected by a sepsis pneumonia leading to acute respiratory depression (ARDS), including death-related hematologic, intestinal, nerve and heart disease [3][4][5][6]. Cellular mimicry, however, has been suggested to contribute to this issue, with SARS-CoV-2 spike glycoprotein antibodies reacting to a similarly structured heptapeptide protein combination and triggering a strong (automatic) immune response against them [5]. Common pathogenetic mechanisms, clinical features and radiation between inflammatory diseases and COVID-19 indicate that SARS-CoV-2 could be a factor in the development of acute autoimmune disease.

Autoimmune diseases start with the immune system losing its ability to tolerate and not differentiate between artificial insemination. Thus, the immune system fights off its gene (RNA in this case). These antibodies are the proteins that bind to RNA and make it hydrolyze to form complex structures precipitating below the endothelium of blood vessels [6].

The anti-DNA antibody is found in many patients associated with autoimmune diseases. However, the chemical composition and processes under autoantibodies remain poorly understood. Researchers [7][8][9] were able to differentiate anti-DNA and only a small number of binding sites of anti-DNA antibodies. Thus, the understanding of protein-RNA binding can provide crucial information on the efficacy and modification of mRNA vaccine targets. Therefore, viral and human RNA's binding sequence and antigenicity should be determined, and viral RNA antibodies should be distinguished from human/human anti RNA antibodies. Sequencing of wet markers and X-ray crystallography require expertise in each body because of their cost and lengthy time. Computerized methods, which are a model of existing data for making novel predictions,

can reduce the field of candidates who may be tested in the lab. Due to the need for hand-crafted functionality, classical machine learning approaches frequently face difficulties processing raw data. In comparison, deep learning techniques can directly extract functional characteristics from the data. However, neural networks in ML have never been an option for long time, although other alternatives have been built up. Combined with many training data, the advent of powerful computer hardware and many fast processors (such as GPUs) have made deep neural networks (DNNs) perform better in many ML systems recently. In this paper, we are proposing a python-based software tool using machine learning approach for testing viral RNA-human antiRNA binding affinity and antigenicity.

II. AUTOIMMUNITY AND ITS MECHANISMS

Autoimmunity is often described as a condition in which the immune system or T cells respond to autoantigens. Autoimmunity promotes autoimmune diseases. Recent research reveals that autoantibodies or active T cells are present even in healthy individuals. The immune system has a variety of mechanisms that suppress the body's response to behavior, and disruption of these mechanisms causes autoimmune diseases [10][11][12].

The formation or activation of CD4 T cells that react to a particular autoantigen are usually thought to result in autoimmunity. Based on different facts, it is now proposed that a specific antigenic factor is the primary cause of independence. Microbial infection can trigger an autoimmune response by molecular modification and polyclonal activation and release a separate autoantigen. Non-infectious factors are also considered to be the beginning of the immune system [13].

Autoimmune diseases are believed to be rooted in a combination of genetic and environmental factors, like many other complex diseases. The simple hypothesis is that polymorphisms cause incomplete control or decrease the limit of lymphocyte activation in different genes. Natural factors cause or enhance the initiation of self-functioning lymphocytes that have escaped control and are ready to respond people. Some of these genes and environmental factors have started to be recognized [14].

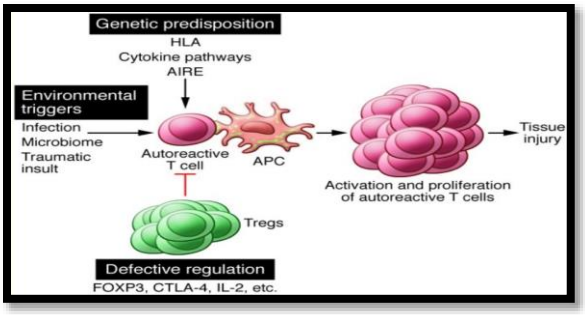

Figure 1 Genetic vulnerability, environmental provocations, and faulty control are accountable for introducing autoimmunity [14].

Genetic polymorphisms can decrease active T cells' activation rate in body-related genes (as well as HLA, cytokines/receptors, and those engaged in mild intolerance). An inflammatory environment that promotes active lymphocytes' development produces natural factors such as inflammation, microbiome, and tissue harm. Tregs usually operate to suppress active T cells, but these cells may be inactive for disability, development, or function and may not control functional T responses. These substances may cause to the escape, development, and proliferation of autoreactive lymphocytes alone or in combination, leading to tissue damage and consecutive therapy (Figure 1) [14][15][16].

III. IMMUNE RESPONSES SIMILARITIES BETWEEN SARS-COV-2 INFECTION AND AUTOIMMUNE DISEASES

Autoimmune disorders are characterized by the existence of autoantibodies and recurrent inflammatory reactions, causing in target organ harm and malfunction because of the lack of immune tolerance and a dysregulated immune system [17]. COVID-19 has immune-mediated injuries as well. The SARS-CoV-2 infection triggers immune responses, affecting vaccine production against this virus [18]. The regulation of SARS-CoV-2 disease relies heavily on T cell immunity. SARS-CoV-2 is protected by antigen-specific CD4+ and CD8+ T cells and neutralizes the response of antibody, while degraded adaptive immune responses can result in poor outcome conditions such as an absence of naive T cells [19]. Lymphopenia is associated in clinical laboratory studies with severe illnesses in COVID-19 patients and may have been the predictor of serious disease and death [20][21][22][23]. Another notable hematological transition is neutrophilia and the associated excess neutrophil extracellular traps, which are parallel in patients with extreme COVID-19 symptoms [24]. In the COVID-19, the immune

response is a two-edged sword with results influenced by the degradation of cytokine imbalance and the activation of the immune cell. Excessive development and release of pro-inflammatory cytokines and chemokines, such as in autoimmune diseases, can lead to severe organ damage in extreme cases. Liu et al. explained the immunopathogenesis similarities between COVID-19 and autoimmune diseases [25].

Table I encapsulates immunopathogenesis similarities between COVID-19 and autoimmune diseases [25].

Items	COVID-19 immunological features like autoimmune diseases
Innate immune cells	Overactivation of monocytes, macrophages, mast cells, and neutrophils. Increased proportion of mature natural killer (NK) cells.
Adaptive immune cells	Decreased T-cell numbers, altered B-cell subsets, dysregulation of T cells and B cells.
Cytokines and chemokines	Increased levels of IL-1, IL-2, IL-6, IL-8, IL-10, IL-17, IL-18, CXCL10, CCL2.
Autoantibodies	ANA, APL, lupus anticoagulant, cold agglutinins, anti-Ro/SSA antibodies, anti-Caspr2 antibody, anti GD1b antibody, anti-MOG antibody
Clinical conditions	Immune-mediated hemolysis, decreased white blood cell counts, cytokine storm syndrome, macrophage activation syndrome, procoagulant condition
Other immunopathogenesis	Increased levels of DAMPs, molecular mimicry

IV. SARS-COV-2 AND MOLECULAR MIMICRY

Autoantibodies are an essential characteristic of autoimmune diseases. The underlying mechanisms, on the other hand, are complex and not completely understood. Infectious pathogens are thought to use molecular mimicry as one of the tools [26]. By exposing antigen epitopes that elicit cross-reactive antibodies, viral infection may disrupt immunologic tolerance. Antigenic mimicry between viral and human proteins has been recorded in several studies. The immune response to Epstein–Barr virus (EBV) in lupus patients is perhaps one of the most well-known molecular mimicry examples in autoimmunity [27]. An abnormal immune response to Epstein–Barr virus Nuclear Antigen-1 (EBNA-1) might be responsible for an autoimmune reaction against the Sm and Ro autoantigen systems [28]. Anti-EBNA-1 antibodies and myelin essential protein cross-reactivity have also been demonstrated in patients with multiple sclerosis [29].

Furthermore, EBNA-1 resembled synuclein, a brain protein linked to multiple sclerosis, and was predicted to bind HLA class II DR2b (HLA-DRB115: 01) [30]. In silico research revealed that a human endogenous retrovirus (HERV) envelope protein has a similar sequence to three myelin proteins that have been connected to an autoimmune response in multiple sclerosis and are predicted to bind to HLA-DRB1: 01. Basavalingappa et al. [31] showed that infection with Coxsackievirus B3 (CVB3) would result in the generation of autoreactive T cells for various antigens.

V. AUTOANTIBODIES IN COVID-19 PATIENTS

Autoantibodies associated with a variety of autoimmune disorders have been found in COVID-19 patients (Table II). Pascolini et al. identifies antinuclear antibodies (ANA), anti cytoplasmic neutrophil antibodies (ANCA), and antiphospholipid (APL) antibodies in 33 COVID-19 patients [32]. According to the findings, 45 percent of the patients had positive autoantibodies, and patients with positive autoantibodies had a poorer prognosis and a slightly higher respiratory rate at admission. Antinuclear antibodies (IgG and/or IgM) were found in 33% of patients, anticardiolipin antibodies (IgG and/or IgM) were found in 24%, and anti2-glycoprotein-I antibodies (IgG and/or IgM) were found in three patients (9 percent). ANCA, on the other hand, was negative in all the patients [32]. Coagulopathy is a dangerous side effect of SARS-CoV-2 infection. A cohort study was recently conducted at Montefiore Medical Center to determine lupus anticoagulant presence in COVID-19 patients. The researchers discovered that COVID-19 patients had a higher rate of lupus anticoagulant positivity than controls who tested negative for COVID-19 reverse transcriptase–PCR.

Furthermore, COVID-19 patients with a positive lupus anticoagulant had a higher thrombosis rate [33]. Amezcua-Guerra et al. [34] also found that patients with severe and essential COVID-19 have a higher prevalence of APL antibodies and that the presence of APL antibodies is linked to a hyperinflammatory state with highly high ferritin, C reactive protein, and IL-6, and with pulmonary thromboembolism. The findings suggest that SARS-CoV-2 can cause autoimmune responses, which may

explain the hypercoagulable condition seen in extreme and critical COVID-19 cases. Liu et al. explained the autoantibodies detected in COVID-19 patients and their clinical significance [25].

Table II describes the autoantibodies detected in COVID-19 patients [25].

Autoantibodies	Clinical significance
ANA	Poor prognosis and a significant higher respiratory rate
APL	Poor prognosis and a significant higher respiratory rate. Possible association with a hyperinflammatory state and thrombosis and thromboembolism
Lupus anticoagulant	A higher rate of thrombosis
Cold agglutinins	Haemolytic anaemia.Complicating laboratory assessment and renal replacement therapy
Anti-Ro/SSA antibodies	Possible association with severe pneumonia
Anti-Caspr2 antibody	Unclear
Anti-GD1b antibody	Unclear
Anti-MOG antibody	Unclear
Red cell bound antibodies	Associated with the severity of anaemia

VI. LITERATURE REVIEW

For both researchers and mechanical researchers, the key functionalities of protein-RNA binding in many biological conditions [35] makes it an essential study field. Advanced measuring methods, both in vivo tests and in vitro, were used on the experimental side. The CLIP method and its findings assess, on a written scale, the binding of protein-RNA in vivo [36][37][38][39]. Several mobile orthogonal events, which have contributed to noise and signal loss, have negatively impacted these values. Consequently, these measurements are not sufficiently precise to provide reliable results for the calculation. Instead, they create a twofold effect: yes (existence of bonds) or no. The identification of about 100 nuclear substances establishes the binding of one of the other viable proteins present in vivo in all CLIP studies. In the sense of protein binding and technical objects and experimental tone, the inherent difficulty of in vivo makes reading the binding proteins-RNA from such data a difficult task [40][41].

RNAcompete, another test form, works in vitro [42][43][44]. The binding of a single protein to 240,000 short RNAs (30-40 nucleotide lengths) is calculated for each RNA-compete test. These tests display low noise-to-signal levels and are consistent enough to generate good measurements of binding parameters or strength, as they do not have distracting cellular processes. 244 of these tests are in the most detailed in vitro database assessed using RNAcompete [45] technology (each for a single protein).

The computer challenges arising from this experimental data are to install models of RNA-binding protein to predict the binding of a given protein with a new RNA transcript. To tackle this challenge, several approaches have been established. All computer systems observe the presence of RNA sequences. Others also noticed a secondary RNA structure. We note that arithmetical methods predict secondary formation based on the sequence itself. Computer-generated predictions are very accurate in the short RNA series [46]. In the same cells currently available with only two proteins, data sets contain RNA-binding and RNA levels [47].

MEMERIS, the first computerized method, uses a wait-and-add algorithm to detect sequencing motifs that may not be abolished in RNA regions and are therefore available for binding [48]. The RNAcontext, made with RNAcompete technology, is a basic model for sequencing and binding preferred structures [49]. The sequence preferences are represented as the weight of the position, thus contributing to the binding in each position independently of the other. In each sense of the house, the property preferences are displayed as the preference vector. The latest method, GraphProt, uses RNA graph structure representation to imitate a binding sequence and structure to obtain rich local graphs [50]. GraphProt takes over seven days to conduct a single RNAcompete test [51]. In order to read and predict the binding of protein-DNA and protein RNA to several data bases, including the RNAcompete and CLIP, DeepBind uses a new approach based on deep learning. It is based solely on the RNA sequence, i.e., without regard to RNA formation [52]. RCK, a recent invention, and state-of-the-art technology, extend RNAcontext through all sequences and structure levels using a k-mer-based model [51]. Each RNA terminal k provides binding points lower than the total of each structure and can thus capture the location dependency within the

binding site. Depending on several data sources reflecting the complexity of mobile space, IDeep faces the challenge of predicting vivo binding. As part of the input [53], it seeks protein binding preferences and solves a different problem. Based on in-depth learning and use of both secondary and tertiary RNA structures, Deepnet-RBP learns RNA binding preferences but was designed to read-only from vivo data [54]. The new approach focused on in-depth learning, the pysster, explores a single structure of RNA [55], respectively. Furthermore, it solves the sequential separation problem and, where possible, does not predict binding capacity. No studies supporting the high-end machine learning mechanism of protein-RNA sequence learning and the order of binding preferences from the highest in vitro data are available today [56].

In order to predict protein binding in DNA-binding hydrolyte antibodies, Clair et al. used long-term, short-term memory (LSTM) models trained directly for FASTA primary sequences. They used the same dataset with CNN models. Although the model of the CNN exceeded the LSTM's primary binding prediction task, an analysis of internal representation of both models indicated that the models with sub-sequences were associated more closely with sites which were known to be involved in binding. These results demonstrate that studying the internal processes of recurring neural network models can be a powerful primary sequence analysis method [1].

Several B-cell epitopes do not produce neutralizing antibodies in B epitope databases (and therefore do not provide protective immunity). This leads to the compilation of well-defined data sets of protective B-cell epitopes [57]. B-cell epitopes are an important step in the development of peptide vaccines. A number of methods have been used to predict continuous B-cell epitopes and most of them were based on the physical and chemical properties of amino acids [58][59]. It is currently hard to say which residue property works better than the others, since the existing methods are not evaluated independently or benchmarked.

The inherent complexity of the immune and recognition processes complicates epitope prediction. Several methods have been developed for the prediction of B-cell epitopes based on the physicochemical properties of the amino acids [60][61]. Classic methods for determination of linear antigenic sequence B-cell epitopes usually rely on the use of susceptibility scales [62][63]. Some methods of predicting linear B-cell epitopes were recently proposed based on machine learning and statistical approaches [64]. The hydrophilic analysis (on 12 proteins) of Hopps and Woods [65] examined the possibility that some antigenic factors may be linked to a sequence of amino acids with loaded and polar residues and a lack of high levels of hydrophobic residues. Parker and others used the modified peptide retention time (HPLC) hydrophilic scales on a reversed-stage column [66]. The known B temperature factors have led to the development of a link between the anti-glycemicity and segmental mobility of carbons of 31 proteins in the known structure, to prevent the mobility of proteins segments. The flexibility scale is used to predict the epitopes of B-cells Methods for predicting epitopes were also developed based on amino acid surface accessibility [65][66]. In 169 experimentally known epitopes, Hopps and Parker derived their own antigenicity scale from residue frequency [65]. Based on the occurrence of amino acids, Pellequer and others derived turn scales at each of the four turning points using a structural database composed of 87 proteins [63]. 70 percent of the known epitopes were correctly predicted by the turn.

Hopp and Woods are most useful in determining portions of the protein sequence involved in molecular surface interactions [67] with their original hydraulicity plotting procedure. When determining any method, it is important to choose the average group length or window. In the location of antigenic protein sites, a window of six residues works optimally [65]. The design of vaccines, immunodiagnostic testing and antibody production plays an important part in mapping B cell epitopes. As B-cell epitopes take time and costly to test the determination of antigenic cells, computational techniques are needed urgently to reliably identify the supposed B-cell epitopes [68].

The best alternative for reducing the number of peptides that can be synthesized for wet laboratory experimentation is the bioinformatics approach to predict the linear B cell epitope in the protein sequence. The prediction of linear B-cell epitopes has previously been based on numerous computing methods and programs based on the hydrostaticity, accessibility, flexibility or secondary structure propensity scales from the 20 natural amino acids. [69]. In this study results of various residue properties commonly used in the B-cell prediction were evaluated. The purpose of the study is to develop a machine learning based software tool that calculates the protein's physiochemical properties from the primary protein sequence and predict secondary protein elements using python and hence can be used as a B-cell epitope from the amino acid sequence

identified in the proteins by means of antigens. Later, we can calculate the binding affinity of those protein sequences.

VII. PROPOSED METHODOLOGY

A. Dataset

Viral proteins are the proteins produced by a virus within the host. When viruses hijack many of their human's mobile machinery, few of their own genes are coded. Consequently, structural parts, such as the viral envelope and capsid are usually viral proteins. In one study, a single sample of bronchoalveolar lavage fluid was sequenced and on 26 December 2019 the virus proteins were found in the novel genome of COVID-19 at the Wuhan Central Hospital. The NCBI (NC 045512) and GenBank (NC 045512.gb) databases contained the coronavirus genome sequence. COVID-19 genome analysis was carried with Biopython and DNA features viewer libraries [70][71].

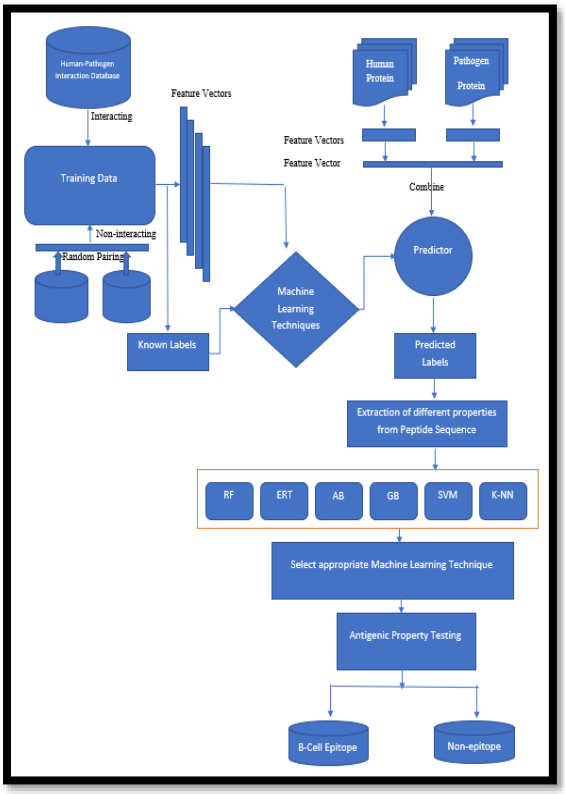

Figure 2 A Machine Learning Approach for Testing COVID RNA-Human AntiRNA Antigenicity and Binding Affinity

B. Classifiers

In order to guarantee the authenticity of a hypothesis unbiased of the choice of machine learning models, we utilized two different supervised classification systems: a linear Vector Support Machine (SVM) [72] and Random Forest Classification (RF) [73]. We used vector support machines (SVMs) as interactions or as non-interactions to classify human-pathogen protein pairs. SVMs have several benefits, including strong theory, optimum margin classification and higher dimension generalization performance. A Random Forest (RF) on the other hand is a machine-learning technique based on ensemble using the algorithm of decision-making trees [74]. RF samples sub-sets of feature spaces are used to train individual decision tree randomly.

C. Feature Extraction

In this study, we employed K-mer to guarantee the authenticity of the hypothesis proposed is unbiased of the choice of the representing function. The structure of K-mer is obtained by counting occurrences of k-mers in a protein sequence and is a widely used narrative of the protein sequence [75]. Here, by combining aminic acids into seven groups based on their hydrophobic and electrostatic properties, we have extracted this representation [76]. By calculating the group occurrences of k-mers, every protein sequence is translated to a characteristic illustration of length 7^k.

D. Model Evaluation

In a fold wise performance evaluation of a predictor, there is always possibility of a check between proteins used as a training and test dataset. This train and test data overlap can provide inflationary performance measurements. We have used the classic K-fold cross validation method, which is used to randomly separate the original data set into K subsets (folds) [77]. K-1 sets are used to train the predictor from these K subsets and the rest of the set is used to test.

E. B-Cell Epitope Prediction for Antigenicity Testing

We extract various characteristics from the peptide sequence and produce several hybrid features with a variety of composition combinations, including amino acid composition (AAC), amino acid index (AAI), dipeptide composition (DPC), chain-transition-distribution (CTD), and physicochemical properties (PCP) [78]. We will then look at six ML algorithms including the vector-assisted support-machine (SVM), random forest (RF), extremely randomized tree

(ERT), AdaBoost (AB), gradient boosting (GB) and k-nearest (k-NN). Finally, to predict the B-cell epitope, we are using our proposed model.

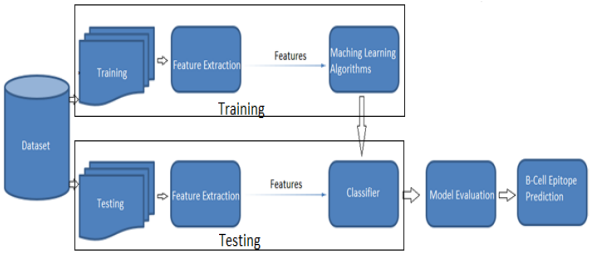

Figure 3 UML Class Diagram for the proposed python-based software tool using Machine Learning approach

VIII. CONCLUSION

The core biologically important protein interaction problems are discussed in this paper. We have highlighted various problems and proposed several learning machines to interact with proteins, binding affinity, and antigenicity. Issues of human-pathogenic interaction prediction (HPI) have been pointed out this research. with assessment scheme and proposed new evaluation scheme with various biologic centric metrics within this area. Several modern sequence-based machine-learning models for the prediction of protein-binding affinities were introduced. We have also established a new method of antigenicity prediction using machine learning algortihms that uses protein structural information and sequence information during training, but only requires sequence information during testing. According to our best knowledge, this is the first step towards combining both antigenicity and binding affinity prediction with ML algorithms in a single software context.

ACKNOWLEDGMENT

I would like to express my deep and sincere gratitude to my research supervisor Dr. Maria M. Larrondo-Petrie, and Dr. Mirjana D. Pavlovic, Florida Atlantic University for giving me the opportunity to do research and providing invaluable guidance throughout

REFERENCES

[1] R. St.Clair, M. Teti, M. Pavlovic, W. Hahn, and E. Barenholtz, “Predicting residues involved in anti-DNA autoantibodies with limited neural networks,” *bioRxiv*. bioRxiv, p. 2020.08.06.240101, Aug. 06, 2020, doi: 10.1101/2020.08.06.240101.

[2] M. Ehrenfeld *et al.*, “Covid-19 and autoimmunity,” *Autoimmunity Reviews*. 2020, doi: 10.1016/j.autrev.2020.102597.

[3] W. Guan *et al.*, “Clinical Characteristics of Coronavirus Disease 2019 in China,” *N. Engl. J. Med.*, 2020, doi: 10.1056/nejmoa2002032.

[4] L. Wang, Y. Wang, D. Ye, and Q. Liu, “Review of the 2019 novel coronavirus (SARS-CoV-2) based on current evidence,” *Int. J. Antimicrob. Agents*, 2020, doi: 10.1016/j.ijantimicag.2020.105948.

[5] G. Zandman-Goddard and Y. Shoenfeld, “Hyperferritinemia in autoimmunity,” *Isr. Med. Assoc. J.*, 2008.

[6] S. Colafrancesco, C. Alessandri, F. Conti, and R. Priori, “COVID-19 gone bad: A new character in the spectrum of the hyperferritinemic syndrome?,” *Autoimmunity Reviews*. 2020, doi: 10.1016/j.autrev.2020.102573.

[7] L. S. Rodkey *et al.*, “DNA hydrolysis by monoclonal autoantibody BV 04-01,” *Appl. Biochem. Biotechnol. - Part A Enzym. Eng. Biotechnol.*, 2000, doi: 10.1385/abab:83:1-3:95.

[8] M. Pavlovic *et al.*, “Highly specific novel method for isolation and purification of lupus anti-DNA antibody via oligo-(dT) magnetic beads,” 2007, doi: 10.1196/annals.1422.023.

[9] A. V. Kozyr, “A novel method for purification of catalytic antibodies toward DNA from sera of patients with lymphoproliferative diseases,” *Biochem. Mol. Biol. Int.*, 1996, doi: 10.1080/15216549600201441.

[10] “mRNA vaccine ‘EMPOWER IAS’ | Empower IAS.” https://empowerias.com/blog/prelims-special-facts/mrna-vaccine-empower-ias (accessed Feb. 02, 2021).

[11] S. Jain, H. Batra, P. Yadav, and S. Chand, “Covid-19 vaccines currently under preclinical and clinical studies, and associated antiviral immune response,” *Vaccines*. 2020, doi: 10.3390/vaccines8040649.

[12] “Comparison of new vaccine approaches – COVID-19 -.” https://www.europeanpharmaceuticalreview.com/article/122275/comparison-of-new-vaccine-approaches-covid-19/ (accessed Feb. 01, 2021).

[13] “(No Title).” https://www.med.or.jp/english/pdf/2004_09/403_406.pdf (accessed Feb. 07, 2021).

[14] M. D. Rosenblum, K. A. Remedios, and A. K. Abbas, “Mechanisms of human autoimmunity,” *J. Clin. Invest.*, 2015, doi: 10.1172/JCI78088.

[15] E. Åhlin *et al.*, “Autoantibodies associated with RNA are more enriched than anti-dsDNA antibodies in circulating immune complexes in SLE,” *Lupus*, 2012, doi: 10.1177/0961203311434938.

[16] “Autoimmune Antibodies: Autoantibodies.” https://emedicine.medscape.com/article/2172244-overview (accessed Feb. 07, 2021).

[17] A. Hejrati *et al.*, “Innate immune response in systemic autoimmune diseases: a potential target of therapy,” *Inflammopharmacology*. 2020, doi: 10.1007/s10787-020-00762-y.

[18] A. Singh, M. Thakur, L. K. Sharma, and K. Chandra, “Designing a multi-epitope peptide based vaccine against SARS-CoV-2,” *Sci. Rep.*, 2020, doi: 10.1038/s41598-020-73371-y.

[19] C. Rydyznski Moderbacher *et al.*, “Antigen-Specific Adaptive Immunity to SARS-CoV-2 in Acute COVID-19 and Associations with Age and Disease Severity,” *Cell*, 2020, doi: 10.1016/j.cell.2020.09.038.

[20] G. Lancman, J. Mascarenhas, and M. Bar-Natan, “Severe COVID-19 virus reactivation following treatment for B cell acute lymphoblastic leukemia,” *J. Hematol. Oncol.*, 2020, doi: 10.1186/s13045-020-00968-1.

[21] S. Setiati *et al.*, "Risk factors and laboratory test results associated with severe illness and mortality in COVID-19 patients: A systematic review," *Acta Med. Indones.*, 2020.

[22] A. Ziadi *et al.*, "Lymphopenia in critically ill COVID-19 patients: A predictor factor of severity and mortality," *International Journal of Laboratory Hematology*. 2021, doi: 10.1111/ijlh.13351.

[23] F. Ciceri *et al.*, "Early predictors of clinical outcomes of COVID-19 outbreak in Milan, Italy," *Clin. Immunol.*, 2020, doi: 10.1016/j.clim.2020.108509.

[24] J. Wang *et al.*, "Excessive Neutrophils and Neutrophil Extracellular Traps in COVID-19," *Front. Immunol.*, 2020, doi: 10.3389/fimmu.2020.02063.

[25] Y. Liu, A. H. Sawalha, and Q. Lu, "COVID-19 and autoimmune diseases," *Curr. Opin. Rheumatol.*, 2020, doi: 10.1097/bor.0000000000000776.

[26] Z. Reyes-Castillo, E. Valdés-Miramontes, M. Llamas-Covarrubias, and J. F. Muñoz-Valle, "Troublesome friends within us: the role of gut microbiota on rheumatoid arthritis etiopathogenesis and its clinical and therapeutic relevance," *Clinical and Experimental Medicine*. 2021, doi: 10.1007/s10238-020-00647-y.

[27] J. B. Harley and J. A. James, "Everyone comes from somewhere: Systemic lupus erythematosus and Epstein-Barr virus induction of host interferon and humoral anti-Epstein-Barr nuclear antigen 1 immunity," *Arthritis and Rheumatism*. 2010, doi: 10.1002/art.27421.

[28] N. R. Jog *et al.*, "Association of Epstein-Barr virus serological reactivation with transitioning to systemic lupus erythematosus in at-risk individuals," *Ann. Rheum. Dis.*, 2019, doi: 10.1136/annrheumdis-2019-215361.

[29] N. R. Jog *et al.*, "Epstein Barr virus nuclear antigen 1 (EBNA-1) peptides recognized by adult multiple sclerosis patient sera induce neurologic symptoms in a murine model," *J. Autoimmun.*, 2020, doi: 10.1016/j.jaut.2019.102332.

[30] R. Ramasamy, F. Mohammed, and U. C. Meier, "HLA DR2b-binding peptides from human endogenous retrovirus envelope, Epstein-Barr virus and brain proteins in the context of molecular mimicry in multiple sclerosis," *Immunol. Lett.*, 2020, doi: 10.1016/j.imlet.2019.10.017.

[31] R. H. Basavalingappa *et al.*, "Viral myocarditis involves the generation of autoreactive T cells with multiple antigen specificities that localize in lymphoid and non-lymphoid organs in the mouse model of CVB3 infection," *Mol. Immunol.*, 2020, doi: 10.1016/j.molimm.2020.06.017.

[32] S. Pascolini *et al.*, "COVID-19 and Immunological Dysregulation: Can Autoantibodies be Useful?," *Clin. Transl. Sci.*, 2020, doi: 10.1111/cts.12908.

[33] M. Reyes Gil, M. Barouqa, J. Szymanski, J. D. Gonzalez-Lugo, S. Rahman, and H. H. Billett, "Assessment of Lupus Anticoagulant Positivity in Patients With Coronavirus Disease 2019 (COVID-19)," *JAMA Netw. open*, 2020, doi: 10.1001/jamanetworkopen.2020.17539.

[34] L. M. Amezcua-Guerra *et al.*, "Presence of antiphospholipid antibodies in COVID-19: case series study," *Ann. Rheum. Dis.*, 2020, doi: 10.1136/annrheumdis-2020-218100.

[35] J. König, K. Zarnack, N. M. Luscombe, and J. Ule, "Protein-RNA interactions: New genomic technologies and perspectives," *Nat. Rev. Genet.*, 2012, doi: 10.1038/nrg3141.

[36] R. B. Darnell, "HITS-CLIP: Panoramic views of protein-RNA regulation in living cells," *Wiley Interdiscip. Rev. RNA*, 2010, doi: 10.1002/wrna.31.

[37] M. Hafner *et al.*, "Transcriptome-wide Identification of RNA-Binding Protein and MicroRNA Target Sites by PAR-CLIP," *Cell*, 2010, doi: 10.1016/j.cell.2010.03.009.

[38] J. Konig *et al.*, "ICLIP - transcriptome-wide mapping of protein-RNA interactions with individual nucleotide resolution," *J. Vis. Exp.*, 2011, doi: 10.3791/2638.

[39] E. L. Van Nostrand *et al.*, "Robust transcriptome-wide discovery of RNA-binding protein binding sites with enhanced

CLIP (eCLIP)," *Nat. Methods*, 2016, doi: 10.1038/nmeth.3810.

[40] S. Kishore, L. Jaskiewicz, L. Burger, J. Hausser, M. Khorshid, and M. Zavolan, "A quantitative analysis of CLIP methods for identifying binding sites of RNA-binding proteins," *Nat. Methods*, 2011, doi: 10.1038/nmeth.1608.

[41] Y. Orenstein, R. Hosur, S. Simmons, J. Bienkoswka, and B. Berger, "Sequence biases in CLIP experimental data are incorporated in protein RNA-binding models," *bioRxiv*, 2016, doi: 10.1101/075259.

[42] K. B. Cook *et al.*, "RNAcompete-S: Combined RNA sequence/structure preferences for RNA binding proteins derived from a single-step in vitro selection," *Methods*, 2017, doi: 10.1016/j.ymeth.2017.06.024.

[43] N. Lambert, A. Robertson, M. Jangi, S. McGeary, P. A. Sharp, and C. B. Burge, "RNA Bind-n-Seq: Quantitative Assessment of the Sequence and Structural Binding Specificity of RNA Binding Proteins," *Mol. Cell*, 2014, doi: 10.1016/j.molcel.2014.04.016.

[44] D. Ray, K. C. H. Ha, K. Nie, H. Zheng, T. R. Hughes, and Q. D. Morris, "RNAcompete methodology and application to determine sequence preferences of unconventional RNA-binding proteins," *Methods*, 2017, doi: 10.1016/j.ymeth.2016.12.003.

[45] D. Ray *et al.*, "A compendium of RNA-binding motifs for decoding gene regulation," *Nature*, 2013, doi: 10.1038/nature12311.

[46] K. J. Doshi, J. J. Cannone, C. W. Cobaugh, and R. R. Gutell, "Evaluation of the suitability of free-energy minimization using nearest-neighbor energy parameters for RNA secondary structure prediction," *BMC Bioinformatics*, 2004, doi: 10.1186/1471-2105-5-105.

[47] R. C. Spitale *et al.*, "Structural imprints in vivo decode RNA regulatory mechanisms," *Nature*, 2015, doi: 10.1038/nature14263.

[48] M. Hiller, R. Pudimat, A. Busch, and R. Backofen, "Using RNA secondary structures to guide sequence motif finding towards single-stranded regions," *Nucleic Acids Res.*, 2006, doi: 10.1093/nar/gkl544.

[49] H. Kazan, D. Ray, E. T. Chan, T. R. Hughes, and Q. Morris, "RNAcontext: A new method for learning the sequence and structure binding preferences of RNA-binding proteins," *PLoS Comput. Biol.*, 2010, doi: 10.1371/journal.pcbi.1000832.

[50] D. Maticzka, S. J. Lange, F. Costa, and R. Backofen, "GraphProt: Modeling binding preferences of RNA-binding proteins," *Genome Biol.*, 2014, doi: 10.1186/gb-2014-15-1-r17.

[51] Y. Orenstein, Y. Wang, and B. Berger, "RCK: Accurate and efficient inference of sequence- and structure-based protein-RNA binding models from RNAcompete data," *Bioinformatics*, 2016, doi: 10.1093/bioinformatics/btw259.

[52] B. Alipanahi, A. Delong, M. T. Weirauch, and B. J. Frey, "Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning," *Nat. Biotechnol.*, 2015, doi: 10.1038/nbt.3300.

[53] X. Pan and H. Bin Shen, "RNA-protein binding motifs mining with a new hybrid deep learning based cross-domain knowledge integration approach," *BMC Bioinformatics*, 2017, doi: 10.1186/s12859-017-1561-8.

[54] S. Zhang *et al.*, "A deep learning framework for modeling structural features of RNA-binding protein targets," *Nucleic Acids Res.*, 2015, doi: 10.1093/nar/gkv1025.

[55] S. Budach and A. Marsico, "Pysster: Classification of biological sequences by learning sequence and structure motifs with convolutional neural networks," *Bioinformatics*, 2018, doi: 10.1093/bioinformatics/bty222.

[56] I. Ben-Bassat, B. Chor, and Y. Orenstein, "A deep neural network approach for learning intrinsic protein-RNA binding preferences," 2018, doi: 10.1093/bioinformatics/bty600.

[57] J. Sollner, R. Grohmann, R. Rapberger, P. Perco, A. Lukas, and B. Mayer, "Analysis and prediction of protective continuous B-cell epitopes on pathogen proteins,"

Immunome Res., 2008, doi: 10.1186/1745-7580-4-1.

[58] R. Meloen, W. Puijk, J. Langeveld, J. Langedijk, and P. Timmerman, "Design of Synthetic Peptides for Diagnostics," *Curr. Protein Pept. Sci.*, 2005, doi: 10.2174/1389203033487144.

[59] A. L. Kazim and M. Z. Atassi, "A novel and comprehensive synthetic approach for the elucidation of protein antigenic structures. Determination of the full antigenic profile of the α-chain of human haemoglobin," *Biochem. J.*, 1980, doi: 10.1042/bj1910261.

[60] S. Saha and G. P. S. Raghava, "BcePred: Prediction of continuous B-cell epitopes in antigenic sequences using physico-chemical properties," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 2004, doi: 10.1007/978-3-540-30220-9_16.

[61] A. S. Kolaskar and P. C. Tongaonkar, "A semi-empirical method for prediction of antigenic determinants on protein antigens," *FEBS Lett.*, 1990, doi: 10.1016/0014-5793(90)80535-Q.

[62] M. Odorico and J. L. Pellequer, "BEPITOPE: Predicting the location of continuous epitopes and patterns in proteins," *J. Mol. Recognit.*, 2003, doi: 10.1002/jmr.602.

[63] J. L. Pellequer, E. Westhof, and M. H. V. Van Regenmortel, "Predicting location of continuous epitopes in proteins from their primary structures," *Methods Enzymol.*, 1991, doi: 10.1016/0076-6879(91)03010-E.

[64] S. Saha and G. P. S. Raghava, "Prediction of continuous B-cell epitopes in an antigen using recurrent neural network," *Proteins Struct. Funct. Genet.*, 2006, doi: 10.1002/prot.21078.

[65] T. P. Hopp and K. R. Woods, "Prediction of protein antigenic determinants from amino acid sequences," *Proc. Natl. Acad. Sci. U. S. A.*, 1981, doi: 10.1073/pnas.78.6.3824.

[66] J. M. R. Parker, D. Guo, and R. S. Hodges, "New Hydrophilicity Scale Derived from High-Performance Liquid Chromatography Peptide Retention Data: Correlation of Predicted Surface Residues with Antigenicity and X-ray-Derived Accessible Sites," *Biochemistry*, 1986, doi: 10.1021/bi00367a013.

[67] T. P. Hopp, "Use of hydrophilicity plotting procedures to identify protein antigenic segments and other interaction sites," *Methods Enzymol.*, 1989, doi: 10.1016/0076-6879(89)78040-X.

[68] Y. El-Manzalawy, D. Dobbs, and V. Honavar, "Predicting protective linear B-cell epitopes using evolutionary information," 2008, doi: 10.1109/BIBM.2008.80.

[69] J. L. Pellequer, E. Westhof, and M. H. V. Van Regenmortel, "Correlation between the location of antigenic sites and the prediction of turns in proteins," *Immunol. Lett.*, 1993, doi: 10.1016/0165-2478(93)90072-A.

[70] P. J. A. Cock *et al.*, "Biopython: Freely available Python tools for computational molecular biology and bioinformatics," *Bioinformatics*, 2009, doi: 10.1093/bioinformatics/btp163.

[71] V. Zulkower and S. Rosser, "DNA features viewer, a sequence annotations formatting and plotting library for python," *bioRxiv*. 2020, doi: 10.1101/2020.01.09.900589.

[72] C. Cortes and V. Vapnik, "Support-Vector Networks," *Mach. Learn.*, 1995, doi: 10.1023/A:1022627411411.

[73] R. K. Barman, S. Saha, and S. Das, "Prediction of interactions between viral and host proteins using supervised machine learning methods," *PLoS One*, 2014, doi: 10.1371/journal.pone.0112034.

[74] L. Breiman, "Random forests," *Mach. Learn.*, 2001, doi: 10.1023/A:1010933404324.

[75] C. Leslie, E. Eskin, and W. S. Noble, "The spectrum kernel: a string kernel for SVM protein classification.," *Pac. Symp. Biocomput.*, 2002, doi: 10.1142/9789812799623_0053.

[76] J. Shen *et al.*, "Predicting protein-protein interactions based only on sequences information," *Proc. Natl. Acad. Sci. U. S. A.*, 2007, doi: 10.1073/pnas.0607879104.

[77] E. Alpaydin, “Design and Analysis of Machine Learning Experiments,” *Introd. to Mach. Learn.*, 2004.

[78] B. Manavalan, R. G. Govindaraj, T. H. Shin, M. O. Kim, and G. Lee, “iBCE-EL: A New Ensemble Learning Framework for Improved Linear B-Cell Epitope Prediction,” *Front. Immunol.*, 2018, doi: 10.3389/fimmu.2018.01695.