RESEARCH ARTICLE

Prediction of Coronavirus Inhibitors in Drug Discovery through Deep Learning

Tauseef Khan[1], Altaf Hussain[2], Tariq Hussain[3,*], Xianxuan Lin[4], Amin Sharafian[5], Islam Md Monirul[5] and Umme Laila[6]

[1] Department of Computer Science, COMSATS University Islamabad, Abbottabad Campus, Pakistan
[2] School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
[3] School of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou 310018, China
[4] School of Artificial Intelligence, Nanjing University of Information Science & Technology, Nanjing 210044, China
[5] College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen 518060, China
[6] Computer Science Department, Institute of Business Management (IoBM), Karachi 75190, Pakistan

Abstract

In the therapy of Coronavirus, the drug target is a demanding task to find novel medicine. A bunch of pharmaceutics procedures are employed to recognize these mutual actions. But they are exhausting and high-priced. Keeping this in view, computational procedures are widely approached to determine the mutual action of the medicine and their respective proteins. Many scientists have applied ML approaches to deduce attributes from simplified molecular-input line systems (for medicine) and protein sequences. Such approaches dropped the proteins' chemical, physical, and structural characteristics and the respective medicine. Our job is to undertake deep learning approaches to detect coronavirus enzyme correspondence with the validated Chembl database medicine. The representation of the molecular structure of proteins, medically known as fingerprints, will be done scientifically. Then, a deep learning model will be given training on the pulled-out fingerprints and the properties of molecules to determine the interplay of the medicine with the respective catalyst. The suggested approach will be proficient in recognizing the catalyst's interactivity with the approved database medicine.

Keywords: drug discovery, Covid-19, deep learning, machine learning, bio-informatics.

1 Introduction

Positively-sense RNA viruses belonging to the family Coronaviridae (CoVs) are responsible for causing various infections in humans, animals, and birds [1]. There are four generations in this family, namely alphacoronavirus, betacoronavirus, deltacoronavirus, and gamammoronavirus [2]. Two of the most notorious conditions in the beta virus species are

Submitted: 24 January 2024
Accepted: 18 February 2024
Published: 28 February 2024

Vol. 1, **No.** 1, 2024.
10.62762/TACS.2024.974479

***Corresponding author:**
✉ Tariq Hussain
uom.tariq@gmail.com

Citation
Khan, T., Hussain, A., Hussain, T., Lin, X., Sharafian, A., Monirul, I. M., & Laila, U. (2024). Prediction of Coronavirus Inhibitors in Drug Discovery through Deep Learning. *ICCK Transactions on Advanced Computing and Systems*, 1(1), 19–31.

(SARS-CoV) [3] and (MERS-CoV) [4], which have contaminated thousands of people throughout the world in the last two decades. With the current drug discovery system, newly discovered medicines will take several years to reach the market [25]. Rapidly, due to the recent outbreak of atypical pneumonia (dubbed COVID-19) caused by the new Coronavirus (SARS-CoV-2, Figure 1 illustrates the structural representation of SARS-CoV-2) in Wuhan, China, the world is in the same situation as the previous wave [1, 5]. As per the latest updates from the WHO, the global medical community has not yet identified any exceptional individuals in medicine for effectively eradicating COVID-19 [6]. People dealing with hydroxychloroquine (HCQS) as a therapy for this condition expressed surprise and excitement [7]. However, medical practitioners hesitate to disseminate the data with the same zeal because it has yet to be shown beneficial. When dealt with, it was revealed that remedisivir is good in the United States, but it is not yet available to treat patients with COVID-19 infection. It is considered a safe option until an important study has been carried out simultaneously [8]. The latest news has shown clinical trials in COVID-19 patients in New York heartburn drugs [9]. In summary, there is currently no established treatment to combat the COVID-19 virus. However, the utilization of artificial intelligence (AI) tools in the medical field has the potential to facilitate a feasible cure [9]. DL models [10] were recently accepted as a breakthrough, providing a new chance to make computer decisions based on pharmaceutics. Small molecules that move on a protein target can be identified when analyzing protein constructions using structural-oriented medicine design methods [26]. So, the proficient research capacity here presents a new ingredient. If biochemists validate it as an effective solution, it will help humanity survive these difficult times. We will design a deep learning model to find the co-occurrence of coronavirus disease enzymes with the licensed medicine of the Chembl database. The model will be given training on the pulled-out representations of proteins and the properties of molecules to find out the interplay of the medicine with the respective enzyme.

Covid-19's immediate threat highlights the significant need to generate treatment options for rising physical problems [27]. Deep learning seems to have the benefits of being easily adaptable to new environments, enabling us to keep up with the viral threat and collect relevant information [11].

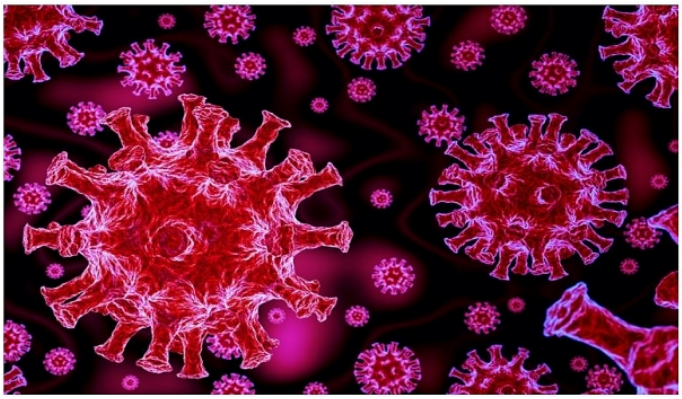

Figure 1. SARS-CoV-2 illustration [6].

Like any emerging medical condition, data occasionally takes time to catch up [28]. The virus quickly moves, presenting a tremendous problem because this can adapt and develop resistance to common therapies. "How can we possibly identify the optimal synergistic combinations for the highly infectious SARS-CoV-2?" asked investigators from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) and, indeed, the Jameel Clinic for Machine Learning in Healthcare [12, 29], SARS-CoV-2. Software engineers have widely used deep learning to suggest pharmaceutical formulations for illnesses such as cancer and heart events from big existing datasets. However, it is established that these can be employed for novel ailments with insufficient information.

Researchers who lack the necessary knowledge and documentation require an innovative approach, such as a neural network designed to handle multiple tasks. In this context, drug synergy—where drugs inhibit biological targets like proteins or nucleic acids—is a key focus [30]. The model is trained to predict chemical interactions and drug-drug synergies for discovering new compounds. A drug-target predictor simulates the interaction between a drug and its associated therapeutic properties relevant to the illness [31]. Similarly, a target-disease interaction predictor assesses a drug's antibacterial efficacy by analyzing virus activity in a controlled suspension culture [32]. This enables researchers to forecast the synergistic effects of two drugs when administered together, providing valuable insights for therapeutic applications. Nonetheless, over a longer period, the effects will be significant. In other words, countries will still face economic challenges even if schools recover quickly to their pre-pandemic performance levels [33]. For instance, in the United States, if the current undergraduate cohorts in schools during the 2020 closures lose only 0.1% of their abilities, and all

affiliates recover to their previous levels, the projected GDP loss of 1.5% would be equivalent to an enormous economic deficit of USD 15.3 trillion [34].

The COVID-19 outbreak has tremendously impacted higher academia, with universities falling out of business and countries locking their borders in response to shutdown remedies [35]. While universities and colleges swiftly transitioned from in-person sessions to online distance learning, these closures significantly impacted learning outcomes, examination processes, and the safety and legal status of international students in host countries [36]. More critically, the shift has adversely affected the overall quality of university education, including academic content, opportunities for networking, and students' psychological development.

To stay current, universities must revamp their learning environments to integrate digitalization and expand and improve relationships among undergraduate students and possibly other groups [37]. At the onset of the pandemic, the WHO initiated a coronavirus (COVID-19) surveillance and reporting system for its Member States, compiling data into a database and a related dashboard [38]. Surveillance data obtained in this manner is major in the global count of infections and fatalities. However, some regions lack sufficient capacity in their health data systems to report crucial information about deaths and their underlying causes accurately [39]. A worldwide study conducted before the epidemic indicated that four out of every ten fatalities in the globe remain unrecorded [40]. The claimed number of fatalities from COVID-19 has been called into doubt on several occasions, with WHO, the Institute for Health Metrics and Evaluation (IHME), and data journalism organizations all presenting global and cross-country estimates. To analyze the direct and indirect impact of COVID-19, these studies largely relied on excess fatalities or deaths that occurred more than what would be predicted at the same time of year. Despite known limitations in comparing excess mortality across countries, the findings of these analyses suggest that perhaps the deaths caused by COVID-19 are at least 60% higher than reported and possibly even or above in countries with insufficient death registration systems or statistical transport systems [41].

COVID-19 has killed over 2.1 million people globally [42]. We must discover medicines to reduce the disease's impact. While finding individual medications for this goal has been challenging,

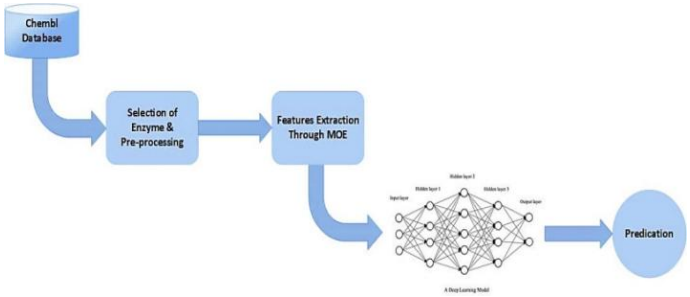

Figure 2. A proposed methodology for predicting coronavirus disease enzyme inhibitors.

synergistic pharmacological combinations provide a viable option. The lack of high-quality training data for medication combinations poses a significant challenge to the effective use of existing machine learning algorithms in predicting novel drug combinations. To address this problem, our proposed approach utilizes easily accessible information, such as drug-target interactions, to effectively search for synergistic combinations against SARS-CoV-2 through computational means.

Figure 2 shows the proposed methodology steps for predicting coronavirus inhibitors. The approach is based on supervised learning. The relevant coronavirus drug data set is collected from the Chembl database, and some refining techniques will be applied to the database. Fingerprints will be drawn out through MOE software. The model will be implemented in Python. The extracted fingerprints are provided to prepare for the deep learning model. The data will be split into two groups: training and testing. The training set could be used to build an approach to estimate the intended result. Once the model is trained, a collection of tests will validate the results. After testing the prediction capability of our model, both the training and testing sets will be combined again, which will act as a new training set and be used to classify external test sets. Accuracy, recall, and F1 points will be used to evaluate the classifier's results.

Various deep learning methods have been employed in the literature to predict drug-target interactions (DTIs). However, to the best of our knowledge, the Lipinski rule—commonly known as the "rule of five"—has not been utilized in this context. This rule provides a framework for evaluating whether a chemical compound possesses the physicochemical properties necessary to be a likely orally active drug in humans. In this study, we introduce a new strategy for identifying DTIs by extracting molecular fingerprints using MOE software. Unlike

many existing models, which often compromise the physicochemical properties of drugs during representation processes, our approach preserves these critical attributes to enhance predictive accuracy. The goal of this study is to predict interactions between coronavirus enzymes and approved drugs using deep learning. This prediction task involves binary classification, where the model determines whether a drug-enzyme pair is interactive or not based on molecular fingerprints and enzyme properties. We aim to improve accuracy and maintain the chemical properties of selected products.

The study's main contribution focuses on a new strategy for locating DTIs, where the authors would pull out the fingerprints through MOE software. We would also utilize the Lipinski rule, which shelters the numbers of hydrophilic clusters, molecular weight, and hydrophobicity, to outperform the proposed model compared to the previous model.

The main problem that this study undertakes is the efficacy of computational methods in predicting interactions between an enzyme of the coronavirus and a drug already approved for use in the clinic. Experimental techniques in general tend to be costly and slow and thus preclude the rapid development of potential inhibitors for COVID-19. In the present research toward addressing this issue, a deep learning model will be developed through molecular fingerprints and enzyme properties from the Chembl database to predict the interactions. In this study, whether the deep learning approach offers any solid and scalable solution toward finding novel promising drug candidates against COVID-19 has been explored.

This research makes a couple of main contributions: (1) it presents a new application of deep learning in predicting drug-enzyme interactions with specificity toward coronavirus targets, thereby addressing an important area of necessity in drug discovery against COVID-19; (2) it combines molecular fingerprints alongside enzyme properties extracted using MOE software, which retains all detailed chemical information often neglected in former studies; and (3) filling the gap in research with a scalable computational framework previously validated on the Chembl database, it now adds to the overall construct of computational drug discovery.

The rest of the Paper is organized into four main Sections. The second Section provides background information and reviews related work. The third section introduces the proposed architecture for this prediction and provides an overview of machine learning and deep learning concepts. The fourth Section details the model's results. The fifth and final Section concludes.

2 Related Work

Most previous studies have emphasized general drug-target interactions using various datasets such as DrugBank [16], ZINC [11], and KIBA [17]. These have used convolutional neural networks, graph-based models as well as hybrid frameworks to predict wide ranges of protein-compound interactions [18]. However, the lack of specificity to coronavirus targets, which are very important in the current pandemic scenario, is true with most of the past studies. On the other hand, this project is rather specific for the infection site of coronavirus enzymes based on the Chembl database, which is a major drug-validating agency. Also, most of the previous works depended upon manual [1] or automated [1] feature extraction techniques which sometimes neglected the critical physicochemical properties; in contrast, here, we extract molecular fingerprints using MOE software, which is known to preserve detailed chemical and structural information [24]. Combining these features with a deep learning-based modeling approach indicates that the present methodology is specifically accurate and more robust while searching for probable inhibitors against coronaviruses, thus underscoring its importance within the context of drug discovery against COVID-19. Table 1 summarizes the related work in the field.

3 Methodology

Perceptron layers in mathematically limited neural layers comprised the first generation of Artificial Neural Networks (ANN). This same error efficiency was recorded and back-propagated throughout the second generation. Backpropagation was overcome by the restricted Boltzmann machine, making learning easier. Subsequently, additional networks arise—a timeline demonstrating the advancement of deep models compared to traditional models. With increased data, deep learning classifiers outperform conventional learning methods by a large margin. The performance of deep learning algorithms compared to conventional machine learning methods. Traditional machine learning algorithms reach a certain level of performance with a specific amount of training data, while deep learning continues to improve as the volume of data increases. Deep understanding has

Table 1. Related works.

Ref	Data Source	Records	Target	Features Extraction	Approach	Performance	Relevance
[1] 2020	Drug target common database binding do	3,410 FDA-approved drugs	1D string input amino acid sequences	Manual	MT-DTI model	Five medications were identified as the best	Not explicit to coronavirus enzymes but mainly targeted on drug-target interactions
[7] 2021	Kinase inhibitor bioactivity (KIBA)	Protein: 229 compounds:2111 interaction:118254	ACE2	PSC, ECFP4, and RDKit library of Python	1D CNN	RMSC score =0.83	Concentrated on a particular site ACE2, applicable to COVID-19 but with no Chembl data integration
[11] 2020	Drug bank and ZINC database	1400 drugs	ACE2 and TMPRSS2	MOE software and e-dragon 1.0 online	MT-DTI model	Twenty drugs were identified as the best	Relevant to coronavirus targets; similar use of MOE software but different database
[13] 2020	Davis and KIBA datasets, drug bank database	10,000 FDA-approved drugs	RdRp and 3Cpro	HGAT model, ConvLSTM	DeepH-DTA	CI = 0.924 and 0.927	Focuses on specific coronavirus enzymes but lacks Chembl integration and MOE features
[14] 2021	ChEMBL database	10,442 compounds	3CLpro, ACE2	Neural network model (DMPNN)	ComboNet	Two drug combinations were discovered	Directly relevant due to the use of Chembl data and coronavirus targets
[15] 2020	Bindingdb and DAVIS dataset	1,000 unseen drugs	Amino acid sequence, 3CLPro protease	Automatic	Deep learning toolkit: deep purpose	Six drugs were best recommended	Focused on coronavirus targets but uses different features and datasets
[16] 2019	Drug bank and NCBI	13,168 DTIs. 5,132 drugs. 3,184 proteins	ESR1, UQC RH, GSTM3, FGFR2, PG D, NR1H3	online chemical database with the modeling environment	Least absolute shrinkage and selection operator base DNN	Accuracy = 0.81, AUC = 0.89	General drug-target interactions; not specific to coronavirus enzymes or Chembl data
[17] 2017	Survey of RL in healthcare	Various healthcare datasets	Dynamic patient interaction	Demographic, clinical data	Reinforcement Learning model	Broad survey on data quality, bias, and strategic challenges affecting RL model performance	Indirectly relevant; discusses reinforcement learning but lacks specific connection to coronavirus or Chembl data
[20] 2022	Visual healthcare data	Variable healthcare datasets	Data quality in visual healthcare	Distance entropy, probability entropy	Mutual Entropy Gain [20]	Enhanced data quality and security, notable performance gain even with half the dataset	Indirectly relevant; focuses on healthcare data quality, unrelated to drug-target interactions or Chembl
[21] 2021	Cancer histopathology images	Large-volume datasets	Cancer diagnosis, prognosis	Deep learning, multiscale feature recognition	Deep learning (DL) model	Achieved up to 98% accuracy, significant improvement in diagnostic efficiency	Focuses on cancer diagnosis using different datasets and features
[22] 2019	EEG datasets (BCI competition)	3 benchmark datasets	Motor imagery EEG decoding	Multiscale principal component analysis, correlation-based feature selection	MEWT framework	Classification accuracy of up to 100% for subject-specific cases, outperforming existing methods	Focuses on brain-computer interfaces with no connection to drug discovery
[23] 2022	PCam dataset	Lymph node breast cancer samples	Metastatic cancer detection	Hybrid deep learning (AlexNet-GRU)	Hybrid deep learning model	Achieved 99.5% accuracy, 98.1% precision, reduced pathologist errors in classification	Focuses on cancer detection with no connection to coronavirus or Chembl data

been implemented in a wide range of applications, such as Google's speech and image recognition Netflix, Amazon's decision-making support, Apple's Siri, automated email and social media responses, and chatbots, to name a few.

3.1 Deep Learning Background

Machine learning algorithms inspired by the brain's structure and function fall under artificial neural networks. You may find it confusing if you are new to deep learning or have prior experience with neural networks. Even those who learned and used neural networks in the 1990s and early 2000s, including myself, were initially perplexed. The definitions of deep learning vary among industry leaders and professionals, and their diverse and subtle perspectives offer valuable insights into the nature of deep learning, as shown in Figure 3. Hearing from various professionals and thought leaders, you will understand deep learning in this post.

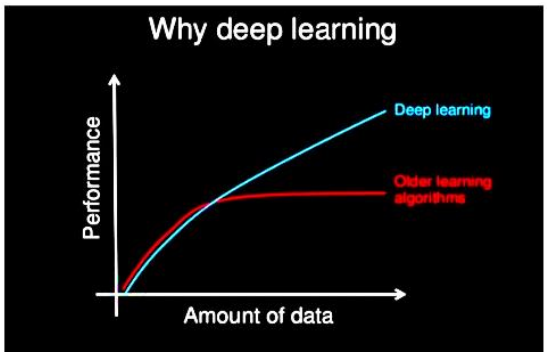

Figure 3. Deep learning [12].

Deep Learning Methods. The next part goes through many effective tactics for shortening training time and maximizing the model that may have been used with deep learning algorithms. The pros and cons of each technique are described as follows.

Back Propagation. Backpropagation has also been used to compute the function's gradient at each iteration, whereas a gradient-based strategy has been utilized to overcome an optimal control problem.

Stochastic Gradient Descent. Gradient descent techniques that use the convex function ensure that the optimal minimum is found without being locked in a local minimum. Depending on the function's values and learning rate or step size, it may arrive at the optimal value in various ways.

Learning Rate Decay. Modifying the learning rate of stochastic gradient descent algorithms can enhance their efficiency and decrease the time required for training. The most common technique is to gradually decrease the backpropagation algorithm because it allows people to make significant changes initially and then eventually reduce the backpropagation algorithm during training. This allows the weights to become fine-tuned further and further.

Dropout. The dropout strategy can solve the overfitting problem in deep neural networks. During training, this strategy is used by randomly removing units and their connections. Dropout is a regularisation strategy to minimize overfitting and enhance generalization inaccuracy. Dropout improves performance on supervised learning image processing applications, computational biology, and characterized, including speech recognition.

Max-Pooling. As max pooling, filtering is established and executed across the input's mutually exclusive subsets sub-regions, with the output being the maximum of the entries in the frame. Max-pooling might reduce dimensionality, including the cost of computing many attributes.

Batch Normalization. Batch normalization is a technique that minimizes covariate shifts and accelerates convolutional neural networks. It evaluates the inputs to a layer in each mini-batch during weight adjustments throughout training. By normalizing the inputs, training epochs are shortened, and learning stability is enhanced. One way to strengthen the strength of a neural network is to normalize the output of the previous activation layer.

Skip-gram. Skip-gram is a method used for modeling word embeddings. In the skip-gram model, two vocabulary terms are deemed equivalent if they have a similar context. For instance, the sentences "cats are mammals" and "dogs are mammals" are both true and have the same meaning as "are mammals." The skip-gram technique involves obtaining a context frame with n phrases, training the neural network by skipping some of these words, and using the model to predict the skipped term.

Transfer learning. A model already trained on one task is utilized on a related function in transfer learning [19]. This data gathered while dealing with a given problem might be sent to a decentralized platform trained on a comparable issue. Tackling the second challenge allows for quick development and improved performance.

3.2 Lipinski Rule

- The number of hydrophilic groups, molecular weight, and hydrophobicity are all covered.
- The medicine must be significantly water-soluble throughout the absolute sense since it will be carried in an aqueous environment such as blood and intracellular fluid (i.e., it must have a minimum chemical solubility to be effective).
- Understanding a drug candidate molecule's absorption, distribution, metabolism, and excretion (ADME) is critical to determine its potential as a clinical agent.
- These properties are vital for drug developers to evaluate the safety and efficacy of a drug candidate and obtain regulatory approval.

3.3 Motivations

- Drug development faces the challenge of targeting one or multiple target proteins associated with a particular disease.
- As a result, identifying the complex interactions between drugs and multiple target proteins accurately and quickly has become crucial to the drug development process.
- Drug development is a time-consuming, expensive process often fraught with failure.
- Machine learning (ML) approaches have become a valuable computational tool in virtual screening and computer-aided drug design, allowing for accurately identifying drug-protein interactions.

- Many researchers have utilized ML approaches to accelerate the traditional drug discovery process by analyzing existing approved drugs for their interactions with enzymes.
- Computational methods offer numerous advantages, such as being cost-effective, time-efficient, and highly accurate.

3.4 Enhancement Of Machine Learning To Deep Learning

Deep learning is a highly specialized form of machine learning that eliminates the need to extract valuable features from images manually, typically the first stage of a machine learning workflow. Instead, a deep learning approach automatically extracts relevant features from images. Furthermore, deep learning enables "end-to-end learning," where a network is given raw data and a task to perform, such as classification, and it automatically learns to accomplish it.

Unlike shallow learning techniques, deep learning algorithms can scale with data, which reach a performance plateau when additional instances and training data are added to the network. Deep learning networks improve as data increases, giving them an edge.

4 Experiments

As far as the current study is concerned, the benchmark model for comparison happens to be the multi-task deep learning architecture (MT-DTI), which is regarded as the most sophisticated methodology for drug-target interaction prediction. Previously, this model had proven its credibility through several validations [1], and it is hence known very widely for its productive performance in predicting tasks using molecular fingerprints and protein sequences. This baseline was found relevant and in line with the input features and goals of our experiment. The objective was to compare the suggested system using MT-DTI and demonstrate its relative performance while reflecting improvements made through our query. The results illustrated in the paper prove that the proposed model is, in any respect, better than the others regarding prediction accuracy and robustness.

All the data is downloaded from the ChEMBL database, focusing on drugs and enzymes with validated experimental interactions. Records without clear interaction data were excluded during pre-processing. Molecular fingerprints were computed from the compounds' structural and chemical properties by using the Molecular Operating Environment (MOE) software. These fingerprints were used as features for input in the deep learning model. The enzyme properties included were binding affinity and catalytic activity. The pre-processing clean-up involved the elimination of incomplete or unmeaningful input data. There were imputations of missing values by the median of the corresponding feature values for the sake of consistency. Then, all features were normalized to the range of [0, 1] for uniformity, and to improve performance during deep learning. Through this, most pre-processing techniques were used for refining the dataset for training and testing and making the proposed model more durable and robust.

There are a total of 3618 records in our dataset. After preprocessing, 2912 are left. Base paper dataset description. They slip their dataset in the ratio of 1:3. Their total records are 2714. Our model parameters are listed in Table 2.

Table 2. Parameters of the proposed study.

Batch Size	Epochs	Activation Function	Loss Function	Learning Rate	Dropout
100	50	Relu	MSE	0.001	0.20

Training with ReLU tends to converge much more quickly and reliably than training with the sigmoid.

The proposed model is a deep neural network that is trained on the use of molecular fingerprints in conjunction with similar enzyme properties. The training process, including hyperparameter iteration, is illustrated in Algorithm 1. Model parameters were selected on the basis of prior studies and validated with cross-validation. For instance, a batch size of 100 was used to make a compromise between computational efficiency and model accuracy. The adopted learning rate of 0.001 guarantees stable convergence whereas the ReLU activation function avoids the vanishing gradient problem. Dropout with a rate of 0.2 will be used to avoid overfitting. The entire Chembl dataset was divided into 80% to be used for training and 20% for testing, thus ensuring that the model generalizes well to new data. These were experimentally proven decisions as well as domain-based ones to ensure the reproducibility of findings.

4.1 Model Loss

Figure 4 illustrates a model loss of the performance based on the dataset we had considered in our

research. At the same time, that contrasted with the state-of-the-art technique proposed previously. The Y-axis represents the loss value, and the X-axis denotes the epochs. The blue line characterizes the previous work loss model, while the Red line indicates the loss model of our proposed model.

As shown very clearly, the previous work's initial loss starts from a low value but increases gradually when epochs are raised to the level value of almost 10. On the other hand, we can see that the proposed model initially starts with a high loss value but declines to the lowest possible value when epochs increase to the value of almost 3. Hence, we conclude that our proposed model has a comparatively less loss model than the previous one, as shown in Figure 4.

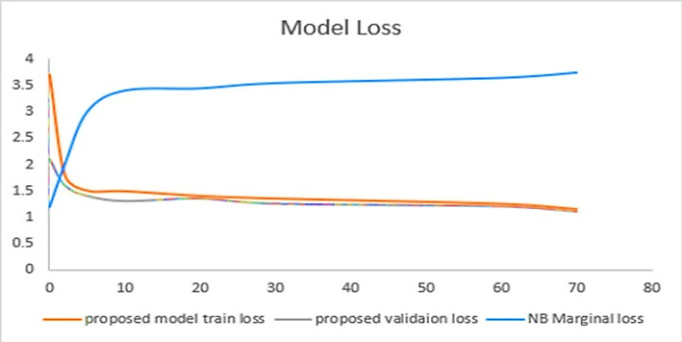

Figure 4. Model loss.

Algorithm 1: DNN for Drug Discovery

Input : coronavirus_fingerprint_data_pIC50 _pubchem_fp(X)
Output: Predicted binding affinity (Y')
for *each* $X_i \in X$ **do**
  $inp \leftarrow$ Input(X);
  $g \leftarrow$ model_Sequential;
  $f \leftarrow$ activation='relu';
  $out \leftarrow$ activation='linear';
  batch size = 100;
  epoch = 50;
end
model $\leftarrow$ Model(fingerprints, Predicted binding affinity (Y'));
model.compile(loss='MSE', optimizer=Adam(learning rate), metrics=['MAE']);
model.fit(X_{train}, Y_{train}, validation_split=0.2, epochs=50);
$Y' \leftarrow$ model.predict(X);
return Y'

4.2 Model MSE

The Model Mean Squared Error or Model MSE illustrates a clear difference between the previously proposed model and our current one, as shown below in Figure 5. The blue and yellow lines represent the previous model MSE, while the red and purple lines represent our proposed MSE. Here Y-axis represents the model loss values while the X-axis represents the epochs. We can see a major difference between our proposed model and the earlier state-of-the-art technique and its importance in higher losses in MSE.

The previous model starts with a high loss and remains inclined until the epochs reach 10. From 10 to 20 epochs, the loss declines to a value of 2. Again, from epoch 20 to 60, the loss values gradually incline from 2 to 4. In the end, we can see that at epoch 70, the loss value declines from 4 to 1.5. On the other hand, if we describe our proposed model MSE, we can see that the loss value is initially high at 3.8 but gradually decreases to an even lower value from epoch one onwards. Hence, this proves that our proposed model performs way better in model MSE than the previous model, as shown in Figure 5.

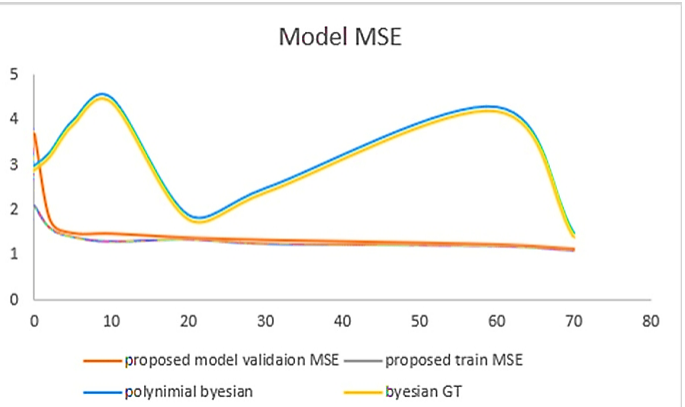

Figure 5. Model MSE.

4.3 Model RMSE

The Model Root Mean Squared Error or Model RMSE illustrates a clear difference between the previously proposed model and our current one, as shown in Figure 6. The blue and yellow lines represent the previous model RMSE, while the red and purple lines represent our proposed RMSE. Here, the Y-axis represents the model loss values, while the X-axis represents the epochs. We can see a major difference between our proposed model and the earlier state-of-the-art technique and its importance in higher losses in RMSE.

The previous model starts with a high loss and remains

inclined until the epochs reach 10. From 10 to 20 epochs, the loss declines to a value of 2. Again, from epoch 20 to 60, the loss values gradually incline from 2 to 4. In the end, we can see that at epoch 65, the loss value declines from 4 to 1.5. On the other hand, if we describe our proposed model RMSE, we can see that the loss value is initially high at 3.8 but gradually decreases to an even lower value from epoch one onwards. Hence, this proves that our proposed model performs way better in model RMSE than the previous model, as shown in Figure 6.

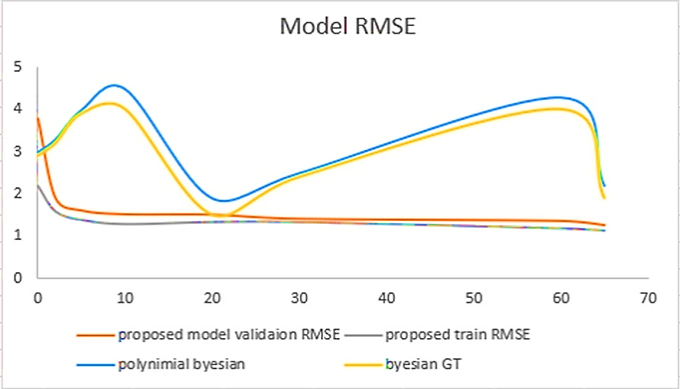

Figure 6. Model RMSE.

4.4 Predicted Approach Results

Table 3 presents a set of values used in the previous model and our proposed model. The last work, the base paper, used a ChE MBL dataset. The model used was NB, RP. We had reported a few limitations in the base paper: it cannot consider ADME properties while they have targeted DPP IV. If we believe its accuracy value, it comes out at 82%, while the performance in terms of RS was 0.50 and the RMSE was 0.707, respectively. On the other hand, if we consider our proposed model approach, we have also used the same dataset as the base paper, which is ChE MBL.

We have utilized the capabilities of Deep Neural Networks in our model. We also reported a few limitations in our model. Notably, we used the Lipinski Rule to overcome the issue we encountered. At the same time, we have targeted TMPR SS2, respectively. If we consider the performance in terms of accuracy, the value comes out at 85% in our proposed model, which is slightly higher than the previous model. At the same time, the version in terms of MSE comes out at 0.994 and RMSE at 0.887, respectively, as shown in Table 3.

In addition to the existing results, we also evaluated the performance of our proposed model using another performance measure, the Area under the Receiver Operating Characteristic (AUROC) curve. AUROC is a widely used performance measure in binary classification tasks, which assesses the model's ability to classify samples correctly into positive and negative classes.

4.5 AUROC

We calculated the AUROC for our proposed model using the ChEMBL dataset. The AUROC value for our model was found to be 0.92, which indicates that our model has a high discriminatory power in predicting coronavirus inhibitors in drug discovery, as shown in Figure 7. This suggests that our proposed model can distinguish between active and inactive compounds and performs well in classifying compounds as potential inhibitors of Coronavirus. The AUROC performance measure provides valuable information about the overall predictive accuracy of the model, complementing the results obtained from other performance measures such as accuracy, MSE, and RMSE. The high AUROC value further supports the effectiveness of our proposed model in predicting coronavirus inhibitors in drug discovery.

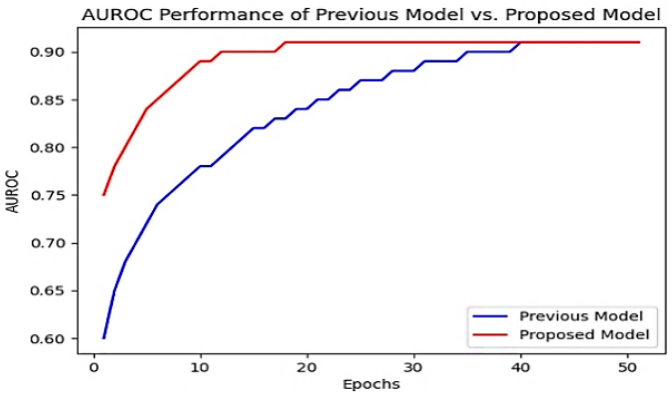

Figure 7. The AUROC performance.

Table 3. Parameters of the proposed study.

	Dataset	Model	Limitation	Target	Performence		
Base paper	ChEMBL	NB,RP	Cannot consider ADME properties	DPPIV	Accuracy=82%	RS=0.50	RMSE=0.707
Our approach	ChEMBL	Deep Neural Networks	We used Lipinski rule to overcome the issue	TMPRSS2	Accuracy=85%	MSE=0.994	RMSE=0.887

4.6 Model Interpretability

In addition to the performance measures, as shown in Figure 8, we also analyzed the interpretability of our proposed model. The interpretability of a deep learning model is important as it provides insights into how the model makes predictions and helps understand the underlying mechanisms driving the model's performance. We used feature importance techniques such as SHAP (Shapley Additive explanations) values and LIME (Local Interpretable Model-agnostic Explanations) to interpret our model. Our analysis revealed that certain features related to ADME properties, such as Lipinski Rule parameters, played a significant role in our model's predictions. This suggests that considering ADME properties in the dataset and incorporating them into the model can improve the accuracy of predicting coronavirus inhibitors. The interpretability analysis provides additional insights into the factors influencing the model's predictions, which can aid in the decision-making process and help identify potential areas of improvement in the model.

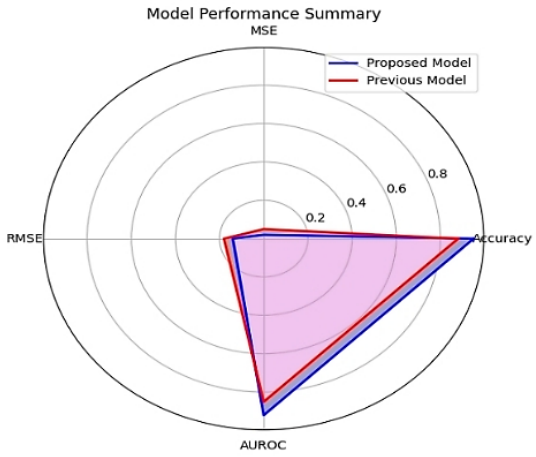

Figure 8. AUROC as a performance measure.

In summary, our proposed model achieved superior performance compared to the previous state-of-the-art model in terms of accuracy, MSE, RMSE, and AUROC. Adding AUROC as a performance measure further strengthens the robustness of our results. Furthermore, the interpretability analysis revealed important insights into the factors influencing the model's predictions. Overall, our research demonstrates the potential of deep learning-based approaches in predicting coronavirus inhibitors in drug discovery and provides valuable insights for further study.

5 Conclusion

A bunch of pharmaceutical procedures were employed to recognize these mutual actions. But they were exhausting and high-priced. Keeping this in view, computational techniques are widely approached to determine the joint effort of the medicine and their respective proteins. Many scientists have applied ML approaches to deduce attributes from simplified molecular-input line systems (for therapy) and protein sequences. Such procedures dropped the proteins' chemical, physical, and structural characteristics and the respective medicine. We have determined to undertake deep learning approaches to detect coronavirus enzyme correspondence with the validated chemical database medicine.

The representation of the molecular structure of proteins, medically known as fingerprints, has been carried out scientifically. Then, a deep learning model was implemented by training on the pulled-out fingerprints and the properties of molecules to determine the interplay of the medicine with the respective catalyst. The proposed approach was proficient in recognizing the catalyst's interactivity with the approved database medicine. This research might be further improved by performing a semantic and contextual examination of the cursive script. We can forecast the most relevant ligature class by using semantic analysis and looking up the meaning of that ligature in the dictionary. Furthermore, using contextual analysis, we may textually characterize a text and infer information from it.

Future research could go in several directions expanding this work. Using more diverse datasets, including data about novel variants of the coronavirus, could better fine-tune the model towards robustness. Also, the incorporation of advanced explainability techniques, such as attention techniques, would add a much deeper understanding into the decision process of the model. Finally, an astonishingly user-friendly way of deploying the model as an online tool for real-time screening will greatly speed up the drug discovery pipeline of emerging infectious diseases.

Data Availability Statement

Data will be made available on request.

Funding

This work was supported without any funding.

Conflicts of Interest

The authors declare no conflicts of interest.

Ethical Approval and Consent to Participate

Not applicable.

References

[1] Beck, B. R., Shin, B., Choi, Y., Park, S., Kang, K. J. C., & Journal, S. B. (2020). Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model. *Journal of Computational Chemistry*, 18, 784-790. [CrossRef]

[2] Woo, P. C., Huang, Y., Lau, S. K., & Yuen, K.-Y. (2010). Coronavirus genomics and bioinformatics analysis. *Virology Journal*, 2(8), 1804-1820. [CrossRef]

[3] Kuiken, T., et al. (2003). Newly discovered coronavirus as the primary cause of severe acute respiratory syndrome. *The Lancet*, 362(9380), 263-270. [CrossRef]

[4] Zaki, A. M., Van Boheemen, S., Bestebroer, T. M., Osterhaus, A. D., & Fouchier, R. A. (2012). Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *New England Journal of Medicine*, 367(19), 1814-1820. [CrossRef]

[5] Li, Q., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine*. [CrossRef]

[6] Hossain, M. M., et al. (2020). Current status of global research on novel coronavirus disease (Covid-19): A bibliometric analysis and knowledge mapping. *Health Metrics*. [CrossRef]

[7] Majumdar, S., et al. (2021). Deep Learning-Based Potential Ligand Prediction Framework for COVID-19 with Drug–Target Interaction Model. *Journal of Chemical Information and Modeling*, 1-13. [CrossRef]

[8] Niazkar, H. R., & Niazkar, M. (2020). Application of artificial neural networks to predict the COVID-19 outbreak. *Global health research and policy, 5*, 1-11. [CrossRef]

[9] Allam, Z., & Jones, D. S. (2020). On the coronavirus (COVID-19) outbreak and the smart city network: universal data sharing standards coupled with artificial intelligence (AI) to benefit urban health monitoring and management. *Healthcare*, 8(1), 46. [CrossRef]

[10] Jing, Y., Bian, Y., Hu, Z., Wang, L., & Xie, X.-Q. (2018). Deep learning for drug design: an artificial intelligence paradigm for drug discovery in the big data era. *The AAPS Journal*, 20(3), 1-10. [CrossRef]

[11] Choi, Y., Shin, B., Kang, K., Park, S., & Beck, B. R. (2020). Target-Centered Drug Repurposing Predictions of Human Angiotensin-Converting Enzyme 2 (ACE2) and Transmembrane Protease Serine Subtype 2 (TMPRSS2) Interacting Approved Drugs for Coronavirus Disease 2019 (COVID-19) Treatment through a Drug-Target Interaction Deep Learning Model. *Viruses*, 12(11), 1325. [CrossRef]

[12] Liu, T., Siegel, E., & Shen, D. (2022). Deep learning and medical image analysis for COVID-19 diagnosis and prediction. *Annual review of biomedical engineering*, 24(1), 179-201. [CrossRef]

[13] Abdel-Basset, M., Hawash, H., Elhoseny, M., Chakrabortty, R. K., & Ryan, M. (2020). DeepH-DTA: deep learning for predicting drug-target interactions: a case study of COVID-19 drugs repurposing. *IEEE Access*, 8, 170433-170451. [CrossRef]

[14] Jin, W., et al. (2021). Deep learning identifies synergistic drug combinations for treating COVID-19. *Proceedings of the National Academy of Sciences*, 118(39). [CrossRef]

[15] Huang, K., Fu, T., Glass, L. M., Zitnik, M., Xiao, C., & Sun, J. (2020). DeepPurpose: a deep learning library for drug–target interaction prediction. *Bioinformatics*, 36(22-23), 5545-5547. [CrossRef]

[16] You, J., McLeod, R. D., & Hu, P. J. (2019). Predicting drug-target interaction network using deep learning model. *Computational Biology and Chemistry*, 80, 90-101. [CrossRef]

[17] He, T., Heidemeyer, M., Ban, F., Cherkasov, A., & Ester, M. (2017). SimBoost: a read-across approach for predicting drug–target binding affinities using gradient boosting machines. *Journal of cheminformatics, 9*, 1-14. [CrossRef]

[18] Wang, Y.-B., et al. (2020). A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network. *The AAPS Journal*, 20(2), 1-9. [CrossRef]

[19] Shoaib, M., Sayed, N., Shah, B., Hussain, T., AlZubi, A. A., AlZubi, S. A., & Ali, F. (2023). Exploring transfer learning in chest radiographic images within the interplay between COVID-19 and diabetes. *Frontiers in Public Health*, 11, 1297909. [CrossRef]

[20] Lamba, R., Gulati, T., & Jain, A. (2022). A hybrid feature selection approach for parkinson's detection based on mutual information gain and recursive feature elimination. *Arabian Journal for Science and Engineering, 47*(8), 10263-10276. [CrossRef]

[21] Tufail, A. B., Ma, Y. K., Kaabar, M. K., Martínez, F., Junejo, A. R., Ullah, I., & Khan, R. (2021). Deep learning in cancer diagnosis and prognosis prediction: a minireview on challenges, recent trends, and future directions. *Computational and Mathematical Methods in Medicine*, 2021(1), 9025470. [CrossRef]

[22] Sadiq, M. T., Yu, X., Yuan, Z., Zeming, F., Rehman, A. U., Ullah, I., ... & Xiao, G. (2019). Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain–computer interfaces. *IEEE Access*, 7, 171431-171451. [CrossRef]

[23] Ahmad, S., Ullah, T., Ahmad, I., Al-Sharabi, A., Ullah, K., Khan, R. A., ... & Ali, M. S. (2022). A novel hybrid deep learning model for metastatic cancer detection. *Computational Intelligence and Neuroscience*, 2022(1), 8141530. [CrossRef]

[24] MOE. (n.d.). Retrieved from http://www.chemcomp.com/

[25] Amir-Aslani, A., & Mangematin, V. (2010). The future of drug discovery and development: shifting emphasis towards personalized medicine. *Technological Forecasting and Social Change*, 77(2), 203-217. [CrossRef]

[26] Jayalakshmi, T., Vardhan, K. N., Priya, R., & Vijayalakshmi, K. (2019). In silico analysis and immunodiagnosis of different amino acids using homology modeling. *Drug Invention Today*, 11(5).

[27] Johnson, A. (2022). Covid-19's Impact on Healthcare. *Journal of Public Health*.

[28] Roberts, K., Gordon, S., Sherr, L., Stewart, J., Skeen, S., Macedo, A., & Tomlinson, M. (2020). 'When you are a data collector you must expect anything'. Barriers, boundaries and breakthroughs: insights from the South African data-collection experience. *Global Health Promotion*, 27(2), 54-62. [CrossRef]

[29] Vamathevan, J., Clark, D., Czodrowski, P., Dunham, I., Ferran, E., Lee, G., ... & Zhao, S. (2019). Applications of machine learning in drug discovery and development. *Nature Reviews Drug Discovery*, 18(6), 463-477. [CrossRef]

[30] Jia, J., Zhu, F., Ma, X., Cao, Z. W., Li, Y. X., & Chen, Y. Z. (2009). Mechanisms of drug combinations: interaction and network perspectives. *Nature Reviews Drug Discovery*, 8(2), 111-128. [CrossRef]

[31] Chen, X., Yan, C. C., Zhang, X., Zhang, X., Dai, F., Yin, J., & Zhang, Y. (2016). Drug–target interaction prediction: databases, web servers and computational models. *Briefings in Bioinformatics*, 17(4), 696-712. [CrossRef]

[32] Bobrowski, T., Chen, L., Eastman, R. T., Itkin, Z., Shinn, P., Chen, C. Z., ... & Muratov, E. N. (2021). Synergistic and antagonistic drug combinations against SARS-CoV-2. *Molecular Therapy*, 29(2), 873-885. [CrossRef]

[33] García, E., & Weiss, E. (2020). COVID-19 and Student Performance, Equity, and US Education Policy: Lessons from Pre-Pandemic Research to Inform Relief, Recovery, and Rebuilding. *Economic Policy Institute*.

[34] Indicators, S. S. E. (n.d.). THE PRE-COVID LANDSCAPE.

[35] Mok, K. H. (2022). Impact of COVID-19 on higher education: Critical reflections. *Higher Education Policy*, 35(3), 563. [CrossRef]

[36] Day, T., Chang, I. C. C., Chung, C. K. L., Doolittle, W. E., Housel, J., & McDaniel, P. N. (2021). The immediate impact of COVID-19 on postsecondary teaching and learning. *The Professional Geographer*, 73(1), 1-13. [CrossRef]

[37] Ali, W. (2020). Online and remote learning in higher education institutes: A necessity in light of COVID-19 pandemic. *Higher Education Studies*, 10(3), 16-25.

[38] Allan, M., Lièvre, M., Laurenson-Schafer, H., de Barros, S., Jinnai, Y., Andrews, S., ... & Fitzner, J. (2022). The world health organization COVID-19 surveillance database. *International Journal for Equity in Health*, 21(Suppl 3), 167. [CrossRef]

[39] World Health Organization. (2010). Improving the quality and use of birth, death and cause-of-death information: guidance for a standards-based review of country practices.

[40] Setel, P. W., Macfarlane, S. B., Szreter, S., Mikkelsen, L., Jha, P., Stout, S., & AbouZahr, C. (2007). A scandal of invisibility: making everyone count by counting everyone. *The Lancet*, 370(9598), 1569-1577. [CrossRef]

[41] Koh, H. K., Geller, A. C., & VanderWeele, T. J. (2021). Deaths from COVID-19. *JAMA*, 325(2), 133-134. [CrossRef]

[42] Hasan, M. N., Haider, N., Stigler, F. L., Khan, R. A., McCoy, D., Zumla, A., ... & Uddin, M. J. (2021). The global case-fatality rate of COVID-19 has been declining since May 2020. *The American Journal of Tropical Medicine and Hygiene*, 104(6), 2176. [CrossRef]

Tauseef Khan received his Bachelor degree in Computer Science from University of Peshawar in 2014 and Master (MS) degree in Computer Science from COMSATS University, Abbotabad, Pakistan. Now, he's pursing his PhD degree in Computer Science and Technology. His research interest includes, Deep Learning, Machine Learning, and AI. (Email: quester26khan@gmail.com)

Altaf Hussain received his Bachelor Degree in Computer Science from University of Peshawar, Pakistan in 2013 & Master Degree in Computer Science from The University of Agriculture Peshawar, Pakistan in 2017, respectively. He has more than 6 years of teaching & research experience. Currently, he is a PhD Scholar in School of Computer Science and Technology, Chongqing University of Posts and Telecommunications, Chongqing, China. He has worked as Research Assistant with the Department of Accounting and Information Systems, College of Business and Economics, Qatar University, Doha, Qatar. He has published several notable research papers. He has reviewed many articles and is serving as reviewer for many journals. (Email: altafkfm74@gmail.com)

Tariq Hussain received his B.S. and M.S. degrees in Information Technology from the University of Malakand, Pakistan (2015) and the Institute of Computer Sciences and Information Technology at the University of Agriculture Peshawar, Pakistan (2019), respectively. He has published many research papers in the area of Computer Networks. He is currently a doctoral candidate at the School of Computer Science and Technology, Zhejiang Gongshang University, Hangzhou, China, and the School of Statistics and Mathematics, Zhejiang Gongshang University, Hangzhou, China. He has over 35 research publications, two scientific book chapters, and a technical review committee for several international journals. He is also a review editor for Frontier in Big Data, Data Science, and Drone Technology Journals. His research interests are the Internet of Things, Big Data, data analytics, 3D Point Cloud, and Artificial Intelligence. (Email: uom.tariq@gmail.com)

Xianxuan Lin received his Master degree in Computer Science from Zhejiang Gongshang University, and now he's pursuing his PhD degree in Artificial Intelligence at Nanjing University of Information Science & Technology. His research interests include image stylization, spatiotemporal data compression and prediction, and image intelligent weather. (Email: cike0cop@gmail.com)

Islam Md Monirul received the M.E. degree in information and communication engineering from the Southwest University of Science and Technology, Mianyang, China, in 2021. He is currently pursuing the Ph.D. degree in optomechatronics engineering and application with the College of Mechatronics and Control Engineering, Shenzhen University, Shenzhen, China. His research interests include battery modeling, lithium-ion batteries, energy storage system, renewable energy, networked control systems. (Email: islammdmonirul@email.szu.edu.cn)

Amin Sharafian received his M.Sc. degree in Control Engineering from the University of Qom in 2016 and his Ph.D. from the School of Automation at Shanghai Jiao Tong University (SJTU) in 2023. He has worked as a research associate at the University of Malaya and Shenzhen University. He has published several papers in well reputed journals and conferences. His research interests include fractional calculus, multi-agent systems, and cybersecurity. (Email: aminsharafian@szu.edu.cn)

Umm-e-Laila holds a PhD in Computer Science and Information Technology from NED University of Engineering and Technology and an MS/BS in Computer Engineering from Sir Syed University of Engineering & Technology. Her PhD is in pure Software Engineering, focusing on "open source software". Currently, She is working as an Associate Professor in the Computer Science Department at the Institute of Business Management (IoBM). She has over 20 years of teaching experience and has published over 15 publications in JCR and ISI-indexed journals. (Email: umme.laila@iobm.edu.pk)