IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES
ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 11, Iss 03, 2022

Exploratory study on the Applications of Deep Learning in Pharmaceutical Drug Discovery Learning

Ramandeep Kaur, Kirna Devi
Guru Kashi University, Talwandi Sabo

Abstract

Using a morphological imaging dataset from Recursion Pharmaceuticals, this exploratory work explores the use of deep learning in pharmaceutical drug development. The dataset, which was created in April 2023, includes 309,522 5-channel images showing how cells were treated in 1738 microplates. Three different cell treatments are included in the research methodology: fake cells, cells containing irradiated SARS-CoV-2 virons, and cells infected with active SARS-CoV-2 that have been treated with drug library candidates. The resulting RxRx19a dataset is an essential tool for studying the effects of compounds on human kidney cells infected with SARS-CoV-2. It is organised by FDA, EMA, and clinical trial chemicals. The cytopathic character of the SARS-CoV-2 virus is shown by electron microscopy data, which advise the utilization regarding persistent renal trade treatment for Coronavirus patients who have intense kidney injury. Additionally, the Rx1 dataset—which is treated with siRNA rather than compounds—is included in the article, facilitating the first training of a classification model. The SARS-CoV-2 dataset's specific viral circumstances are used to fine-tune the model training strategy, which is a two-step process that uses the heterogeneous siRNA dataset. Experiments show that DenseNet is better than other deep neural network designs at classifying siRNA pictures, and on the SARS-CoV-2 dataset, a cascade transfer learning model performs better than other models. The model is then used to rank the effectiveness of various drugs in treating COVID-19, thereby finding possible research prospects. This thorough investigation highlights the potential of deep learning to improve drug discovery, especially when it comes to viral illnesses such as SARS-CoV-2.

Keywords: *Deep learning, Pharmaceutical drug discovery, cascade transfer learning, DenseNet, Artificial intelligence.*

1. INTRODUCTION

The goal of drug discovery is to identify secure and efficient treatments for diseases that affect people. Drug development is a time-consuming and expensive process that involves everything from target discovery to meticulous clinical testing. Drug discovery is the study of how different medications interact with the body and what changes the body must undergo for a medication to have a therapeutic effect [1]. Physiology-based and target-based techniques are two of the various methodologies that make up drug discovery strategies. This tactic is predicated on knowledge about the target and ligand. Specifically, drug sensitivity and response, drug-drug interaction, drug-drug similarity, and drug-target interactions were the subjects we focused on in this respect [2]. To improve prognosis and pathogenesis interactions, many drug combinations are needed for some diseases, like cancer, or pandemic scenarios like COVID-19. Medication creation remains an expensive and labor-intensive

procedure despite all the recent advancements in the pharmaceutical industry [3]. Consequently, a number of computational algorithms are put out to expedite the process of discovering new drugs.

Large pharmaceutical companies have also shifted towards AI in response to the advancement of DL techniques, forgoing antiquated, ineffective practices in an effort to boost patient profit while simultaneously improving their own [4]. Even with DL's remarkable performance, drug discovery is still a crucial and difficult endeavour, and there is still need for study into developing a number of algorithms that enhance drug discovery [5].

1.1. Deep Learning Methods Used In Drug Discovery

A great place to start if you're interested in leading edge research and development is with deep learning algorithms. One of the main tenets of deep learning is the translation of artificial neural networks (ANNs), which were initially developed in the 1950s, from theoretical and expected applications to practical algorithms [6]. With DL techniques, one can learn through abstraction to represent multidimensional data. The application of DL methods has proven beneficial for lead compound development, drug activity prediction, and target discovery. The DL foundations are frequently utilised in NN systems to build systems that can comprehend, generate, and recognise complex data [7].

1.2. DL applications in drug discovery problems

Applications of deep learning (DL) in drug discovery are essential to the transformation of the pharmaceutical sector. DL models use sophisticated neural network topologies to help with multiple aspects of drug discovery, such as target selection, virtual screening, and molecular property prediction [8]. These models perform exceptionally well in virtual high-throughput screening, generative molecular design, and quantitative structure-activity relationship (QSAR) prediction. Additionally, DL is useful for predicting clinical outcomes, interpreting biological pictures, and understanding cellular responses. These applications enable personalised medicine and enhance patient stratification in clinical trials [9]. Additionally, by utilising multi-omics data and natural language processing for knowledge extraction from scientific literature, DL models support drug repurposing, adverse event anticipation, and toxicity prediction [10]. The subject of drug discovery is in need of continuous research and improvement due to obstacles including data quality, interpretability, and regulatory considerations, despite the unparalleled prospects it presents to improve efficiency and success rates [11].

2. LITERATURE REVIEW

Arshed et al. (2022) [12] offered a cutting-edge deep learning framework that incorporates drug chemical substructure analysis to predict the negative effects of several drugs. Their research, which was published in the International Journal of Innovative Science and Technology, focuses on using deep learning capabilities to improve the accuracy of drug side effect predictions. This makes a significant contribution to the field of drug safety evaluation, which is constantly changing.

Nag et al. (2022) [13] examined the deeper field of deep learning instruments for pharmaceutical research and discovery. Their study, which was published in 3 Biotech, demonstrates the many uses for deep learning and how it could completely transform the drug development process at different phases. The study explores how to anticipate pharmacological qualities, optimise molecular structures, and find new therapeutic candidates more quickly by using sophisticated computational tools.

Ren et al. (2022) [14] provide a de novo cell-drug sensitivity prediction model using graph regularised matrix factorization based on deep learning. This novel method incorporates matrix factorization with graph regularisation, showing promise in precisely forecasting cellular susceptibilities to different medications.

Song et al. (2022) [15] introducing Deep Fusion, a deep learning-based multi-scale feature fusion technique created especially for drug-target interaction prediction. This model demonstrates its potential to improve drug-target interaction prediction accuracy by integrating various parameters at many scales.

3. RESEARCH METHODOLOGY

3.1. Data Collection

We utilised a morphological imaging dataset of 309,522 5-channel pictures that Recursion Pharmaceuticals supplied in order to maximise the effectiveness of our research. Recursion created this dataset in April 2023, which included cell treatment in 1738 microplates. 5% paraformaldehyde was used in the fixation procedure. 0.25% Triton was then used to permeabilize the material, and several dyes were used for staining.

3.2. Cell Treatments

Three distinct cell therapies were given: lighted and deactivated SARS-CoV-2 virons, cells with farce cells as a control, and cells contaminated with dynamic SARS-CoV-2 at a proportion of 4 virons to 10 cells. A wide range of potential compounds from a medication library, including those approved by the FDA and EMA and those undergoing SARS clinical studies, were used to treat these cells. RxRx19a was the dataset that was produced.

3.3. Drug Library and Compound Treatments

The RxRx19a dataset contained the medications, which were divided into three categories: FDA approved, EMA authorised, and those undergoing clinical studies for SARS. Every chemical was given six times at six different concentrations. This dataset is an essential tool for researching how different substances affect human kidney cells infected with SARS-CoV-2.

3.4. SARS-CoV-2 Virus Characteristics

Patients who contracted the virus had virus-like particles inside their kidney cells as a result of the SARS-CoV-2 virus's targeting of human kidney cells. Direct tubular damage via cytotoxicity and a cytopathic character were demonstrated by electron microscope

investigations. The utilization of ceaseless renal substitution treatment for Coronavirus patients with intense kidney injury was directed by this comprehension.

3.5. Creation of RxRx19a Dataset

Recursion made use of VERO cells from African green monkey kidney cells and human renal cortical epithelial cells (HRCE). The RxRx19a dataset was different from the RxRx19 dataset in that it had 1138 classes of siRNA, a 512 × 512 image resolution with 6 channels, and a 384-well plate density.

3.6. siRNA Dataset

Compared to RxRx19, the Rx1 dataset had differences in plate density, image resolution, channels, and classes due to its higher size. siRNA was used to treat the cells in place of the target chemicals. This dataset, which included 1139 types of siRNA applications, could be used to train the classification classifier initially since it shared commonalities with the SARS-CoV-2 dataset.

3.7. Model Training Strategy

Using the bigger siRNA dataset, we first trained the classification model according to our plan. The SARS-CoV-2 dataset's viral/mock cells were then used to retrain the model. Using the diversity of the siRNA dataset, this two-step method created a strong initial model that was then fine-tuned with particular viral circumstances to increase the model's efficacy in identifying features associated to SARS-CoV-2.

4. EXPERIMENTAL RESULTS AND DISCUSSION

MATLAB software was used to implement the research on a computer server equipped with a 256 GB RAM Nvidia DGX Workstation. Using DenseNet161, we first developed a model for classifying siRNA pictures. We then contrasted this model with other pretrained models, including VGG10, AlexNet, and GoogleNet. The classification outcomes of several models employing siRNA image datasets are displayed in Table 1. The table makes it evident that DenseNet generated the best accuracy performance.

Table 1: Results of experiments using several pre-trained models on the siRNA dataset

Model	Accuracy
VGG16	82.4
GoogleNet	84.3
VGG19	84.5
AlexNet	78.6
DenseNet	97.6

The experimental outcomes for a variety of pre-trained models on the siRNA dataset are shown in Table 1, along with accuracy values for each model. With an accuracy of 97.6%, DenseNet performed the best out of all the models, demonstrating its greater capacity to classify siRNA pictures. By contrast, the accuracy of VGG16 was 82.4%, that of GoogleNet

ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 11, Iss 03, 2022

was 84.3%, that of VGG19 was 84.5%, and that of AlexNet was 78.6%. When compared to the other models that were evaluated, these findings demonstrate the clear benefit of using DenseNet for siRNA picture classification, as well as its effectiveness in recognising complex patterns in the dataset.

Particularity measures the exactness of a test's invalid expectation, while responsiveness goes about as a norm for a test's positive expectation precision. Within the framework of our investigation, a model's sensitivity refers to its capacity to identify viral cells that are active, whereas its specificity pertains to its capacity to identify mock or control cells. The model's capacity to prevent false positives and false negatives is indicated by the F1 score, which statistically quantifies the balance between sensitivity and specificity. Last but not least, the Kappa score is a statistical tool for gauging agreement across all examined cases on a range of less than or equal to 1, where a score of less than 0 denotes minimal agreement and a score of 1 complete agreement. Table 2 shows that the cascade model fared better compared to other broadly utilized deep neural organization structures. Our cascade transfer learning model outflanked the other two notable deep learning models for the RxRx19 dataset with regards to the predetermined measurements. It ought to be referenced that before being retrained on the SARS-CoV-2 dataset, the vgg19 and GoogleNet models were pre-prepared on the ImageNet dataset as opposed to the siRNA dataset.

Table 2: Results of experiments using several pre-trained models on the SARS-CoV-2 dataset

Deep Learning Model	Sensitivity	Specificity	F1-Score	Kappa
vgg19	0.88	1.00	0.92	0.76
GoogleNet	0.86	0.99	0.86	0.72
Cascade Transfer Learning	0.99	1.01	0.99	0.89

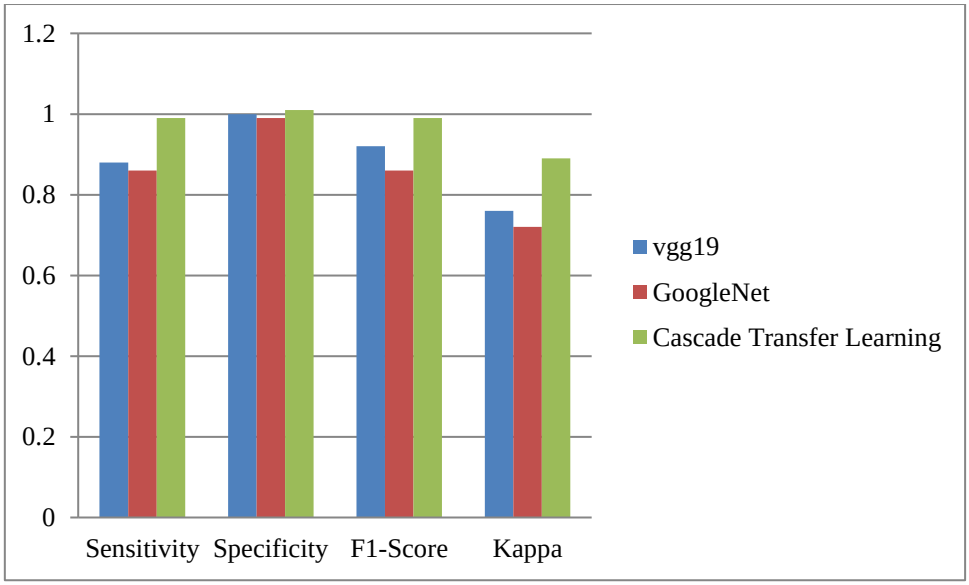

Figure 1: Visual depiction of experimental outcomes using several trained models on the SARS-CoV-2 dataset

ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 11, Iss 03, 2022

Key execution measures are given by Table 2, which shows the exploratory discoveries for different pre-prepared models on the SARS-CoV-2 dataset. With a responsiveness of 0.99, particularity of 1.01, F1-score of 0.99, and Kappa score of 0.89, the cascade transfer learning model performed very well. Nearly, GoogleNet showed a responsiveness of 0.86, particularity of 0.99, F1-score of 0.86, and Kappa score of 0.72, though vgg19 showed a responsiveness of 0.88, explicitness of 1.00, F1-score of 0.92, and Kappa score of 0.76. The discoveries of this study show that the cascade transfer learning model displayed better execution thought about than vgg19 and GoogleNet in totally surveyed measurements. This highlights the model's ability to reliably identify features in the SARS-CoV-2 dataset and its potential for efficient implementation in the context of viral cell classification.

The viability of the drugs used to treat Coronavirus was then positioned utilizing the cascade transfer learning model, as shown in the last block of Figure 5, with a score of under 0.5 demonstrating the drug's true capacity as a lead. Tables 3 and 4 present the results of these tests.

Table 3: List of substances labelled as fake that have low probability scores

Compound	Probability
GS-441524	0.10
Remdesivir (GS-5734)	0.10
CX-4945	0.15
Aloxistatin	0.20
Calcipotriene	0.23

A list of chemicals designated as fake is given in Table 3, along with the low probability scores that correspond to them. Notably, probability values of 0.10 were obtained by Remdesivir (GS-5734) and GS-441524, indicating a significant probability of being categorised as mock compounds. With a likelihood score of 0.15, CX-4945 comes in second, followed by 0.20 for Aloxistatin and 0.23 for Calcipotriene. These low probability ratings highlight the possibility that these compounds should not be further investigated in drug effectiveness studies for COVID-19 treatment, since they strongly suggest that they may not be effective against the target.

Table 4: List of Compounds that Are Classified as Active Viruses and Have High Probability Scores

Compound	Probability
Sertaconazole	1.00
PKC 412	0.99
L-Adrenaline	0.99
Isoetharine	0.98
desoximetasone	0.97

Compounds designated as active viruses are included in Table 4 along with the high probability ratings that correspond to them. Sertaconazole, in particular, has an exceptional

ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 11, Iss 03, 2022

probability score of 1.00, suggesting a high chance of being a successful antiviral medication. In a similar vein, PKC 412 and L-adrenaline both had high ratings of 0.99, with isoetharine and desoximetasone scoring 0.98 and 0.97, respectively, following closely behind. These compounds have a strong potential to display antiviral activity, as indicated by their raised likelihood scores. This suggests that these compounds could be worth further exploration as potential treatments for COVID-19.

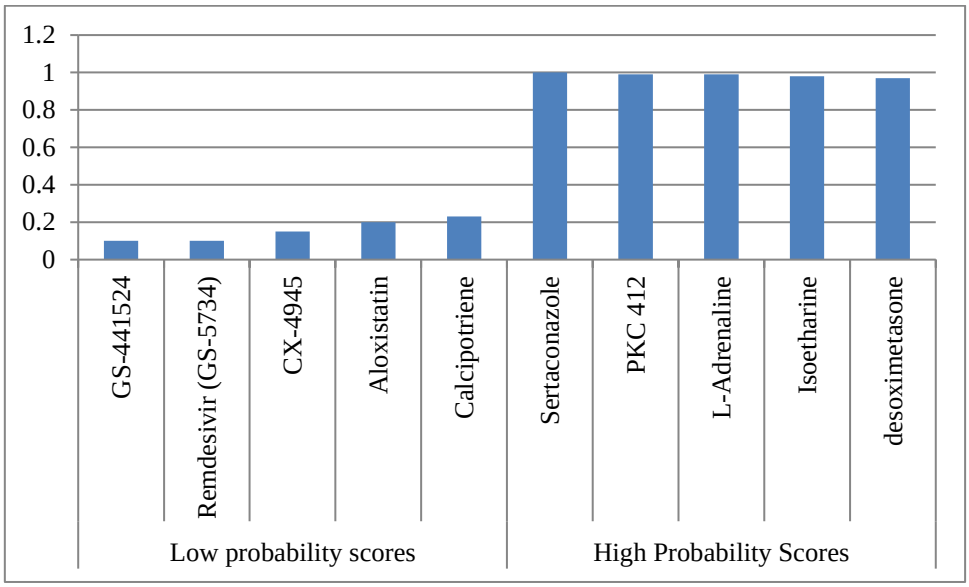

Figure 2: A list of chemicals with varying likelihood scores represented graphically

5. CONCLUSION

This exploratory work has demonstrated a thorough research technique and offered important insights into the applications of deep learning in pharmaceutical drug development. The study investigated cell treatments, drug library usage, and the development of the RxRx19a dataset, focusing on SARS-CoV-2 virus features, using a morphological imaging dataset from Recursion Pharmaceuticals. A two-step model training technique was made possible by the integration of the Rx1 dataset, highlighting the significance of varied datasets in boosting the efficacy of the model. The experimental results, especially Table 1, demonstrated how well the DenseNet model outperformed other pre-trained algorithms in precisely categorising siRNA pictures. The remarkable results of the cascade transfer learning model in terms of sensitivity, specificity, F1-score, and Kappa score were further evaluated using the SARS-CoV-2 dataset, as shown in Table 2. As displayed in Tables 3 and 4, the utilization of this model to rank the viability of mixtures for Corona virus treatment gave smart data to conceivable prescription competitors. All things considered, this work shows how deep learning can help with medication discovery, especially when it comes to viral illnesses like SARS-CoV-2. **REFERENCES**

[1] Dr. D. P. Kothari. (2022). Development of Multimedia Signal Processing and Its Technology. Acta Energetica, (02), 36–43.

[2] Kimber, T.B.; Chen, Y.; Volkamer, A. Deep learning in virtual screening: Recent applications and developments. Int. J. Mol. Sci. 2021, 22, 4435.

[3] Mouchlis, V.D.; Afantitis, A.; Serra, A.; Fratello, M.; Papadiamantis, A.G.; Aidinis, V.; Lynch, I.; Greco, D.; Melagraki, G. Advances in de novo drug design: From conventional to machine learning methods. Int. J. Mol. Sci. 2021, 22, 1–22.

[4] Pouryahya M, Oh JH, Mathews JC, Belkhatir Z, Moosmüller C, Deasy JO, Tannenbaum AR (2022) Pan-cancer prediction of cell-line drug sensitivity using network-based methods. Int J Mol Sci 23:1074.

[5] Rolf Bracke, & Nouby M. Ghazaly. (2022). An Exploratory Study of Sharing Research Energy Resource Data and Intellectual Property Law in Electrical Patients. Acta Energetica, (01), 01–07.

[6] Saberian, M.S.; Moriarty, K.P.; Olmstead, A.D.; Nabi, I.R.; Jean, F.; Libbrecht, M.W.; Hamarneh, G. DEEMD: Drug Efficacy Estimation against SARS-CoV-2 based on cell Morphology with Deep multiple instance learning. arXiv 2021, arXiv:2105.05758.

[7] Thafar MA, Alshahrani M, Albaradei S et al (2022) Afnity2Vec: drug–target binding afnity prediction through representation learning, graph mining, and machine learning. Sci Rep 12:4751.

[8] Townshend, R.J.L.; Powers, A.; Eismann, S.; Derry, A. ATOM3D: Tasks On Molecules in Three Dimensions. arXiv 2021, arXiv:2012.04035.

[9] Yan C, Duan G, Zhang Y, Wu F-X, Pan Y, Wang J (2022) Predicting drug–drug interactions based on integrated similarity and semi-supervised learning. IEEE/ACM Trans Comput Biol Bioinf 19(1):168–179.

[10] Zhang C, Lu Y, Zang T (2022) CNN-DDI: a learning-based method for predicting drug–drug interactions using convolution neural networks. BMC Bioinf 23:88.

[11] Zhou Y, Zhang Y, Lian X, Li F, Wang C, Zhu F, Qiu Y, Chen Y (2022) Therapeutic target database update 2022: facilitating drug discovery with enriched comparative data of targeted agents. Nucleic Acids Res 50:1398–1407

[12] Arshed MA, Mumtaz S, Riaz O, Sharif W, Abdullah S (2022) A deep learning framework for multi drug side efects prediction with drug chemical substructure. Int J Innovat Sci Technol 4(1):19–31.

[13] Nag S, Baidya ATK, Mandal A et al (2022) Deep learning tools for advancing drug discovery and development. 3 Biotech 12:110.

[14] Ren S, Tao Y, Yu K et al (2022) De novo prediction of Cell-Drug sensitivities using deep learning-based graph regularized matrix factorization. Pacif Symp Biocomput.

ISSN PRINT 2319 1775 Online 2320 7876
 Journal Volume 11, Iss 03, 2022

[15] Song T, Zhang X, Ding M, Rodriguez-Paton A, Wang S, Wang G (2022) DeepFusion: a deep learning based multi-scale feature fusion method for predicting drug–target interactions. Methods 204:269–277.