Interdisciplinary Sciences: Computational Life Sciences (2021) 13:44–60
https://doi.org/10.1007/s12539-020-00405-4

ORIGINAL RESEARCH ARTICLE

A Novel Protein Mapping Method for Predicting the Protein Interactions in COVID-19 Disease by Deep Learning

Talha Burak Alakus[1] · **Ibrahim Turkoglu**[2]

Received: 6 August 2020 / Revised: 23 November 2020 / Accepted: 28 November 2020 / Published online: 12 January 2021

Abstract

The new type of corona virus (SARS-COV-2) emerging in Wuhan, China has spread rapidly to the world and has become a pandemic. In addition to having a significant impact on daily life, it also shows its effect in different areas, including public health and economy. Currently, there is no vaccine or antiviral drug available to prevent the COVID-19 disease. Therefore, determination of protein interactions of new types of corona virus is vital in clinical studies, drug therapy, identification of preclinical compounds and protein functions. Protein–protein interactions are important to examine protein functions and pathways involved in various biological processes and to determine the cause and progression of diseases. Various high-throughput experimental methods have been used to identify protein–protein interactions in organisms, yet, there is still a huge gap in specifying all possible protein interactions in an organism. In addition, since the experimental methods used include cloning, labeling, affinity purification mass spectrometry, the processes take a long time. Determining these interactions with artificial intelligence-based methods rather than experimental approaches may help to identify protein functions faster. Thus, protein–protein interaction prediction using deep-learning algorithms has been employed in conjunction with experimental method to explore new protein interactions. However, to predict protein interactions with artificial intelligence techniques, protein sequences need to be mapped. There are various types and numbers of protein-mapping methods in the literature. In this study, we wanted to contribute to the literature by proposing a novel protein-mapping method based on the AVL tree. The proposed method was inspired by the fast search performance on the dictionary structure of AVL tree and was used to verify the protein interactions between SARS-COV-2 virus and human. First, protein sequences were mapped by both the proposed method and various protein-mapping methods. Then, the mapped protein sequences were normalized and classified by bidirectional recurrent neural networks. The performance of the proposed method was evaluated with accuracy, f1-score, precision, recall, and AUC scores. Our results indicated that our mapping method predicts the protein interactions between SARS-COV-2 virus proteins and human proteins at an accuracy of 97.76%, precision of 97.60%, recall of 98.33%, f1-score of 79.42%, and with AUC 89% in average.

Keywords COVID-19 · AVL tree · Protein mapping · Deep learning · SARS-COV-2

1 Introduction

The first corona virus incident occurred in Wuhan, China in December 2019, and spread rapidly to every region of the world [1, 2]. The disease caused by the SARS-COV-2 virus is called COVID-19 (Corona Virus Disease-2019). Most of corona viruses affect animals, yet they can also be transmitted to humans due to the genomic nature of humans [3]. The SARS-COV virus that emerged in 2002 and 2003 affected approximately 8000 people and had a mortality rate of 10% [4]. Similarly, the MERS-COV virus appeared in 2012 and a total of 2500 confirmed cases were observed [4]. The mortality rate of this virus has been observed as 36%

✉ Talha Burak Alakus
talhaburakalakus@klu.edu.tr

Ibrahim Turkoglu
iturkoglu@firat.edu.tr

[1] Faculty of Engineering, Department of Software Engineering, Kirklareli University, 39000 Kirklareli, Turkey

[2] Faculty of Technology, Department of Software Engineering, Firat University, 23119 Elazig, Turkey

[5]. These types of corona viruses reveal Acute Respiratory Distress Syndrome (SARS), which causes infection in the lungs and cause the death. It has been observed that the new type of corona virus spreads faster, is more difficult to control and has higher pandemic potential [5, 6]. Although 80% of infected people gently overcome this disease, COVID-19 can be fatal for people with medical problems such as diabetics, cancer, chronic respiratory failure, and heart failure [7]. Due to these reasons, it is vital to develop new strategies to counteract the SARS-COV-2 virus, to have knowledge of how the virus contacted the host during infection, and to develop new drugs or to reuse existing drugs.

So far, no medication has been developed for SARS-COV, MERS-COV, and SARS-COV-2 [8]. However, clinical trials are being conducted for any treatment, and both RNA and protein sequences are used effectively [9, 10]. One of these methods is based on protein–protein interactions. Thanks to these interactions, drug targets can be determined and drugs that can provide appropriate treatment can be identified [11, 12]. In addition to these, protein interactions are used effectively in determination of protein functions [13], diagnosis of cancer cells [14], phylogenetic analysis [15]. When determining interactions between proteins, two types of methods are generally applied; experimental and computational. Experimental methods consist of cloning, tandem affinity purification, nuclear magnetic resonance. All these experimental methods produce a large amount of data, and time and laboratory equipment are required to process these data [16]. In addition, the results of protein interactions vary according to the experimental methods used. Furthermore, since experimental methods are sensitive to both the environment and operational processes, false-positive and false-negative results may occur [17]. For these reasons, recently, computational methods have been preferred more than experimental methods and their popularity has increased [18, 19].

There are many computational methods have been proposed in the literature as complementary to experimental methods to predict interactions between proteins. These methods typically perform binary classification and predict whether the protein pairs interact or not. Prediction is usually performed based on protein domain information, gene expressions, gene neighborhood, protein structure information, and phylogenetic profiles. However, if there is no specific prior knowledge, these methods cannot be implemented. To perform the prediction process by computational methods, the interaction information of the protein pairs must be known beforehand. In computational methods, first genomic sequences are mapped and then protein interactions are predicted by classifying them with machine-learning and deep-learning approaches. Compared to experimental methods, computational methods are time efficient and can analyze the protein interactions with less equipment. Furthermore, with the recent development of technology, protein sequence information can be obtained easily. Experimental results have shown that amino acid sequences alone are sufficient in predicting protein interactions [17].

In this work, a novel numerical mapping method was proposed to predict the protein interactions between SARS-COV-2 and human, and the performance of the proposed method was validated on proteins belonging to COVID-19 disease. In the first part of the study, SARS-COV-2 and human proteins were mapped with both the proposed method and various mapping methods. The protein sequences of COVID-19 disease and the human genome were obtained from the BioGRID data set. Then, the mapped genomic data were normalized and classified with DeepBiRNN (Bidirectional Recurrent Neural Networks). The performance of numerical mapping methods was determined by accuracy, precision, recall, f1-score, and AUC (Area Under Curve) values. The experimental results showed that the proposed mapping method predicted the protein interactions between SARS-COV-2 virus proteins and human proteins at an accuracy of 97.76%, precision of 97.60%, recall of 98.33% and with AUC 89% in average. It has been observed that the proposed method is at least as effective and successful as other mapping methods. In some cases, it has even achieved the best evaluation results. Analyzing protein–protein interactions is crucial to understand the biological activities of organisms. In this way, protein functions and protein families can be designated. Using computational methods rather than experimental methods provides faster acquisition of protein function, protein family and protein interaction information. We have previously stated that experimental methods are both costly and time consuming. Thanks to the computational methods, we have proposed and existed in the literature, these problems have been avoided. In addition, by proposing a novel protein numerical mapping method, we have provided an alternative to mapping method that are scarce in this field. The proposed mapping method is an algorithm-based method, and it was used for the first time in the field of protein mapping. The main difference of the proposed mapping method from other methods used in the study is that it belongs to a different category. One of the methods is character based, the other one is signal based and the another one is physicochemical based. Information about the methods is given in the following sections. However, as can be seen, the main difference of the proposed mapping method is that the mapping process is performed in an algorithm-based manner.

The main contributions of the study can be summarized as follows:

- A novel numerical mapping method has been proposed to map protein sequences.

- To the best of our knowledge, for the first time in this study, protein interactions between SARS-COV-2 and humans were predicted and validated with computational methods.
- It has been observed that computational methods can be as successful as experimental methods.
- The proposed mapping method is aimed to be used in other protein studies by contributing to the literature.
- It has been observed that an algorithm-based mapping method is at least as successful as the methods of other categories.

The rest of the paper is organized as follows: Sect. 2 provides the related works based on protein interactions in the literature. Section 3 elicits the data used in this study. In addition, the proposed numerical method is explained in detail. In addition, in this section, information about other protein-mapping methods is given. Section 4 shows the interaction results of SARS-COV-2 and human protein pairs with both the proposed method and other mapping methods. Furthermore, the performance of each protein-mapping method is compared in this section. In the last section of the study, the importance and usage areas of the proposed method are discussed.

2 Related Works

In this section, studies conducted to determine the interactions between proteins are examined. In study [8], researches aimed to find potential drug targets by identifying protein interactions for COVID-19. A numerical method was not used in the study, and all results were obtained with the experimental methods. The interactions between SARS-COV-2 and human protein pairs and potential drug targets were designated by cloning, labeling, and affinity purification mass spectrometry. According to the results of the study, it was determined that 26 COVID-19 proteins interact with 332 human proteins in total.

Machine-learning algorithms are used effectively in protein interaction studies. In study [20], authors used radial-based functional neural networks to determine the protein–protein interaction sites. Protein sequences were converted to the numerical representations at the first stage, and frequency values of each amino acid were calculated. Then, by calculating the relative entropy, a total of 1000 features were obtained. Classification performance was determined by f1-score, and accuracy values, and these values were calculated as 99%, and 80%, respectively. In study [21], the interactions between proteins were specified using protein signatures. Helicobacter pylori protein data from human and mice were used in the study. Protein sequences were mapped by the protein signature method and classified using SVM (Support Vector Machine). The proposed method has been tested on three different species: *Helicobacter pylori*, *Escherichia coli*, and *Saccharomyces cerevisiae*. The performance of the method was measured with accuracy, specificity, and sensitivity.

In some cases, where protein sequences are numerous, machine-learning algorithms are not effective and key features cannot be obtained [16]. For this reason, besides machine learning, deep-learning models are also applied in this field. In study [22], the interactions between proteins were determined using CNN (Convolutional Neural Network), and LSTM (Long-Short Term Memory) models by applying primary protein sequences. Motifs, semantic and long-short term relationships between proteins were specified and features were collected. Then fivefold cross-validation was performed and average accuracy was achieved as 98.78%. In study [16], it was aimed to designate the interactions between protein pairs using LSTM deep-learning model. In the study, protein sequences were mapped with protein signature and Prot2Vec (Protein2Vector) method. These converted genomic data were later classified by the LSTM deep-learning model and the performance of the two methods was compared. ROC (Receiver Characteristic Curve), log loss, and accuracy metrics were used, and the Prot2Vec method was the most effective of these two numerical mapping methods. In study [20], protein interactions were specified using SNN (Siamese Neural Network) and the performance of the network was tested on four different datasets. Protein sequences were converted to the numbers with both protein signatures and Prot2Vec methods. The success of the SNN was measured by AUC scores and the average value was observed as 83.25%.

3 Data and the Proposed Method

3.1 The Protein Data

The genomic structure of the SARS-COV-2 virus, causing COVID-19 has been investigated and the proteins of the virus have been identified in the literature. The virus consists of four structural: S (surface), M (membrane), E (envelope), N (nuclecapsid), and six non-structural (orf3a, orf3b, orf6, orf7a, orf7b, and orf8) genes [23]. The structure of the genome of the virus is given in Fig. 1.

In this study, non-structural proteins were used and the interaction information between COVID-19, and human protein pairs were obtained from BioGRID dataset. The reason for using non-structural proteins in the study is that these proteins are thought to be necessary for the replication of viral genomes [24]. Similarly, non-structural proteins important for viral RNA synthesis and for antagonizing host antiviral immunity [25]. Therefore, predicting or determining

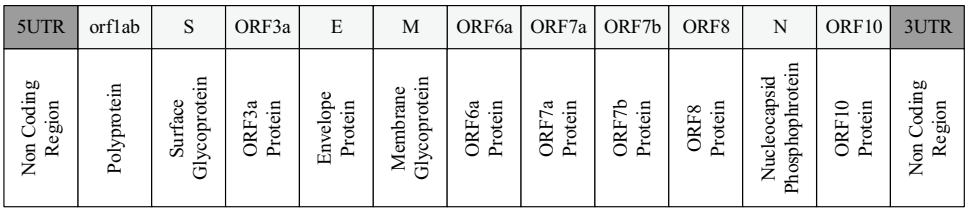


Fig. 1 Genomic structure of SARS-COV-2 virus

the interaction network of non-structural proteins is key to understanding protein interactions. Table 1 shows the interacting COVID-19 and human proteins and total protein numbers.

Since no interaction of the orf7b protein was observed [8], the orf7b protein was not considered in this study. The protein sequences were obtained from the NCBI dataset. A total of 3201 protein sequences were considered in the study. The same protein sequences were not used multiple times, but due to the small number of data, similar proteins were used in the study with the BLAST (Basic Local Alignment Search Tool) algorithm. The main reason for the lack of data is that COVID-19 is a new disease. This reveals the main limitation of our study. We tried to overcome this problem using the BLAST algorithm. Using the BLAST algorithm, we included protein sequences with a similarity rate of 90% and above.

3.2 Protein Mapping Modules

In this study, as a novel protein numerical mapping method is proposed, the performance of the proposed method is compared with various numerical mapping models. For this, we used EIIP (Electron–Ion Interaction Potential), CPNR (Complex Prime Number Representation) and hydrophobicity methods in this study. The EIIP method was first proposed to determine protein–DNA interactions [26]. With this method, genomic sequences were first converted into signals. Then, the signals were converted again and the power spectrum values of these signals were obtained. Fourier transform method was used for all these transformation processes. Finally, the power spectrum values obtained were assigned to each amino acid and the amino acids were mapped. EIIP method is one of the most frequently used methods in the literature [27, 28]. The CPNR method was first proposed in the comparison of protein functions [29]. In this method, amino acid codes were divided to codon number and each amino acid code was assigned to a specific prime number. The main purpose in the development of the method is to eliminate the degeneration problem that occurs in the EIIP method. There are a few studies in the literature performed with the CPNR method [30, 31]. The hydrophobicity method was proposed based on the hydrophilic and hydrophobic tendencies of the polypeptide chains of proteins [32]. It is generally used in the prediction of protein interactions and classification of protein functions [33, 34].

The biggest difference of the method we propose from these methods is the mapping process. In summary, it differs categorically. Since the mapping process in the EIIP method is performed depending on the signal, this method is a signal-based mapping method [35]. In the CPNR method, the mapping process is not based on a specific protein information (structure, function, etc.). Therefore, the mapping process of this method can be expressed as character based. Finally, the hydrophobicity method is a physicochemical-based method. In summary, mapping is performed based on the chemical information of the proteins [33]. The method we have proposed is an algorithm-based method. As far as we know, there are no numerical mapping methods in this category in the literature. In our method, as in the CPNR

Table 1 Interacting COVID-19 and human proteins and the total number of these proteins

COVID-19 proteins	Interacting human proteins	Total number of proteins
orf3a	ALG5, ARL6IP6, CLCC1, HMOX1, SUN2, TRIM59, VPS11, VPS39	8
orf3b	STOML2	1
orf6	MTCH1, NUO98, RAE1	3
orf7a	HEATR3, MDN1	2
orf8	ADAM9, ADAMTS1, CHPF2, CHPF, CISD3, COL6A1, DNMT1, EDEM3, EMC1, ERLEC1, ERO1LB, ERP44, FBXL12, FKBP7, FKBP10, FOXRED2, GDF15, GGH, HS6ST2, HYOU1, IL17RA, INHBE, ITGB1, KDELC1, KDELC2, LOX, MFGE8, NEU1, NGLY1, NPC2, NPTX1, OS9, PCSK6, PLAT, PLD3, PLEKHF2, PLOD2, POFUT1, PUSL1, PVR, SDF2, SIL1, SMOC1, STC2, TM2D3, TOR1A, UGGT2	47

method, there is information acquisition depending on the amino acid character (codes). In other words, the mapping process is not based on a specific protein information. However, unlike CPNR, the mapped process is based on a specific algorithmic structure.

3.3 A Novel Protein Mapping Method Based on AVL Tree

In computer science, algorithms are highly used to build a program. Algorithm can be described as the set of instructions to be followed to solve a problem. Once an algorithm is defined and modeled for a given problem, it is required to determine the how much time or space the algorithm will need. An algorithm that solves the problem in years or that requires thousands of gigabytes of main memory is not useful. Thus, it is essential to model a smooth algorithm for time and space efficient program. The performance of algorithms is determined with the time and space complexity analysis [36]. Analysis depends on many environmental and internal conditions such as operating system, processor, and hardware. Yet these are not considered while analyzing and only the execution time of an algorithm is calculated. It is well known that when the input size increases the best performance is obtained from the constant $O(c)$ time [36, 37]. This is followed by $O(logN)$, $O(N)$, $O(NlogN)$, $O(N^c)$, and $O\left(c^N\right)$, respectively. For large amount of data like in bioinformatics, it is difficult to obtain a linear complexity. Therefore, a tree structure is needed since average running time of a tree is $O(logN)$ in many cases [36, 38].

There are many tree structures, including binary tree, AVL tree, Red–Black tree, and N-ary tree in computer science. These trees are important in terms of data structures and algorithm analysis. Today, tree structures are used in most of the computer-based applications. Tree structures are generally preferred for storing data with a hierarchical structure. In addition, trees are frequently used in search operations. Each tree has its own advantages and disadvantages. In this study, we developed a protein numerical mapping method based on the AVL tree within existing tree structures. We explained the advantages and why we use the AVL tree at the end of this section.

AVL tree is a kind of binary tree with a balance condition. A tree in which no node can have more than two children called a binary tree and it is useful in algorithm analysis applications. The balance of an AVL tree is determined the height of nodes. Basically, the left and right subtrees need to be at the same height. In detail, the differences between heights of right and left subtrees for each node must be less than or equal to 1.

In nature, there are 20 amino acids. We add these amino acid codes to AVL tree in alphabetical order. For instance, first we add Alanine (A) in the AVL tree since it is the first amino acid in alphabetical order and we insert Arginine (R), Asparagine (N), and so on. The final status of the proposed method AVL tree can be seen in Fig. 2.

According to Fig. 2, root node is the Asparagine (N) node. When we add nodes to the trees, the balance of the tree is corrupted. Yet we provide the balance with AVL tree insertion rules. The detailed information about insertion and deletion rules of AVL can be seen in [36]. Histidine (H) and Serine (S) are the siblings and children of root node N. Glutamic acid (E)–Lysine (K), and Glutamine (Q)–Tryptophan (W) amino acid pairs come later which are left subtree and right subtree of H and S nodes, respectively. Cysteine (C) is the left child of node E and have two children which are Alanine (A), and Aspartic acid (D). The right child of E is Glycine (G) which has only one child called Phenylalanine (F). Isoleucine (I), and Leucine (L) are the siblings and children of K node. Methionine (M) is a grandchild of K node and is found right subtree of the L. Both Q and W nodes have two children including Proline (P), Arginine (R), Threonine (T), and Tyrosine (Y). Only the T node has child Valine (V). After the insertion process, we calculate the depth values of each node. Figure 3 depicts the depth values of each node visually.

We can see that only the root node has a 0 depth value. All other nodes have different depth values than root value. After calculation of the depth values, we convert amino acid sequences to the numerical representations. Let say

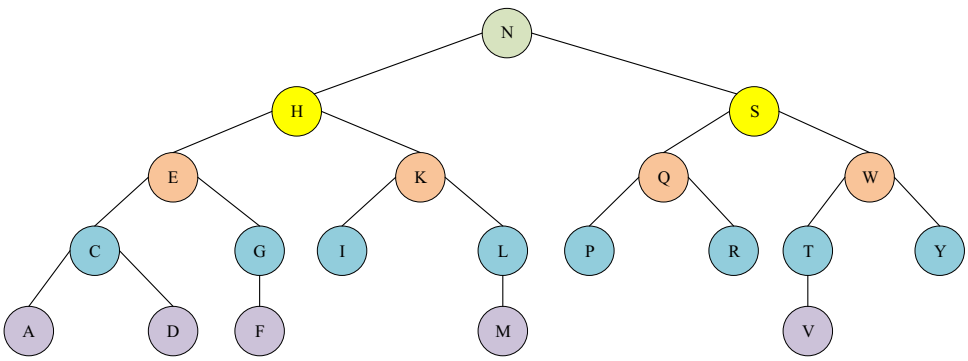


Fig. 2 Insertion of amino acid codes to the AVL tree

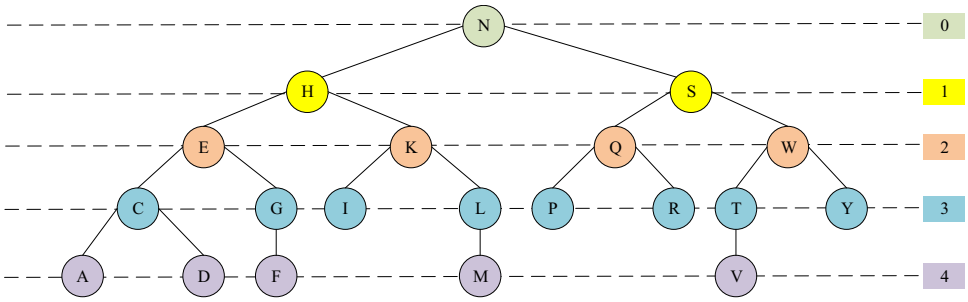


Fig. 3 Depth values of each amino acid

we have a protein sequence $S(n) = MHIILKFDASHN$. The numerical representation of this sequence is calculated $S(n) = 413332444110$.

The performance of the proposed method is determined by DeepBiRNN deep-learning model by comparing the interaction results with other protein-mapping methods. This mapping method demonstrated high-performance for finding protein interactions between COVID-19 and human. We select an AVL tree since the execution time is good and the time complexity is $O(logN)$ in most of the cases. It is well known fact that there are some other trees running in $O(logN)$ time such as binary tree and splay tree. In binary tree, there is no balance condition. In splay tree, a balance exists yet, insertion and deletion operations take $O(NlogN)$ time which requires more execution time than the AVL tree [39]. In addition, AVL tree is logical, however, splay tree is heuristic [40]. AVL trees are often compared consistently with Red–Black trees. However, the biggest advantage of the AVL tree over Red–Black trees or even other trees is that it has a fast search process [37]. As we have stated before, the biggest advantage of this tree is its balance condition. An unbalanced tree means that the operations take longer which causes in time intensive lookup applications. For these reasons, AVL tree was chosen as the tree structure. The experimental design of the study is given in Fig. 4.

In the first stage that can be seen from Fig. 4, proteins belonging to human were obtained from NCBI dataset. Proteins belonging to the virus were not used in the study. Since it is known which protein of the virus interacts with which human protein. For example, the orf3b protein interacts only with STOML2. While STOML2 was considered as a positive interaction, other proteins were considered as a negative interaction. This approach has also been used for other protein interactions. Subsequently, proteins belonging to human were mapped using both the proposed numerical mapping method and other mapping methods. All mapping operations were done for each sequence one by one. The mapped protein sequences were then normalized. The z-score was used for normalization. Thus, the mean value was subtracted from each value in the dataset and divided by the standard deviation of the whole dataset. After the normalization process, interactions were classified with the DeepBiRNN deep-learning model. Prediction studies based on protein interactions are generally based on two scenarios: interaction either exists or does not exist. In other words, binary classification is made. Since it is known which proteins belonging to the virus interact with which proteins belonging to humans, the output of the interacting proteins is determined as 1 and the others as 0. After performing binary classification with DeepBiRNN, the performance of each mapping method was determined by accuracy, precision, recall, f1-score and AUC scores.

4 Application Results and Discussion

Numerous methods have been proposed for the diagnosis of COVID-19 disease so far and many studies have been carried out in the literature. When the studies were examined in detail, it was observed that the majority were made based on X-ray images or CT images. Deep learning was frequently used in these studies, and it came to the fore as an effective method. These references can be cited as examples to studies conducted with deep learning [3, 41, 42]. The successes achieved in these studies and the popularity of deep learning have led us to use the deep-learning method for COVID-19 disease. However, in this study, rather than the diagnosis of the disease, the internal structure of the virus was considered and the proteins of the virus were analyzed. For this, a protein–protein interaction study was carried out and a novel numerical mapping method was proposed. To find out the effect of the proposed mapping method and other mapping methods on predicting the protein interactions, in this study we applied DeepBiRNN as a deep learning. BiRNNs connect two hidden layers of opposite directions to the same output. With this way, the output layer can get information from past (backward) and future (forward) states simultaneously. It is introduced the increase the amount of input information available to network. The main difference between BiRNN and RNN (Recurrent Neural Network) models is how information is obtained and stored. In RNN, the future input cannot be obtained from the current state [43–45]. On the other hand, BiRNN does not require their input data to be fixed. The detailed information about BiRNN can be seen

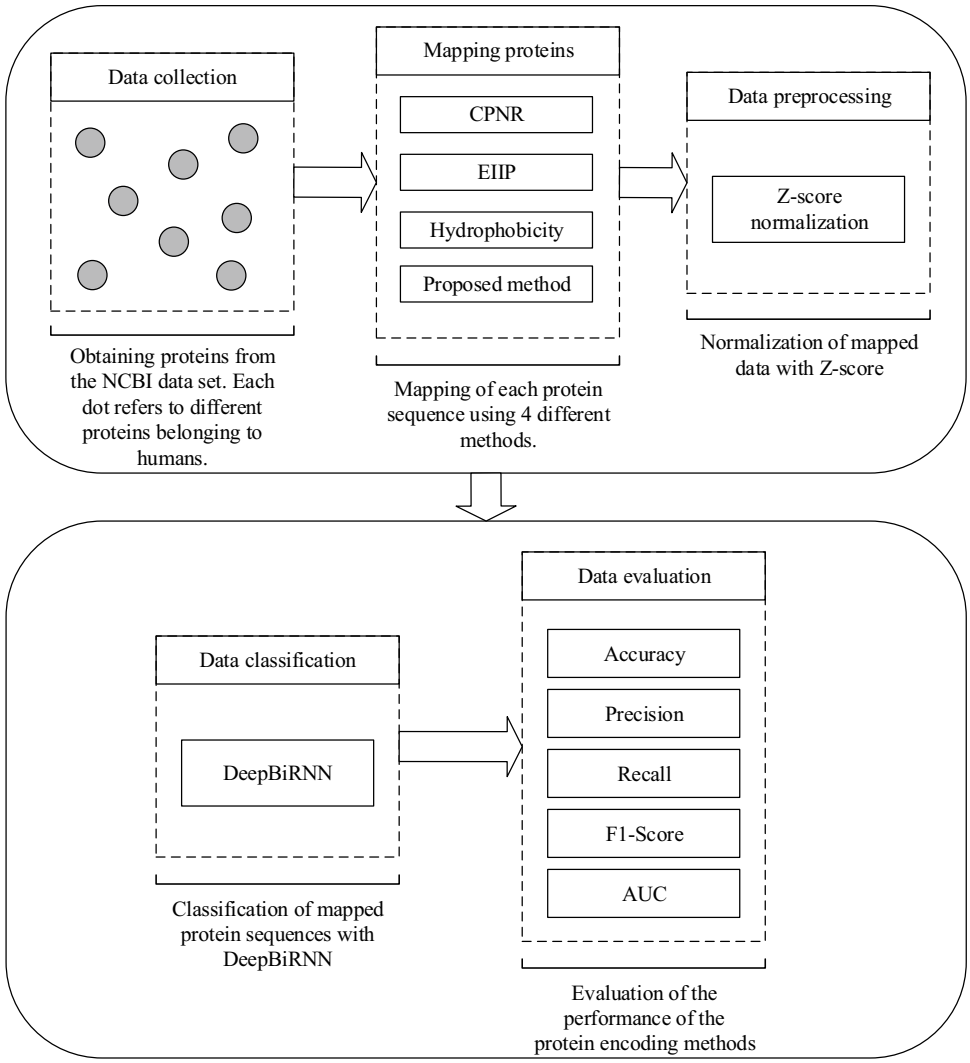
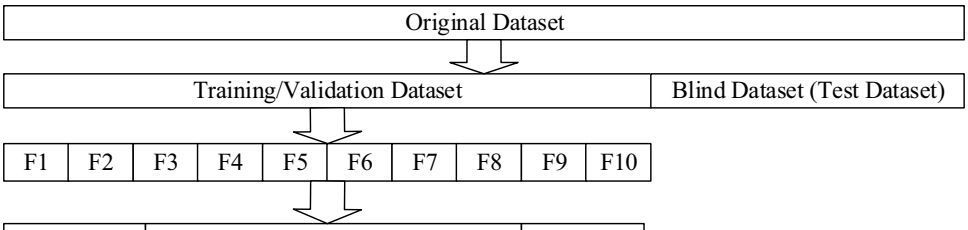


Fig. 4 Experimental design of the study

in [39, 42]. The parameters of designed DeepBiRNN model were determined by trial and error approach and can be summarized as follows:

- In the first layer, a total of 64 BiRNN units were used. Then output values were calculated with ReLU activation function.
- BiRNN was used again in the second layer. The number of units in this layer is determined as 32. ReLU was again used as the activation function.
- BiRNN was used again in the third layer. The number of units in this layer is determined as 16. ReLU was again used as the activation function.
- Then the Flatten function was applied to flatten the data in matrix format.
- Batch normalization was performed to prevent changes in data distribution.
- After flatten and batch normalization, dropout was performed and 25% of the data was forgotten.
- Later, fully connected layers were designed and 512 neurons were used in the first fully connected layer.
- Classification was carried out in the second fully connected layer, and the interactions were predicted with the sigmoid function.
- SGD (Stochastic Gradient Descent) was applied as the optimizer and the learning rate and momentum values were designed to be 0.0001 and 0.9, respectively.
- For model loss, binary crossentropy was used and the model was complied with a value of 500 epochs.

We split the original dataset for training, testing and validation to evaluate the performance of mapping methods. Using the testing dataset as a blind dataset, we aimed to determine the performance of mapping methods. Only 15% of the original dataset was considered as a blind dataset (test data). The remaining 85% (70% for training and 15% for validation) was used in training and validation with tenfold cross-validation process. After the training and validation,



Iteration #	Iteration #	Test Set
Iteration 1	F1...F9	F10
Iteration 2	F1...F8, F10	F9
Iteration 3	F1...F7, F9, F10	F8
Iteration 4	F1...F6, F8...F10	F7
Iteration 5	F1...F5, F7...F10	F6
Iteration 6	F1...F4, F6...F10	F5
Iteration 7	F1...F3, F5...F10	F4
Iteration 8	F1, F2, F4...F10	F3
Iteration 9	F1, F3...F10	F2
Iteration 10	F2...F10	F1

Fig. 5 Validation process of the study. While 85% of the original dataset consists of training and validation data, the remaining 15% is the blind dataset. After completing the iteration process, the performances of the mapping methods were determined with the blind dataset

Table 2 Average interaction prediction results of orf3a protein with other human proteins (all values were obtained by averaging the tenfold cross-validation process)

Mapping methods	Accuracy	Precision	Recall	f1-score	AUC score
CPNR	0.8398	0.7541	0.8911	0.8106	0.92
EIIP	0.8144	0.7200	0.8394	0.7737	0.91
Hydrophobicity	0.8019	0.7263	0.8488	0.7788	0.89
The proposed method	0.8107	0.8177	0.8169	0.8163	0.89

Table 3 Average interaction prediction results of orf3b protein with other human proteins (all values were obtained by averaging the tenfold cross-validation process)

Mapping methods	Accuracy	Precision	Recall	f1-score	AUC score
CPNR	0.8065	0.6429	0.6953	0.6652	0.91
EIIP	0.6796	0.4770	0.4918	0.4843	0.67
Hydrophobicity	0.7472	0.7926	0.6406	0.7064	0.79
The proposed method	0.9065	0.8429	0.8953	0.8652	0.93

Table 4 Average interaction prediction results of orf6 protein with other human proteins (all values were obtained by averaging the tenfold cross-validation process)

Mapping methods	Accuracy	Precision	Recall	F1-score	AUC score
CPNR	0.7790	0.6629	0.7678	0.7081	0.90
EIIP	0.4109	0.4680	0.4888	0.4781	0.43
Hydrophobicity	0.6921	0.7225	0.6539	0.6827	0.81
The proposed method	0.8523	0.7291	0.7231	0.7176	0.93

the performance of the mapping methods was determined on the blind dataset. Information on the dataset used in this study can be obtained from https://drive.google.com/drive/folders/1emQV3B8fRQNgxVWFPuyNKqSCxxDFot2z?usp=sharing. In Fig. 5, the diagrammatic scheme of the validation phase of the PPI (protein–protein interaction) used in the study is given.

Protein sequences were mapped with the proposed numerical mapping and other mapping methods. In the study, the interaction of non-structural COVID-19 proteins with other human proteins was predicted. Taking into account the data provided as a result of biochemical experiments in Table 1, how much these interactions are compatible with the proposed and other mapping methods is given in Tables 2, 3, 4, 5, 6.

As can be seen from Table 2, all mapping methods produced over 80% results. The best protein interaction of orf3a protein with eight other human proteins was predicted with

CPNR-mapping method with accuracy of 83.98%. The proposed mapping method was at least as effective as the CPNR- and EIIP-mapping methods. In the study performed by experimental methods, the interaction result of orf3a protein was obtained as 92% on average [8]. The interactions between orf3a and other human proteins were validated with the accuracy of 81.07% with the proposed method. ROC and PR (Positive Rate) plots of the protein-mapping methods are given in Fig. 6.

According to the study of [8], orf3b protein only interacted with the STOML2 human protein and the interaction score was determined 92.9%. The interaction prediction

Table 5 Average interaction prediction results of orf7a protein with other human proteins (all values were obtained by averaging the tenfold cross-validation process)

Mapping methods	Accuracy	Precision	Recall	f1-score	AUC score
CPNR	0.8457	0.8182	0.8121	0.8133	0.90
EIIP	0.3840	0.4052	0.4385	0.4198	0.40
Hydrophobicity	0.8145	0.7939	0.7402	0.7147	0.85
The proposed method	0.8891	0.8181	0.9310	0.8709	0.95

Table 6 Average interaction prediction results of orf8 protein with other human proteins (all values were obtained by averaging the tenfold cross-validation process)

Mapping methods	Accuracy	Precision	Recall	f1-score	AUC score
CPNR	0.6409	0.6203	0.5927	0.6015	0.75
EIIP	0.6085	0.5729	0.4329	0.4914	0.78
Hydrophobicity	0.8002	0.7377	0.6528	0.6926	0.91
The proposed method	0.8078	0.7567	0.6605	0.7011	0.93

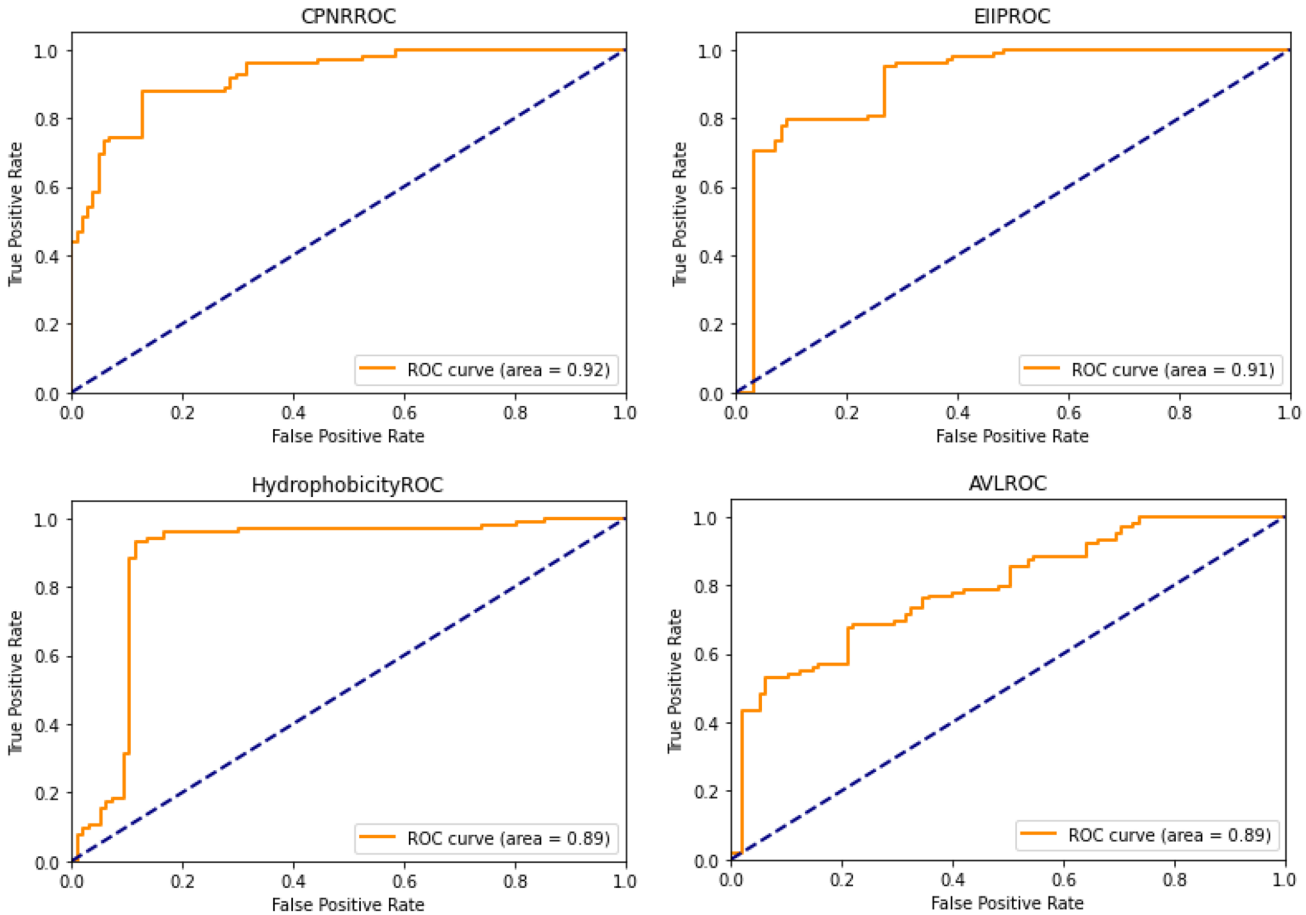


Fig. 6 ROC and PR plots of protein-mapping methods that predicted the interaction between orf3a protein and human proteins

results of orf3b protein with numerical methods are given in Table 3.

The best evaluation values were obtained with the proposed method. With the proposed method, an accuracy performance of 90.65% was achieved and this success was supported by the AUC score. The closest effective result to the proposed method was obtained with the CPNR-mapping method. The AUC scores were close to each other, although the accuracy was about 10% less from the proposed method. It was observed that the interaction values performed by numerical methods validated the experimental results. Only the accuracy of the interaction result with the EIIP-mapping method was lower than 70.00%. ROC and PR plots of the protein-mapping methods are given in Fig. 7.

According to the study of [8], it was determined that orf6 protein only interacted with three human proteins. The interaction prediction results of orf6 protein with other human proteins are given in Table 4.

As can be seen in Table 4, the proposed method has achieved the best evaluation results for all but recall. While determining the interactions of orf6 protein, 85.23% accuracy, and 0.93 AUC score were obtained with the proposed AVL mapping method. Unlike other mapping methods, the EIIP method has been very unsuccessful in predicting this protein. All of the evaluation criteria fell below 50%. As with other previous protein interaction prediction, CPNR provided the closest performance to the proposed method. ROC and PR plots of the protein-mapping methods are given in Fig. 8.

In study [8], the orf7a protein interacted only with MDN1, and HEATR3 human proteins with 88.55% interaction score. Table 5 provides the interaction prediction results of orf7a protein.

According to the results in Table 5, the proposed method was observed as the most successful method that predicts the best interaction between orf7a and other human proteins. The accuracy value of the proposed AVL method was determined as 88.91%. While the CPNR and hydrophobicity methods were at least as successful as the proposed method, the EIIP method was again ineffective. Except for EIIP, all methods produced an accuracy of more than 80% and an AUC score, while these values were below 50% with EIIP. Figure 9 shows the ROC and PR plots of the protein-mapping methods.

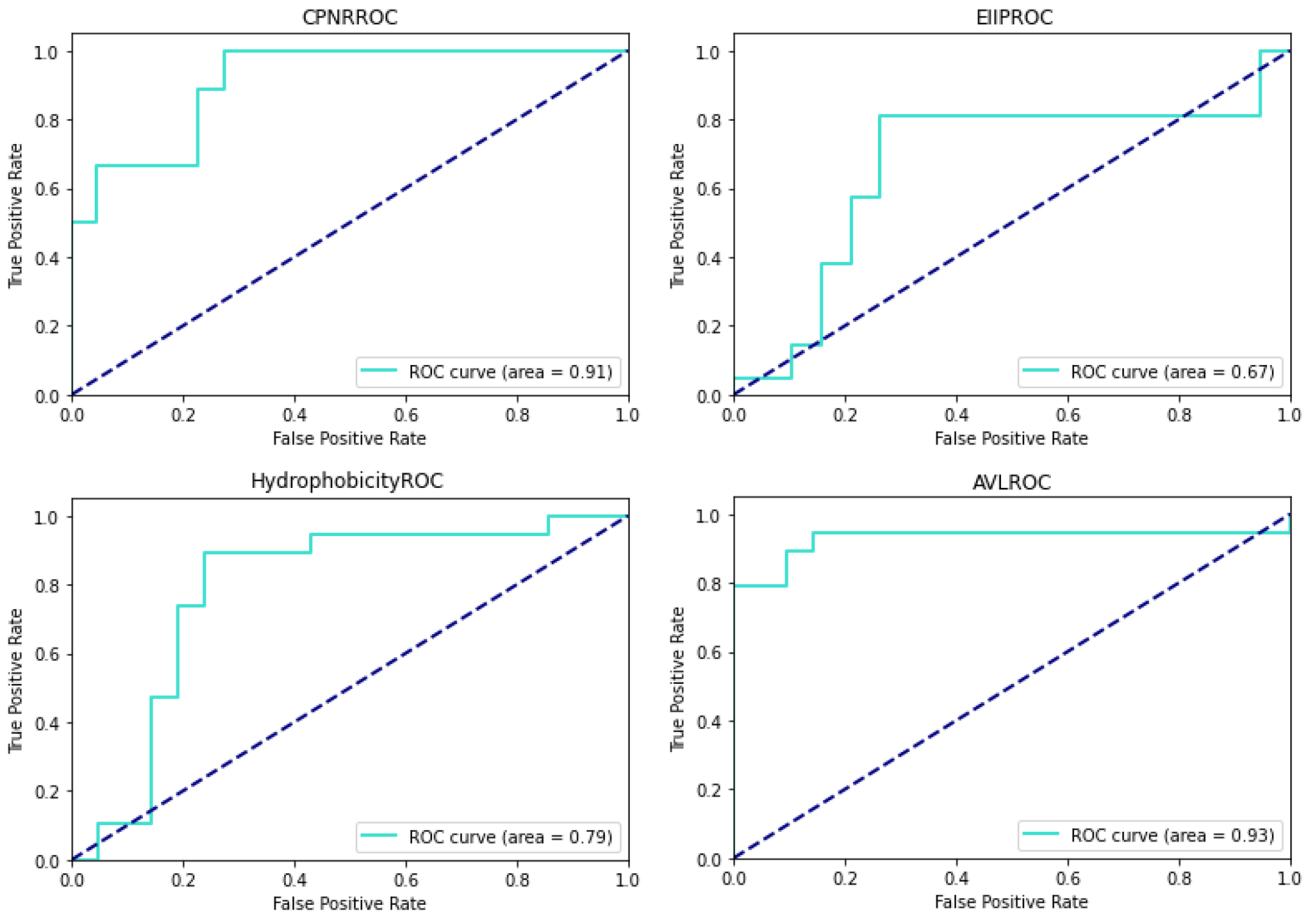


Fig. 7 ROC and PR plots of protein-mapping methods that predicted the interaction between orf3b protein and human proteins

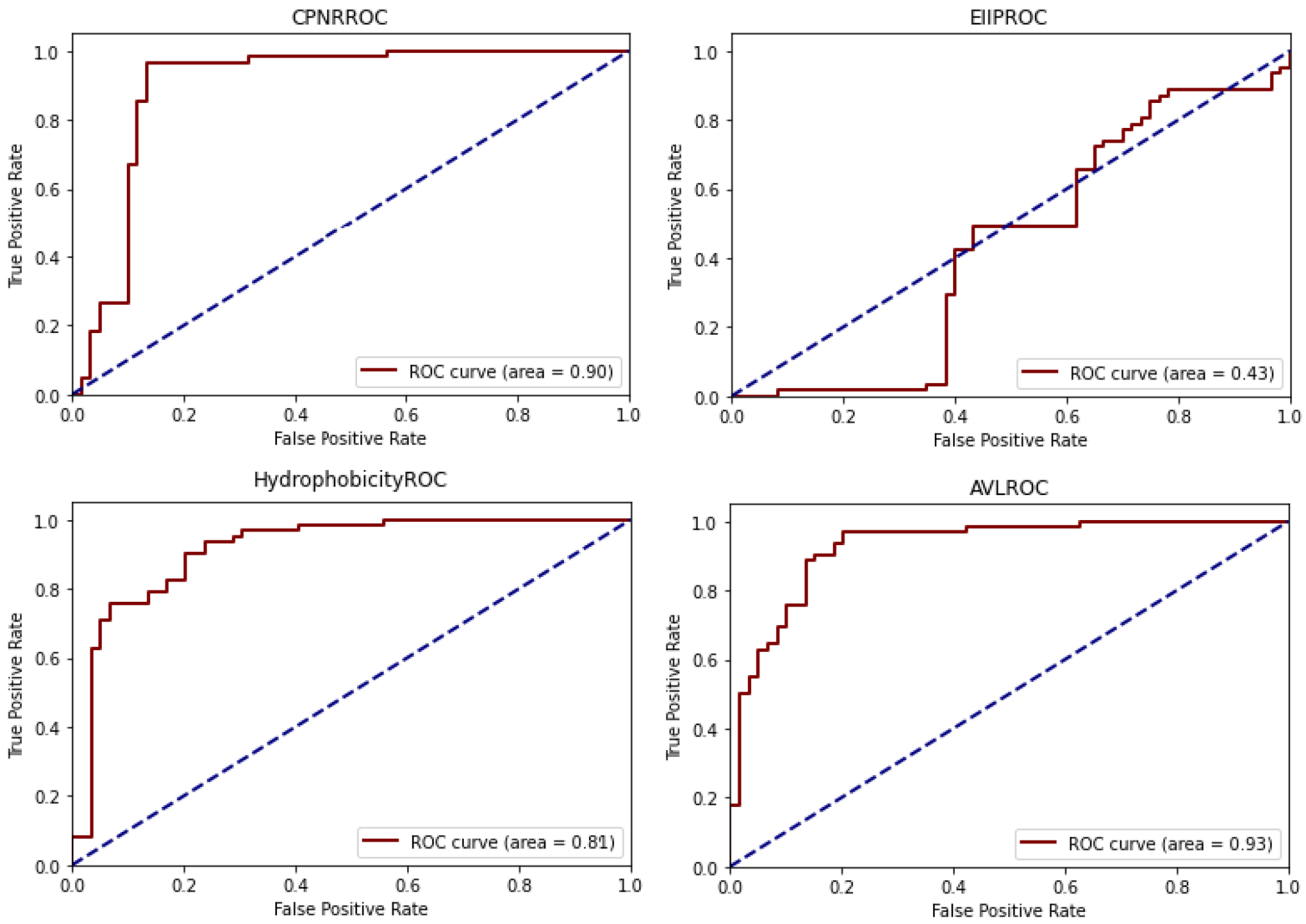


Fig. 8 ROC and PR plots of protein-mapping methods that predicted the interaction between orf6 protein and human proteins

The interaction network of orf3a, orf3bb, orf6, and orf7a proteins is given in Fig. 10.

Orf8 has the most protein interaction among the COVID-19 proteins. The interaction score of this protein, which interacts with 47 human proteins in total, has been determined as an average of 87.78% according to the study [8]. The evaluation results of protein-mapping methods to predict protein interactions of orf8 are shown in Table 6.

The best interaction network of orf8 protein was predicted by the proposed method with the accuracy of 80.78%, and AUC score of 0.93. When all evaluation criteria were examined, it was observed that the best result was again obtained with AVL. Unlike other interactions, the second best accuracy performance for this protein was achieved with hydrophobicity. Hydrophobicity produced an accuracy and AUC result close to the proposed method and was at least as effective as the proposed method. In Fig. 11, the ROC and PR plots of the protein-mapping methods are given. In addition, the interaction network of the orf8a protein is shown in Fig. 12. Table 7 shows the average prediction accuracy results of all protein-mapping methods.

According to the comparison results given in Table 7, the best interaction accuracies between proteins were obtained by the proposed AVL method in average. In addition, all numerical mapping methods validated the results of the experimental method used in study [8].

When the results are examined in general, it is seen that all mapping methods make a successful prediction. However, prediction accuracy varies according to the methods used. The EIIP method is generally effective and works well in determining proteins with the same function information [46]. Moreover, with the EIIP method, the same or very similar representations can be obtained from two proteins that are functionally different from each other which, causes degeneration to occur [47]. Since COVID-19 is a new disease, the functions of its proteins are not fully known. EIIP may have been the most unsuccessful method, as we do not have protein function information. However, despite this, it achieved considerable success. Since hydrophobicity is a method developed based on the physicochemical knowledge of proteins, it is effective in studies where physicochemical information is at the forefront [33]. Physicochemical

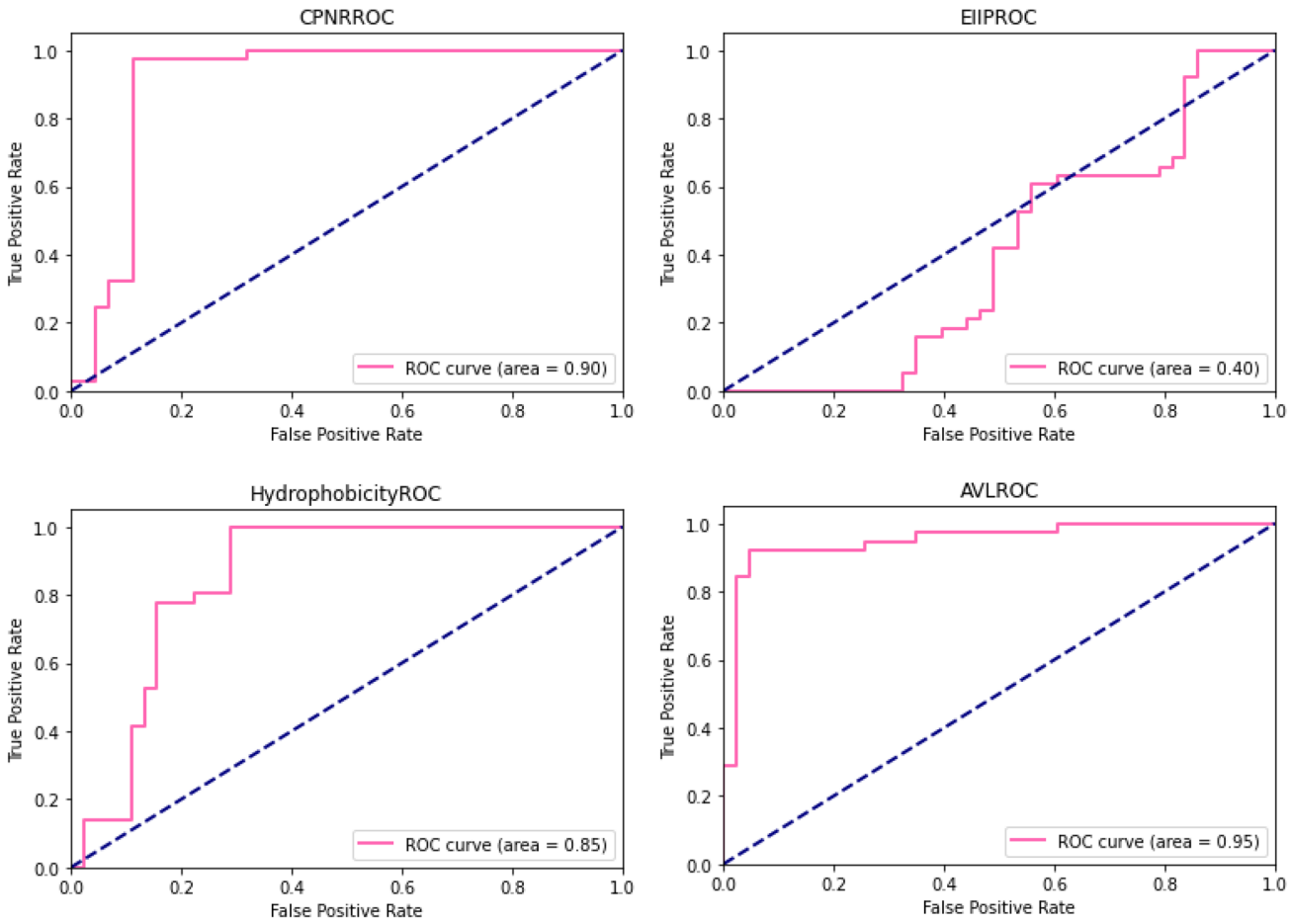


Fig. 9 ROC and PR plots of protein-mapping methods that predicted the interaction between orf7a protein and human proteins

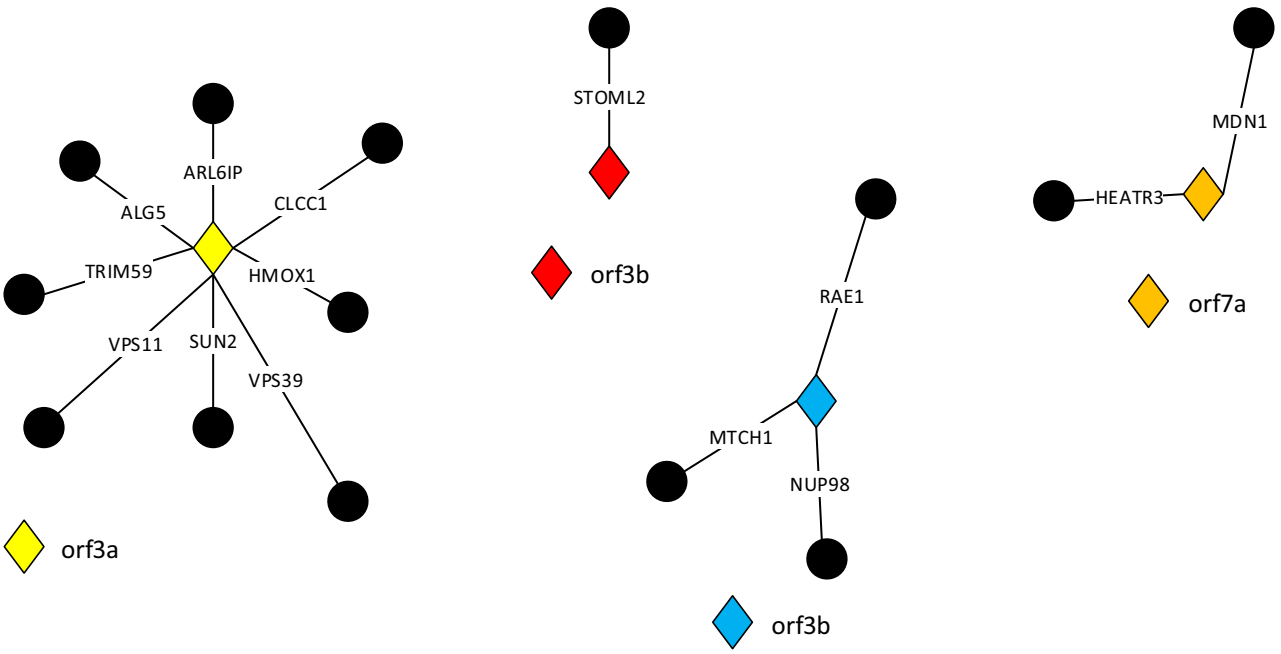


Fig. 10 The interaction network of orf3a, orf3bb, orf6, and orf7a proteins

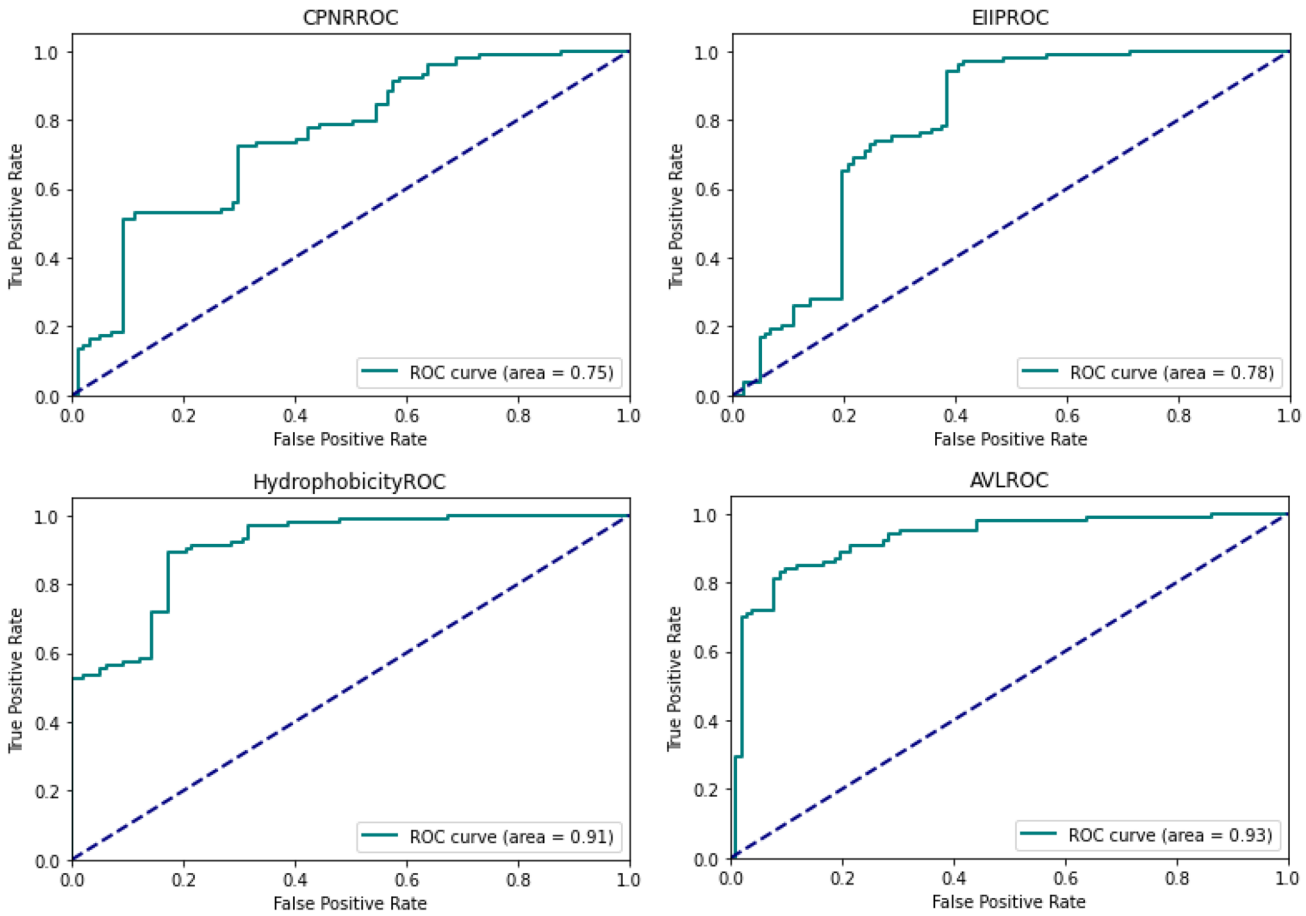


Fig. 11 ROC and PR plots of protein-mapping methods that predicted the interaction between orf8 protein and human proteins

knowledge of amino acids plays an important role in protein-folding studies. Yet, since the protein folding-based physicochemical properties are unknown, providing an effective physicochemical properties encoding approach is still an open problem. This method may have been less effective than other methods, as non-structural genes are used rather than the physicochemical information of the virus that causes COVID-19. This method can produce more effective results in physicochemical-based studies (classification of protein families, coevolution analysis, etc.). The second most effective method in the study was CPNR. This may be because the CPNR method is fault-tolerant to point mutations [48]. Another reason for this success may be that the protein sequences in the CPNR method are mapped without relying on a specific structure, protein and physicochemical information, as in the proposed method. Unlike the EIIP, CPNR, and hydrophobicity methods, it is no surprise that the proposed method works best. One of the biggest reasons for this is that the mapping process in the proposed method is not based on a specific structure, chemical or function information. The fact that the CPNR method also maps with a similar approach and gives the second best result supports this success. The reason why the proposed method succeeds the CPNR method may be that it uses an algorithm-based approach rather than a character-based approach. By proposing an algorithm-based method, we aimed to map amino acids according to the tree structure frequently used in computer science, rather than gene expression. For this purpose, we tried to propose an effective method that can be used without the need for any specific information about proteins. When looking at the results, there is a 1% difference between the proposed method and the CPNR in average. This shows that our method is at least as effective as CPNR.

The advantages of the study can be listed as follows:

- Although the number of data is small, both the proposed numerical mapping method and other mapping methods performed a successful prediction process. The number of data is important in studies conducted with deep learning. However, success has also been achieved with a small dataset.
- By predicting protein interactions of the SARS-COV-2 virus, drug studies may gain momentum. Protein inter-

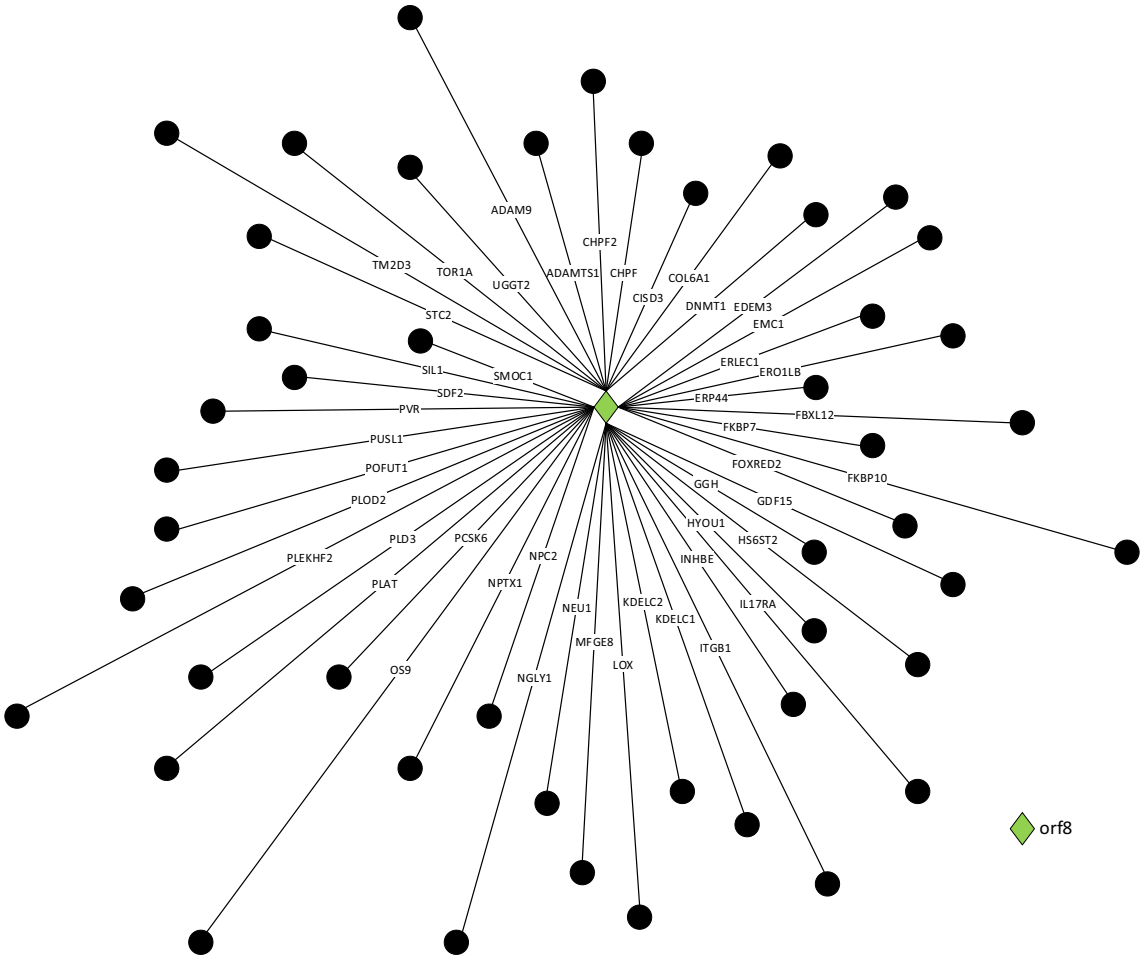


Fig. 12 The interaction network of orf8 protein

Table 7 Average protein interaction accuracy results of all protein-mapping methods

Mapping method	Interaction accuracy
CPNR	78.24%
EIIP	57.95%
Hydrophobicity	77.12%
The proposed method	85.33%

actions are important in determining potential drugs or using existing drugs [8].

- Similarly, by predicting interactions, detailed information about host cells can be obtained.
- In addition, by predicting interactions, information about the functions of proteins belonging to SARS-COV-2 can be collected.

The disadvantages of the study can be listed as follows:

- At the time of the study, protein interactions were determined for five different non-structural genes. In future studies, interactions between proteins belonging to the SARS-COV-2 virus and human proteins may increase or decrease. Accordingly, the results obtained in this study may differ.
- In some cases, machine-learning algorithms can produce better results than deep-learning algorithms. Deep learning algorithms are mostly preferred to avoid feature extraction. However, combining feature selection with clustering and classification can produce better results. Clustering and machine learning were not used in this study. These approaches should also be carried out. Some text-based clustering studies can produce effective results [49, 50].
- In this study, optimization process was not performed. Maybe the results obtained can be improved with certain optimization algorithms. In this context, it is necessary to examine and evaluate certain optimization algorithms. Optimization algorithms are used in forecasting studies

of COVID-19 disease [51]. It is necessary to determine whether these methods used in the process of forecasting COVID-19 disease will have an effect in this area.

- To predict these methods with artificial intelligence, certain prior knowledge is required. These preliminary information is also obtained by experimental methods. Currently, the inability to determine protein interactions without prior knowledge is one of the biggest problems in this field.
- In deep-learning studies, the number of data is generally expected to be large. However, the number of data may not be sufficient because COVID-19 is a new disease. We recommend that researchers consider this situation.

5 Conclusion

In this study, a novel protein-mapping method was proposed to predict the interactions of non-structural proteins belonging to COVID-19. Orf3a, orf3b, orf6, orf7a, and orf8 proteins were considered in the study, and their interactions with other human proteins were designated. A bidirectional recurrent neural network deep-learning model was used to identify the interactions. In the first part of the study, proteins were collected from the NCBI dataset and protein sequences were mapped using both the proposed AVL method and three different mapping methods. Then, mapped protein sequences were normalized, and classified with the developed DeepBiRNN model. The performance of the numerical mapping methods was measured by accuracy, precision, recall, f1-score, and AUC scores. Orf3a protein interacted with eight different human proteins and these interactions were validated with the accuracy of 81.07% using the proposed AVL method. Orf3b protein only interacted with the STOML2 human protein and this interaction was validated with the accuracy of 90.65% by the proposed method. Orf6 protein interacted with three different proteins in total, generating the protein interaction network. The best evaluation results were determined by the proposed method with the accuracy of 85.23%, precision of 72.91%, recall of 72.31%, f1-score of 71.76%, and AUC score of 0.93. It was validated that orf7a protein interacted with two proteins. The proposed method achieved the second best performance. Orf8 protein interacted with a total of 47 human proteins and formed a complex protein interaction network. With the proposed method, the best accuracy result was obtained. With the proposed method, the best interaction accuracy was obtained with an average of 85.33% accuracy. At the end of the study, it was observed that the proposed method is at least as effective as other methods and even more successful in some cases. One of the main reasons for this success may be that the amino acid codes in the protein sequence are dynamically placed on the tree and expressed in an algorithmic manner. Using the dictionary structure of the algorithm and considering the balance structure of the AVL tree caused the method to be more stable and robust. In this way, we have shown that an algorithm-based mapping method is as successful as the methods of other categories. As can be understood from the experimental results, protein interactions were successfully predicted without performing feature extraction with the proposed scheme. We also found that our mapping method can be an effective tool to accurately predict potential protein–protein interactions. In general, these results show the feasibility and superiority of the proposed algorithm-based mapping method in the PPI study. Different algorithm-based methods may be proposed in the future, and comparing these methods with methods belonging to other categories will reveal the performance of algorithm-based approaches in more detail. In addition, using more deep-learning strategies or performing protein prediction in different areas can provide more robust information about the performance of the proposed mapping method. For this reason, the working area of the proposed method can be extended to other areas and can be used effectively in the following areas;

- In determining the drug-target interactions,
- In drug therapy and drug development studies,
- In identification and classification of protein families,
- In phylogenetic analysis studies,
- In determining viral–host protein interactions,
- In predicting cancer–protein interactions,
- In identification of protein functions,
- In prediction of protein structure networks.

In the future studies, the proposed method will be used and tested on mentioned different protein studies.

Acknowledgements We would like to thank Dr. Askin Sen, assistant professor at Firat University, Faculty of Medicine, Department of Medical Genetics, for guiding us about the concept and terminology of genetics.

Compliance with Ethical Standards

Conflict of interest There is no conflict of interest in this study.

References

1. Fan W, Zhao S, Yu B, Chen Y, Wang W, Song Z, Hu Y et al (2020) A new coronavirus associated with human respiratory disease in China. Nature 579:265–269. https://doi.org/10.1038/s41586-020-2008-3
2. Sahin AR, Erdogan A, Agaoglu PM, Dineri Y, Cakırcı AY, Senel ME, Okyay RA, Tasdogan AM (2020) 2019 Novel coronavirus (COVID-19) outbreak: a review of the current literature. Eurasian J Med Oncol 4(1):1–7. https://doi.org/10.14744/ejmo.2020.12220

3. Ozturk T, Talo M, Yildirim EA, Baloglu UB, Yildirim O, Acharya UR (2020) Automated detection of COVID-19 cases using deep neural networks with X-ray images. Comput Biol Med 121:103792. https://doi.org/10.1016/j.compbiomed.2020.103792
4. Wit E, Doremalen N, Falzarano D, Munster VJ (2016) SARS and MERS: Recent insights into emerging coronaviruses. Nat Rev Microbiol 14:523–534. https://doi.org/10.1038/nrmicro.2016.81
5. Gates B (2020) "Responding to COVID-19: A once in a century pandemic? N Engl J Med 382:1677–1679. https://doi.org/10.1056/NEJMp2003762
6. Anderson RM, Heesterbeek H, Klinkenberg D, Hollingsworth TD (2020) How will country-based mitigation measures influence the course of the COVID-19 epidemic? Lancet 395(10228):931–934. https://doi.org/10.1016/S0140-6736(20)30567-5
7. World Health Organization (WHO), https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen
8. Gordon DE, Jang GM, Bouhaddou JM, Xu J, Obernier K, White KM, O'Meara MJ et al (2020) A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. Nature. https://doi.org/10.1038/s41586-020-2286-9
9. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC (2011) The clinicaltrials.gov results database: update and key issues. N Engl J Med 364:852–860. https://doi.org/10.1056/NEJMsa1012065
10. Sheahan TP, Sims AC, Leist SR, Schafer A et al (2020) Comparative therapeutic efficacy pf remdesivir and combination lopinavir, ritonavir, and interaction beta against MERS-CoV. Nat Commun 11:222. https://doi.org/10.1038/s41467-019-13940-6
11. Goncearenco A, Li M, Simonetti FL, Shoemaker BA, Panchenko AR (2017) Exploring protein-protein interactions as drug targets for anti-cancer therapy with in silico workflows. Methods Mol Biol, p 1647. https://doi.org/10.1007/978-1-4939-7201-2_15
12. Chene P (2006) Drugs targeting protein-protein interactions. Chem Med Chem 1(4):400–411. https://doi.org/10.1002/cmdc.200600004
13. Rao VS, Srinivas K, Sujini GN, Kumar GNS (2014) Protein-protein interaction detection: methods and analysis. Int J Proteom 2014:147648. https://doi.org/10.1155/2014/147648
14. Ruffalo M, Bar-Joseph Z (2019) Protein interaction disruption in cancer. BMC Cancer 19. https://doi.org/10.1186/s12885-019-5532-5
15. Jothi R, Kann MG, Przytycka TM (2005) Predicting protein-protein interaction by searching evolutionary tree automorphism space. Bioinformatics 21:241–250. https://doi.org/10.1093/bioinformatics/bti1009
16. Alakus TB, Turkgolu I (2019) Prediction of protein-protein interactions with LSTM deep learning modes, Proceedings in 3rd International Symposium on Multidisciplinary Studies and Innovative Technologies –ISMSIT, Ankara, Turkey, 2019. https://doi.org/10.1109/ISMSIT.2019.8932876
17. Wang L, Wang H, Liu SR, Song KJ (2019) Predicting protein-protein interactions from matrix based protein sequence using convolutional neural network and feature-selective rotation forest. Sci Rep 9.https://doi.org/10.1038/s41598-019-46369-4
18. Chen KH, Wang TF, Hu YJ (2019) Protein-protein interaction prediction using a hybrid feature representation and a stacked generalization scheme. BMC Bioinform 20(1):2019. https://doi.org/10.1186/s12859-019-2907-1
19. Sarkar D, Saha S (2019) Machine-learning techniques for the prediction of protein-protein interactions. J Biosci 44(104). https://doi.org/10.1007/s12038-019-9909-z
20. Chen Y, Xu J, Yang B, Zhao Y, He W (2012) A novel method for prediction of protein interaction sites based on integrated RBF neural networks. Comput Biol Med 42(4):402–407. https://doi.org/10.1016/j.compbiomed.2011.12.007
21. Martin S, Roe D, Faulon J (2005) Predicting protein-protein interaction using signature products. Bioinformatics 21(2):218–226. https://doi.org/10.1093/bioinformatics/bth483
22. Li H, Gong X, Yu H, Zhou C (1923) Deep neural network based predictions of protein interactions using primary sequences. Molecules 23(8). https://doi.org/10.3390/molecules23081923
23. Khailany RA, Safdar M, Ozaslan M (2020) Genomic characterization of a novel SARS-CoV-2. Gene Rep 19:100682. https://doi.org/10.1016/j.genrep.2020.100682
24. Dimitrova M, Imbert I, Kieny MP, Schuster C (2003) Protein-protein interactions between Hepatitis C virus nonstructural proteins. J Virol 77(9):5401–5414. https://doi.org/10.1128/JVI.77.9.5401-5414.2003
25. Song J, Liu Y, Gao P, Hu Y, Chai Y et al (2018) Mapping the nonstrcutural protein interaction network of porcine reproductive and respiratory syndrome virus. J Virol 92(24):112–118. https://doi.org/10.1128/JVI.01112-18
26. Veljkovic N, Glisic S, Prljic J, Perovic V, Botta M, Veljkovic V (2008) Discovery of new therapeutic targets by the informational spectrum method. Curr Protein Pept Sci 9(5):493–506. https://doi.org/10.2174/138920308785915245
27. Sencanski M, Sumonja N, Perovic V, Glisic S, Veljkovic N, Veljkovic V (2019) Application of information spectrum method on small molecules and target recognition. arXiv, 1907.02713., 2019.
28. Kasparek J, Maderankova D, Tkacz E (2014) Protein hotspot prediction using S-transform. Inf Technol Biomed 3:327–336. https://doi.org/10.1007/978-3-319-06593-9_29
29. Chen D, Wang J, Yan M, Bao FS (2016) A complex prime numerical representation of amino acids for protein function comparison. J Comput Biol 23(8):669–677. https://doi.org/10.1089/cmb.2015.0178
30. Mary GA, Babu GA, Rao GAR (2018) Identification of hotspots in protein sequences using CPNR and DWT. Int J Adv Res Comput Sci 9(3):219–223. https://doi.org/10.26483/ijarcs.v9i3.6108
31. Alakus TB, Turkoglu I (2020) A novel Fibonacci hash method for protein family identification by using recurrent neural networks, Turkish J Electr Eng Comput Sci, Accepted article, 2020. Doi: https://doi.org/10.3906/elk-2003-116
32. Kyte J, Doolittle RF (1982) A simple method for displaying the hydropathic character of a protein. J Mol Biol 157(1):105–132. https://doi.org/10.1016/0022-2836(82)90515-0
33. Jing X, Dong Q, Hong D, Lu R (2019) Amino acid encoding methods for protein sequences: a comprehensie review and assessment. IEEE/ACM Trans Comput Biol Bioinform, early access. https://doi.org/10.1109/TCBB.2019.2911677
34. Yin C, Yau ST (2017) A coevolution analysis for identfying protein-protein interactions by Fourier transform. PLOS One 12(4). https://doi.org/10.1371/journal.pone.0174862
35. Cadet F, Fontaine N, Vetrivel I, Chong MNF, Savriama O, Cadet X, Charton P (2018) Application of fourier transform and proteochemometrics principles to protein engineering. BMC Bioinform 19(1). https://doi.org/10.1186/s12859-018-2407-8
36. Weiss MA (2013) Data structures and algorithm analysis in C++. London
37. Nagaraj N, Balasubramanian K, Dey S (2013) A new complexity measure for time series analysis and classification. Eur Phys J Special Topics 222:847–860. https://doi.org/10.1140/epjst/e2013-01888-9
38. Nasar AA (2016) The history of algorithmic complexity. CUNY Academic Works. https://academicworks.cuny.edu/cgi/viewcontent.cgi?article=1073&context=bm_pubs. Accessed 12 Nov 2020
39. Thareja R (2014) Data structures using C. New Delhi, India.
40. Koffman EB, Pat W (2016) Data structures abstraction and design using java. Wiley, River Street, NJ
41. Islam Z, Islam M, Asraf A (2020) A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19)

using X-ray images. Inf Med Unlock 20:100412. https://doi.org/10.1016/j.imu.2020.100412
42. Jagannatha AN, Yu H (2016) Bidirectional RNN for medical event detection in electronic health records. Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 473–482, San Diego, California, 2016. https://doi.org/10.18653/v1/N16-1056
43. Schuster M, Paliwal KK (1997) Bidirectional recurrent neural networks. IEEE Trans Signal Process 45(11):2673–2681. https://doi.org/10.1109/78.650093
44. Toraman S, Alakus TB, Turkoglu I (2020) Convolutional capsnet: a novel artificial neural network approach to detect COVID-19 disaese from X-ray images using capsule networks. Chaos, Solutions Fractals, 140. https://doi.org/10.1016/j.chaos.2020.110122
45. Khan A, Sohail A, Zahoora U, Qurshi AS (2020) A survey of the recent architectures of deep convolutional neural networks. Artif Intell Rev. https://doi.org/10.1007/s10462-020-09825-6
46. Cosic I, Pirogova E (2007) Bioactive peptide design using the resonant recognition model. Nonlinear Biomed Phys 1(1). https://doi.org/10.1186/1753-4631-1-7
47. Yau SST, Wang J, Niknejad A, Lu C, Jin N, Ho YK (2003) DNA sequence representation without degeneracy. Nucleic Acid Re 31(12):3078–3080. https://doi.org/10.1093/nar/gkg432
48. Lehmann J, Libchaber A (2008) Degeneracy of the genetic code and stability of the base pair at the second position of the anticodon. RNA 14(7):1264–1269. https://doi.org/10.1261/rna.1029808
49. Abualigah LM (2019) Feature selection and enhanced krill herd algorithm for text document clustering, studies in computational ıntelligence, 816. https://doi.org/10.1007/978-3-030-10674-4
50. Abualigah LM, Khader AT, Hanandeh ES (2018) A new feature selection method to improve the document clustering using particle swarm optimization algorithm. J Comput Sci 25:456–466. https://doi.org/10.1016/j.jocs.2017.07.018
51. Alqanees MAA, Ewees AA, Fan H, Abualigah L, Elaziz MA (2020) Marine predators algorithm for forecasting confirmed cases of COVID-19 in Italy, USA, Iran and Korea. Int J Environ Res Publ Health 17(10). https://doi.org/10.3390/ijerph17103520