DeepH-DTA: Deep Learning for Predicting Drug-Target Interactions: A Case Study of COVID-19 Drug Repurposing

MOHAMED ABDEL-BASSET[1], **HOSSAM HAWASH**[1],
MOHAMED ELHOSENY[2,3], **(Senior Member, IEEE)**,
RIPON K. CHAKRABORTTY[4], **(Member, IEEE)**,
AND MICHAEL RYAN[4], **(Senior Member, IEEE)**

[1]Faculty of Computers and Informatics, Zagazig University, Zagazig 44519, Egypt
[2]Department of Computer Science, College of Computer Information Technology, American University in the Emirates, Dubai 503000, United Arab Emirates
[3]Faculty of Computers and Information, Mansoura University, Mansoura 35516, Egypt
[4]Capability Systems Centre, School of Engineering and IT, University of New South Wales Canberra, Canberra, ACT 2612, Australia

Corresponding author: Mohamed Abdel-Basset (analyst_mohamed@zu.edu.eg)

ABSTRACT The rapid spread of novel coronavirus pneumonia (COVID-19) has led to a dramatically increased mortality rate worldwide. Despite many efforts, the rapid development of an effective vaccine for this novel virus will take considerable time and relies on the identification of drug-target (DT) interactions utilizing commercially available medication to identify potential inhibitors. Motivated by this, we propose a new framework, called DeepH-DTA, for predicting DT binding affinities for heterogeneous drugs. We propose a heterogeneous graph attention (HGAT) model to learn topological information of compound molecules and bidirectional ConvLSTM layers for modeling spatio-sequential information in simplified molecular-input line-entry system (SMILES) sequences of drug data. For protein sequences, we propose a squeezed-excited dense convolutional network for learning hidden representations within amino acid sequences; while utilizing advanced embedding techniques for encoding both kinds of input sequences. The performance of DeepH-DTA is evaluated through extensive experiments against cutting-edge approaches utilising two public datasets (Davis, and KIBA) which comprise eclectic samples of the kinase protein family and the pertinent inhibitors. DeepH-DTA attains the highest Concordance Index (CI) of 0.924 and 0.927 and also achieved a mean square error (MSE) of 0.195 and 0.111 on the Davis and KIBA datasets respectively. Moreover, a study using FDA-approved drugs from the Drug Bank database is performed using DeepH-DTA to predict the affinity scores of drugs against SARS-CoV-2 amino acid sequences, and the results show that that the model can predict some of the SARS-Cov-2 inhibitors that have been recently approved in many clinical studies.

INDEX TERMS Deep learning, drug-target interaction, SARS-CoV-2.

I. INTRODUCTION

One of the preliminary stages of drug discovery is the determination of innovative candidate drug compounds that interact with particular target proteins. Through in vivo and in vitro studies, several high-throughput experiments have been conducted to identify the novel compounds with the anticipated interactive characteristics [1]. However, expensive costs and chronological order requirements make it impracticable to scan immense volumes of targets and mixtures. Consequently, the identification of novel drugs takes an extraordinarily long time [2].

At present, the compound database (i.e., PubChem, ChEMBL) contains over 105 million compound candidates, and more than 250 million bioactivities in both data sets (combined) [3], [4]. On the other hand, the recent number of FDA-approved drugs is about 10000, according to Drug-Bank [5]. Additionally, only a small number of proteins in the human proteome are targeted by recognized drugs. According to current statistics, knowledge of the drug–target (DT) space is still incomplete and requires a novel approach to enable broader investigation [8].

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

A. RESEARCH MOTIVATION

Recognizing drug-target interactions (DTI) is a critical phase in the process of discovering and developing new drugs that enable the repurposing of prevailing drugs and singles out the novel interactive partners for approved drugs. Consequently, DTI has attracted much research attention.

Until recently, the task of modeling DTI has been addressed as a binary classification problem ignoring a vitally significant section of characteristics regarding protein-ligand interactions, specifically the binding affinity scores which represent interactivity strength between DT pairs. Such scores are regularly quantified with measures such as half-maximal inhibitory concentration (IC50) which relies on the attentiveness of the ligand and target, dissociation constant (K_d), and inhibition constant (K_i) [6]. Lower values of IC50, K_i, and K_d indicate strong binding affinity. K_d and K_i values are typically used to compute the negative logarithm of the dissociation or inhibition constants and denoted as pK_d or pK_i [7], [29]. In DTI binary classification studies, dataset construction is a significant stage, since the selection of the non-binding instances directly influences the performance of the model [6], [8], [10]. Recently, four datasets have been widely used in several DTI studies in which pairs of DTs with unknown binding evidence are considered as non-binding instances. Recently, DTI studies that depend on affinity information databases have offered a new representative binary dataset formed with a selected binding affinity threshold value.

As explained in [7], [11], [12], [27], [30], treating DTI as a binary classification problem has two main drawbacks. First, the true-negative interactions and unidentified scores that are not discriminated against. Second, the binary associations are broadly known to be very intelligible, while it is more instructive to harness a continuous value that estimates the binding strength between a drug molecule and the target sequences which is articulated in terms of the beforementioned measures. Accordingly, researchers have been motivated to address the DTI as a regression problem. This, in turn, offers a number of advantages. First, it avoids the impact of selecting a negative sample on the deep learning approach and can deliver additional applied and valuable information [27]. Second, it allows the development of more accurate models, as well as the construction of a more realistic database. Third, a regression-based model has the benefit of forecasting an approximate value of the strength of the DT interaction which, in turn, can be significantly advantageous for the reduction of the enormous compound search-space in the process of drug discovery.

In early DTI systems, conventional machine-learning (ML) approaches have been utilized, such as support vector machine (SVM) and naïve Bayesian (NB) [9]. The performances of these approaches primarily depend on the surface-level features captured from drug data and protein sequences. However, adding additional shallow features does not lead to increase performance because of the probable intensification of features. Therefore, to guarantee the effective recognition of compound-target interactions, the typical procedure is to extract a vast quantity of such shallow features ignoring whether they are finally exploited for identification of interactions, and then feature selection techniques—such as Principle Component Analysis (PCA)—are adapted to form the critical DT features into a uniform vector space [6]. Such traditional learning-based DTI schemes, however, are unable to perform well for modeling complex interactions. Currently, deep learning-based DTI models have gained increased attention due to their ability to automatically learn and extract feature representations through the numerous internal hidden layers [28], [29].

As a result of remarkable performance in such applications as computer vision and speech synthesis, deep learning has become widely used in bioinformatics as well as in quantitative structure-activity relationship (QSAR) studies in drug discovery [1], making use of efficient data representations using non-linear transformations that smooth the learning process of embedded hidden patterns. A small number of studies adopted deep neural networks (DNN) for predicting DTI binary class employing various inputs of proteins and drugs. In particular, the convolutional neural network (CNN) architectures are broadly utilized for modeling DTI characteristics [7]. Despite their advantages, CNN-based approaches are inefficient in that they only capture invariant local patterns and do not capture the long-term dependencies [10]. Extracting the global information of protein sequences and drug compounds will not only improve the efficiency of DTI but will also support the detection of complex interactions. Recurrent neural network (RNN) architectures are proposed for sequential data, where the current data element state is calculated depending on the preceding one or the upcoming one. Mainly, long short-term memory (LSTM) models are talented to capture and remember longer sequences compared to simple RNN or Gated Recurrent Unit (GRUs) which make it the best choice to learn sequential dependencies within molecule sequences or amino acid sequences [6].

B. SARS-COV-2

Since late December 2019, human beings across the world continents have been subject to viral infection and mass transmission of a novel coronavirus that has caused widespread infection in birds, mammals, and humans [15]. The virus was identified as non-partitioned positive-strand ribonucleic acid (RNA) belonging to the Coronavirinae species. Scientifically, it is called severe acute respiratory syndrome coronavirus 2 (SARS-COV-2) [16]. The Coronavirinae species primarily comprises four genera: Alphacoronavirus, Betacoronavirus, Deltacoronavirus, and Gammacoronavirus [17]. The Betacoronavirus genus contains two notorious infectious coronaviruses namely: the Middle East respiratory syndrome coronavirus (MERS-CoV) [18] and severe acute respiratory syndrome coronavirus (SARS-CoV), which have recorded a lot of infection cases that exceed 10,000 cases, with death

rates of 37% and 10% respectively [19]. Such rapid infection rates lead to an urgent demand for treatment to inhibit, if not prohibit, SARS-CoV2 prevalence [17]–[19]. Unfortunately, contemporary drug development cannot accomplish this task with sufficient speed and considerable time is required to develop a new drug and deliver it to the market. Such delays leave the world facing very high death rates due to the recent uncommon pneumonia identified as coronavirus disease 2019 (COVID-19) caused by SARS-CoV-2 [20].

SARS-CoV-2 is clinically identified as single-stranded RNA that belongs to Betacoronavirus. It encompasses genes encoding 3C-like proteinase (3Cpro), 2-O-ribose methyltransferase (2OMT), RNA-dependent RNA polymerase (RdRp), nucleocapsid phosphoprotein, envelope protein, spike protein, nucleocapsid phosphoprotein, and many other proteins, depending on the obtained genome sequences of SARS-CoV-2 [21]. The standard clinical symptoms of COVID-19 include dry cough, high fever and fatigue [49]–[55]. The replication process of SARS-CoV-2 entails several phases following the host cell entrance: 1) the genomic RNA (gRNA) translated onto polyproteins; 2) transformation of polyproteins with viral 3Cpro into smaller replicase-transcriptase proteins; 3) replication of gRNA with the replicase-transcriptase complex that comprises of RdRp, helicase, 30-to-50 exonuclease, endoRNAse, and 2OMT; and 4) the viral components assembly. These replication-related proteins are considered the main targets of next-entry remediation drugs to stamp out viral replication [22]. Despite significant international efforts, there are no novel drugs or vaccines for the treatment of SARS-CoV-2. COVID-19 patients rely solely on their immune systems and any available, but less-effective, drugs.

C. MAIN CONTRIBUTIONS

In this study, we introduce a novel deep-learning-based DTI framework, called deepH-DTA, which geometrically exploits existing the topological structure of drug molecules as input features, along with the corresponding molecular fingerprints. Although several studies incorporated structural representations of molecules for predictive tasks under various settings for drug development or discovery, none investigated twofold prediction of chemical interactions between protein sequences of target and homogeneity of drug candidate compounds. To this end, we propose a novel DTI prediction framework that utilizes a HGAT [23] for efficient modeling of interactions of various targeted topological representation of drugs. Simultaneously, we introduce two layers of bidirectional ConvLSTM [24] to capture spatio-sequential characteristics of drug sequences encoded in a simplified molecular-input line-entry system (SMILES) format [25]. The spatio-sequential sequences capture both positional features of input SMILES and the long-term dependency representation within input sequences.

To address this shortcoming that LSTM is unable to capture the spatial correlation within long term sequences, which means that ConvLSTM is the best choice. We introduce an optimized and applied approach for predicting drug-target affinity with superior performance over cutting-edge studies.

Case Study: The proposed DeepH-DTA is applied to recognize the commercially available antiviral drugs that have the potential to act as a suppressor for the viral components of SARS-CoV-2. Furthermore, the aim is to enable our model to learn effectively the interactions between drugs and SARS-COV-2. We adopt a comprehensive set of commercially available antiviral drugs from different heterogeneity that could potentially hinder the reproduction cycle of SARS-CoV-2, providing guidance to scientists looking to develop an effective drug.

The remainder of the paper is structured as follows: Section 2, reviews the most relevant literature and associated works on the identification of potential inhibitors of SARS-COV-2. Illustrative explanation of the proposed frameworks, as well as construction principles, are presented in Section 3. The recommended experimental configurations, the dataset employed, and the results obtained in comparison with current studies are discussed in Section 4. Section 5 provides potential inhibitors for SARS-COV-2. Section 6 present the managerial implication of the current study. Section 7 presents some limitations of the current study. Finally, Section 6 draws some conclusions o and explains the intended future research directions.

II. RELATED WORK

A. DRUG-TARGET INTERACTIONS

Adopting deep learning for prediction drug-target affinities (DTA) has been a useful technique that plays a significant role in several vaccine discoveries and development problems. In [26], He *et al.* proposed a gradient boosting algorithm for estimating drug affinities based on the handcrafted features created from drugs and target information. Pahikkala *et al.* [27] introduced a Kronecker Regularized approach for minimizing a cost function using a similarity matrix calculated via the Pubchem clustering server. However, these approaches rely heavily on the nature of feature engineering adopted. In contrast, Tsubaki *et al.* [28] introduced a deep-learning approach for modeling DT, but did not exploit the topological structure of chemical molecules. In [29], Ozkirimli *et al.* introduced two-stream CNN for modeling DTI and predicting affinity scores; albeit employing traditional word embedding and ignoring contextual information in input sequences. In [7], Ozkirimli *et al.* introduce a methodology for predicting DT binding affinities using CNN over word representation of protein and compound sequences demonstrating that most essential binding information implanted in the protein domain. Zhao *et al.* [13] proposed a generative adversarial network (GAN) to learn beneficial patterns within labeled and unlabeled sequences and utilized convolutional regression to forecast binding affinity score. However, they did not address GAN training on a small dataset. Similarly, Zhao *et al.* [14] introduced a CNN architecture accompanied by attentional mechanisms to determine which protein sequences are more significant

for a drug and which drug SMILES sequences are more vital for a protein during its affinity estimation. However, all of these approaches transform drug compounds into a corresponding string representation that is not an effective method to characterize molecules. Utilizing such strings leads to the loss of molecular structure information, which in turn could weaken the predictive performance of the model and the operational relevance of the learned hidden space.

On the other hand, Nguyen *et al.* [11] exploited the topological drug information through graph networks, namely graph convolutional network (GCN), graph attention network (GAN), and graph isomorphism network (GIN). Similarly, Wang *et al.* [12] utilized these networks to learn structural information of drug as well as protein dipeptide frequency of word frequency encoding to predict affinity score using different graph networks. Moreover, Lin *et al.* [30] proposed a novel deep learning approach for predicting drug-target affinities. However, these approaches use conventional CNN to process protein sequences which suffer from information losses in deeper layers and also unable to exploit the dependencies between various convolutional channels. More, they did not consider exploiting spatio-sequential information within SMILES sequences. The graph networks adopted in these studies did not consider the topological heterogeneity of drug molecules, which that they fail to generalize for heterogeneity-based applications.

B. SARS-COV-2 DRUG REPURPOSING
In addressing SARS-COV-2 inhibition, Beck *et al.* [31] adopted the pre-trained deep-molecule transformer architecture introduced in [32] depending on the mechanism of self-attention for identification of potential effective antiviral inhibitor against RdRp and 3Cpro of SARS-COV-2. However, they did not address the impact of SARS-COV or MERS-COV antiviral drugs on the SARS-COV-2. In [33], Hu *et al.* proposed a two-stage deep-learning model, where the shared layers were designated for joint representation modeling and the task-specific-layers were utilized for learning the weights of the specific blocks. However, they incorporated a small number of drugs in their experiments. Additionally, Ge *et al.* [34] introduced a data-driven approach that combines machine learning and statistical analysis for mining related disease data to predict the candidate's antiviral inhibitor drug against SARS-CoV-2 by employing graph convolution model to exploit network topology data for capturing and calculating nodes' pattern information to build a heterogeneous knowledge graph. In [35], Tang *et al.* proposed a novel fragment-based drug design architecture based on a new deep Q-learning network for producing and predicting lead compounds targeting 3CLpro of SARS-CoV-2. Besides, in [36], Zhavoronkov *et al.* employed three generative chemistry approaches for generating drug-like compounds (i.e., ligand-based generation, homology modelling-based generation, and crystal-derived pocked-based) as a potential antiviral for SARS-CoV-2. Nguyen *et al.* [37] showed that the binding sites corresponding to protease inhibitor of

SARS-CoV-2 and SARS-CoV are very similar, and adopted a new deep generative network complex to learn affinity scores for pairs of drug-protein interaction to identify the optimal antiviral suppressors for COVID-19 spread. However, such a generative approach is computationally exhaustive, and the generated compounds have no clinical interpretation and require more biological investigation and optimization. Zhang *et al.* [38] adopted a full CNN for modeling protein-ligand interactions, conducting virtual screening for commercially existing drugs, and forecast the probable tripeptide lists for 3Cpro of SARS-CoV-2.

To sum up, the current studies of DTI can be divided into four groups: first, ML-based studies that often depend on feature engineering techniques; second, deep learning approaches that use the string representation of both drug and target which make them unable to capture the structure information of drug molecules; third, the graph networks used to learn drug representation within drugs from different nodes separately without addressing sequential characteristics of drug sequences. This in turn motivated us to propose a novel multi-path architecture that is able to learn both sequential and topological representation of input molecules, and also capture information from nodes and their connecting meta-paths.

III. METHODOLOGY
In this section we provide an explanation of our proposed three-channel approach called Deep heterogeneous learning framework for DTA (DeepH-DTA), and also introduce a detailed visualization of the proposed approach in Fig.1. Our model primarily comprises four major blocks: (1) the first upper module is introduced to learn protein structure representation using Dense Net augmented with SE operation; (2) simultaneously, a heterogeneous graph network is introduced to learn the topological representation of drug molecules (the middle module Fig.1); (3) the sequential characteristics of SMILES representation of input compounds is learned through bidirectional ConvLSTM architecture; and (4) the extracted representation is concatenated and fed into the output layer for affinity score calculation. The detailed explanation of the model's implementation is discussed in the following subsections.

A. LEARN TARGET SEQUENCES
In this section, amino acid sequences are encoded into embedding vector representation through the Polypeptide Frequency of Word Frequency approach [12] used for extracting protein features. Similar to bioinformatics calculation of term frequency (TF), the polypeptide frequency (F) can be computed with equation (1),

$$F = (v_1, v_2, v_3 \cdots, v_{25^n})^T \qquad (1)$$

where v_i denotes the count of the i-th feature, and n represents the numeral of 25 (as introduced in [12]) remains to exist in the polypeptide, hence 25^n diverse polymers are shaped by desiccation intensification. The calculation of v_i is expressed
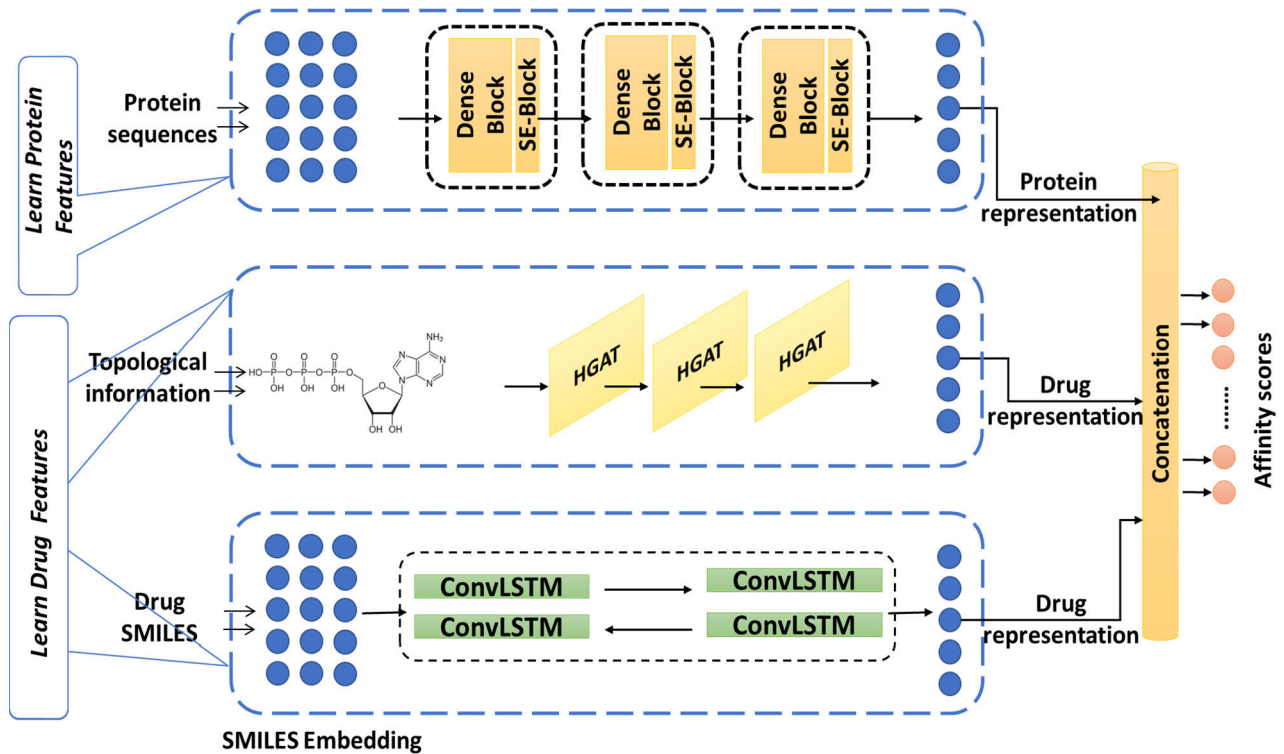
FIGURE 1. Illustration of the proposed DeepH-DTA. The upper module takes a fixed length of protein sequence as an input to produce corresponding protein information consisting of three dense blocks each followed by the SE block. The middle module processes the topological drug information to generate drug representation using improved heterogeneous GAT network. Concurrently, the lower module operates on the SMILES representations of drug to using Bidirectional ConvLSTM. Finally, the output of the three modules is concatenated to produce the final prediction score.

in equation (2),

$$v_i = \frac{n^u}{\sum_{u=1}^{25^n} n_u} = \frac{n_u}{L-1} \qquad (2)$$

where n^u denotes the times of incidence of the u-th dipeptide pattern across the protein sequence, and L is the protein sequence length. The inverse document frequency (IDF) calculated to raise the significant weight of TF is formulated in equation (3), where N represents the number of protein sequences, and W_i denotes the count of protein sequences incorporating the i-th polypeptide.

$$\text{IDF} = \log\left(\frac{N}{W_i}\right), \quad (1, 2, 3, \cdots, 25^n) \qquad (3)$$

Then the polypeptide frequency of word frequency can be calculated using equation (4),

$$\text{WF} = (\text{wf}_1, \text{wf}_2, \text{wf}_3 \cdots, \text{wf}_{25^n}). \qquad (4)$$

where wf_i represents the frequency of the i-th polypeptide of word frequency and calculated with equation (5),

$$\text{wf}_i = \frac{w_i}{N} \times \frac{p_i}{L-1} \qquad (5)$$

where p_i represents the number of occurrences of the ith peptide in the present protein sequence.

Subsequently, the concealed relationships corresponding to the polypeptide frequency of word frequency are attained through the proposed convolution model inspired by the DenseNet architecture [39], and three dense convolutional

blocks are introduced to learn protein features. In each block, the collective knowledge of preceding convolutions is used as input for the current convolution, and a simple dropout layer is added after the first and the second layer of each block to avoid overfitting. Meanwhile, dense collective learning raises the channel count, so a transition is added between dense blocks to manage model complexity and to minimize the number of channels by using the 1×1 convolutional layer.

Squeeze and Excite (SE) Block: Exploiting channel dependencies has been shown to enhance convolution model performance [40]. Thus, we attached spatial squeezing and channel excitation operations at the end of each block. Each of the F filters convolve along their receptive field, which restrains the calculated convolutional output $X = [x_1, \cdots, x_F]$ from making use of correlation information outside of this region, where $x_f \in \mathbb{R}^{W \times H}$ passed as input to squeeze and excite (SE) module to be combined using global average pooling (GAP) to produce a channel descriptor of the entire context of input channels. Hence, the spatial squeeze of the f-th channel is calculated using the spatial squeeze function F_{sq} as expressed in equation (6),

$$Z_f = F_{\text{sq}}(x_f) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{w} x_f(i, j) \qquad (6)$$

where $x_f(i, j)$ denotes the spatial position of the f-th channel with width W and height H. For clarification, the input feature map is compressed by GAP to yield Z_f.

An excitation operation is then applied to detect the channels' nonlinear interaction and also to capture a non-mutually exclusive association using two fully connected layers (FCL), where the pooled vector of features is encoded to the dimension of $1 \times 1 \times \frac{F}{r}$, and then encoded to $1 \times 1 \times F$ using a simple gating operation with a sigmoid activation as formulated in equation (7),

$$s = F_{\text{ex}}(Z, W) = \sigma(W_2 \& (W_1 Z)) \tag{7}$$

where & denotes the rectified linear unit function (Relu activation), $W_1 \in \mathbb{R}^{F \times \frac{F}{r}}$, and $W_2 \in \mathbb{R}^{F \times \frac{F}{r}}$ represents the parameters of the first FCL, respectively, and r denotes reduction threshold used for complexity reduction and ease generalization; we achieved higher results with $r = 2$. After that, a dimensionality-increasing layer is adopted in the second FCL to establish the dimension to the output's channel. The output of the SE block is generated, the output U is computed and rescaled activations as:

$$\tilde{x}_{\text{f}} = F_{\text{scale}}(x_f, s_f) = s_f \cdot x_f \tag{8}$$

where $\tilde{X}_f = [x_1, x_2 \ldots \ldots x_c]$ and $F_{\text{scale}}(x_f, s_c)$ represents the channel-wise production of feature maps $x_f \in \mathbb{R}^{H \times W}$ with the scalar value s_f.

B. LEARNING DRUG FEATURES

1) TOPOLOGICAL LEARNING

A vital indication for the estimation of DTA is to effectively exploit molecular structure information to reveal the interconnection between atoms in the drug [11], [12]. Thus, to accomplish this, we transformed the SMILES chemical molecules molecule graph representation $G = (V, E)$, using RDKit[1], where each node $v_i \in V$ denotes the i-th atom, and $e_{ij} \in E$ represents the chemical bond between the i-th and the j-th atoms. Graph attention networks (GAT) have shown their superiority for modeling graph representation in many studies [30]. However, it could be observed that the correlation between nodes in the generated compounds' heterogeneous graph can have different semantics reflected in meta-paths, owing to the complication of the heterogeneous graph where every two objects (nodes) are linked via various semantic information paths, which are called meta-paths. Hence, adopting GAT is ineffective for such heterogeneous molecular graphs, since traditional GAT performs attention at the node level only and cannot exploit meta-path semantic relations. So, it cannot preserve the graph meta-path architectural information when embedding the network into a low dimensional space; thus, the learned embeddings could be applied to other downstream tasks.

To this end, inspired by the heterogeneous graph attention network (HAN) [23], we propose hierarchical attention schemes, in which we first perform attention at node level to learn the weights of neighbors N^{Φ} of meta-path Φ, and combine them to obtain the embedding of the semantic-specific node. Then, the difference among meta-paths is computed via semantic-level attention to find the ideal weighted mixture of the semantic-level node embedding for the targeted task.

Owing to node heterogeneity, each type of node has diverse feature spaces. Therefore, for projecting features of each type of node into the same feature space, we compute the node-type transformation matrix M_{ϕ_i} as expressed in equation (9).

$$h'_i = M_{\phi_i} \cdot h_i \tag{9}$$

where h'_i and h_i respectively denote the original and projected features of node i, and ϕ_i represents node type. Then, the self-attention mechanism is adopted to capture the weights in between node pairs (i, j) with meta-path Φ. So, the relative importance of node $j \in N_i^{\Phi}$ for the node, i represents the node-level attention e_{ij}^{Φ} computed with equation (10) where att_{node} designates the neural network that accomplishes the node-level attention, which subsequently was normalized with the softmax function to obtain weight coefficient $\propto_{ij}^{\Phi}$ as formulated in equation (11),

$$e_{ij}^{\Phi} = \text{att}_{\text{node}}(h'_i, h'_j; \Phi) \tag{10}$$

$$\propto_{ij}^{\Phi} = \text{softmax}(e_{ij}^{\Phi}). \tag{11}$$

Subsequently, the embedding (meta-path) corresponding to node i can be combined by projected features of the neighbors with the respective coefficients as depicted in equation (12),

$$Z_i^{\Phi} = \sigma \sum_{j \in N_i^{\Phi}} \propto_{ij}^{\Phi} \cdot h'_j \tag{12}$$

where Z_i^{Φ} represents the node i learned embedding on the meta-path Φ via its neighbors. However, such attention only learns on the type of semantic information due to utilizing a single meta-path for calculating attention weight $\propto_{ij}^{\Phi}$. Additionally, we observe that the scale-free nature of heterogeneous graphs causes a high variance of graph data. To tackle this problem, we expand node-level attention to multi-head attention to preserve the stability of the training process. We apply the node-level attention w to K times, and the learned embeddings are concatenated to obtain the semantic-specific embedding as formulated in equation (13),

$$Z_i^{\Phi} = ||_{k=1}^k \sigma \left(\sum_{j \in N_i^{\Phi}} \propto_{ij}^{\Phi} \cdot h'_j \right) \tag{13}$$

where $||$ denotes concatenation, and $Z = \{Z_{\Phi_0}, Z_{\Phi_1}, \cdots, Z_{\Phi_P}\}$ represents grouped semantic embedding generated from node-level attention on the P meta-path set $\{\Phi_0, \Phi_1, \ldots, \Phi_P\}$.

Generally speaking, each node in the graph encompasses several types of semantic information and semantic-specific node embedding that represent a single aspect of node t. Also, achieving collective learning of node embedding necessitates the fusion of numerous meta-paths' semantics. This problem is tackled by applying semantic-level attention that is able to capture the importance of various meta-paths and exploits them for the targeted task. Given the input of P sets generated from node attention, the learned weights of each meta-path $\{\beta_{\Phi_0}, \beta_{\Phi_1}, \ldots, \beta_{\Phi_P}\}$ can be expressed with equation (14),

$$\{\beta_{\Phi_0}, \beta_{\Phi_1}, \ldots, \beta_{\Phi_P}\} = \text{att}_{\text{sem}}\{Z_{\Phi_0}, Z_{\Phi_1}, \cdots, Z_{\Phi_P}\}. \tag{14}$$

We perform a nonlinear transformation on semantic-level embedding and then determine its significance by measuring the resemblance of both transformed embedding and vector q of semantic-level attention. We also calculate the importance of each meta-path as the average of all the semantic-specific node embedding as depicted in equation (15), which is subsequently normalized with equation (16) to obtain β_{Φ_i} that represents the influence of the meta-path Φi for the molecule graph. The higher value of β_{Φ_i} denotes higher importance.

$$w_{\Phi_i} = \frac{1}{|v|} \sum_{i \epsilon v} q^T \cdot \tanh(W \cdot Z_i^{\Phi} + b) \qquad (15)$$

$$\beta_{\Phi_i} = \frac{\exp(w_{\Phi_i})}{\sum_1^P \exp(w_{\Phi_i})} \qquad (16)$$

Then we can calculate final embedding Z using the beforementioned semantic-specific embeddings by taking the computed weights as a parameter, as shown in equation (17)

$$Z = \sum_1^P (\beta_{\Phi_i} \cdot Z_i^{\Phi}). \qquad (17)$$

2) SEQUENTIAL LEARNING
In this section, we adopt SMILES to represent the chemical structure of drug compounds in the form of a line notation of atoms and covalent bonds. For instance, the sequence of atoms and covalent bonds is denoted as "CC1=C2C=C(C=CC...". The generated SMILES sequence needs to be encoded to be learned with later deep-learning layers. Several studies adopted one-hot encoding for SMILES tokens [39] but this encoding method ignores the contextual value of the symbols and is therefore unable to expose the operativity of the tokens in the surrounding context. To address this issue, we employed Smi2Vec [41], a method analogous to Word2Vec [42], to encode the tokens in the SMILES sequences. In which fixed-length SMILES symbols detach into a discrete atom, that mapped by finding corresponding embeddings from the pre-trained dictionary or producing a random value if no embedding exists.Finally, atom embedding vectors are aggregated to form the final embedding matrix.

Bi-Directional ConvLSTM (BConvLSTM): The embedding representation of drug SMILES X^{drug} is now passing into a BConvLSTM layer. The critical shortcoming of the conventional LSTM architecture is that these networks do not consider the spatial association since it utilizes full connections in state-to-state transitions and input-to-state transitions. To tackle this problem, ConvLSTM [24] was introduced, which make use of convolution operations to replace the full connection between various gates. It comprises an input gate I^t, a forget gate F^t, and a memory cell C^t, and an output gate O^t, while the operation of ConvLSTM can be formulated with equations (18-22),

$$I_t = \sigma(W_{xi} * X_t + W_{hi} * H_{t-1} + W_{ci} * C_{t-1} + b_i) \qquad (18)$$

$$F_t = \sigma(W_{xf} * X_t + W_{hf} * H_{t-1} + W_{cf} * C_{t-1} + b_f) \qquad (19)$$

$$C_t = f_t \circ C_{t-1} \tanh(W_{xf} * X_t + W_{hc} * C_{t-1} + b_c) \qquad (20)$$

$$O_t = \sigma(W_{xo} * X_t + W_{ho} * H_{t-1} + W_{ci} \circ C_{t-1} + b_c) \qquad (21)$$

$$H_t = (O_t \circ \tanh(C_t)) \qquad (22)$$

where $*$ and $\circ$ designate the convolution and Hadamard operation, correspondingly. Here, X_t denotes the input tensor (SMILEs embedding), C_t and H_t represent the memory and the hidden cell tensor, respectively, W and b denote 2D Convolution kernels and bias terms belonging to each cell. For convenience, we eliminate the subscript and superscript from the parameters.

In our architecture, we process the input X_t in both forward and backward directions using two ConvLSTMs and then calculate a decision of the current input depending on dual dependencies from both directions. This fully exploits the information in SMILES sequences and so might be effective to improve overall learning performance. Each of the forward and backward ConvLSTMs are regarded as separate ConvLSTM with two sets of parameters for each direction. So, we can calculate the BConvLSTM output with equation (23),

$$Y_t = \tanh(W_y^{\vec{H}} * \vec{H}_t + W_y^{\overleftarrow{H}} * \overleftarrow{H}_t + b) \qquad (23)$$

where $\vec{H}_t$, $\overleftarrow{H}_t$ denotes the tensors of hidden state tensors in both the forward and backward units respectively, b represents the bias term, and $Y_t \in \mathbb{R}^{F_t \times W_t \times H_t}$ designates the computed spatio-sequential output. Further, tanh implements the hyperbolic tangent for combining states in both directions.

C. OUTPUT LAYER
In this part of the network, the final feature representations produced from each channel were concatenated and fed into three FCLs. We build the FCLs with 1024, 768, and 512 nodes for each layer in respective order. After each layer we introduce a regularization dropout layer (0.1) to evade over-fitting by keeping activation for some neurons in the preceding layer. Additionally, we adopt a Rectified Linear Unit (ReLU) as an activation function. Finally, model training attempts to minimize the cost function. Here, we employ mean squared error (MSE) for measuring model loss using equation (24),

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (P_i - Y_i)^2 \qquad (24)$$

where P denotes the predicted affinities vector, Y represents the actual outputs, and n is the total number of samples.

IV. EXPERIMENTS
Given the set of DT pairs and the corresponding affinity s in a training data, the DeepH-DTA is trained to minimize the objective function presented in equation (24). For a generalized learning purpose, we have arbitrarily divided the dataset into six similar chunks, wherein a single chunk is designated as the self-governing test set. The other chunks of the data are employed to specify the hyper-parameters (as presented in the next section) by means of 5-fold cross validation. To assure performance robustness, we evaluate the model on leave out the test set and utilize the other five sets of 5-fold

cross validation for training the proposed deepH-DTA using the parameters presented in Table 1. Our experiments were performed on Windows 10 (4.2GHz Intel(R) Xeon(R) and Nividia Quadro (4GB)). Implementation details along with the train and test folds (and source codes) of the datasets can be accessed in the link: https://github.com/Hawash-AI/deepH-DTA. The DeepH-DTA takes the protein sequence as a first input. The input molecule is two-folded: a SMILES representation of the molecule and a graph representation of the molecule generated by RDKit. However, the stereochemistry of SMILES (i.e. where some of SMILES representations are not stereospecific). So, in this work, we eliminate stereochemistry in the SMILES input; since the number of relevant cases represent an unimportant percentage of the entire data. Additionally, the graph representation identifies bonds and atoms to account for stereochemistry.

A. MODEL HYPERPARAMETERS

Overall implementation of our model conducted using Pytorch library, we initialize bias value with zeros and adopt random weights initialization. Depending on the highest results obtained from several experiments, we discover the optimal hyperparameter setting for our architecture, which is analyzed and discussed in the next section. Table 1 summarizes the optimal hyperparameters of our model.

TABLE 1. The hyperparameters of DeepH-DTA.

Hyperparameters	Optimal values
No. dense blocks	3
No. conv layer (each block)	3
Filters (each block)	[6,9,12]
dropout	0.1
squeeze-and-excitation (threshold)	2
HGAT depth	4
No. BConvLSTM layers	100
Proteins sequence Length	1000
SMILES sequence Length	1000
FCLs	1024; 768; 512
epochs	200
batch size	256
dropout	0.1
optimizer	Adam
learning rate (lr)	0.00001

B. DATASETS

To assess the performance of the proposed approach, we adopt two broadly used benchmark datasets for DTI, namely the DAVIS dataset [43] and KIBA dataset [44]. The Davis dataset encompasses protein samples belonging to the kinase family and their inhibitors along with corresponding dissociation constant (K_d) values. In this paper, for numerical

stability, we transform the K_d values in the DAVIS dataset into log space, $pK_d = -\log_{10} \frac{K_d}{1 \times 10^9}$, as proposed by Wang *et al.* [12]. On the other hand, the KIBA dataset integrates various sources of inhibitor bioactivities for optimizing following consistency between K_i, K_d, and IC$_{50}$ by applying their statistical information. Table 2 provides a summary of both data sets and adopted splits for our model.

TABLE 2. Summary of experimental datasets.

Data Sets	No. Comps	No. Proteins	No. Inters	Split	No. samples
DAVIS	68	422	30,056	Train	20,037
				Validation	5,009
				Test	5,010
KIBA	2111	229	118,254	Train	78,836
				Validation	19,709
				Test	19,709

Comps= (compounds), Inter=(interactions)

C. EVALUATION MATRICES

We use several metrics used for evaluating the performance of our model, which are reliable with those used in previous studies. The computation of these metrics is as follows.

- Concordance Index (CI): used to measure whether the order of estimated binding affinity scores of couples of drugs–the target is identical to the order of true values, and we handle statistical significance using a paired *t*-test with 95% confidence interval (the larger value of CI indicate better model performance). The calculation of CI is in accordance with equation (25),

$$CI = \frac{1}{Z} \sum_{\delta_i > \delta_j} h(b_i - b_j) \qquad (25)$$

where b_i, b_j represents the prediction score for the higher affinity δ_i and lower affinity δ_j, respectively, Z denotes a normalization constant, and the step function h(x) can be formulated with equation (26)

$$h(x) = \begin{cases} 1, & \text{if } x > 0 \\ 0.5, & \text{if } x = 0 \\ 0, & \text{if } x < 0. \end{cases} \qquad (26)$$

- Mean Squared Error (MSE): represents the average of differences between predicted and actual output values (the smaller, the better).
- *R* squared r_m^2: denotes the external prediction performance of the model. Meanwhile, the model is acceptable only when $r_m^2 > 0.5$, and $r_m^2 = r^2 \times \left(1 - \sqrt{r^2 - r_0^2}\right)$, where r^2 and r_0^2 designate the squared correlation coefficient parameters for the predicted and actual values with and without intercept.
- The area under the precision curve (AUPC): a widely adopted measure for binary classification studies. In an attempt to measure the AUPR for our model, we transformed prediction datasets into binary datasets specifying the threshold for binding affinity for each one. Thus,

we select pK_d values of 7 and 12.1 as a threshold for the Davis and KIBA datasets, respectively. We choose these values based on because of its proven optimality and wide adoption in previous studies [7], [29], [13].

D. RESULTS AND COMPARISONS

For demonstrating the competitiveness of our model, we conduct an end-to-end comparison with the cutting-edge approaches (either machine- or deep-learning approaches) adopted for predicting affinity scores, and we conducted the comparative experiments under the same conditions. In Table 3 and Table 4, we provide the average obtained CI, MSE, r_m^2, and AUPC corresponding to each study on the Davis and KIBA datasets, respectively. It can be noted that machine-learning models such as KronRLS and SimBoost show worse performance compared to other deep-learning approaches. This is owing to their dependence on similarity matrices between drugs and targets as well as hand-crafted features. On the other hand, deep-learning techniques that automatically capture feature representation show great performance improvement.

TABLE 3. Model comparison with cutting edge approaches on the DAVIS dataset.

Models	CI (std)↑	MSE↓	r_m^2 (std)↑	AUPC (std)↑
KronRLS [27]	0.869 (0.001)	0.379	.407 (0.005)	0.661 (0.010)
SimBoost [26]	0.873 (0.002)	0.282	0.644 (0.006)	0.709 (0.008)
String Representation Based Approaches				
DeepDTA [29]	0.878 (0.004)	0.261	0.630 (0.017)	0.714 (0.010)
MT-DTI (wo-FT) [32]	0.875 (0.003)	0.268	0.633 (0.013)	0.700 (0.011)
MT-DTI [32]	0.887 (0.001)	0.245	0.665 (0.014)	0.730 (0.014)
DeepCPI [28]	0.867	0.293	0.607	0.705
WideDTA [7]	0.886 (0.003)	0.262	0.633 (0.011)	0.711 (0.012)
GANsDTA [13]	0.881 (0.005)	0.276	0.653 (0.015)	0.653 (0.017)
Attention-DTA [14]	0.887 (0.005)	0.245	0.657 (0.024)	0.746 (0.024)
Graph Representation Based Approaches				
GAT [11]	0.892 (0.003)	0.232	0.662 (0.010)	0.728 (0.016)
GIN [11]	0.893 (0.003)	0.229	0.649 (0.013)	0.720 (0.016)
GIN [12]	0.899 (0.003)	0.220	0.623 (0.011)	0.726 (0.015)
DeepGS [30]	0.880 (0.005)	0.252	0.686 (0.012)	0.763 (0.012)
DeepH-DTA*	**0.924 (0.001)**	**0.195**	**0.725 (0.009)**	**0.801 (0.010)**

The * denote the proposed architecture

First, this paper considers a few recent textual representation approaches such as: DeepDTA [29], MT-DTI [32], Deep-CPI [28], WideDTA [7], GANsDTA [13], and

TABLE 4. Model comparison with cutting edge approaches on the KIBA dataset.

Models	CI (std)↑	MSE↓	r_m^2 (std)↑	AUPC (std)↑
KronRLS [27]	0.782 (0.001)	0.411	0.342 (0.001)	0.635 (0.004)
SimBoost [26]	0.836 (0.001)	0.222	0.629 (0.007)	0.760 (0.003)
String Representation Based Approaches				
DeepDTA [29]	0.863 (0.002)	0.194	0.673 (0.009)	0.788 (0.004)
MT-DTI (wo-FT) [32]	0.844 (0.001)	0.220	0.584 (0.002)	0.789 (0.004)
MT-DTI [32]	0.882 (0.001)	0.220	0.584 (0.003)	0.789 (0.006)
DeepCPI [28]	0.852 (0.002)	0.211	0.657 (0.004)	0.782 (0.005)
WideDTA [7]	0.875 (0.001)	0.179	0.675 (0.005)	0.788 (0.008)
GANsDTA [13]	0.866 (0.001)	0.224	0.775 (0.008)	0.753 (0.007)
Attention-DTA [14]	0.882 (0.004)	0.162	0.735 (0.003)	0.829 (0.005)
Graph Representation Based Approaches				
GAT [11]	0.889 (0.001)	0.139	0.671 (0.005)	0.781 (0.006)
GIN [11]	0.891 (0.004)	0.139	0.684 (0.004)	0.801 (0.005)
GIN [12]	0.901 (0.002)	0.129	0.680 (0.003)	0.799 (0.004)
DeepGS [30]	0.860 (0.003)	0.193	0.684 (0.002)	0.801 (0.005)
DeepH-DTA*	**0.927 (0.003)**	**0.111**	**0.799 (0.004)**	**0.861 (0.002)**

The * denote the proposed architecture

Attention-DTA [14]. Among these approaches, Attention-DTA and MT-DTI yielded best results with CI of 0.887, MSE of 0.245 on the Davis dataset; also, on the KIBA dataset, they both achieved CI of 0.882 and MSE of 0.220 and 0.162 respectively. This explains the effectiveness of the attention convolutional operation in learning sequential drug and target information in the case of Attention-DTA [14]; and also, the efficiency of the pre-trained BERT representation presented in MT-DTI.

Second, graph network approaches [11], [12], [16] can effectively capture topological relationships of drug molecules, which enable further performance improvement. Amongst them, the GIN [12] shows a higher CI value of 0.899 and lower MSE of 0.222 on the Davis dataset; and 0.901 of CI and 0.129 of MSE on the KIBA dataset. Meanwhile, DeepGS [30] yield the least performance with CI of 0.880 and 0.860 and MSE of 0.252 and 0.193, respectively, on the Davis and KIBA datasets.

It can be observed that the proposed DeepH-DTA has a robust performance on both datasets, achieving 0.924 (0.025 improvement), 0.195 (reduced by 0.025), 0.725 (0,039 improvement), and 0.801(0.038 improvement) for CI, MSE, r_m^2, and AUPC, respectively, for the Davis dataset. For the KIBA dataset, we achieved 0.927 for CI (0.026 improvement), 0.111 for MSE (reduced by 0.018), 0.799 for r_m^2 (0.024 improvement), and 0.861 for AUPC (0.032 improvement).

For the Davis dataset, the proposed DeepH-DTA outperforms the traditional ML techniques [26], [27] by 5% of CI and with statistical significance (*p*-value of 0.00009 for both). Further, DeepH-DTA outperforms WideDTA [7] by 4% of CI score with statistical significance (*p*-value, 0.002). Also, it outperforms the best graph-based approach GIN [12] by 2.5% of CI with statistical significance (*p*-value, 0.01). On the other hand, for the KIBA dataset, it outperforms both of these techniques by > 9% with statistical significance (*p*-value around 0.0005 for both [26] and [27]). For the string-based approach, DeepH-DTA outperforms WideDTA [7] by 5.2% of CI score with statistical significance (*p*-value, 0.004). Further, it outperforms the best graph-based approach GIN [12] by 2.6% of CI with statistical significance (*p*-value, 0.009).

In order to further verify model performance, we note that DeepH-DTA attains 2.5% of MSE lower than the lowest existing MSE in GIN [12] with statistical significance (*p*-value, 0.007) on the Davis dataset. Also, it achieves 1.8% of MSE lower than the lowest current MSE in GIN [12] with statistical significance (*p*-value, 0.01) on KIBA dataset. This indicates the superiority of our proposed approached compared to the most recent studies for predicting DTA. Accordingly, we observe that our model outperforms existing deep-learning methods on four measures, which can be explained due to several factors:

1) In comparison with DeepCPI, our model cooperatively exploits the drugs' topological structures along with following characteristics of chemical context, which in turn significantly improves the performance.

2) Compared with both Deep-DTA architectures, we adopted the HGAT architecture to learn the structural information of the drug, and employed innovative embedding methods to obtain extra contextual information for both drugs and protein sequences.

3) Compared with graph-based approaches [11], [12], [30], the proposed dense network with squeeze-and-excitation operation models protein sequence information more effectively compared to traditional CNN. Further, utilizing HGAT allows better exploitation of semantic information in meta-path data. Also, Bi-ConvLSTM allows for better exploitation of spatio-sequential representation from SMILEs sequences.

Generally, the obtained results and comparisons demonstrate that our model achieves competitive performance outperforms against these baselines methods in all metrics.Moreover, Fig.2 and Fig.3 present the scatter plots of the proposed model predicted affinity score against the actual measured value on the DAVIS and KIBA datasets. The model achieves better performance when the estimated affinity scores are close to the original scores, and hence the instances should appear close to the red line. With regard to the DAVIS dataset, it can be observed that the greater number of the pK_d scores are found in the range of [5], [6] along the x-axis, principally because the pK_d score of 5 establishes more than half of the dataset. Additionally,

there is a crowded area of KIBA scores lying in the range [11], [14] along the x-axis, which shows similar behavior to the Davis dataset. Principally, for both datasets, the data instances are close to the red regression line which, in turn, demonstrates that the proposed architecture has a competitive prediction performance.
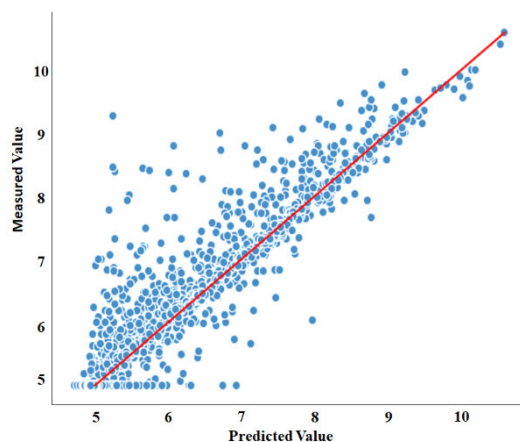

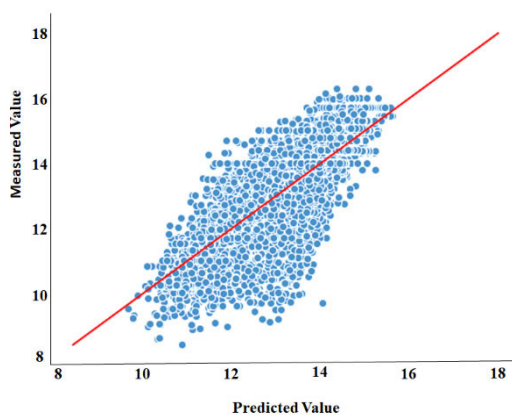**FIGURE 2.** Predictions from our model against measured (real) binding affinity values for Davis dataset (*pK_d*).

FIGURE 3. Predictions from our model against measured (real) binding affinity values for the KIBA dataset (KIBA score).

1) IMPACT OF DENSE & SE BLOCK

In Fig.4 we show the result of model implementation using a different CNN architecture—namely: traditional CNN (CNN), Residual CNN (ResNet), and Dense CNN (DenseNet)—and it can be observed that employing squeezed-excited operation after each dense block improves model performance compared to other architectures due to exploiting multi-channel dependency and hence capturing interrelationships of protein features. It could be noted that the Dense Net architecture attains 0.017 higher CI than traditional CNN and 0.009 higher than Residual implementation. This explains the effectiveness of collective learning of dense networks in learning protein sequences. Additionally, the proposed Dense net with SE block attains a further 0.016 improvement.
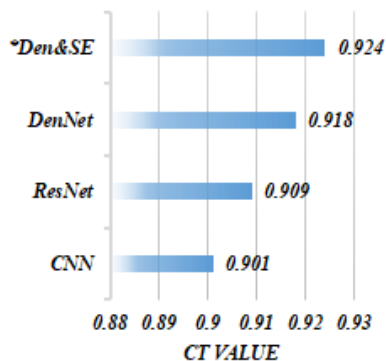
FIGURE 4. The CI value attained by implementing DeepH-DTA using different CNN implementations on the Davis dataset.

2) IMPACT OF HGAT

In order to demonstrate the efficiency of the proposed graph neural network (GNN), we implemented different versions of our architecture utilizing various types of a graphical network, particularly GCN [60], GAT [61], and hybrid architecture (GCAT). The GCN consist of a novel variant of CNN that effectively operates on graph data, whereas the GAT performs similar operations by applying self-attentional layers to attend to the features of the node's neighborhoods. The corresponding results are shown in Fig.5 which shows that the proposed HGAT adopted in our model significantly improves model performance by 14% owing to applying semantic attention mechanism on meta-path.
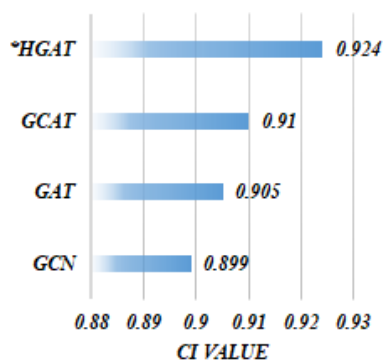
FIGURE 5. The CI value attained by implementing DeepH-DTA using different types of GN on the Davis dataset.

3) IMPACT OF ConvLSTM

Further, in an attempt to verify our hypothesis about the effectiveness of ConvLSTM in capturing spatio-sequential information from input SMILES. We evaluate the performance of the proposed DeepH-DTA on DAVIS dataset using different versions of RNNs as presented in Fig.6. It could be noted that simple RNN attains the lowest CI value with 0.886, and 0.008 improvements are achieved when using GRU. Also, an extra improvement with 2.3% could be observed when using LSTM while attaining the maximum CI value when ConvLSTM is employed to implement DeepH-DTA with 0.924 of CI outperforming the LSTM performance by 1.1%.

This experiment demonstrates the effectiveness of using ConvLSTM for modeling the SMILES string input.Furthermore, in Fig.7 and Fig.8, we present model training progress in terms of CI values corresponding to the Davis and KIBA dataset correspondingly. On both datasets, we observe rapid validation convergence after 100 epochs. It could be noted that the model validation CI is higher than training CI at first 25 epochs on the Davis dataset, and show similar behavior with the early 50 epochs on the KIBA dataset. The training CI values maintain a higher value than validation CI. Fig.9 and Fig.10 display the training progress in terms of MSE loss on the Davis and KIBA datasets. On the Davis dataset, we observe early convergence after 120 epochs; meanwhile, start confluence after 125 epochs of training on the KIBA dataset. Our model always has training MSE lower that validation MSE on both datasets. However, it shows the opposite behavior on the first 35 epochs of training on KIBA dataset. Finally, we observe that the progress of validation MSE on the Davis dataset is more stable than the KIBA dataset.
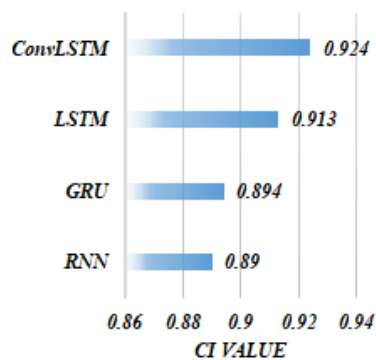

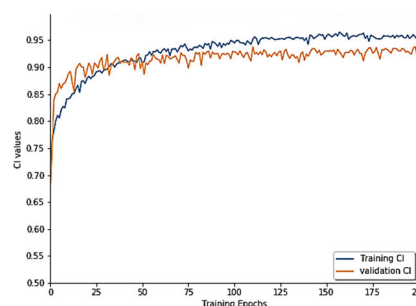**FIGURE 6.** The CI value attained by implementing DeepH-DTA using different types of RNN on the Davis dataset.

FIGURE 7. Model's CI training progress on Davis dataset.

E. COMPUTATIONAL COMPLEXITY

The most advantageous characteristic of deep learning approaches is that they could be executed and trained using the Graphics Processing Unit (GPU). Concerning time complexity, GPU-based deep-learning approaches show a significant reduction in time complexity compared to the case when running on a traditional CPU. Gawehn *et al.* [62] discussed and introduced several strategies for employing GPU to accelerate drug discovery systems. For further verification
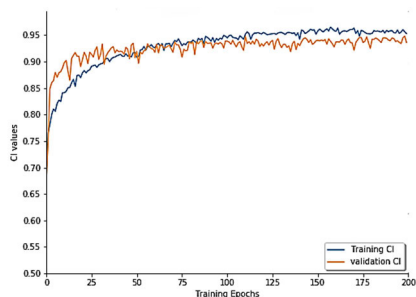
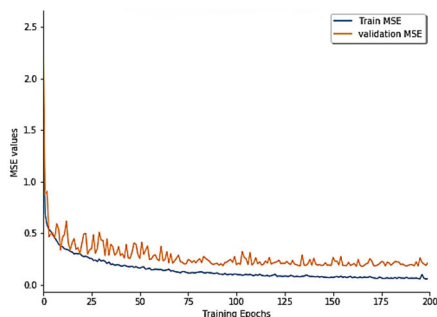**FIGURE 8.** Model's CI training progress on the KIBA dataset.

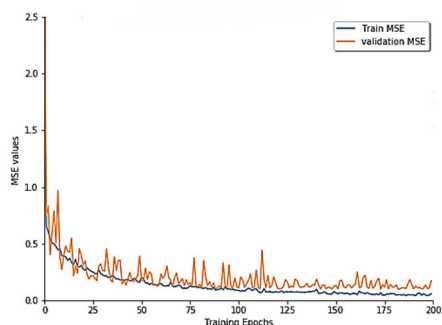**FIGURE 9.** Model's MSE training progress on the Davis dataset.

FIGURE 10. Model's MSE training progress on the KIBA dataset.

of the effectiveness of the proposed DeepH-DTA, the computational complexity needs to be addressed. In this regard, we compare the execution time (time in seconds/epoch) of the DeepH-DTA against the before mentioned graph-representation-based approaches, as presented in Table 5. It can be noted that the GIN based approaches [11], [12] consume the least execution time on both datasets. Further, the proposed DeepH-DTA consumes comparable execution time to the GAT [11], and the DeepGS [30]. This could be explained due to the time taken to calculate attention at meta-paths and the time consumed for spatio-sequential learning using ConvLSTM. Compared to the attained performance improvements, this slight increase in execution time indicates the effectiveness and the ability to integrate the DeepH-DTA in real-life scenarios.

Furthermore, DeepH-DTA does not necessitate matrix factorization or resemblance matrices, hence it offers further scalability compared to the SimBoost and the KronRLS.

TABLE 5. Comparison between average execution time (second / epoch).

Dataset / Models	GAT [11]	GIN [11]	GIN [12]	DeepGS [30]	**DeepH-DTA***
DAVIS	391s	198s	209s	404s	426s
KIBA	1431s	706s	692s	1530s	1556s

Given *m* represents the number of protein sequences and *n* represents the number of compounds, meanwhile SimBoost and KronRLS and require the resemblance matrices, which have space and time complexity of $O(n^2 + m^2)$. SimBoost includes matrix factorization and thereby represents additional expense. On the other hand, in every epoch of DeepH-DTA training procedures, the time complexity only be contingent on the number of DT pairs in the training set which, in the highest situation, is $O(\max(n, m))$, and $O(nm)$ in the worst situations. There is no clearly formulated inter-relationship between the count of epochs while waiting for convergence and *n* or *m*, thus the count of epochs could not be investigated hypothetically. Nevertheless, it is notable that, in practice, such count is probably sub-linear in *n* and *m* or even autonomous from *n* and *m*, so the epochs count can be statistically set to a slight constant if we aim to realize comparatively primitive results, whereas the SimBoost and KronRLS firmly necessitate a minimum of $O(n^2 + m^2)$ time to attain any results.

V. SARS-COV-2 DRUG REPURPOSING
A. MODELING DRUG AND SARS-COV-2 INTERACTIONS
In this section, we apply our proposed model for predicting binding affinity scores for commercially existing drugs, and SARS-CoV-2 proteins in order to identify the best inhibitors that can suppress virus spread and provide scientists with a start point for developing new vaccines. For this purpose, we collected several amino acid sequences from the Protein Data Bank (PDB) database and the National Center for Biotechnology Information (NCBI), as listed in Table 6.

TABLE 6. Model comparison with cutting edge approach on KIBA dataset.

Proteins	Identifier
3C-like proteinase	**PDB ID: 6WQF**
RNA-dependent RNA polymerase (RdRp)	**NCBI: YP_009725307.1**

Then the proposed DeepH-DTA is trained on two public databases that are manually combined: namely, the journal curated Binding DB [45] and the Drug Bank database (interactions: 26167; drugs: 7591; target: 4187) [5] with three types of efficiency scores as in the KIBA dataset. The DeepH-DTA is trained for 75 epochs using the same hyperparameters presented in Table 1 and under the same experimental conditions discussed in the previous section. To attain rapid convergence, the DeepH-DTA parameters were initialized by transferring the learned parameters from the KIBA dataset. Further, we average the consistence-score procedure [46]
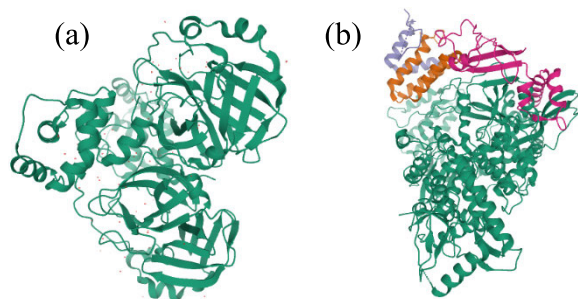
FIGURE 11. (a) 3D View of the PDB ID: 6WQF. Fig.11. (b) 3D View of the NCBI: YP_009725307.1.

to integrate these scores and keep their Pearson correlation score above 0.9. Since the aggregated data involves extensive heterogeneity of molecules and proteins, the proposed DeepH-DTA has inherent superiority for modeling the interactions between antiviral medications and protein sequences of SARS-COV-2. After that, the DeepH-DTA predictions were filtered out for FDA-approved drugs with the highest binding to target viral proteins. Moreover, we included Remdesivir and Ivermectin because therapeutic potential to COVID-19 has been proposed recently in [48], [49], and we also included drugs from clinical trials.

B. FINDINGS AND DISCUSSIONS

We exploit the advantage of heterogonies for modeling DTI to predict affinity scores of 3,001 FDA-approved drugs against $3C_{pro}$, RdRp, helicase, 30-to-50 exonuclease, endoRNAse, and 2OMT of SARS-CoV-2. For a better understanding of these genes, please refer to section 1(B).

Table 7 and Table 8 present the top inhibitor list for SARS-CoV-2 main-protease and RdRp proteins, respectively. Both tables provide the commercial drug name, corresponding SMILES format, models' predicted affinity scores (K_d in nM), and the clinical evidence for this prediction if exist (clinical approval means the research study that proves the effectiveness of the certain drug against COVID-19). In Table 7, we observe that SARS-CoV-2 main-protease was estimated to bind with Cilostazol (K_d: 53.13 nM), Baricitinib (K_d: 59.27 nM), Fluconazole (K_d: 64.34 nM), Itraconazole (K_d: 70.35), Quercetin (K_d: 79.24 nM), Rabeprazole (K_d: 85.26 nM), Grazoprevir (K_d: 79.24 nM) and other drugs with a prediction affinity of over 100 nM. Additionally, we present Structural graphical formulas of some of suggested drugs in Fig.12.

On the other hand, in Table 8, we introduce the top estimated affinities with RdRp; it can be observed that RdRp of SARS-COV-2 bind with Sirolimus (K_d: 8.13 nM), Ivermectin (K_d: 9.94 nM), Methylprednisolone (K_d: 11.26 nM), Abacavir sulfate (K_d: 16.92 nM), Rifaximin (K_d: 20.03nM), Ritonavir (K_d: 23.26 nM), and Metoprololtartrate (K_d: 28.06 nM) and some other drug candidates. Also, we present Structural graphical formulas of some suggested drugs in Fig.13. From the results obtained on both tables with the lowest K_d the value represents the drugs with the highest

binding affinity against SARS-CoV-2, which can help clinical researchers to investigate these drugs or use them as a starting point to develop a new vaccine.

Several studies used DTI prediction as a tool for drug repurposing to discover novel utilization of current drugs. Accordingly, we make use of our proposed approach to enable controlling the explosion of SARS-CoV-2. Recently, numerous studies have recognized encouraging drug nominees that can assist in inhibiting various aspects of SARS-CoV-2. For example, Baricitinib and Methotrexate revealed inhibitory impacts against SARS-CoV-2 [49], [50]. Also, Grazoprevir, Bortezomib, Asunaprevir, Ritonavir demonstrated its effectiveness for SARS-Cov-2 in multiple in-vitro studies [51]–[56]. Moreover, Chloroquine has been shown as an effective inhibitor in [47], [58]. Nevertheless, these clinical studies depend on prior experienced knowledge that enables the selection of specific drugs that have some inhibitory possessions on similar coronaviruses. In contrast, the proposed architecture was pre-trained on several binding databases without domain experience.

Toxicity Information: For the Cilostazol (100 mg twice/daily) the signs of an acute overdose can be a severe headache, diarrhea, hypotension, tachycardia, and a potential increase in heart rate. The public side effects of Baricitinib (2 mg/day) are herpes simplex and zoster infections, cholesterol and creatinine elevations, neutropenia, fatigue, diarrhea nausea, and symptoms of upper respiratory tract infection. Besides, side effects of both of the Itraconazole and Fluconazole (100-400 mg/day) embrace headache, vomiting, nausea, and it exhibits rare yet serious cases of serious hepatic toxicity [4]. Quercetin (500-1000 mg/day) has similar toxicity as Itraconazole plus abdominal discomfort however it did not report any Hepatic toxicity [57]. The Rabeprazole (120 mg/day) shows some rare side effects including hypersensitivity reactions, hypomagnesemia, bone fractures for lung in case of lung use, lupus erythematosus, and acute interstitial nephritis. Further, Grazoprevir (100 mg/day) has been reported to cause mild effects including fatigue, headache and nausea; and multiple hypersensitivity reactions have been reported for Abacavir sulfate (300 mg/day) which occurred in association with anaphylaxis, liver and renal failure, hypotension, fever, rash, fatigue, GI symptoms such as nausea, vomiting, diarrhea, and abdominal pain. The recommended dosage of Bortezomib (1.3 mg/m^2) differs by indication, tolerance, and hepatic function, and it shows fatal outcomes when the patient follows the administration of more than twice the recommended therapeutic dose; and include the acute onset of symptomatic hypotension and thrombocytopenia [52]. Further, Extrahepatic manifestations due to Rifampin (60 mg/day) hepatotoxicity such as fever, rash, arthralgias, edema, and eosinophilia are uncommon as is autoantibody formation. Several contrary effects of either Lopinavir (800 mg/day) or Ritonavir (50-100 mg twice /day) may arise including hepatotoxicity, pancreatitis, and hyperlipidemia and lipodystrophy [55], [56]. The tetraethylene glycol (2000 mg/day) caused minimal skin irritation and was

TABLE 7. DTI prediction results of FDA approved antiviral drugs 3C-like proteinase of SARS-CoV-2.

Small molecules	SMILES format	Predicted K_d in nM	Clinical approved	DOSE
Cilostazol	O=C1CCC2=C(N1)C=CC(OCCCCC1=NN=NN1C1CCCCC1)=C2	53.13	-	100 mg twice/daily
Baricitinib	CCS(=O)(=O)N1CC(C1)(CC#N)N2C=C(C=N2)C3=C4C=CNC4=NC=N3	59.27	Clinical trials, [49],[50]	2 mg once / daily
Fluconazole	OC(CN1C=NC=N1)(CN1C=NC=N1)C1=C(F)C=C(F)C=C1	64.34	-	100 - 400 mg/ daily
Itraconazole	CCC(C)N1N=CN(C1=O)C1=CC=C(C=C1)N1CCN(CC1)C1=CC=C(OC[C@H]2CO[C@@](CN3C=NC=N3)(O2)C2=CC=C(Cl)C=C2Cl)C=C1	70.35	-	100 - 400 mg/ daily
Quercetin	OC1=CC2=C(C(O)=C1)C(=O)C(O)=C(O2)C1=CC=C(O)C(O)=C1	79.24	Clinical trial, [57]	500 -1000 mg/daily
Rabeprazole	[Na+].COCCCOC1=CC=NC(CS(=O)C2=NC3=CC=CC=C3[N-]2)=C1C	85.26	Clinical trial	120 mg/daily
Grazoprevir	COC1=CC2=NC3=C(CCCCC[C@@H]4C[C@H]4OC(=O)N[C@H](C(=O)N4C[C@@H](C[C@H]4C(=O)N[C@@]4(C[C@H]4C=C)C(=O)NS(=O)(=O)C4CC4)O3)C(C)(C)C)N=C2C=C1	98.47	[51]	100 mg/day
Abacavir (sulfate)	OS(O)(=O)=O.NC1=NC2=C(N=CN2[C@@H]2C[C@H](CO)C=C2)C(NC2CC2)=N1.NC1=NC2=C(N=CN2[C@@H]2C[C@H](CO)C=C2)C(NC2CC2)=N1	107.62	-	300 mg twice /day
Bortezomib	CC(C)C[C@H](NC(=O)[C@H](CC1=CC=CC=C1)NC(=O)C1=CN=CC=N1)B(O)O	117.23	[52]	1.3 mg/12 hour
Metoprolol tartrate	[H][C@](O)(C(O)=O)[C@@]([H])(O)C(O)=O.COCCC1=CC=C(OCC(O)CNC(C)C)C=C1.COCCC1=CC=C(OCC(O)CNC(C)C)C=C1	129.23	-	160 mg/day
Rifabutin	CO[C@H]1\C=C\O[C@@]2(C)OC3=C(C)C(O)=C4C(O)=C(NC(=O)\C(C)=C/C=C/[C@H](C)[C@H](O)[C@@H](C)[C@@H](O)[C@@H](C)[C@H](OC(C)=O)[C@@H]1C)C1=C(N=C5C=C(C)C=CN15)C4=C3C2=O	145.19	-	60 mg/day
Ritonavir	CC(C)C1=NC(=CS1)CN(C)C(=O)NC(C(C)C)C(=O)NC(CC2=CC=CC=C2)CC(C(CC3=CC=CC=C3)NC(=O)OCC4=CN=CS4)O	163.36	[54-56]	50 to 100 mg twice/day
Tetraethylene glycol	C(COCCOCCOCCO)O	171.14	Clinical trials	2000 mg/day
Adenosine monophosphate	C1=NC(=C2C(=N1)N(C=N2)C3C(C(C(O3)COP(=O)(O)O)O)O)N	178.32	-	500 mg/day
Lopinavir	CC(C)[C@H](N1CCCNC1=O)C(=O)N[C@H](C[C@H](O)[C@H](CC1=CC=CC=C1)NC(=O)COC1=C(C)C=CC=C1C)CC1=CC=CC=C1	183.14	[55-56]	800 mg/day
Chloroquine phosphate	OP(O)(O)=O.CCN(CC)CCCC(C)NC1=CC=NC2=CC(Cl)=CC=C12	189.35	[47]	500 mg / week
Atazanavir	COC(=O)N[C@H](C(=O)N[C@@H](CC1=CC=CC=C1)[C@@H](O)CN(CC1=CC=C(C=C1)C1=CC=CC=N1)NC(=O)[C@@H](NC(=O)OC)C(C)(C)C)C(C)(C)C	195.57	-	400 mg / day
Remdesivir	CCC(CC)COC(=O)[C@H](C)N[P@](=O)(OC[C@H]1O[C@@](C#N)([C@H](O)[C@@H]1O)C1=CC=C2N1N=CN=C2N)OC1=CC=CC=C1	201.13	[51]	200 mg / day

not a skin sensitizer when tested in humans. Also, an overdose of Adenosine monophosphate (500 mg/day) could cause local erythema, slight flushing, dizziness, diuresis, and palpitation.

Moreover, the chloroquine (500 mg/week) overdose can trigger an acute attack with drowsiness, visual disturbances, and serum aminotransferase elevations, occasionally resulting in jaundice [47]. Hydroxychloroquine [58] (400 mg/day) does not cause this reaction and appears to have partial beneficial effects in porphyria with an exception for patients with porphyria cutanea tarda; where relatively high doses can trigger an acute hepatic injury with sudden onset of fever and marked serum enzyme elevations with increased excretion of porphyrins. Furthermore, the overdose of Atazanavir

(400 mg/day) can cause several forms of liver injury including transient serum enzyme elevations, indirect hyperbilirubinemia, idiosyncratic acute liver injury, and exacerbation of underlying chronic viral hepatitis. Meanwhile, hepatic artery thrombosis has been stated to be known with sirolimus (2 mg/day) therapy after liver transplantation, but this suggestion is still controversial. Also, an overdose of Ivermectin (15 mg/day) could cause some adverse effects including muscle or joint pain, dizziness, fever, headache, skin rash, and fast heartbeat [48]. Rifaximin (550 mg/day) shows some adverse effects include peripheral edema, muscle spasms, and gastrointestinal upset. Yet there is no evidence that an overdose can cause liver injury. The Digoxin (0.1 mg/day) toxicity

TABLE 8. DTI prediction results of FDA approved antiviral drugs RdRp of SARS-CoV-2.

Small Molecules	SMILES format	Predicted K_d in nM	Clinical approved	Dose
Sirolimus	CC1CCC2CC(C(=CC=CC=CC(CC(C(=O)C(C(C(=CC(C(=O)CC(OC(=O)C3CCCCN3C(=O)C(=O)C1(O2)O)C)C)CC4CCC(C(C4)OC)O)C)C)O)OC)C)C)OC	8.13	Clincal trials, [19]	2 mg / day
Ivermectin	CO[C@H]1C[C@@H](O[C@@H](C)[C@@H]1O)O[C@H]1[C@H](C)O[C@H](C[C@@H]1OC)O[C@H]1[C@@H](C)\C=C\C=C2/CO[C@@H]3[C@H](O)C(C)=C[C@@H](C(=O)O[C@H]4C[C@@H](C\C=C1/C)O[C@@]1(CC[C@H](C)[C@@H](C(C)C)O1)C4)[C@]23O.CC[C@@H](C)[C@H]1O[C@@]2(CC[C@@H]1C)O[C@@H]1C\C=C(C)\[C@@H](O[C@@H]3O[C@@H](C)[C@@H](O[C@@H]4O[C@@H](C)[C@@H](O)[C@@H](OC)C4)[C@@H](OC)C3)[C@@H](C)\C=C\C=C3/CO[C@@H]4[C@H](O)C(C)=C[C@@H](C(=O)O[C@@H](C1)C2)[C@]34O	9.94	[48]	15 mg/day
Methylprednisolone	[H][C@@]12CC[C@](O)(C(=O)CO)[C@@]1(C)C[C@H](O)[C@@]1([H])[C@@]2[H]C[C@H](C)C2=CC(=O)C=C[C@]12C	11.26	Clincal trials	60 mg/day
Abacavir (sulfate)	OS(O)(=O)=O.NC1=NC2=C(N=CN2[C@@H]2C[C@H](CO)C=C2)C(NC2CC2)=N1.NC1=NC2=C(N=CN2[C@@H]2C[C@H](CO)C=C2)C(NC2CC2)=N1	16.92	-	300 mg twice /day
Rifaximin	CO[C@H]1\C=C\O[C@@]2(C)OC3=C(C2=O)C2=C(C(O)=C3C)C(=O)C(NC(=O)\C(C)=C/C=C/[C@H](C)[C@H](O)[C@@H](C)[C@@H](O)[C@@H](C)[C@H](OC(=O)[C@@H]1C)C1=C1NC3(CCN(CC3)CC(C)C)N=C21	20.03	-	550 mg three / day
Ritonavir	CC(C)[C@H](NC(=O)N(C)CC1=CSC(=N1)C(C)C)C(=O)N[C@H](C[C@H](O)[C@H](CC1=CC=CC=C1)NC(=O)OCC1=CN=CS1)CC1=CC=CC=C1	23.26	[55-56]	50 to 100 mg twice/day
Metoprolol Tartrate	[H][C@](O)(C(O)=O)[C@@]([H])(O)C(O)=O.COCCC1=CC=C(OCC(O)CNC(C)C)C=C1.COCCC1=CC=C(OCC(O)CNC(C)C)C=C1	28.06	-	160 mg/day
Digoxin	[H][C@]12CC[C@]3([H])[C@]([H])(H)(C[C@@H](O)[C@]4(C)[C@@H](CC[C@]34O)C3=CC(=O)OC3)[C@@]1(C)CC[C@@H](C2)O[C@H]1C[C@H](O)[C@H](O[C@H]2C[C@H](O)[C@H](O[C@H]3C[C@H](O)[C@H](O[C@@H](C)O3)[C@@H](C)O2)[C@@H](C)O1	32.01	-	0.1 mg /day
Atazanavir	COC(=O)N[C@H](C(=O)N[C@@H](CC1=CC=CC=C1)[C@@H](O)CN(CC1=CC=C(C=C1)C1=CC=CC=N1)NC(=O)[C@@H](NC(=O)OC)C(C)(C)C)C(C)(C)C	35.67	-	200 mg / day
Ciclesonide	CC(C)C(=O)OCC(=O)C12C(CC3C1(CC(C4C3CCC5=CC(=O)C=CC45C)O)C)OC(O2)C6CCCCC6	36.27	Clinical trials	1.6 mg /day
Dexamethasone	[H][C@@]12C[C@@H](C)[C@](O)(C(=O)CO)[C@@]1(C)C[C@H](O)[C@@]1(F)[C@@]2([H])CCC2=CC(=O)C=C[C@]12C	38.02	-	47.59 mg/day
N-Acetyl-beta-D-glucosamine	CC(=O)NC1C(C(C(OC1O)CO)O)O	39.89	Clinical trials	4170 mg/day
Daclatasvir	COC(=O)N[C@@H](C(C)C)C(=O)N1CCC[C@H]1C1=NC=C(N1)C1=CC=C(C=C1)C1=CC=C(C=C1)C1=CN=C(N1)[C@@H]1CCCN1C(=O)[C@H](NC(=O)OC)C(C)C	42.23	Clinical trials. [50]	60 mg/ day
Rifabutin	CO[C@H]1\C=C\O[C@@]2(C)OC3=C(C)C(O)=C4C(O)=C(NC(=O)\C(C)=C/C=C/[C@H](C)[C@H](O)[C@@H](C)[C@@H](O)[C@@H](C)[C@H](OC(C)=O)[C@@H]1C)C1=C(N=C5C=C(C)C=CN15)C4=C3C2=O	49.12	-	60 mg/day
Remdesivir	CCC(CC)COC(=O)[C@H](C)N[P@](=O)(OC[C@H]1O[C@](C#N)([C@H](O)[C@@H]1O)C1=CC=C2N1N=CN=C2N)OC1=CC=CC=C1	51.13	[51]	400 mg / day
Lopinavir	CC(C)[C@H](N1CCCNC1=O)C(=O)N[C@H](C[C@H](O)[C@H](CC1=CC=CC=C1)NC(=O)COC1=C(C)C=CC=C1C)C1=CC=CC=C1	53.04	[55-56]	800 mg/day
Chloroquine	CCN(CC)CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl	58.21	[58]	500 mg / week
Leflunomide	CC1=C(C=NO1)C(=O)NC1=CC=C(C=C1)C(F)(F)F	62.75	Clinical trials. [50],[59]	-
Hydroxychloroquine	CCN(CCCC(C)NC1=C2C=CC(=CC2=NC=C1)Cl)CCO	69.21	[58]	400 mg/day

may be established by indications of nausea, vomiting, visual changes, in addition to arrhythmia. Older age, lower body weight, and decreased renal function or electrolyte abnormalities lead to an increased risk of digoxin toxicity [4].

VI. LIMITATIONS OF THIS STUDY
Despite the superiority of the proposed DeepH-DTA, it still suffers from some shortcomings that limit realizing the most optimal performance. First, the marginal improvement in CI
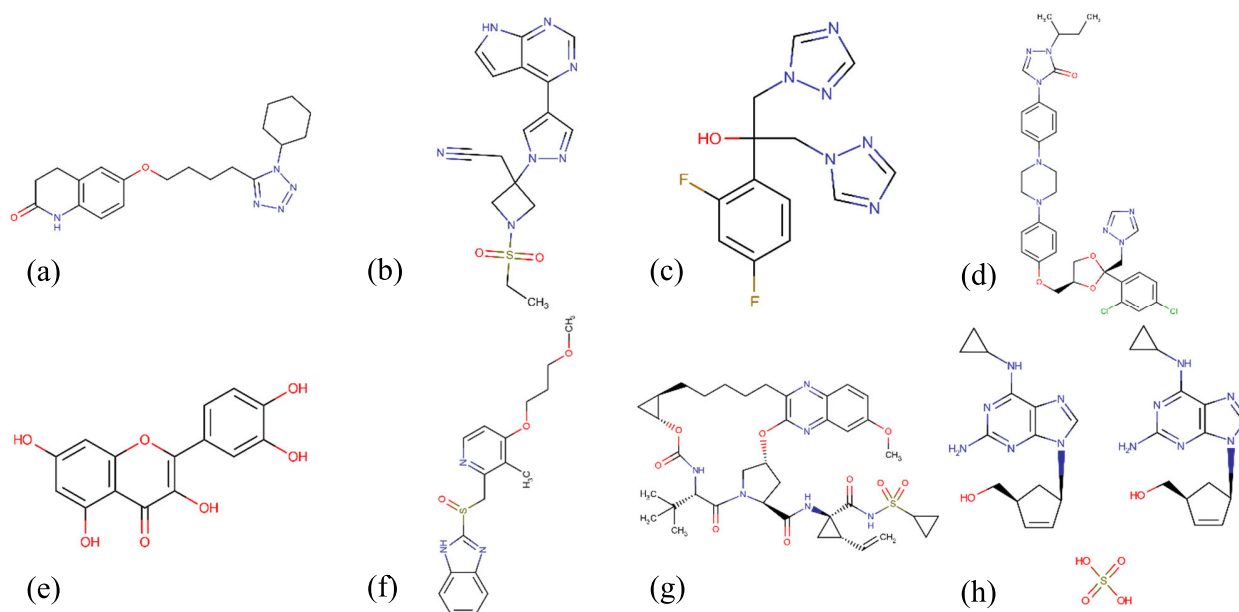
FIGURE 12. Structural formulas of candidate drugs against SARS-CoV-2 3C-like proteinase (a) Cilostazol, (b) Baricitinib, (c) Fluconazole, (d) Itraconazole, (e) Quercetin, (f) Rabeprazole, (g) Grazoprevir, (h) Abacavir (sulfate).
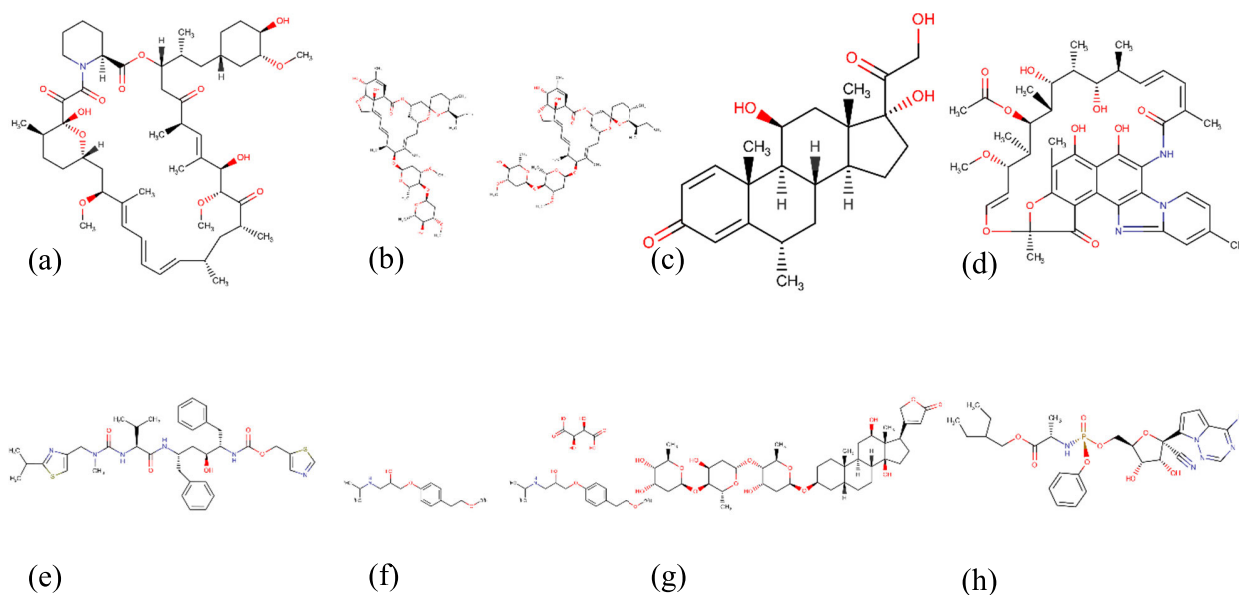

FIGURE 13. Structural formulas of candidate drugs against SARS-CoV-2 RdRp (a) Sirolimu, (b) Ivermectin,(c) Methylprednisolone, (d) Rifaximin, (e) Ritonavi, (f) Metoprolol Tartrate, (g) Digoxin, (h) Remdesivir.

measure (0.025, 0.026) could be reasoned by the fact that the model considers learning the representations of proteins sequences and drug molecules separately, then merges these representations for final decisions. This could be handled by learning the interaction patterns between proteins sequences and drug molecules. Second, the DeepH-DTA the semantic representation of input sequences that have been shown effective in many sequential data. Transformer models [63], [64] could be employed to generate more informative sequential data representation. Third, the execution time required by HGAT is high, as presented in Table 5. Fourth, some of the

predictions for SARS-Cov-2 also need to be confirmed in vitro, in vivo, and in an inclusive series of scientific trials for effectiveness and safety.

VII. MANAGERIAL IMPLICATIONS
This section provides a summary of the results and how our model could be useful in real-life situations. In this study, we introduce an efficient and applicable deep learning approach (DeepH-DTA) drug-target affinity prediction that is able to support clinical staff to discover the most effective inhibitor against newly discovered diseases like COVID-19.

The major advantages of the proposed architecture are its capability to exploit the topological and sequential representation of drug molecules. Second, it is not restricted to the specific data used in this paper. In other words, it is possible to apply deepH-DTA to various drug repurposing problems as shown with COVID-19 data.

VIII. CONCLUSION AND FUTURE DIRECTIONS

We introduce a novel deep-learning framework for predicting DT binding affinity using target protein sequences and various heterogeneity drugs. We use squeezed-excited dense convolutional networks to capture hidden representations of proteins sequence. We adopt a modified HGAT network for topological modeling information of heterogeneous chemical molecules, while BConvLSTM exploited the spatio-sequential description of SMILES encoded molecules. The generated representations are concatenated and passed for final FCL, where the final affinity value is estimated. From comparative experiments with recent approaches, we conclude that our model outperforms the state-of-the-art approaches. However, our model construction did not show binding locations within raw sequence data, which provide clinically interpretable results. Moreover, we applied our model for estimating the binding affinities between SARS-CoV-2 and FDA drugs for predicting optimal antiviral inhibitors, and we find that some of our models predicted output had been approved for studies or clinical trials.

In future work, we are intended to extend our model to negative samples of DT pairs in binary classification-based DTI. Further, we intend to exploit the generative approach along with heterogeneous graph networks for drug repurposing. Additionally, we will adapt our approach to addressing multi-target interactions and we will explore recent advances in the language model for generating contextual embedding for protein sequences.

REFERENCES

[1] P. Saraon et al., "A drug discovery platform to identify compounds that inhibit EGFR triple mutants," Nature Chem. Biol., vol. 16, no. 5, pp. 577–586, May 2020.

[2] A. D. Hingorani, V. Kuan, C. Finan, F. A. Kruger, A. Gaulton, S. Chopade, R. Sofat, R. J. MacAllister, J. P. Overington, H. Hemingway, S. Denaxas, D. Prieto, and J. P. Casas, "Improving the odds of drug development success through human genomics: Modelling study," Sci. Rep., vol. 9, no. 1, Dec. 2019, Art. no. 18911.

[3] D. Mendez et al., "ChEMBL: Towards direct deposition of bioassay data," Nucleic Acids Res., vol. 47, no. D1, pp. D930–D940, Jan. 2019.

[4] S. Kim, J. Chen, T. Cheng, A. Gindulyte, J. He, S. He, Q. Li, B. A. Shoemaker, P. A. Thiessen, B. Yu, L. Zaslavsky, J. Zhang, and E. E. Bolton, "PubChem 2019 update: Improved access to chemical data," Nucleic Acids Res., vol. 47, no. D1, pp. D1102–D1109, Jan. 2019.

[5] D. S. Wishart et al., "DrugBank 5.0: A major update to the DrugBank database for 2018," Nucleic Acids Res., vol. 46, no. D1, pp. D1074–D1082, Jan. 2018.

[6] Y.-B. Wang, Z.-H. You, S. Yang, H.-C. Yi, Z.-H. Chen, and K. Zheng, "A deep learning-based method for drug-target interaction prediction based on long short-term memory neural network," BMC Med. Informat. Decis. Making, vol. 20, no. S2, Mar. 2020, Art. no. 49, doi: 10.1186/s12911-020-1052-0.

[7] H. Öztürk, E. Ozkirimli, and A. Özgür, "WideDTA: Prediction of drug-target binding affinity," 2019, arXiv:1902.04166. [Online]. Available: http://arxiv.org/abs/1902.04166

[8] A. S. Rifaioglu, E. Nalbat, V. Atalay, M. J. Martin, R. Cetin-Atalay, and T. Doğan, "DEEPScreen: High performance drug-target interaction prediction with convolutional neural networks using 2-D structural compound representations," Chem. Sci., vol. 11, pp. 2531–2557, Jan. 2020.

[9] N. S. Madhukar, P. K. Khade, L. Huang, K. Gayvert, G. Galletti, M. Stogniew, J. E. Allen, P. Giannakakou, and O. Elemento, "A Bayesian machine learning approach for drug target identification using diverse data types," Nature Commun., vol. 10, no. 1, Dec. 2019, Art. no. 5221.

[10] S. Hu, C. Zhang, P. Chen, P. Gu, J. Zhang, and B. Wang, "Predicting drug-target interactions from drug structure and protein sequence using novel convolutional neural networks," BMC Bioinf., vol. 20, no. S25, Dec. 2019, Art. no. 689, doi: 10.1186/s12859-019-3263-x.

[11] T. Nguyen, H. Le, and S. Venkatesh, "GraphDTA: Prediction of drug-target binding affinity using graph convolutional networks," BioRxiv, 2019, doi: 10.1101/684662.

[12] X. Wang, Y. Liu, F. Lu, H. Li, P. Gao, and D. Wei, "Dipeptide frequency of word frequency and graph convolutional networks for DTA prediction," Frontiers Bioeng. Biotechnol., vol. 8, p. 267, Apr. 2020, doi: 10.3389/fbioe.2020.00267.

[13] L. Zhao, J. Wang, L. Pang, Y. Liu, and J. Zhang, "GANsDTA: Predicting drug-target binding affinity using GANs," Frontiers Genet., vol. 10, p. 1243, Jan. 2020, doi: 10.3389/fgene.2019.01243.

[14] Q. Zhao, F. Xiao, M. Yang, Y. Li, and J. Wang, "AttentionDTA: Prediction of drug–target binding affinity using attention model," in Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM), Nov. 2019, pp. 64–69, doi: 10.1109/BIBM47256.2019.8983125.

[15] M. Kandeel and M. Al-Nazawi, "Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease," Life Sci., vol. 251, Jun. 2020, Art. no. 117627, doi: 10.1016/j.lfs.2020.117627.

[16] H. Zhang, J. M. Penninger, Y. Li, N. Zhong, and A. S. Slutsky, "Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: Molecular mechanisms and potential therapeutic target," Intensive Care Med., vol. 46, no. 4, pp. 586–590, Apr. 2020.

[17] M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N.-H. Wu, A. Nitsche, and M. A. Müller, "SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor," Cell, vol. 181, no. 2, pp. 271–280, 2020, doi: 10.1016/j.cell.2020.02.052.

[18] Middle East Respiratory Syndrome Coronavirus (MERS-CoV), World Health Org., Geneva, Switzerland, 2020.

[19] Y. Zhou, Y. Hou, J. Shen, Y. Huang, W. Martin, and F. Cheng, "Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2," Cell Discovery, vol. 6, no. 1, Dec. 2020, Art. no. 14, doi: 10.1038/s41421-020-0153-3.

[20] R. J. Vidmar, "On the use of atmospheric plasmas as electromagnetic reflectors (online source style)," IEEE Trans. Plasma Sci., vol. 21, no. 3, pp. 876–880, Aug. 1992. [Online]. Available: http://www.halcyon.com/pub/journals/21ps03-vidmar

[21] T. S. Fung and D. X. Liu, "Human coronavirus: Host-pathogen interaction," Annu. Rev. Microbiol., vol. 73, no. 1, pp. 529–557, Sep. 2019.

[22] Y.-I. Kim et al., "Infection and rapid transmission of SARS-CoV-2 in ferrets," Cell Host Microbe, vol. 27, no. 5, pp. 704–709, 2020, doi: 10.1016/j.chom.2020.03.023.

[23] X. Wang, H. Ji, C. Shi, B. Wang, Y. Ye, P. Cui, and P. S. Yu, "Heterogeneous graph attention network," in Proc. World Wide Web Conf., 2019, pp. 2022–2032, doi: 10.1145/3308558.3313562.

[24] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in Proc. Adv. Neural Inf. Process. Syst., 2015, pp. 802–810.

[25] D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," J. Chem. Inf. Model., vol. 28, no. 1, pp. 31–36, Feb. 1988.

[26] T. He, M. Heidemeyer, F. Ban, A. Cherkasov, and M. Ester, "SimBoost: A read-across approach for predicting drug–target binding affinities using gradient boosting machines," J. Cheminform., vol. 9, no. 1, Dec. 2017, Art. no. 24, doi: 10.1186/s13321-017-0209-z.

[27] T. Pahikkala, A. Airola, S. Pietila, S. Shakyawar, A. Szwajda, J. Tang, and T. Aittokallio, "Toward more realistic drug-target interaction predictions," Briefings Bioinf., vol. 16, no. 2, pp. 325–337, Mar. 2015.

[28] M. Tsubaki, K. Tomii, and J. Sese, "Compound–protein interaction prediction with end-to-end learning of neural networks for graphs and sequences," Bioinformatics, vol. 35, no. 2, pp. 309–318, Jan. 2019.

[29] H. Öztürk, A. Özgür, and E. Ozkirimli, "DeepDTA: Deep drug–target binding affinity prediction," *Bioinformatics*, vol. 34, no. 17, pp. i821–i829, Sep. 2018.

[30] X. Lin, "DeepGS: Deep representation learning of graphs and sequences for drug-target binding affinity prediction," 2020, *arXiv:2003.13902*. [Online]. Available: http://arxiv.org/abs/2003.13902

[31] B. R. Beck, B. Shin, Y. Choi, S. Park, and K. Kang, "Predicting commercially available antiviral drugs that may act on the novel coronavirus (SARS-CoV-2) through a drug-target interaction deep learning model," *Comput. Struct. Biotechnol. J.*, vol. 18, pp. 784–790, Mar. 2020, doi: 10.1016/j.csbj.2020.03.025.

[32] B. Shin, S. Park, K. Kang, and J. C. Ho, "Self-attention based molecule representation for predicting drug-target interaction," 2019, *arXiv:1908.06760*. [Online]. Available: http://arxiv.org/abs/1908.06760

[33] F. Hu, J. Jiang, and P. Yin, "Prediction of potential commercially inhibitors against SARS-CoV-2 by multi-task deep model," 2020, *arXiv:2003.00728*. [Online]. Available: http://arxiv.org/abs/2003.00728

[34] Y. Ge *et al.*, "A data-driven drug repositioning framework discovered a potential therapeutic agent targeting COVID-19," *BioRxiv*, 2020, doi: 10.1101/2020.03.11.986836.

[35] B. Tang, F. He, D. Liu, M. Fang, Z. Wu, and D. Xu, "AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2," *BioRxiv*, 2020, doi: 10.1101/2020.03.03.972133.

[36] A. Zhavoronkov, V. Aladinskiy, A. Zhebrak, B. Zagribelnyy, V. Terentiev, D. S. Bezrukov, D. Polykovskiy, and Y. Yan, "Potential COVID-2019 3C-like protease inhibitors designed using generative deep learning approaches," Insilico Med. Hong Kong, Tech. Rep., 2020, vol. 307, doi: 10.26434/chemrxiv.12301457.v1.

[37] D. D. Nguyen, K. Gao, R. Wang, and G.-W. Wei, "Machine intelligence design of 2019-nCoV drugs," *BioRxiv*, 2020, doi: 10.1101/2020.01.30.927889.

[38] H. Zhang, K. M. Saravanan, Y. Yang, M. T. Hossain, J. Li, X. Ren, Y. Pan, and Y. Wei, "Deep learning based drug screening for novel coronavirus 2019-nCov," *Interdiscipl. Sci., Comput. Life Sci.*, vol. 12, no. 3, pp. 368–376, Sep. 2020, doi: 10.1007/s12539-020-00376-6.

[39] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[40] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[41] Z. Quan, X. Lin, Z.-J. Wang, Y. Liu, F. Wang, and K. Li, "A system for learning atoms based on long short-term memory recurrent neural networks," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2018, pp. 728–733.

[42] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 3111–3119.

[43] M. I. Davis, J. P. Hunt, S. Herrgard, P. Ciceri, L. M. Wodicka, G. Pallares, M. Hocker, D. K. Treiber, and P. P. Zarrinkar, "Comprehensive analysis of kinase inhibitor selectivity," *Nature Biotechnol.*, vol. 29, no. 11, pp. 1046–1051, Nov. 2011.

[44] J. Tang, A. Szwajda, S. Shakyawar, T. Xu, P. Hintsanen, K. Wennerberg, and T. Aittokallio, "Making sense of large-scale kinase inhibitor bioactivity data sets: A comparative and integrative analysis," *J. Chem. Inf. Model.*, vol. 54, no. 3, pp. 735–743, Mar. 2014.

[45] T. Liu, Y. Lin, X. Wen, R. N. Jorissen, and M. K. Gilson, "BindingDB: A Web-accessible database of experimentally determined protein-ligand binding affinities," *Nucleic Acids Res.*, vol. 35, pp. D198–D201, Jan. 2007.

[46] G. V. Paolini, R. H. B. Shapland, W. P. van Hoorn, J. S. Mason, and A. L. Hopkins, "Global mapping of pharmacological space," *Nature Biotechnol.*, vol. 24, no. 7, pp. 805–815, Jul. 2006.

[47] M. Wang, R. Cao, L. Zhang, X. Yang, J. Liu, M. Xu, Z. Shi, Z. Hu, W. Zhong, and G. Xiao, "Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) *in vitro*," *Cell Res.*, vol. 30, no. 3, pp. 269–271, Mar. 2020.

[48] L. Caly, J. D. Druce, M. G. Catton, D. A. Jans, and K. M. Wagstaff, "The FDA-approved drug ivermectin inhibits the replication of SARS-CoV-2 *in vitro*," *Antiviral Res.*, vol. 178, Jun. 2020, Art. no. 104787, doi: 10.1016/j.antiviral.2020.104787.

[49] E. G. Favalli, M. Biggioggero, G. Maioli, and R. Caporali, "Baricitinib for COVID-19: A suitable treatment?" *Lancet Infectious Diseases*, vol. 20, no. 9, pp. P1012–P1013, 2020, doi: 10.1016/S1473-3099(20)30262-0.

[50] M. P. Lythgoe and P. Middleton, "Ongoing clinical trials for the management of the COVID-19 pandemic," *Trends Pharmacological Sci.*, vol. 41, no. 6, pp. 363–382, Jun. 2020, doi: 10.1016/j.tips.2020.03.006.

[51] A. A. Elfiky, "Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study," *Life Sci.*, vol. 253, Jul. 2020, Art. no. 117592, doi: 10.1016/j.lfs.2020.117592.

[52] D. C. Hall and H.-F. Ji, "A search for medications to treat COVID-19 via in silico molecular docking models of the SARS-CoV-2 spike glycoprotein and 3CL protease," *Travel Med. Infectious Disease*, vol. 35, May 2020, Art. no. 101646, doi: 10.1016/j.tmaid.2020.101646.

[53] P. Calligari, S. Bobone, G. Ricci, and A. Bocedi, "Molecular investigation of SARS–CoV-2 proteins and their interactions with antiviral drugs," *Viruses*, vol. 12, no. 4, p. 445, Apr. 2020, doi: 10.3390/v12040445.

[54] B. Shah, P. Modi, and S. R. Sagar, "In silico studies on therapeutic agents for COVID-19: Drug repurposing approach," *Life Sci.*, vol. 252, Jul. 2020, Art. no. 117652, doi: 10.1016/j.lfs.2020.117652.

[55] H. Stower, "Lopinavir–Ritonavir in severe COVID-19," *Nature Med.*, vol. 26, no. 4, Apr. 2020, Art. no. 465, doi: 10.1038/s41591-020-0849-9.

[56] Z. Zhu, Z. Lu, T. Xu, C. Chen, G. Yang, T. Zha, J. Lu, and Y. Xue, "Arbidol monotherapy is superior to lopinavir/ritonavir in treating COVID-19," *J. Infection*, vol. 81, no. 1, pp. e21–e23, Jul. 2020, doi: 10.1016/j.jinf.2020.03.060.

[57] C. Sargiacomo, F. Sotgia, and M. P. Lisanti, "COVID-19 and chronological aging: Senolytics and other anti-aging drugs for the treatment or prevention of corona virus infection?" *Aging*, vol. 12, no. 8, pp. 6511–6517, Mar. 2020, doi: 10.18632/aging.103001.

[58] D. Jakhar and I. Kaur, "Potential of chloroquine and hydroxychloroquine to treat COVID-19 causes fears of shortages among people with systemic lupus erythematosus," *Nature Med.*, vol. 26, no. 5, p. 632, May 2020, doi: 10.1038/s41591-020-0853-0.

[59] P. C. Robinson and J. Yazdany, "The COVID-19 global rheumatology alliance: Collecting data in a pandemic," *Nature Rev. Rheumatol.*, vol. 16, no. 6, pp. 293–294, Jun. 2020, doi: 10.1038/s41584-020-0418-0.

[60] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," 2016, *arXiv:1609.02907*. [Online]. Available: http://arxiv.org/abs/1609.02907

[61] W. Song, Z. Xiao, Y. Wang, L. Charlin, M. Zhang, and J. Tang, "Session-based social recommendation via dynamic graph attention networks," in *Proc. 12th ACM Int. Conf. Web Search Data Mining*, Jan. 2019, pp. 555–563.

[62] E. Gawehn, J. A. Hiss, J. B. Brown, and G. Schneider, "Advancing drug discovery via GPU-based deep learning," *Expert Opinion Drug Discovery*, vol. 13, no. 7, pp. 579–582, Jul. 2018, doi: 10.1080/17460441.2018.1465407.

[63] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, "BioBERT: A pre-trained biomedical language representation model for biomedical text mining," *Bioinformatics*, vol. 36, pp. 1234–1240, Sep. 2019.

[64] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov, and Q. V. Le, "XLNet: Generalized autoregressive pretraining for language understanding," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 5753–5763.

MOHAMED ABDEL-BASSET received the B.Sc., M.Sc., and Ph.D. degrees in operations research from the Faculty of Computers and Informatics, Zagazig University, Egypt. He is currently the Head of the Department of Computer Science, Faculty of Computers and Informatics, Zagazig University. He is also working on application of multiobjective and robust meta-heuristic optimization techniques. He has published more than 100 articles in international journals and conference proceedings. His current research interests include optimization, operations research, data mining, computational intelligence, applied statistics, decision support systems, robust optimization, engineering optimization, multiobjective optimization, swarm intelligence, evolutionary algorithms, and artificial neural networks. He serves as an/a editor/reviewer for different international journals and conferences.

HOSSAM HAWASH received the B.Sc. degree from the Department of Computer Science, Faculty of Computers and Informatics, Zagazig University, Egypt. His research interests include optimization, deep learning algorithms, swarm intelligence, evolutionary algorithms, and artificial neural networks.

RIPON K. CHAKRABORTTY (Member, IEEE) received the B.Sc. and M.Sc. degrees in industrial and production engineering from the Bangladesh University of Engineering and Technology, in 2009 and 2013, respectively, and the Ph.D. degree in computer science from the University of New South Wales (UNSW Australia), Canberra, in 2017. He is currently a Lecturer in system engineering and project management with the School of Engineering and Information Technology, UNSW Australia. He has written two book chapters and over 50 technical journal and conference papers. His research interests include operations research, optimization problems, project management, supply chain management, and information systems management.

MOHAMED ELHOSENY (Senior Member, IEEE) received the B.Sc. and master's degrees from Mansoura University, in 2006 and 2010, respectively, and the Ph.D. degree in computers and information (Scientific Program) with the University of North Texas and Mansoura University, in 2015. He is currently an Assistant Professor with Mansoura University. He is an ACM Distinguished Speaker. He serves as an Associate Editor for the IEEE JOURNAL OF BIOMEDICAL HEALTH INFORMATICS, IEEE ACCESS, *Scientific Reports*, and *Remote Sensing*. He serves as the Editor in Chief for the *International Journal of Smart Sensor Technologies and Applications* (IGI Global). He also serves as the EiC for two book series published by Taylor&Francis and one book series by Springer.

MICHAEL RYAN (Senior Member, IEEE) received the bachelor's, master's, and Doctor of Philosophy degrees in engineering from the University of New South Wales Canberra. He has completed two years formal Engineering Management Training, U.K. He is currently the Director of the Capability Systems Centre, University of New South Wales Canberra. He also lectures and regularly consults a range of subjects, including communications systems, systems engineering, requirements engineering, and project management. He is the author or coauthor of 12 books, three book chapters, and over 250 technical articles and reports. He is a Fellow of Engineers Australia, a Fellow of the International Council on Systems Engineering, and a Fellow of the Institute of Managers and Leaders. He serves as the Co-Chair for the Requirements Working Group, International Council on Systems Engineering (INCOSE).

· · ·