Analysis of Protein-Ligand Interactions of SARS-CoV-2 Against Selective Drug Using Deep Neural Networks

Natarajan Yuvaraj, Kannan Srihari*, Selvaraj Chandragandhi, Rajan Arshath Raja, Gaurav Dhiman, and Amandeep Kaur

Abstract: In recent time, data analysis using machine learning accelerates optimized solutions on clinical healthcare systems. The machine learning methods greatly offer an efficient prediction ability in diagnosis system alternative with the clinicians. Most of the systems operate on the extracted features from the patients and most of the predicted cases are accurate. However, in recent time, the prevalence of COVID-19 has emerged the global healthcare industry to find a new drug that suppresses the pandemic outbreak. In this paper, we design a Deep Neural Network (DNN) model that accurately finds the protein-ligand interactions with the drug used. The DNN senses the response of protein-ligand interactions for a specific drug and identifies which drug makes the interaction that combats effectively the virus. With limited genome sequence of Indian patients submitted to the GISAID database, we find that the DNN system is effective in identifying the protein-ligand interactions for a specific drug.

Key words: Deep Neural Network (DNN); coronavirus; protein-ligand interactions; deep learning; clinical healthcare system

- Natarajan Yuvaraj is with St. Peter's Institute of Higher Education and Research, Chennai 600054, India, and also with Information Communication Technology Academy, Chennai 600096, India. E-mail: yraj1989@gmail.com.
- Kannan Srihari is with Department of Computer Science and Engineering, SNS College of Engineering, Coimbatore 641107, India. E-mail: harionto@gmail.com.
- Selvaraj Chandragandhi is with Department of Computer Science and Engineering, Jagannath Educational Health and Charitable Trust College of Engineering and Technology, Coimbatore 641105, India. E-mail: chandragandhi09@gmail. com.
- Rajan Arshath Raja is with B. S. Abdur Rahman Crescent Institute of Science and Technology, Chennai 600048, India, and also with Information Communication Technology Academy, Chennai 600096, India. E-mail: arshathraja.ru@gmail.com.
- Gaurav Dhiman is with Department of Computer Science, Government Bikram College of Commerce, Patiala 147001, India. E-mail: gdhiman0001@gmail.com.
- Amandeep Kaur is with Department of Computer Science and Engineering, Sri Guru Granth Sahib World University, Fatehgarh 140406, India. E-mail: sweetchintoo2008@gmail.com.
- * To whom correspondence should be addressed.
  Manuscript received: 2020-06-04; accepted: 2020-06-18

1 Introduction

The Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has increased the mortality across globe. The prevention of anti-drugs is essential in reducing the mortality rate at a quicker rate so as to effectively combat the virus globally[1].

On one hand, with the advent of computationally intensive methods, the protein sequencing is made less expensive and faster. On the other hand, high-cost experiments and technical difficulties have led to poor accuracy in finding the protein-ligand interaction, where the smaller structural details are found[2]. The primary protein structure (amino acid sequence) and its binding residue help in direct determination of tertiary protein structure. Therefore, the binding residues are found easily through the protein properties, but it fails to reveal the complex nature of protein structure and binding residues from the primary structures[3].

In the field of bioinformatics, the utilization of deep learning algorithm in recent years helps in the prediction of binding residues using complex structural mappings.

It further varies from predicting the RNA-protein binding, secondary protein structure, and compound interaction of proteins and contact mapping[4]. The conventional computation methods using structural sequencing for the protein-ligand interaction assist in predicting the binding residues[5]. The deep learning architectures fit well to work on 3D-structure templates for binding and mapping. It extensively reduces the cost associated with template matching in distributed environment. The other major problem involves the higher order of 3D protein structure leading to poor utilization of 3D-structural data, thereby limiting the prediction of protein binding residues. With deep learning as the protein structure prediction model, we model the present study to find the relative drugs most suitable for treating the SARS-CoV-2 viral infection. Hence, in this paper, a Deep Neural Network (DNN) model is designed to find the protein-ligand interactions of selective drugs, which include ketoamide, lopinavir, nelfinavir, remdesivir, and ritonavir[6]. We analyze the protein-interaction behavior of COVID-19 with all the five drugs using DNN training and testing module.

The main contributions of this work involve the following:

(1) We develop a deep learning model that involves in finding the potential drugs against the formation of protein-ligand interactions of SARS-CoV-2 virus from the asymptomatic Indian patients based on the limited datasets collected.

(2) The study is supported by testing the patients against various drugs reported in Ref. [6], which include ketoamide, lopinavir, nelfinavir, remdesivir, and ritonavir.

(3) The optimal protease inhibitors are studied using the DNN model to find the negative dock value. Among all the selective inhibitors, the DNN with its repeated iteration confirms which of the utilized drugs is allowed to report the most effective binding value in comparison with the past binding results from Ref. [6].

The outline of the paper is presented below: Section 2 provides the related work. Section 3 discusses the proposed DNN protein-ligand interaction system. Section 4 evaluates the entire work. Section 5 provides the result and discussion. Section 6 concludes the work with possible directions of future scope.

2 Related Work

Upon reviewing various literatures on finding the COVID-19 drugs, we found that most of the existing methods[3–10] were operated on screening the patients using the CT and X-ray images of chest regions to detect the COVID-19 symptoms. In addition, with the image screening process, the authors in Ref. [11] used the nucleic acid testing (and polymerase chain reaction[12]) of SARS-CoV-2 for final identification of COVID-19 related symptoms. The authors in Ref. [13] aimed to develop a deep trained neural network model for finding the effective protein-ligand interaction from three different datasets, which include homology modelling and virtual screening on chimdiv database and tri-peptide database. The process involves alignment of protein sequence, virtual screening against four small compounds, and tri-peptides.

There exist several deep learning algorithms[14] on predicting the protein structure using deep learning algorithms, which include feed-forward neural networks[15], deep neural networks[16], recurrent neural networks[17], and convolutional neural network[18]. These methods show the protein structure annotations that help in describing the structure of proteins based on the extraction of protein sequence with local and global conformations to guide the protein folding[19]. The introduction of deep learning model produces higher representations by reducing the input dimensionality and aids in complex protein structures. With such motivation, the present study predicts the protein structure of SARS-CoV-2 patients using DNN, thereby providing insights on drug development process.

3 Material and Method

Dataset: The RNA sequences are collected from GISAID database, where they are translated into protein sequences for building the 3D model for possible DNN prediction. The RNA strands are collected from Indian citizens to study the effect of drugs on the SARS-CoV-2 virus.

Drugs considered: The study reports five different drugs including ketoamide, lopinavir, nelfinavir, remdesivir, and ritonavir. The binding energy of each drug with bond length is given as an input to the DNN model for drug prediction based on protein-ligand interaction, as shown in Table 1.

Definition of ligand binding residues: Protein-ligand interactions were studied extensively in Refs. [20–24] including drug interaction, carbohydrate recognition, and DNA binding. In addition, the drug discovery process considers the ligand binding regions prediction as an essential component, and DNN is used for such

**Table 1 Binding energy and bond length of COVID-19 protease inhibitors.**

Ligand	Binding energy (kcal·mol^{-1})	Bond length (Å)
Ketoamide	−5.80	2.8
Lopinavir	−6.08	2.7
Nelfinavir	−7.54	3.0
Remdesivir	−5.51	2.7
Ritonavir	−5.96	2.7

Note: 1 cal≈4.18 J; Å = 0.1 nm.

prediction purposes. To reduce the adverse effects of the drugs utilized for the study, we find the location of the binding sites based on predicting the distance between the atoms or residues that facilitates in optimizing the drug discovery process with selective structural features. These features highly influence the binding selectivity for limiting the adverse effect of drugs. There exist several methods on predicting the ligand binding sites using machine learning methods[25–28] with amino acids residues as their residues[29–31]. In the present study, the understanding on ligand interaction with protein membrane is limited due to unavailability of literatures as compared with globular proteins.

In this paper, we carry out the protein-ligand interaction to study the interaction of drugs on protein structure, which is analyzed by DNN model over several iterations. We assume the ligand atoms are intact with residual atoms. If the distance existing between these two kinds of atoms is less than the cut-off level (4.5 Å), then the atoms are said to be in binding state. From the dataset of several residues, we collect protein sequence of SARS-CoV-2 and then identify the sequence as a ligand binding residue.

Description of DNN: The architecture of DNN is given in Fig. 1. DNN is used frequently in several applications. The performance of DNN network entirely depends on the optimal selection and arrangement of network layer. The network training for DNN is carried out using the learning task. A requirement-based
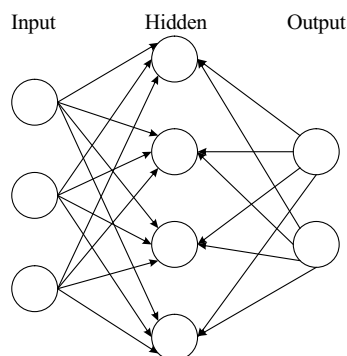
Fig. 1   Architecture of DNN.

pre-trained network is modified and the new protein sequence dataset is formed into transfer learning.

The general neural network architecture shown in Fig. 1 consists of an input layer, a hidden layer, and an output layer. Each layer is selected and arranged on the basis of the size of output, when a DNN network is designed from scratch. Initial convolutions are made based on the various filter sizes. The convolution layer extracts the protein sequences from the 3D models. The feature maps are formed and then sub-sampled using pooling layer, and then the size of feature map reduces for propagation over upcoming layers.
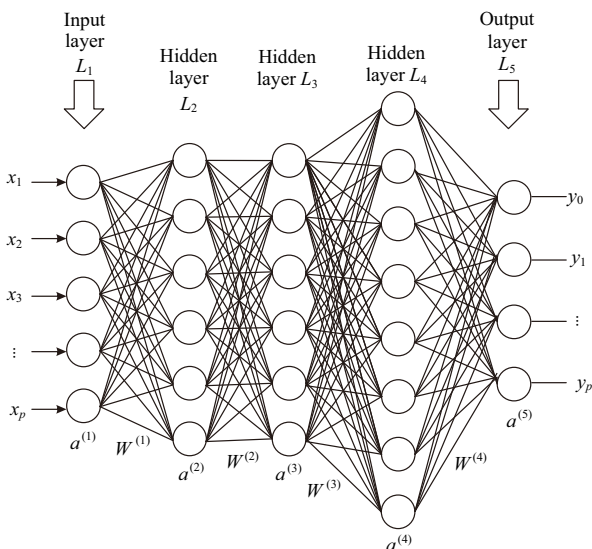
The activation function in DNN is a Rectified Linear Unit (ReLU). To reduce the simplicity of computations, it operates more like a linear activation function than other activation functions. The rectified linear unit as an activation function exploits improvements in training the multi-layered networks with reduced complexity in contrast with a nonlinear activation function. The nonlinearity of the DNN model is added in order to perform more effectively at training for the network.

The DNN layer is completely connected and it includes an input layer, multiple hidden layer, and an output layer. The neural network functions consistently, as each DNN neuron layer is connected to next layer of a neuron. The result of classification is calculated using the connected DNN layer based on cut-off value (as discussed earlier) calculated with softmax layer for finding the distance between the residuals. Here, after classifying scores are obtained, the 3D image is classified into different classes.

Artificial Neural Networks (ANNs) consist of several artificial neuron nodes that emulate human brain neurons. In contrast to the biological neurons, there is only one kind of link between a neuron and others. The neurons collect input data, which are simple to be operated. These operations lead to the transmission of other neurons. The activation function determines if the result has been passed. The activation function plays an important role for both extraction and classification of features.

A multi-layer neural network (Fig. 2) could be considered stacked. As name suggests, when DNN contains over a hidden layer and the system moves to a layer, it is called the Multi-Layer Perceptron (MLP). These neural networks can be classified and forecast. When the DNN is used as classification, the input and output nodes match input and output classes.

The input layer $x = \{x_1, x_2, \ldots, x_p\}$ (with activation

Fig. 2   **Feed-forward DNN with 3 hidden layers.**

Fig. 3   Feature extraction of protein structure using artificial neural network.

function $a^{(1)}$ of weight $W^{(1)}$) consists of four elements that can be used to classify one result from another and identify a pattern in the input entities. For this purpose, we design a series of hidden units ($a^{(2)}$, $a^{(3)}$, and $a^{(4)}$) with activation functions and weights ($W^{(2)}$, $W^{(3)}$, and $W^{(4)}$), respectively. Active unit pattern is utilized by ReLU and other nonlinear activations.
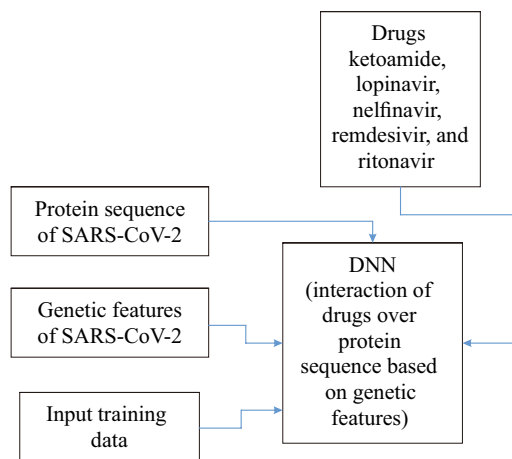
As shown in Fig. 2, the output layer $y = \{y_1, y_2, \ldots, y_p\}$ is activated in hidden layer $a^{(5)}$. Weight and distortion are adjusted for hidden nodes to activate. Initially, weights and biases are randomly changed. Then we train our network for the tens of thousands of inputs. The test uses retrospective error propagation to change biases and weights in order to activate hidden neurons at suitable values.

The set or kernel of functions is the set of weights and partialities, which is used to identify a specific output. DNN will be assumed to be the key to the solution when the pattern for discrimination is complex enough to prevent traditional statistical and numerical approaches.

Pretrained network. The pre-trained network is amended to match the output requirements, which is one of the most common methods of learning. There are a lot of pre-trained user training networks. These pre-trained networks tend to face a range of training challenges when they are trained over millions. A comparison of Alexnet, ResNet-18, and GoogLeNet is discussed. Each of the three kinds of pre-trained networks brings together two kinds of layers: convolutional and pooling layers.

4   Proposed Method

In this section, as shown in Fig. 3, the prediction of protein structure from the repository is carried out to virtually screen the drugs for finding suitable curable drug. The complexes of protein-ligand are considered as positive and cross-docking is considered as negative for the input training datasets. The DNN is a fully connected network that has the ability to learn more abstract features from the input training data.

The DNN is modelled as a distance measure to predict the actual distance between the residual pairs within the protein-ligand interaction. It directly estimates the candidate structure's accuracy and generation of protein structure. The DNN operates as a template modelling that sets main protease of SARS-CoV-2 as its template to perform the prediction of sequence alignment.

The DNN is responsible for sequentially analyzing the SARS-CoV-2 proteins obtained from the RNA sequences, where the sequences are initially translated from the amino acid sequences. The DNN predicts the proteins by focusing entirely on the ligand binding region and S protein regions.

Then the virtual screening against all the selected drugs was performed based on the homology model. This model consists of the ligand binding region from the protease of SARS-CoV-2 as its template. The cut-off distance between the residual pair is set as 1 nm based on the template setting. The ligand database with millions of components is used as input feature by the DNN algorithm to perform virtual screening of drugs. The scalability of the model is regularised by considering the mean and standard deviation of the datasets used during normalization of input training data. With the features collected from the input training sets, the validation of DNN model is carried out with docking simulation. The validation provides the results

of compound list that can bind potentially with the protein-ligand structure. The compounds are then ranked and the one with high validation score is considered as an inhibitor for validation of suitable drug to battle the SARS-CoV-2.

Depending on the input training datasets, the validation is carried out by the DNN model. It finds the accuracy of drug interacting with protein sequences based on the genetic features and input training data. The results of validation are thus presented in the following section.

5   Result and Discussion

The protein sequence dataset[32] is used for the design and evaluation on diagnostic validation of drugs interaction over SARS-CoV-2 virus. The synthesis of protein sequence from ProteinNet[32] and its interaction behaviour with the drug used in the present study are noted at each iteration while training the DNNs. The simulations of DNN are conducted on Matlab simulator on a high-end computing machine.

The validation of the trained DNN is carried out by comparing the test results with existing methods. The protein synthesis interaction in the validation is carried out by the test dataset and the classified output is finally obtained from DNN.

If the accuracy of the prediction is not obtained well, then the network is altered until the prediction results are obtained to be precise. This helps to achieve higher validation accuracy of drug interaction with the protein synthesis with training progress. The performance of deep learning algorithm is presented via confusion tables.

The performance of the proposed method is tested in terms of accuracy, precision, recall, and F1-score as in the following equations:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \qquad (1)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \qquad (2)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \qquad (3)$$

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \qquad (4)$$

where TP represents true positive, TN represents true negative, FP represents false positive, and FN represents false negative. The performance of different pre-trained systems is thoroughly analyzed as an early stage in the development of a DNN for the classification of protein

synthesis. The performance is comparable by three pre-worked networks: Alexnet, GoogLeNet and ResNet-18. In test datasets, the network is initially loaded, altered, and trained. The networks tend to achieve a better performance comparison with other performance parameters once the training has been completed.

The DNN is set with same options to compare them with parameters and same dataset for specific network architectures. The networks are trained up to 40 iterations and then the protein data are shuffled and validated over multiple iterations.

In this paper, we present the frequency of validation to be five iterations. The set of options can also accurately trace the training progress and display the validation accuracy after all three iterations. The modified pretrained network is trained for the same number of epochs. Figure 4 presents the classification accuracy of the three networks.

After training the DNN, the performance of network is evaluated under different performance metrics and confusion matrix determines all the performance of target classes, i.e., drugs interaction. The performance of DNN is then compared with other algorithms that include Deep Belief Network (DBN), ANN, Feed-Forward Neural Network (FFNN), and Back Propagation Neural Network (BPNN). The performance of the proposed DNN AlexNet is chosen for final simulation of finding the accuracy, precision, recall, and F1-score. The results are evident as shown in Tables 2–6.

The results show that the proposed DNN is effective in finding the drug interaction with protein sequence. Out of all the five drugs, the results show that remdesivir drug is effective in interacting with the protein sequence
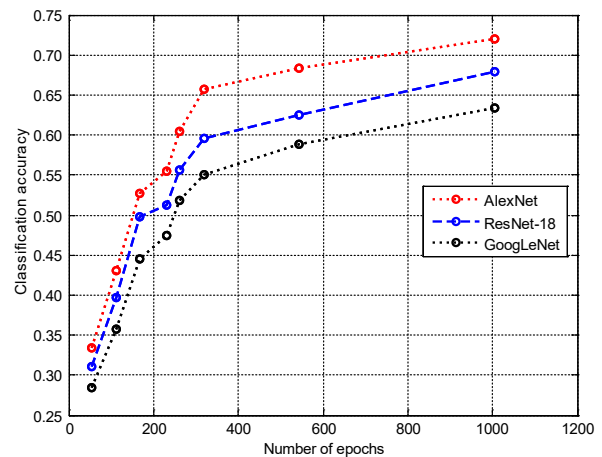
Fig. 4   Classification accuracy of training models of DNN with three pre-trained networks.

Table 2　Performance of ketoamide drug interaction with SARS-CoV-2 virus over Indian datasets.

Algorithm	Accuracy	Precision	Recall	F1-score
DNN	0.9208	0.9389	0.6064	0.7356
DBN	0.9210	0.9394	0.6100	0.7369
ANN	0.9211	0.9395	0.6152	0.7421
FFNN	0.9220	0.9398	0.6313	0.7529
BPNN	0.9222	0.9399	0.6363	0.7597

Table 3　Performance of lopinavir drug interaction with SARS-CoV-2 virus over Indian datasets.

Algorithm	Accuracy	Precision	Recall	F1-score
DNN	0.9460	0.9568	0.6968	0.6256
DBN	0.9442	0.9564	0.6786	0.6103
ANN	0.9436	0.9559	0.6751	0.6057
FFNN	0.9429	0.9558	0.6662	0.6024
BPNN	0.9414	0.9542	0.6589	0.5844

Table 4　Performance of nelfinavir drug interaction with SARS-CoV-2 virus over Indian datasets.

Algorithm	Accuracy	Precision	Recall	F1-score
DNN	0.9665	0.9762	0.9642	0.8958
DBN	0.9587	0.9761	0.9638	0.8956
ANN	0.9452	0.9760	0.9571	0.8854
FFNN	0.9403	0.9758	0.9553	0.8854
BPNN	0.9382	0.9758	0.9544	0.8643

Table 5　Performance of remdesivir drug interaction with SARS-CoV-2 virus over Indian datasets.

Algorithm	Accuracy	Precision	Recall	F1-score
DNN	0.9969	0.9984	0.9443	0.9155
DBN	0.9968	0.9982	0.9438	0.9151
ANN	0.9961	0.9973	0.9364	0.9015
FFNN	0.9961	0.9973	0.9351	0.9012
BPNN	0.9954	0.9963	0.9270	0.8806

Table 6　Performance of ritonavir drug interaction with SARS-CoV-2 virus over Indian datasets.

Algorithm	Accuracy	Precision	Recall	F1-score
DNN	0.9822	0.6963	0.9999	0.8197
DBN	0.9820	0.6913	0.9998	0.8129
ANN	0.9811	0.6752	0.9995	0.8021
FFNN	0.9810	0.6700	0.9994	0.7969
BPNN	0.9808	0.6664	0.9989	0.7956

and hence it can be treated as an effective drug for neutralizing the SARS-CoV-2 virus. The remdesivir has shown increased molecular interaction due to multiple active site residues in protease structure. There exists a single active site residue in protease structure for remaining ligand. This helps the protease inhibitors to show more molecular interactions by the remdesivir than other compounds.

6　Conclusion

In this paper, we design a DNN model for the potential identification of drugs against the protein-ligand interaction of SARS-CoV-2 virus, specifically in Indian patients. The results of DNN are accurate and it is quick in finding the antidote among the selective drugs using virtual drug screening process. The DNN assists the drug screening process by generating possible compound and tripeptide list on asymptotic Indian patients. The rich antiviral property in the nucleic acid analogue of adenosine and nucleotide analogues used in remdesivir is found to be effective in combating the viral property of SARS-CoV-2 infections. The experiments on protein-ligand interaction provide possible cure by the remdesivir drug than other selective drugs. The results are accurate with repeated iterations with the negative dock value and binding energy value. Depending on the molecular interaction with protease inhibitors, it is found that the remdesivir drug provides the instant relief against the COVID-19 symptoms and possible cures against the viral infection. In the future, the study may be developed with a possible combination of reinforcement learning and deep learning algorithm to deeply analyze the peptide structure from protein-ligand interaction.

References

[1]　A. K. Singh, A. Singh, A. Shaikh, R. Singh, and A. Misra, Chloroquine and hydroxychloroquine in the treatment of COVID-19 with or without diabetes: A systematic search and a narrative review with a special reference to India and other developing countries, *Diabetes Metab. Syndr. Clin. Res. Rev.*, vol. 14, no. 3, pp. 241–246, 2020.

[2]　J. Y. Yang, A. Roy, and Y. Zhang, Protein-ligand binding site recognition using complementary binding-specific substructure comparison and sequence profile alignment, *Bioinformatics*, vol. 29, no. 20, pp. 2588–2595, 2013.

[3]　Y. F. Cui, Q. W. Dong, D. C. Hong, and X. Wang, Predicting protein-ligand binding residues with deep convolutional neural networks, *BMC Bioinformatics*, vol. 20, no. 1, p. 93, 2019.

[4]　S. Wang, S. Q. Sun, Z. Li, R. Y. Zhang, and J. B. Xu, Accurate de novo prediction of protein contact map by ultra-deep learning model, *PLoS Comput. Biol.*, vol. 13, no. 1, p. e1005324, 2017.

[5]　J. Y. Yang, A. Ro, and Y. Zhang, BioLiP: A semi-manually curated database for biologically relevant ligand-protein interactions, *Nucl. Acids Res.*, vol. 41, no. D1, pp. D1096–D1103, 2012.

[6]　D. Mothay and K. V. Ramesh, Binding site analysis of potential protease inhibitors of COVID-19 using AutoDock, *VirusDis.*, vol. 31, no. 2, pp. 194–199, 2020.

[7]　L. Huang, R. Han, T. Ai, P. X. Yu, H. Kang, Q. Tao,

and L. M. Xia, Serial quantitative chest CT assessment of COVID-19: Deep-learning approach, *Radiol. Cardiothorac. Imaging*, vol. 2, no. 2, p. e200075, 2020.

[8] L. Li, L. X. Qin, Z. G. Xu, Y. B. Yin, X. Wang, B. Kong, J. J Bai, Y. Lu, Z. H. Fang, Q. Song, et al., Artificial intelligence distinguishes COVID-19 from community acquired pneumonia on chest CT, https://doi.org/10.1148/radiol.2020200905, 2020.

[9] I. D. Apostolopoulos and T. A. Mpesiana, Covid-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks, *Phys. Eng. Sci. Med.*, vol. 431, pp. 635–640, 2020.

[10] C. Butt, J. Gill, D. Chun, and B. A. Babu, Deep learning system to screen coronavirus disease 2019 pneumonia, *Appl. Intell.*, doi:10.1007/s10489-020-01714-3.

[11] C. S. Zheng, X. B. Deng, Q. Fu, Q. Zhou, J. P. Feng, H. Ma, W. Y. Liu, and X. G. Wang, Deep learning-based detection for COVID-19 from chest CT using weak label, https://doi.org/10.1101/2020.03.12.20027185, 2020.

[12] D. S. Li, D. W. Wang, J. P. Dong, N. N. Wang, H. Huang, H. W. Xu, and C. Xia, False-negative results of real-time reverse-transcriptase polymerase chain reaction for severe acute respiratory syndrome coronavirus 2: Role of deep-learning-based CT diagnosis and insights from two cases, *Korean J. Radiol.*, vol. 21, no. 4, pp. 505–508, 2020.

[13] H. P. Zhang, K. M. Saravanan, Y. Yang, M. T. Hossain, J. X. Li, X. H. Ren, Y. Pan, and Y. J. Wei, Deep learning based drug screening for novel coronavirus 2019-nCov, *Interdiscip. Sci. Comput. Life Sci.*, doi: 10.1007/s12539-020-00376-6.

[14] M. Torrisi, G. Pollastri, and Q. Le, Deep learning methods in protein structure prediction, *Comput. Structural Biotechnol. J.*, vol. 18, pp. 1301–1310, 2020.

[15] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, *Nature,* vol. 323, no. 6088, pp. 533–536, 1986.

[16] A. W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. L. Qin, A. Žídek, A. W. R. Nelson, A. Bridgland, et al., Improved protein structure prediction using potentials from deep learning, *Nature*, vol. 577, no. 7792, pp. 706–710, 2020.

[17] J. L. Elman, Finding structure in time, *Cogn. Sci.*, vol. 14, no. 2, pp. 179–211, 1990.

[18] S. Wang, J. Peng, J. Z. Ma, and J. B. Xu, Protein secondary structure prediction using deep convolutional neural fields, *Sci. Rep.,* vol. 6, no. 1, p. 18 962, 2016.

[19] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature,* vol. 521, no. 7553, pp. 436–444, 2015.

[20] H. J. Hinz, Thermodynamics of protein-ligand interactions: Calorimetric approaches, *Annu. Rev. Biophys. Bioeng.*, vol. 12, no. 1, pp. 285–317, 1983.

[21] T. Simonson, G. Archontis, and M. Karplus, Free energy simulations come of age: Protein-ligand recognition, *Acc. Chem. Res.*, vol. 35, no. 6, pp. 430–437, 2002.

[22] R. Perozzo, G. Folkers, and L. Scapozza, Thermodynamics of protein-ligand interactions: History, presence, and future aspects, *J. Recept. Signal Transduct. Res.*, vol. 24, nos. 1 & 2, pp. 1–52, 2004.

[23] M. Hendlich, A. Bergner, J. Günther, and G. Klebe, Relibase: Design and development of a database for comprehensive analysis of protein-ligand interactions, *J. Mol. Biol.*, vol. 326, no. 2, pp. 607–620, 2003.

[24] L. Y. Lian, I. L. Barsukov, M. J. Sutcliffe, K. H. Sze, and G. C. Roberts, Protein-ligand interactions: Exchange processes and determination of ligand conformation and protein-ligand contacts, *Meth. Enzymol.,* vol. 239, pp. 657–700, 1994.

[25] G. Mirceva and A. Kulakov, Improvement of protein binding sites prediction by selecting amino acid residues' features, *J. Struct. Biol.*, vol. 189, no. 1, pp. 9–19, 2015.

[26] C. H. Yan and Y. F. Wang, A graph kernel method for DNA-binding site prediction, *BMC Syst. Biol.*, vol. 8, no. S4, p. S10, 2014.

[27] S. C. Izidoro, R. C. de Melo-Minardi, and G. L. Pappa, GASS: Identifying enzyme active sites with genetic algorithms, *Bioinformatics*, vol. 31, no. 6, pp. 864–870, 2015.

[28] B. Park, J. Im, N. Tuvshinjargal, W. Lee, and K. Han, Sequence-based prediction of protein-binding sites in DNA: Comparative study of two SVM models, *Comp. Methods Programs Biomed.*, vol. 117, no. 2, pp. 158–167, 2014.

[29] G. J. Bartlett, C. T. Porter, N. Borkakoti, and J. M. Thornton, Analysis of catalytic residues in enzyme active sites, *J. Mol. Biol.*, vol. 324, no. 1, pp. 105–121, 2002.

[30] M. J. J. M. Zvelebil and M. J. E. Sternberg, Analysis and prediction of the location of catalytic residues in enzymes, *Protein Eng. Des. Sel.*, vol. 2, no. 2, pp. 127–138, 1988.

[31] C. Taroni, S. Jones, and J. M. Thornton, Analysis and prediction of carbohydrate binding sites, *Protein Eng. Des. Sel.*, vol. 13, no. 2, pp. 89–98, 2000.

[32] M. AlQuraishi, ProteinNet: A standardized data set for machine learning of protein structure, *BMC Bioinform.*, vol. 20, no. 1, p. 311, 2019.

Natarajan Yuvaraj received the BEng and MEng degrees in computer science engineering from Anna University, Tamil Nadu, India in 2010 and 2012, respectively. He is pursuing the PhD degree in computer science engineering at St. Peter's Institute of Higher Education and Research, Chennai, India. From 2012 to 2015, he was an assistant professor at various educational institutes. Currently, he is working as a deputy manager of research & development department at Information Communication Technology Academy, Tamil Nadu, India. His research interest includes data mining, wireless sensor networks, and mobile computing. He has filed 5 Indian patents, wrote 2 text books, and published 4 book chapters in Elsevier, Springer, CRC Taylor and Francis, and John Wiley and Sons. He has also published 40 research papers in SCI, Scopus, and international and national conferences.

Kannan Srihari received the MEng and PhD degrees from Anna University, Tamil Nadu, India in 2009 and 2013, respectively. He is currently working as an associate professor at Department of Computer Science and Engineering, SNS College of Engineering, India. He published over 60 papers in international journals. His research area includes semantic search engines, microprocessor, FPGA, big data, and cloud computing. He has guided 6 PhD students and currently he is guiding 4 scholars.

Selvaraj Chandragandhi received the BTech and MTech degrees from Anna University, India in 2008 and 2013, respectively. She is working as an assistant professor at Jagannath Educational Health and Charitable Trust College of Engineering and Technology. She has published 5 Scopus papers. Her research area is IoT, bio medical, and block chain.

Rajan Arshath Raja received the bachelor of engineering degree in electronics and communication engineering from Anna University in 2012. He received the master of engineering degree in communication systems from Sree Sastha Institute of Engineering and Technology in 2014. He is currently pursuing the PhD degree in wireless communication at B. S. Abdur Rahman Crescent Institute of Science and Technology. He is currently working as a senior research associate of research & development department at Information Communication Technology Academy, India. His research interest includes the wireless communication, machine learning, and artificial intelligence. He has filed 2 Indian patents, and published 4 book chapters in Elsevier, Springer, CRC Taylor and Francis, and John Wiley and Sons. He also published 10 research papers in SCI, Scopus, and international and national conferences.
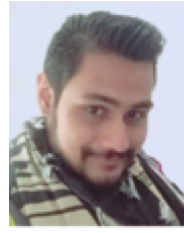
Gaurav Dhiman received the master and PhD degrees in computer engineering from Thapar Institute of Engineering & Technology, Patiala, India in 2015 and 2019, respectively. He is currently working as an assistant professor at Department of Computer Science, Government Bikram College of Commerce, Patiala, India. He was selected as the outstanding reviewer from knowledge-based systems (Elsevier). He has published more than 50 peer-reviewed research papers (indexed in SCI-SCIE) and 5 international books. He is also serving as the guest editor of more than five special issues in various peer-reviewed journals. He is also with editorial board members of *Current Chinese Computer Science* (Bentham Science), *Current Dissertation*s (Bentham Science), and *Journal of Fuzzy Logic and Modeling in Engineering* (Bentham Science). His research interest includes single- and multi-objectives optimization (bio-inspired, evolutionary, and quantum), soft computing (type-1 and type-2 fuzzy sets), power systems, and change detection using remotely sensed high-resolution satellite data. His research articles can be found in *Knowledge-Based Systems* (Elsevier), *Advances in Engineering Software* (Elsevier), *Applied Intelligence* (Springer), *Journal of Computational Science* (Elsevier), *Engineering with Computers* (Springer), *Applied Soft Computing* (Elsevier), etc.

Amandeep Kaur received the master and PhD degrees in computer engineering from Thapar Institute of Engineering & Technology, Patiala, India in 2012 and 2020, respectively. She is currently working as an assistant professor at Department of Computer Science and Engineering, Sri Guru Granth Sahib World University, India. Her research interest includes single- and multi-objectives optimization (bio-inspired, evolutionary, and quantum), soft computing (type-1 and type-2 fuzzy sets), power systems, and change detection using remotely sensed high-resolution satellite data.