IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 March 2024, accepted 21 March 2024, date of publication 2 April 2024, date of current version 12 April 2024.

*Digital Object Identifier 10.1109/ACCESS.2024.3382936*

RESEARCH ARTICLE

# Img2Side: A Transfer Learning Based Model for Predicting Drug Side Effects Using 2D Chemical Structural Images

**MUHAMMAD ASAD ARSHED[1], MUHAMMAD IBRAHIM[1], SHAHZAD MUMTAZ[1,2], TENVIR ALI[1], AND GYU SANG CHOI[3], (Member, IEEE)**

[1]Faculty of Computing, The Islamia University of Bahawalpur, Bahawalpur 63100, Pakistan
[2]School of Natural and Computing Sciences, University of Aberdeen, Aberdeen, AB24 3FX Scotland, U.K.
[3]School of Computer Science and Engineering, Yeungnam University, Gyeongsan, Gyeongbuk 38541, South Korea

Corresponding author: Gyu Sang Choi (Castchoi@ynu.ac.kr)

This work was supported by the 2024 Yeungnam University Research Grant.

**ABSTRACT** Drug Side Effects (DSE) are inconvenient and inadvertent retorts of the drugs. DSEs impact on public health and healthcare can prove costly. These DSEs can be an important factor in the failure/acceptance of drugs. Every approved drug should be either free from DSEs or these should be minor and reported properly. The drug discovery process should be capable of predicting and preventing these effects in advance. Previously, proposed studies for the prediction/prevention of DSEs utilized the features of 1D drug chemical structures or Natural Language Processing (NLP). Both these techniques required a complex transformation process. In this research authors have proposed a deep learning model, specifically using a transfer learning approach to predict DSEs directly from 2D chemical structure images, eliminating the need for the hefty transformation process of the NLP domain. For this study, a unique dataset is prepared that associates each image (taken from PubChem) with its specific side effects (SIDER). The results are evaluated using Accuracy, Precision, Recall and F-measure. The proposed model showed its dominance with an Accuracy of 73%, Precision of 83%, Recall of 73%, and an F1 score of 75%. The achieved results of the proposed model are compared against established transfer learning models like VGG16, DenseNet121 and some previously used traditional machine learning models like SVM and KNN. The collected results indicate a significant advancement in predicting drug side effects and offer a promising avenue for streamlining the drug development process.

**INDEX TERMS** Drug side effects, drug 2D chemical structure images, transfer learning, fine tuning, pretrained models, deep learning.

## I. INTRODUCTION

DSEs (adverse reactions) are undesirable and unintended responses to the drug. These DSEs can be from minor problems like headaches to situations that put your life in danger, such as liver damage. Symptoms like dry mouth, upset stomach, fever, vomiting, and drowsiness are considered common DSEs, whereas severe DSEs can cause death, disability, or congenital disabilities. It can happen due to many reasons like using a new drug, stopping a drug that you have been using for a while, and increasing or decreasing its dose [1].

Drug development is a demanding, complex, and costly process that needs expertise, research skills, human effort, chemical understanding, and a bunch of dollars. The major reason for drug development failure is primarily due to DSEs [2]. Consequently, it is necessary to identify and address protective flaws in drug development. Approximately two million people are affected by the adverse effects of drugs worldwide [3], [4]. Drugs are immediately withdrawn (e.g., Rofecoxib) from the market due to their unaccepted

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

DSEs [5]. According to a study, one million people lose their lives every year and DSEs are ranked as the fourth leading cause of death in the US [6].

A drug is launched in the market after the initial stage of 10-15 years and the entire process (from development to launch) can cost a lot of time and money. The average cost for the development of a single drug is approximately $2.6 billion [7]. Furthermore, the study of DSEs is necessary to save cost and time as well as to speed up the drug discovery process. For early prediction of DSEs, efficient and fast methods need to be designed as the existing clinical methods are not efficient.

The pharmacological activity of drug molecules can be regulated with the modification of the drug structure. The physiochemical properties and ultimately effects of a drug are determined by its chemical structure.

### A. NLP-BASED DSEs PREDICTION

Machine learning (ML) models can perform effectively in terms of time and cost for early DSEs identification as compared to clinical methods [8]. Numerous computational techniques have been proposed to identify DSEs. In some studies, a binary classifier has been used for each side effect [6], [9], [10] whereas in some studies SEs and drugs are considered as a pair and uniform classifiers used for the classification task [11], [12]. A single drug has four main features i.e., Enzyme, Pathway, Chemical Substructure, and Target. Various ML methods have been proposed by researchers for predicting DSEs. ML methods based on drug chemical information are designed to predict drug-related DSEs using drug chemical structure features. Researchers have recently developed a deep learning framework based on a neural fingerprint technique to represent adverse effects and chemical structures [13]. DSEs were predicted after the identification of alternative targets of drugs in the study [14]. The ability to predict DSEs can be enhanced with the combination of biological and chemical features. The integration of target and chemical structures enhances the ability to predict the DSEs of drugs [15]. In another study, the combined information of gene expression and chemical structures was used to determine adverse reactions (ADRs) [16]. Recently knowledge graph embedding has been used to predict DSEs, graph embedding method was proposed DDTE (Drug-Disease-Target Embedding) for embedding heterogeneous networks with different data types in the study [17].

Zhao et al. proposed a similarity-based method for predicting DSEs using heterogeneous information [12]. In their study, they converted the problem into binary classification after considering drug and DSEs as pairs. Pairs were based on a similarity concept, whereas a pair represented five features, and each feature derived from the drug property. They applied the Random Forest algorithm as a prediction engine. The results indicated a strong correlation between DSEs and drug fingerprints. The extraction of information from several drug heterogeneous networks and drug features provided the base of the network embedding method [11]. Additionally, pairing of drug features and DSEs was used to represent a single sample. The association of drug-side effect, a predictor was proposed in the study [18] which was based on semi-supervised learning and multiple kernel learning (MKL) approaches. This model could combine many input sources to enhance prediction performance.

Furthermore, Yen Lee et al. [19] proposed a hybrid deep learning model, this proposed model consists of a graph convolutional neural network (GCNN) model and bi-directional long short-term memory (Bi-LSTM) for efficient learning of drug features and association of DSEs. Dimitri and Lió [20] proposed a machine learning-based model under the name DrugClust. DrugClust model combined two techniques for DSEs prediction. The model pipeline clustered the drugs based on their features and predicted their DSEs by applying Bayesian scores. Whereas biological validation of resulting clusters can be verified using enrichment analysis.

Jiang et al. [21] introduced an approach grounded in the premise that drugs sharing similar structures are prone to exhibit comparable SEs. They developed a network representation encapsulating both drug structures and SEs, employing a path-based algorithm to discern noteworthy associations between drugs and their corresponding SEs. Liang et al. [22] presented a dual source data approach, incorporating drug-drug and drug-side effect similarities. Calculations for drug-drug similarities were rooted in chemical structure similarities, while drug-side effect similarities were derived from their occurrence in clinical trials. Employing a transductive matrix co-completion algorithm on the resultant matrix, they predicted missing entries, representing potential DSEs. The algorithm harnessed the inherent similarities between drugs and DSEs to predict absent entries within the matrix. A summarized literature review is presented in Table 1. The review indicates a significant relationship between chemical structures and DSEs, with existing studies predominantly focusing on NLP techniques.

### B. PROPOSED DSEs PREDICTION WITH DRUG CHEMICAL 2D-STRUCTURE IMAGES

The proposed research has been based on 2D structure images. It is imperative to predict the SEs of drugs at an early stage to launch a useful medicine in the market such as pandemic situation e.g., COVID-19. Typically, a drug has four features (Structure, Pathway, Enzyme, and Target). Out of these features, the most reliable feature is chemical structure because it is a persistent and efficient feature to predict early SEs. Initially, DSEs prediction is approached as a binary classification problem. In this study, a transfer-learning-based CNN model is proposed to predict DSEs using 2D chemical structure image. Various Online tools are available that can design a drug molecule structure. Once proposed study is implemented in such a tool, it can predict DSEs based on the molecule's structure as well.

**TABLE 1.** NLP based literature summary.

| Study | Model Name | Method/Approach | Description | Result's |
|---|---|---|---|---|
| [4] | NDDSA | Network and Domain-Based Algorithm | Based on the assigned side effect score model can predict the DSEs of a new drug. | AUPRC of Liu's Datasets (41.2 % and 39.1 %) |
| [16] | - | Machine Learning Classification Approach | Machine learning model based on drugs' chemical structure and gene expression features. | - |
| [11] | - | Random Forest | Network embedding method with the random forest as a prediction engine. | Average Matthews correlation coefficients (64 % for balanced and 64.1% for unbalanced datasets) |
| [18] | - | MKL Algorithm | The model was based on semi-supervised learning and multiple kernel learning approaches. | AUPRC (66.8 %, 67.3%, and 67.0% on three benchmark data sets respectively) |
| [20] | DrugClust | Machine Learning Approach | This model used clustering for drug features and Bayesian scores for DSEs. | - |
| [22] | Co-Completion Algorithm | The authors utilized similarities between drugs (based on chemical structure) and DSEs (based on clinical trial occurrences) to predict potential DSEs. This was achieved through a transductive matrix co-completion algorithm. | The study provides a comprehensive approach to evaluating drug safety and facilitating drug development. It achieves this by comparing similarities between drugs and their side effects, allowing for the prediction of potential adverse effects. | - |
| [23] | XGBoost | Examined data from 10,064 individuals, collecting information on their age, history of COVID-19, gender, smoking status, level of education, symptoms, and type of vaccination. | The study aimed to predict DSEs of the coronavirus vaccine using machine learning methods. | Accuracy (58 to 74%) |
| [24] | Random Forest Classifier | Analyzed DSEs using data from WebMD.com, which was sourced from health and medical forums, and applied machine learning models. | The study aimed to predict DSEs based on descriptions from health and medical forums. | Accuracy (62.4%) |
| [25] | - | Lexicon-Based Approach | Conducted a systematic analysis of 10,822 tweets discussing 74 medications to assess whether they could indicate potential signals for ADRs. | F1 (57.8%) |

The scope of this research is to develop a predictive model for DSEs using 2D chemical structures and deep learning techniques. The research focuses on leveraging deep learning models such as MobileNetV2 model to analyze 2D chemical

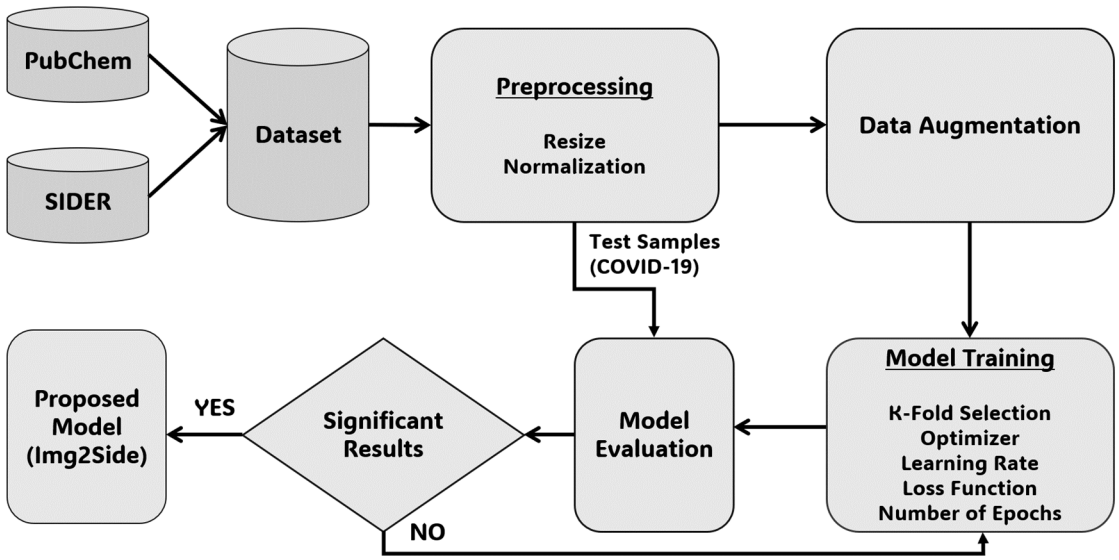



**FIGURE 1.** Abstract diagram of the proposed study.

structures and predict potential DSEs of drugs. The ultimate goal of the project is to develop a reliable and accurate tool that can assist healthcare professionals and researchers in predicting potential DSEs of drugs based on their 2D chemical structures, thereby improving drug safety and patient care. The research does not cover the real-life experimental validation of DSEs prediction. Additionally, the prediction of DSEs in this research is done without the use of genetic information.

### C. MAJOR CONTRIBUTIONS

The major contributions of the proposed study are:

- Preparation of a unique dataset, label each 2D structure image with its associated side effect (SIDER- Side Effect Resource information).
- This study is the first of its kind (as per our knowledge) to predict fever DSEs using 2D chemical structure images.
- A fully tweaked transfer-learning-based model to serve the purpose of proposed study.
- To boost the model performance, a data augmentation technique was used to increase the dataset size with rotation, zoom, and flipping of the original dataset images.
- The proposed model has reduced the hefty transformation process contrary to the NLP domain which employs smiles to fingerprints and features extraction process.
- The proposed model is also tested for predicting the DSEs of COVID-19 drugs and achieved an F1 score of 75%.

## II. PROPOSED METHODOLOGY

This section comprises dataset description, dataset preprocessing, methodologies, and network details. The abstract diagram of the proposed study can be seen in Fig. 1.

### A. DATASET AND LABELING

In this study, drug information and associated DSEs are extracted from the open-source database SIDER (Version 4.1) [26] maintained by the European Molecular Biology Laboratory (EMBL). SIDER contains information about marketed medicines and their associated DSEs. The database encloses information on 1,430 drugs (I did not observe any information with the drug name 'x' so the actual number of drugs considered is 1,429), 5,868 ADRs, and 1,39,756 Drug-ADR associations this database is connected with PubChem [27]. The 2D chemical structure images collected from PubChem [27]. In this study, fever is considered as one of the DSEs. SIDER has 790 drugs (768 unique) that can cause different types of fever (body temperature increased, fever neonatal, fever chills, unknown origin fever, and herpes labialis) as DSEs.

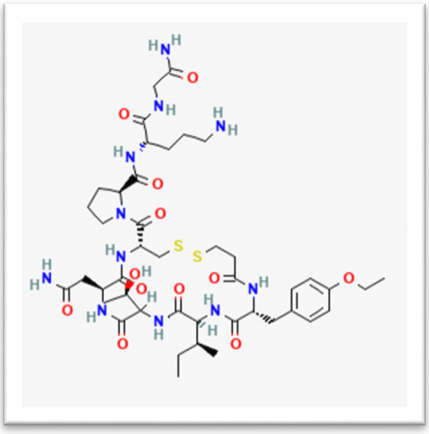
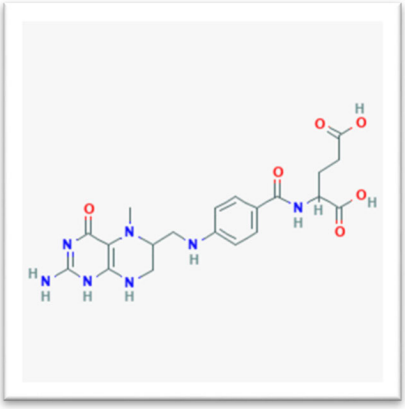
- **Fever:** A very common problem or condition that is experienced by almost all people in their lives. The average normal temperature of the body is 98.6° Fahrenheit [28] and continuous fever is the major reason for being disconnected from routine work. Human often feel fever after taking high-potency medicines. 2D chemical drug structure images collected and labeled as 1 and 0 i.e., a binary classification problem. Whereas, 1 represents the side effect of fever and 0 represents those drugs that have no side effect. Table 2 describes samples and their associated labels.

In addition to that COVID-19 2D chemical structure images extracted from DrugBank [29] exhibiting fever as DSEs. This step was taken to test the generalization and robustness of the proposed model.

### B. DATASET PRE-PROCESSING

The 2D drug's chemical structure images are of 300 × 300 dimension. The images are resized to 224 × 224

**TABLE 2.** Prepared dataset samples, counts, and associated label.

| Drug Name | Drug Chemical Structure Image | Label |
|---|---|---|
| atosiban |  | 1 |
| 5-methyltetrahydrofolate |  | 0 |
| Total 0 Class Images before Pre-Preprocessing | | 651 |
| Total 1 Class Images before Pre-Preprocessing | | 745 |
| Total 0 Class Covid-Medicine Images before Pre-Preprocessing | | 10 |
| Total 1 Class Covid-Medicine Images before Pre-Preprocessing | | 23 |

dimension and normalized pixel (0-255) intensity values between 0-1. Some images that only contain elements names were removed. Table 3 depicts some of the removed elements from the dataset.

#### 1) DATA AUGMENTATION

Data augmentation is a technique used to increase the amount of data samples for deep learning models. The goal of data augmentation is to create additional data samples that are similar to the original ones but with small variations. This helps to mitigate over-fitting, which occurs when a model performs well on the training data but poorly on new, unseen data [30]. It can be performed in several ways, including rotation, zoom, flipping, and adding noise to the original data samples. These operations are applied to the original data to generate new, augmented data samples that can be used for deep learning models. The augmented data samples help the model learn to recognize patterns in the data that are invariant to small transformations, which can improve its generalization performance. Table 4 has lists of pipeline parameters and their values for the augmentation process that applied in this study.

Table 5 shows the original and augmented dataset samples as well as class-wise counts after preprocessing and augmentation.

### C. PRE-TRAINED MODELS

Pretrained models are deep learning models that have already been trained on a large dataset and are made available for

**TABLE 3.** Prepared dataset samples, counts, and associated labels.

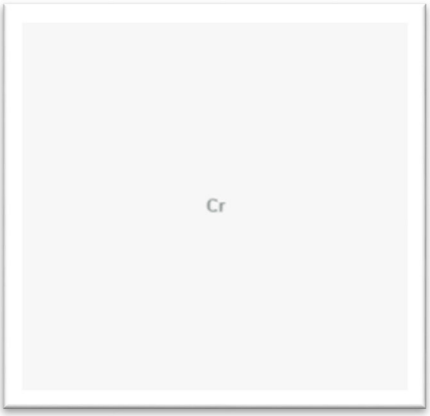
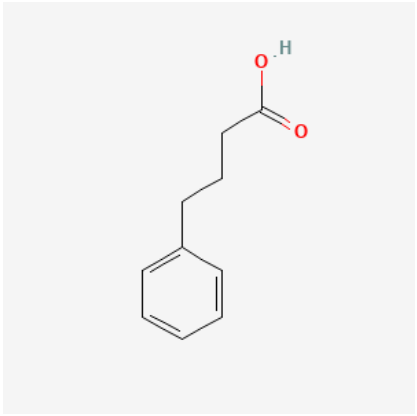
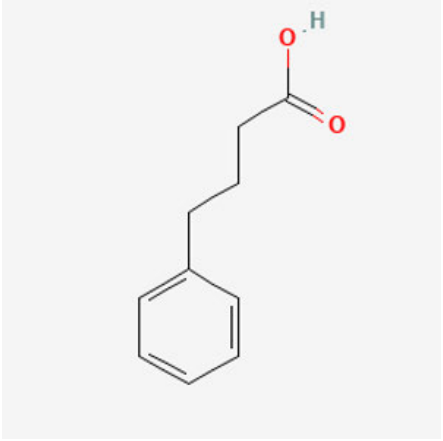
| Drug Name | Drug Chemical Structure Image | Label |
|---|---|---|
| chromium |  | 0 |

**TABLE 4.** Augmentation pipeline parameters.

| Data Augmentation Parameters | Sub Parameters & Value(s) |
|---|---|
| Rotation | Probability (0.3), Max_left_rotation=5, Max_right_rotation=5 |
| Zoom | Probability (0.1), Min_factor=1.0, Max_factor=1.2 |
| Flipping | Flip_Left_Right (Probability of 0.1) |

**TABLE 5.** Prepared dataset samples, counts, and associated labels after pre-processing and augmentation.

| Drug Name | Drug Chemical Original 2D Structure Image | Drug Chemical Augmented 2D Structure Image | Label |
|---|---|---|---|
| 4-PBA |  |  | 0 |
| | Total 0 Class Images after Pre-Preprocessing and Augmentation | | 1242 |
| | Total 1 Class Image after Pre-Preprocessing and Augmentation | | 1490 |
| | Total 0 Class Covid-Medicine Images after Pre-Preprocessing | | 7 |
| | Total 1 Class Covid-Medicine Images after Pre-Preprocessing | | 23 |

others to use. These models are used as a starting point for solving specific tasks, such as image classification, natural language processing, and many others, rather than starting the training process from scratch [31].

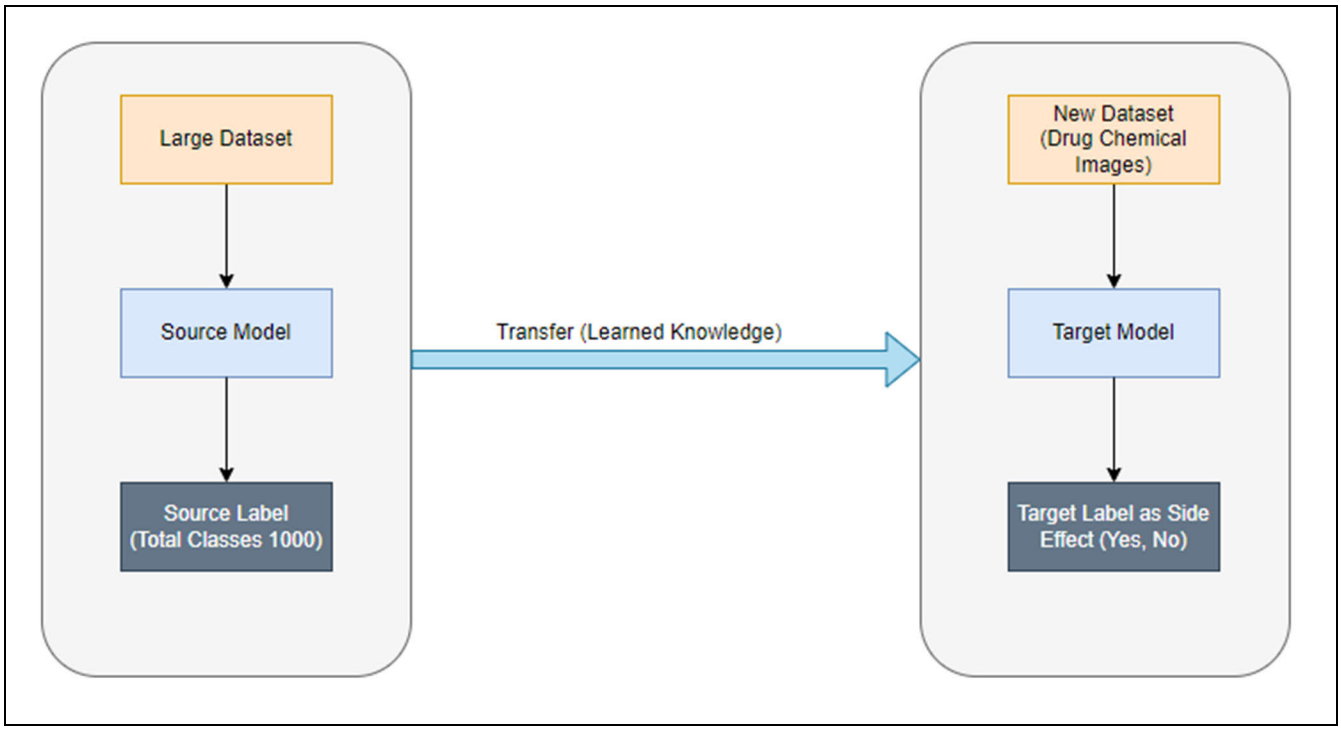



**FIGURE 2.** Transfer learning concept.

The idea behind using pre-trained models in proposed study is that model has already learned important features and patterns from the vast amounts of data that they were trained on, and these features can be fine-tuned for a specific task. This can save a significant amount of time and resources, as it is often more efficient to fine-tune a pretrained model than to train a model from scratch.

#### 1) TRANSFER LEARNING

Transfer learning is a deep learning technique that is being used to improve the performance of a task after learning another task. In the transfer learning approach, for the training of a new model, a pre-trained (such as DenseNet121) model is used as a base model for different but related tasks as can be seen in Fig. 2. The idea is that the pre-trained model has already learned useful representations of data that can be reused and adapted for the new task.

Transfer learning has gained popularity in recent years due to the rise of deep learning and the availability of large pre-trained models, such as BERT, and VGG16.

The transfer learning approach is considered to the numerous benefits as it can reduce the amount of data needed to train a new model, which is particularly useful when labelled data is scarce or expensive and it can also speed up the training process and improve the performance of the new model [32].

#### 2) FINE-TUNING

Fine-tuning [33] is a technique that involves taking a pre-trained model and adapting it to a new task or dataset by updating its parameters through additional training. This process enables the model to learn the specifics of the new task while leveraging the knowledge it has already acquired during the pre-training phase. Fine-tuning involves taking the pre-trained model and updating its parameters on a new, smaller dataset that is specific to the task at hand.

Fine-tuning has several benefits in deep learning. It can save time and resources compared to training a new model from scratch and often leads to better performance on the new task due to the transfer of knowledge from pre-training. However, there are also some challenges associated with fine-tuning.

#### 3) PROPOSED MobileNetV2 PRE-TRAINED MODEL

MobileNetV2 [34] is a deep learning architecture that was developed by Google specifically for mobile and embedded devices. It is an improvement over the original MobileNet architecture, which was designed to be lightweight and efficient while still achieving good accuracy on image classification tasks.

The architecture of MobileNetV2 consists of a series of blocks, each of which contains a combination of depth wise separable convolutions, pointwise convolutions, and pooling layers. It also includes skip connections and residual blocks, which are used to improve the performance of the network. MobileNetV2 has achieved good results on benchmark datasets such as ImageNet and COCO and is widely used in computer vision applications on mobile and embedded devices.

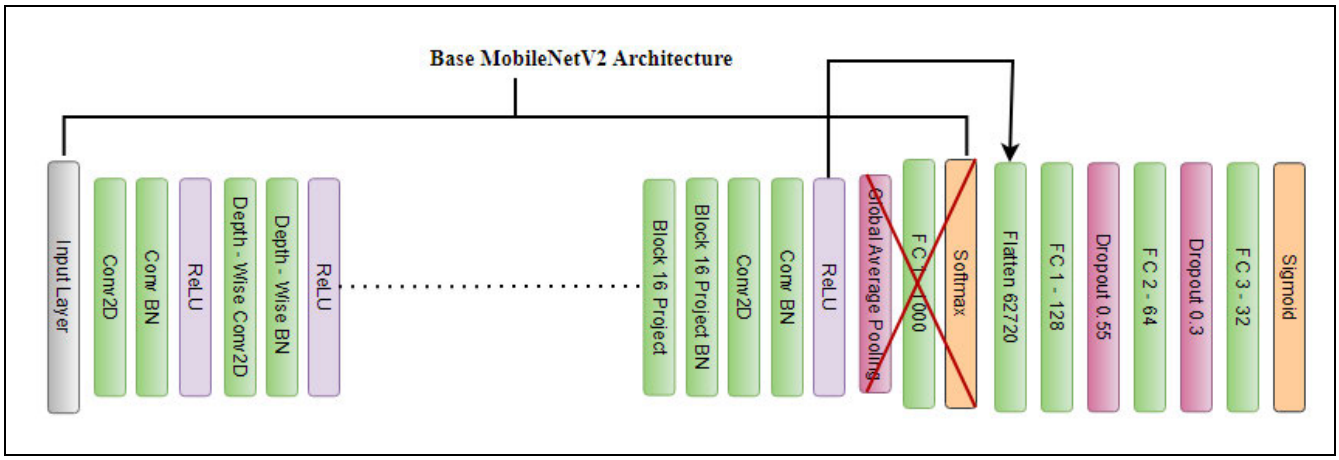



**FIGURE 3.** Img2Side network architecture based on fine-tuned MobileNetV2 model.

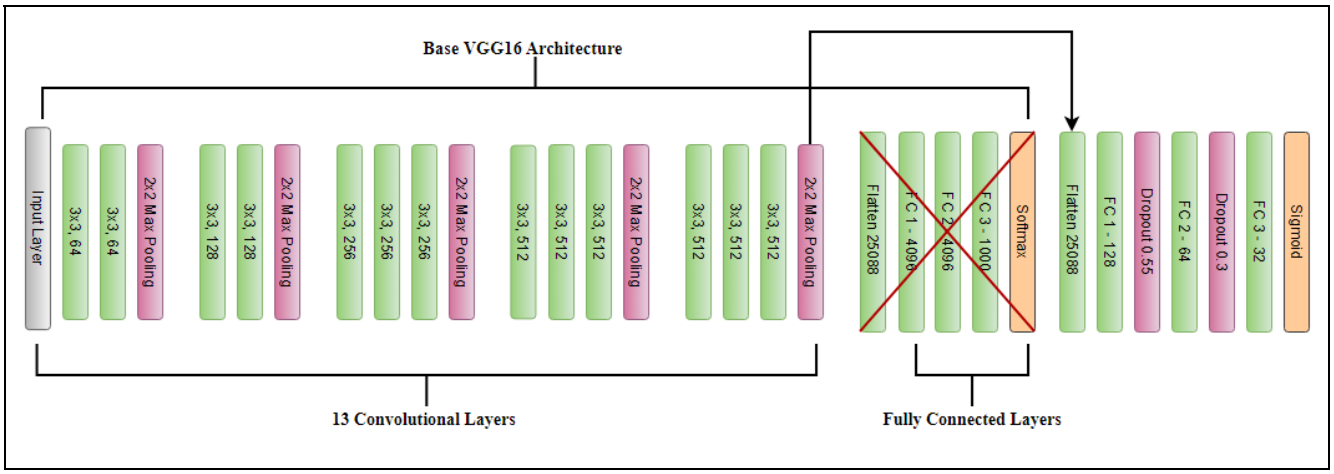



**FIGURE 4.** Fine-tuned VGG16 model architecture.

It is also commonly used as a feature extractor in transfer learning, where the pre-trained weights of MobileNetV2 are used as a starting point to fine-tune the network on a new task with a smaller dataset. All the pre-trained layers are frozen in proposed model and additionally three fully connected (FC) layers of 64, 32, and 16 neurons respectively. Whereas ReLU [35] is an activation function between FC layers and sigmoid [36] is an output activation function. In order to avoid overfitting, additional dropout layers (0.55, 0.3) are added after each FC layer except the last FC layer, see Fig. 3.

#### 4) VGG16 PRETRAINED MODEL

VGG16 [37], [38] is a CNN-based architecture and its name is due to its developer name "Visual Geometry Group (VGG)" and 16 refers to weight layers. VGG16 is one of the most widely used deep learning models for image classification tasks and is known for its simplicity and good performance. The VGG16 architecture uses small convolutional filters ($3 \times 3$) and a lot of them, which gives the network the ability to learn fine-grained features of the input image. Additionally, it uses a deep network of many layers, which allows it to learn high-level abstract features that are more informative for classification.

VGG16 has been trained on a large dataset of images, and the weights of its layers can be used as pre-trained weights for transfer learning, which is a technique used to fine-tune a pre-trained network on a new task with a smaller dataset. This allows the network to adapt to new tasks and improve its performance without having to start from scratch. This model is considered due to its simplicity and high performance, see Fig. 4 for fine-tuned VGG16 network.

#### 5) DenseNet-121 PRETRAINED MODEL

DenseNet121 [39] is a variant of the DenseNet architecture. The architecture of DenseNet121 is based on the idea of densely connected convolutional networks. This connectivity pattern enables DenseNet121 to reuse features learned in earlier layers, which leads to better gradient flow, reduces overfitting, and improves accuracy. Additionally, DenseNet121 employs batch normalization after every

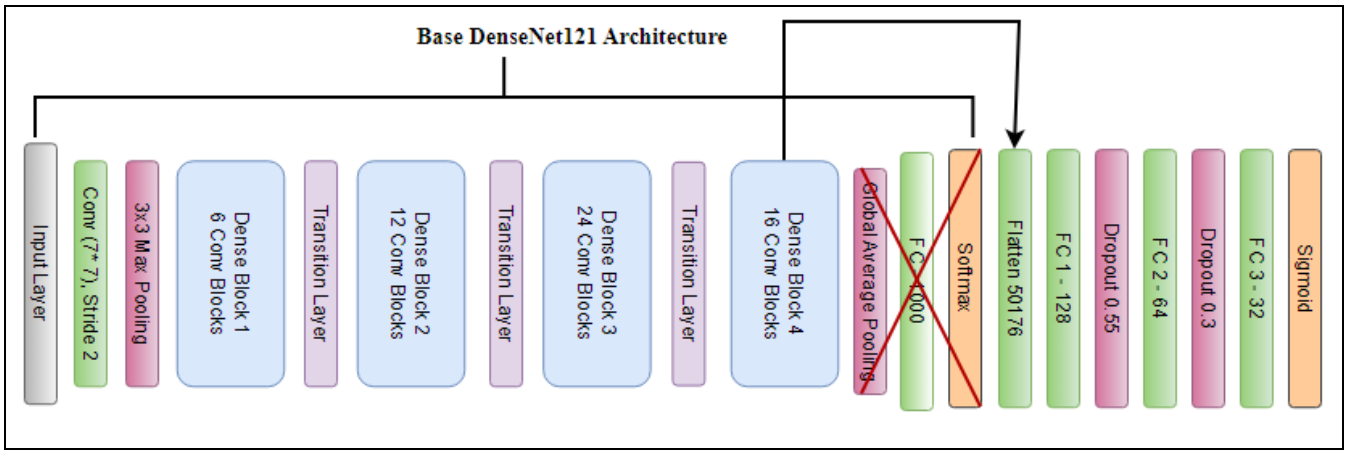



**FIGURE 5.** Fine-tuned DenseNet-121 model architecture.

convolutional layer, which speeds up training and improves accuracy.

DenseNet121 uses a global average pooling layer that reduces the number of parameters and helps to prevent overfitting. DenseNet121 has achieved state-of-the-art results on various image recognition tasks, including the ImageNet dataset. Overall, DenseNet121 is a powerful deep neural network architecture for image classification tasks, and its unique connectivity pattern and efficient use of parameters make it a promising technique for applications in computer vision.

To compare these three pre-trained models, FC layers, activation functions, and dropout layers are used with the same configuration as described in MobileNetV2, see Fig. 5 for fine-tuned DenseNet121 network.

## III. EXPERIMENT RESULTS AND DISCUSSION

A detailed discussion has been provided in Table 2 to Table 5 as a proposed methodology, which includes all the steps such as data scraping, preprocessing, and augmentation to artificially increase the data samples. In this study, three pretrained architectures VGG16, MobileNetV2, and DenseNet121 considered. All three architectures were trained and hyper-tuned for better performance on prepared dataset. In this section, we have evaluated the performance of all three architectures based on evaluation metrics. A detailed discussion of network training and the performance of the proposed model is described in this section.

### A. EVALUATION METRICS

In deep learning, the evaluation of models is a crucial step in understanding how well a model performs. The evaluation matrices help in this regard for the model's efficacy evaluation [40].

- **Accuracy:** Accuracy is a percentage of correctly predicted class from the total prediction without the consideration of class labels. The formula to calculate the accuracy is mentioned in Equation 1, that based on True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN).

$$Accuracy = (TP + TN)/(TP + FP + TN + FN) \quad (1)$$

- **Precision:** Precision measures the accuracy of positive predictions. The formula to calculate the accuracy is mentioned in Equation 2, in which it is observable that precision is calculated with counts of true positive predictions (TP) divided by the sum of true and false positive predictions (TP+FP).

$$P = TP/(TP + FP) \quad (2)$$

- **Recall:** Recall (also known as true positive rate (TPR) as well as sensitivity) is used to measure the model's ability to find all positive instances. The formula to calculate the recall is mentioned in Equation 3, in which it is observable that recall is calculated with counts of true positive predictions (TP) divided by the sum of true positive and false negative predictions (TP+FN).

$$R = TP/(TP + FN) \quad (3)$$

- **F1 Score:** The F1 balances both precision and recall and gives a single score to the model's performance & and it is the harmonic mean of precision (P) and recall (R). The formula to calculate the F1 is mentioned in Equation 4.

$$F1 = 2(PR)/(P + R) \quad (4)$$

- **Binary Cross Entropy:** The Binary cross entropy (BCE) is a loss function commonly used in machine learning for binary classification tasks, where the goal is to predict a binary output (e.g., true/false, 0/1, yes/no, etc.) based on input features. The BCE loss measures the difference between the predicted probability distribution and the true probability distribution. It calculates the cross-entropy between the true labels and the predicted probabilities.

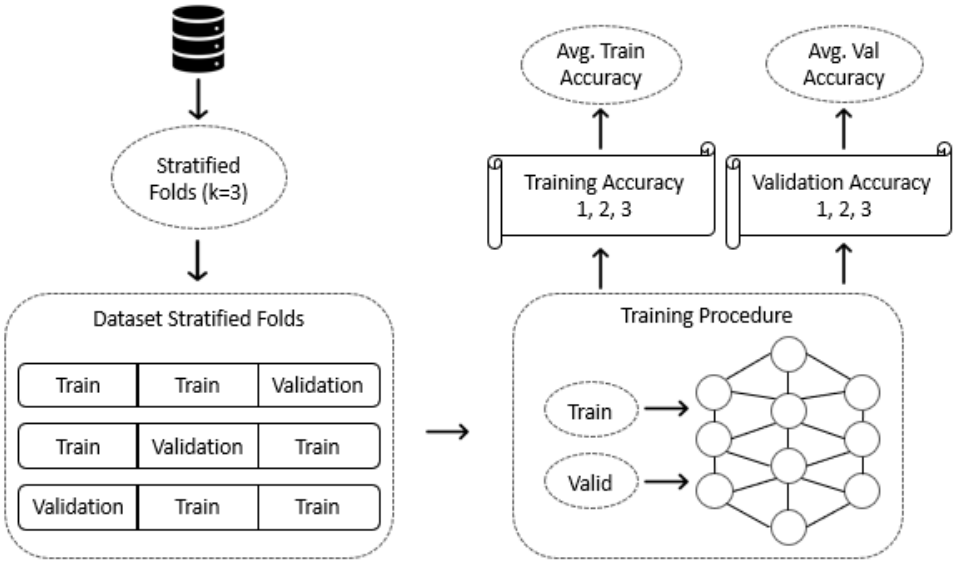



**FIGURE 6.** Stratified-K-fold process for proposed study.

### B. STRATIFIED-KFOLD

In this study, the Stratified-K-Fold considered for cross-validation. It divides data into k-folds, where k is a user-specified number. It's called ''stratified'' because it tries to ensure that the distribution of target classes is roughly the same in each fold as in the complete dataset. Stratified-K-Fold has used to ensure that the model should learn and validate for each class with equal proportions. The detailed process of Stratified-K-Fold for the 2D drug chemical structure images dataset is summarized in Algorithm 1.

In this study, for Stratified-K-Fold K=3 is used which means data is divided into 3 folds in such a way that the proportions of the target class in each fold are similar to the proportions of the target class in the complete dataset. The average scores of 3-fold training and validation are reported in this study, the process of Stratified-K-Fold for this study is elaborated in Fig. 6.

### C. NETWORK TRAINING

A transfer learning approach has been applied to reduce the training computational power. The architecture of the proposed framework is elaborated in Figure 6. The architecture consists of 3 $\times$ 3 convolution layers with 64 filters for convolutional operations with 2 $\times$ 2 max pooling. The 2D drug chemical structure images are resized to 224 $\times$ 224 from 300 $\times$ 300 to make it consistent with MobileNetV2 architecture input. The data augmentation process is applied with three parameters rotation, zoom, and flipping for model robustness and better generalization. The input images are normalized to convert the pixel values [0,255] to a specific range [0,1]. Dropout layers with dropout rates 0.55, and 0.3 are added respectively after FC layers to avoid overfitting. The optimizer used in this study is Adam with a learning rate of 0.0001. Additionally, a reduced learning rate with patience 3 and early stopping with patience 5 has also applied in the study, see Fig. 7.

The model was trained for 50 epochs with Stratified-K-Fold (K=3). The training and validation data are divided into folds, whereas for testing purposes Covid-19 drugs 2D chemical structure images considered to measure the efficacy of the proposed model. The hyperparameters and their tuned values are described in Table 6.

### D. RESULTS AND DISCUSSION

The ultimate goal of proposed model is to identify fever as a side effect of a drug from the dataset, which is extracted by using the proposed Img2Side model. The results are achieved from the 2D drug chemical structure images that used after the preprocessing, augmentation, and normalization. To validate the proposed model properly, Stratified-K-Fold considered with Fold=3 in this study. The highest training accuracy in each split was 0.8192, 0.8145, and 0.8513 respectively with the highest average accuracy of 0.8283. Furthermore, the highest validation accuracy achieved in each fold was 0.8131, 0.7967, and 0.8516. Whereas the average highest validation accuracy recorded was 0.8205.

Img2Side was trained for 50 epochs. Furthermore, to avoid overfitting, the early stopping with patience 5 and reduced the learning rate with patience 3 considered. Fig. 8-10 clearly shows that model training stopped before the completion of 50 epochs as the early stopping criteria is satisfied in each fold i.e., in Fold 1, Fold 2, and Fold 3.

To validate the efficiency of data augmentation, proposed model also trained without data augmentation. Fig. 11 illustrates that the model outperformed when using the original + augmented dataset, as opposed to only the original dataset. Furthermore, it is observable that the training accuracy of 0.5792 is less than the validation accuracy of

**Algorithm 1** Stratified-K-Fold for Proposed Study

**Input:** Drug 2D Chemical Structures Dataset

**Step 1:** Split the dataset into 3 folds

***Repeat***

For fold i=1 to 3 **do**

**Step 2:** Select fold I as the validation and the remaining folds as the training set.

**Step 3:** Fit the model on the training set

**Step 4:** Evaluate for validation set during training

**Step 5:** Store the evaluation scores in list S

***End-for***

**Step 6:** Find the average performance with S

**Output:** Average Performance of the Model

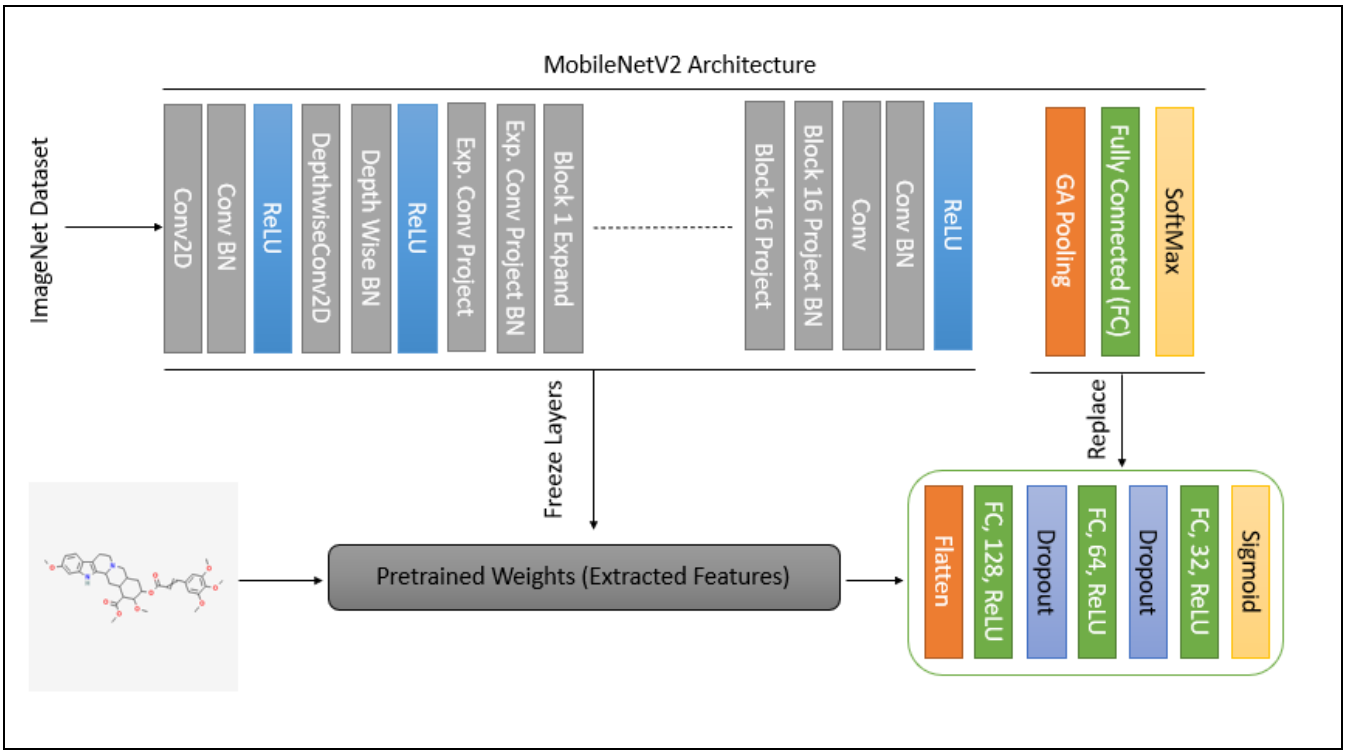



**FIGURE 7.** Transfer learning-based proposed architecture for DSEs prediction using pretrained MobileNetV2 model.

0.5928, which indicates symptoms of underfitting in the model.

### 1) ROBUSTNESS OF PROPOSED Img2Side MODEL

COVID-19, also known as coronavirus disease 2019, is a viral respiratory illness caused by the SARS-CoV-2 virus. The disease was first identified in Wuhan, China, in December 2019, and has since spread globally, leading to a pandemic. Drug discovery has played a critical role in the fight against COVID-19. Several drugs have been repurposed or developed specifically to treat COVID-19, including remdesivir, dexamethasone, and monoclonal antibodies. These drugs work by targeting different aspects of the virus or the body's response

**TABLE 6.** Prepared dataset samples, counts, and associated labels.

| Parameters | Values |
|---|---|
| Train Generator Shuffle | True |
| Batch Size | 32 |
| Epochs | 50 |
| Learning Rate | 0.0001 |
| Reduce Learning Rate | Yes |
| Patience for Reduce Learning Rate | 3 |
| Early Stopping Patience | 5 |
| Stratified-K-Fold K Value | 3 |
| Optimizer | Adam |
| FC Layers Activation Function | ReLU |
| FC Layers Neurons | 128, 64, 32 |
| Output Layer Neurons | 2 |
| Output Activation Function | Sigmoid |
| Dropout between FC Layers | 0.55, 0.3 |
| Compile Loss | Binary Cross Entropy |

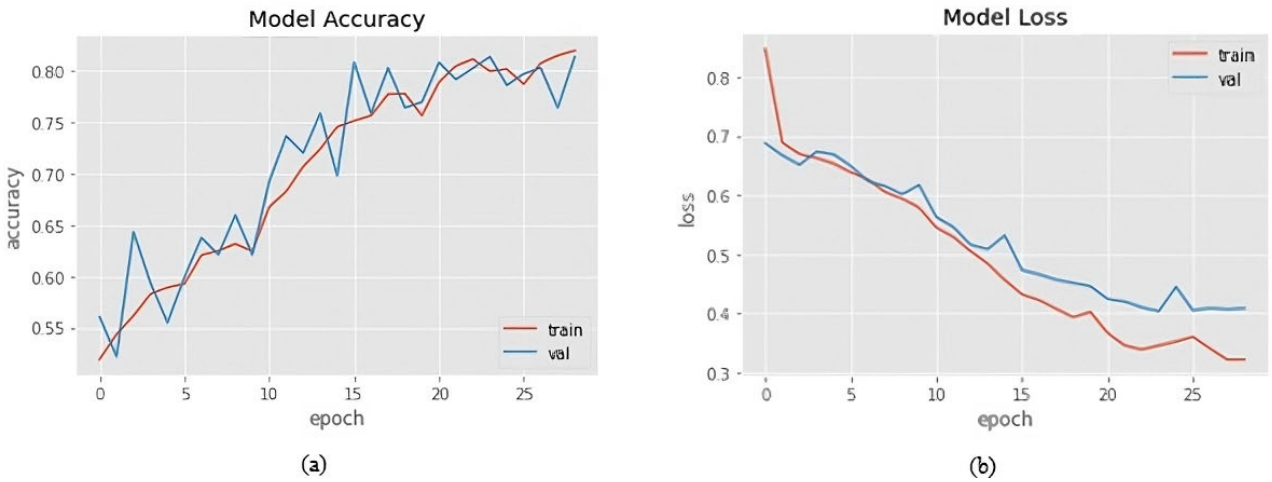



**FIGURE 8.** Model accuracies and losses graph for Fold-1.

to the virus. However, like all drugs, COVID-19 treatments can have DSEs. For example, ritonavir has been associated with fever, while dexamethasone can cause increased blood sugar levels, weight gain, and mood changes.

It is essential to carefully weigh the potential benefits of these treatments against their potential DSEs and to closely monitor patients receiving these medications. Additionally, ongoing research and clinical trials are exploring new treatments and strategies to manage COVID-19 and minimize its impact on public health. In light of these facts, suggested model's generalizability and robustness tested using COVID-19's 2D chemical structure. In this research, the

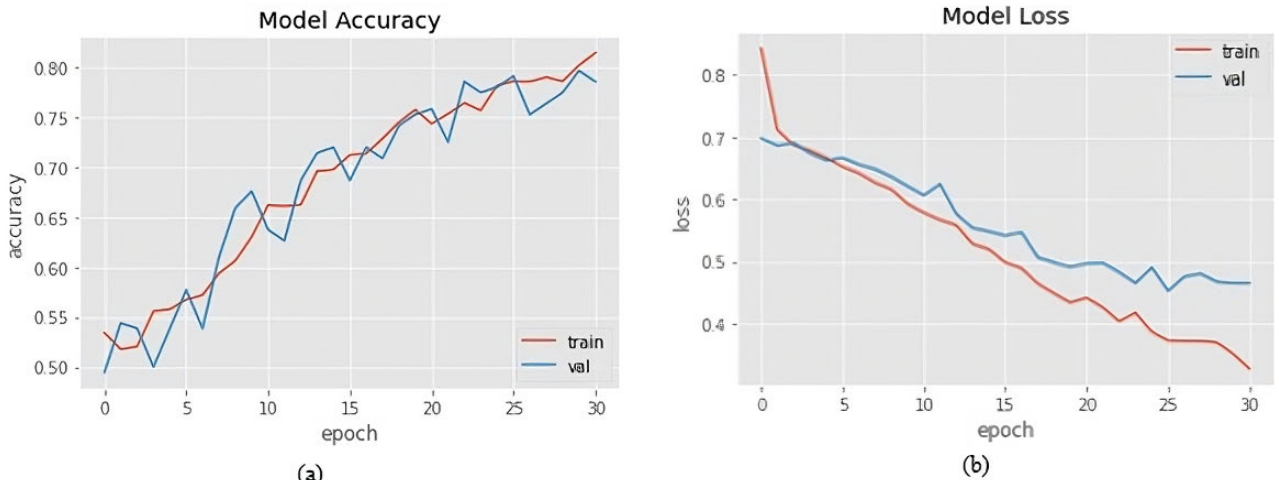



**FIGURE 9.** Model accuracies and losses graph for Fold-2.

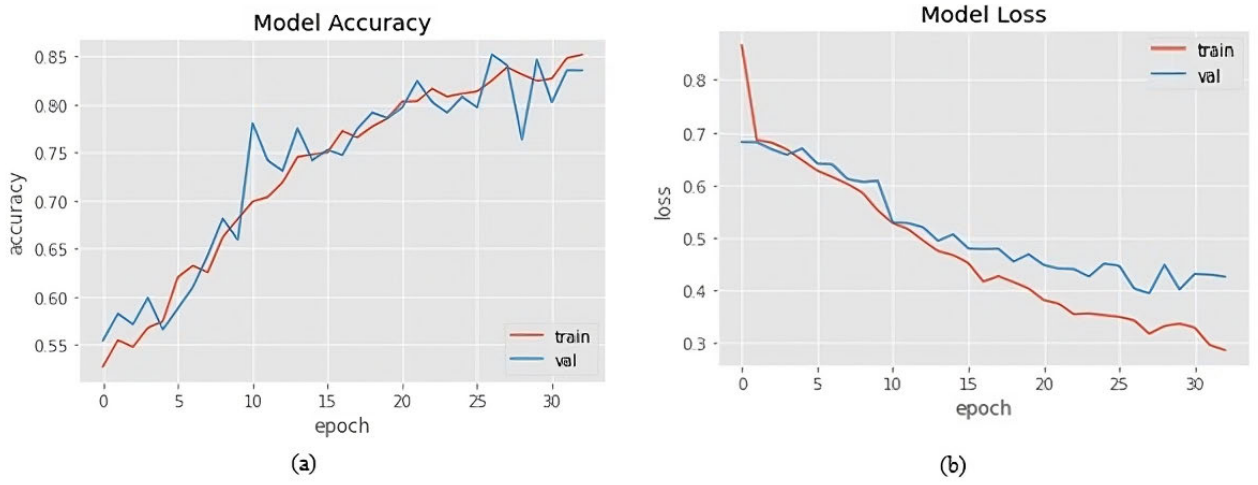



**FIGURE 10.** Model accuracies and losses graph for Fold-3.

COVID-19 drugs extracted from the DrugBank source [26] and proposed model validated using the 30 COVID-19 drugs' 2D chemical structures for fever as a side effect (Table 7).

Table 8 depicts the performance of the proposed Img2Side model. It is observed that the proposed model has shown the highest percentage for important classification measures i.e., precision, recall, F1, and accuracy. It is evident from the classification report that the Img2Side model achieved accuracy, precision, recall, and F1 of 73%, 83%, 73%, and 75% respectively. In Table 8, the 'Support' column denotes the number of samples for each respective class. Out of a total of 30 samples, 7 belong to Class 0 (No Fever), and 23 belong to Class 1 (Fever). The variation in per-class scores is attributed to the imbalanced number of samples during testing.

Furthermore, the overall comparison for the proposed Img2Side against VGG19 and Dense-Net121 is shown in Fig. 12. This clearly shows that proposed model outperformed other variants. An important point to note here is that fully connected layers, layer activation functions, and

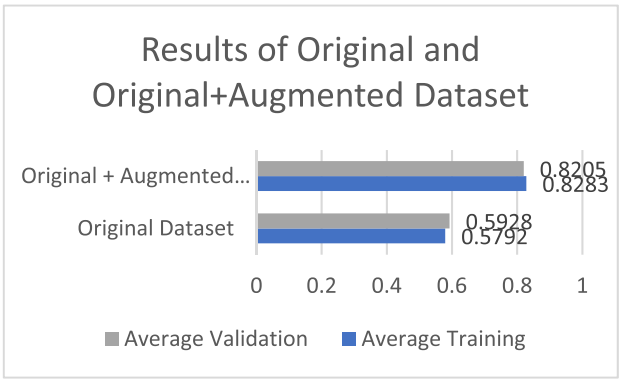



**FIGURE 11.** Results of original and original + augmented dataset.

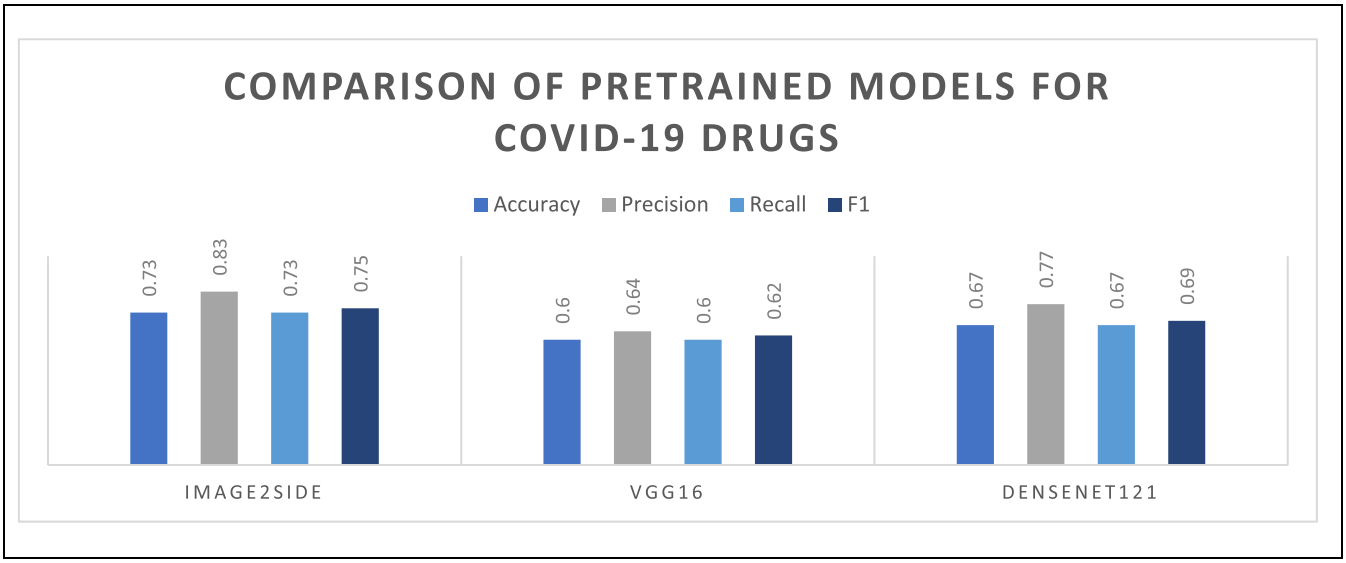



**FIGURE 12.** Img2Side model comparison with pretrained variants (with data augmentation and pretrained weights) on basis of accuracy and weighted average of (Precision, Recall, and F1) for Covid-19 Drugs.

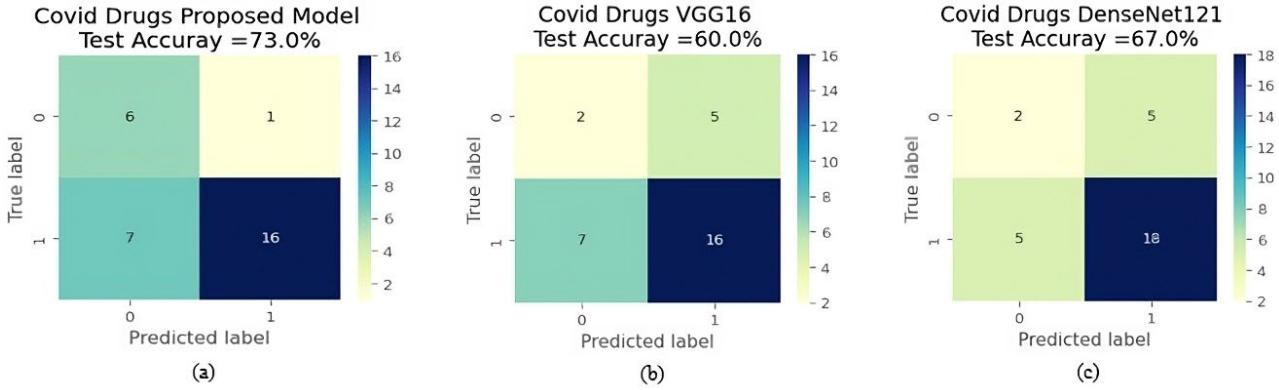



**FIGURE 13.** Img2Side model comparison with pretrained variants (with data augmentation and pretrained weights) on basis of confusion matrix for Covid-19 Drugs.

dropout layer configurations were maintained consistently across all three models.

The comparison based on the confusion matrix of the proposed Img2Side transfer-learning-based model is illustrated in Fig. 13. Furthermore, the confusion matrix comparison also confirmed that proposed model has better performance than other models.

Moreover, to demonstrate the robustness of employing pre-trained ImageNet weights, experiments conducted without pretraining, i.e., without using pre-trained weights. The fine-tuned architectures of Image2Side, VGG16, and DenseNet121 considered and theses models are trained from the ground up for predicting DSEs in this study. Fig. 14 illustrates that these models exhibited inferior performance when trained without pre-trained weights. The Image2Side and DenseNet121 models are overfitted as training accuracy 0.7819 of Image2Side and 0.8367 is too much greater than validation and test accuracy [41]. The VGG16 model is underfitted as the training accuracy of 0.5424 too much less than the test accuracy of 0.77.

Typically, authors have not found any study that predicts DSEs purely based on 2D chemical structure images. However, some research has been done based on Canonical Smiles, which is solely related to the NLP domain. For the use of deep learning approach robustness, the proposed model compared with machine learning models (ML) as well as with variations in terms of data augmentation and pre-trained weights. Proposed Image2Side model is well generalized when it is used with data augmentation and pre-trained weights, see Table 9.

The practical applicability of this work lies in its potential to revolutionize drug development processes. By predicting

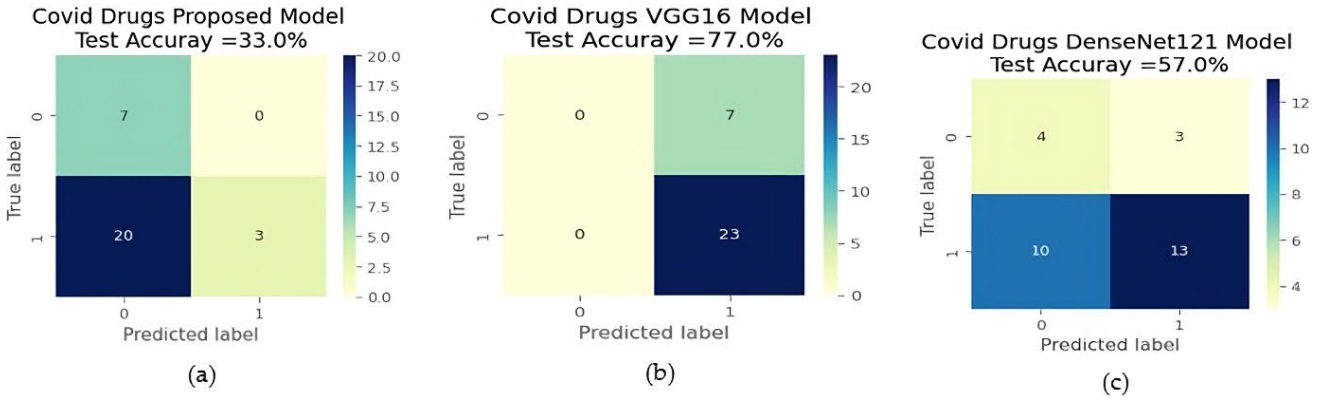



**FIGURE 14.** Img2Side model comparison (with data augmentation and without pretrained weights) on basis of confusion matrix for Covid-19 Drugs.

**TABLE 7.** Covid-19 treatment drugs.

| Covid-19 Treatment Drugs | | |
|---|---|---|
| Bromhexine | Budesonide | Ivermectin |
| Chloroquine | Celecoxib | Losartan |
| colchicine | Chlorpromazine | Montelukast |
| dipyridamole | Darunavir | Nitazoxanide |
| methylprednisolone | Dexamethasone | Quetiapine |
| rivaroxaban | Famotidine | Ribavirin |
| tranexamic acid | Fondaparinux | Ritonavir |
| argatroban | Heparin | Ruxolitinib |
| azithromycin | Hydroxychloroquine | Simvastatin |
| bicalutamide | Ibuprofen | Sofosbuvir |

**TABLE 8.** Proposed Image2Side model performance (with data augmentation and pretrained weights) for Covid-19 treatment drugs.

| | Precision | Recall | F1 | Support |
|---|---|---|---|---|
| 0 (No Fever) | 0.46 | 0.86 | 0.60 | 7 |
| 1 (Fever) | 0.94 | 0.70 | 0.80 | 23 |
| Accuracy | | | **0.73** | **30** |
| Macro Avg | 0.70 | 0.78 | 0.70 | 30 |
| Weighted Avg | 0.83 | 0.73 | 0.75 | 30 |

DSEs early in the development stage, pharmaceutical companies can screen candidate drugs more effectively, saving time, money, and potentially improving public health outcomes. The model's ability to achieve remarkable accuracy, even with COVID-19 drugs, suggests its versatility and reliability in predicting a wide range of DSEs.

In real-time, this model could be integrated into drug development pipelines or used by regulatory agencies to evaluate the safety of new drugs. By analyzing 2D chemical structure images, the model could rapidly predict potential DSEs, allowing for proactive measures to mitigate risks and enhance drug safety.

**TABLE 9.** Performance comparison between different models with variations of parameters.

| Model | Data Augmentation | Pretrained Weights | Average Train Accuracy | Average Validation Accuracy | Covid Test Accuracy | Covid Weighted Precision | Covid Weighted Recall | Covid Weighted F1 | Model Nature |
|---|---|---|---|---|---|---|---|---|---|
| **Img2Side (Proposed)** | **Yes** | **Yes** | **0.8283** | **0.8205** | **0.73** | **0.83** | **0.73** | **0.75** | **Generalized** |
| VGG16 | Yes | Yes | 0.6159 | 0.6981 | 0.60 | 0.64 | 0.60 | 0.62 | Underfitted |
| DenseNet121 | Yes | Yes | 0.6712 | 0.6996 | 0.67 | 0.67 | 0.67 | 0.67 | Underfitted |
| Img2Side | Yes | No | 0.7819 | 0.5567 | 0.33 | 0.83 | 0.33 | 0.22 | Overfitted |
| VGG16 | Yes | No | 0.5424 | 0.5201 | 0.77 | 0.59 | 0.77 | 0.67 | Underfitted |
| DenseNet121 | Yes | No | 0.8367 | 0.7014 | 0.57 | 0.69 | 0.57 | 0.60 | Overfitted |
| Image2Side | No | Yes | 0.5792 | 0.5928 | 0.70 | 0.71 | 0.70 | 0.71 | Underfitted |
| VGG16 | No | Yes | 0.5388 | 0.5672 | 0.77 | 0.59 | 0.77 | 0.67 | Underfitted |
| DenseNet121 | No | Yes | 0.5732 | 0.6073 | 0.67 | 0.57 | 0.67 | 0.61 | Underfitted |
| SVM | Yes | No | 0.5113 | 0.4904 | 0.26 | 0.82 | 0.27 | 0.15 | Overfitted |
| KNN | Yes | No | 0.7452 | 0.5479 | 0.53 | 0.68 | 0.53 | 0.57 | Overfitted |

## IV. CONCLUSION

In this study, Img2Side model is proposed that based on a pre-trained transfer learning MobileNetV2 architecture to predict DSEs using 2D chemical structures. The proposed study addresses limitations in the prediction of DSEs encountered in clinical trials. The Img2Side model improved accuracy observed through the implementation of dropout layers, a reduced learning rate approach, and an early stopping technique. Different approaches have applied and concluded that the Image2Side model is effective and well-generalized when it is used with the combination of data augmentation and pre-trained weights. Achieving a test accuracy of 73%, weighted precision of 0.83%, weighted recall of 0.73% and weighted F1 of 0.75% for COVID-19 drugs, proposed Img2Side architecture holds promise for aiding clinical trials in making early predictions of DSEs. While proposed framework has demonstrated commendable results, ongoing refinement of the proposed architecture is possible.

In the sense of limitation and future work, this study focus has been on individual DSEs; however, future endeavors aim to extend proposed study to predict multiple DSEs associated with a single drug based on its 2D chemical structure images. Proposed Img2Side model witnesses the materials of modern machine learning methods' ability to change the prediction process related to DSEs. The model will be further developed and enhanced, with the belief that it will continue to have a positive impact on streamlining and safeguarding the drug development process for researchers and patients. Additionally, there is a need to increase the dataset and consider other drug features along with 2D structure.

## ABREVATIONS

| | |
|---|---|
| ADR | Adverse Reactions |
| BCE | Binary Cross Entropy |
| Bi-LSTM | Bi-Directional Long Short-Term Memory |
| DDTE | Drug-Disease-Target Embedding |
| DSEs | Drug Side Effects |
| EMBL | European Molecular Biology Laboratory |
| FC | Fully Connected |
| GCNN | Graph Convolutional Neural Network |
| MKL | Multiple Kernel Learning |
| ML | Machine Learning |
| NLP | Natural Language Processing |
| VGG | Visual Geometry Group |

## CONFLICTS OF INTEREST

The authors declare that there are no conflicts of interest regarding the publication of this manuscript. Any affiliations, financial involvement, or relationships with organizations or entities that might pose a conflict of interest with the subject matter discussed in this work are hereby disclosed.

## REFERENCES

[1] *Finding and Learning About Side Effects (adverse Reactions) | FDA*. Accessed: Jul. 22, 2022. [Online]. Available: https://www.fda.gov/drugs/information-consumers-and-patients-drugs/finding-and-learning-about-side-effects-adverse-reactions

[2] S. L. Chan, X. Ang, L. L. Sani, H. Y. Ng, M. D. Winther, J. J. Liu, L. R. Brunham, and A. Chan, "Prevalence and characteristics of adverse drug reactions at admission to hospital: A prospective observational study," *Brit. J. Clin. Pharmacol.*, vol. 82, no. 6, pp. 1636–1646, Dec. 2016, doi: 10.1111/bcp.13081.

[3] Y. Ding, J. Tang, and F. Guo, "Identification of drug-side effect association via multiple information integration with centered kernel alignment," *Neurocomputing*, vol. 325, pp. 211–224, Jan. 2019.

[4] S. Shabani-Mashcool, S.-A. Marashi, and S. Gharaghani, "NDDSA: A network- and domain-based method for predicting drug-side effect associations," *Inf. Process. Manage.*, vol. 57, no. 6, Nov. 2020, Art. no. 102357.

[5] S. Miles, "Rofecoxib," in *Proc. xPharm Compr. Pharmacol. Ref.*, 2007, pp. 1–6, doi: 10.1016/B978-008055232-3.62555-3.

[6] E. Pauwels, V. Stoven, and Y. Yamanishi, "Predicting drug side-effect profiles: A chemical fragment-based approach," *BMC Bioinf.*, vol. 12, no. 1, pp. 1–13, May 2011, doi: 10.1186/1471-2105-12-169.

[7] A. Mullard, "New drugs cost US $2.6 billion to develop," *Nature Rev. Drug Discovery*, vol. 13, p. 877, 2014.

[8] D. A. Nguyen, C. H. Nguyen, and H. Mamitsuka, "A survey on adverse drug reaction studies: Data, tasks and machine learning methods," *Brief. Bioinform.*, vol. 22, no. 1, pp. 164–177, Jan. 2021, doi: 10.1093/BIB/BBZ140.

[9] M. Liu, Y. Wu, Y. Chen, J. Sun, Z. Zhao, X.-W. Chen, M. E. Matheny, and H. Xu, "Large-scale prediction of adverse drug reactions using chemical, biological, and phenotypic properties of drugs," *J. Amer. Med. Inform. Assoc.*, vol. 19, no. e1, pp. e28–e35, Jun. 2012, doi: 10.1136/amiajnl-2011-000699.

[10] S. Jamal, S. Goyal, A. Shanker, and A. Grover, "Predicting neurological adverse drug reactions based on biological, chemical and phenotypic properties of drugs using machine learning models," *Sci. Rep.*, vol. 7, no. 1, Apr. 2017, doi: 10.1038/s41598-017-00908-z.

[11] X. Zhao, L. Chen, Z.-H. Guo, and T. Liu, "Predicting drug side effects with compact integration of heterogeneous networks," *Current Bioinf.*, vol. 14, no. 8, pp. 709–720, Dec. 2019, doi: 10.2174/1574893614666190220114644.

[12] X. Zhao, L. Chen, and J. Lu, "A similarity-based method for prediction of drug side effects with heterogeneous information," *Math. Biosciences*, vol. 306, pp. 136–144, Dec. 2018, doi: 10.1016/j.mbs.2018.09.010.

[13] S. Dey, H. Luo, A. Fokoue, J. Hu, and P. Zhang, "Predicting adverse drug reactions through interpretable deep learning framework," *BMC Bioinf.*, vol. 19, no. S21, Dec. 2018, Art. no. 476, doi: 10.1186/s12859-018-2544-0.

[14] S. Mizutani, E. Pauwels, V. Stoven, S. Goto, and Y. Yamanishi, "Relating drug–protein interaction network with drug side effects," *Bioinformatics*, vol. 28, no. 18, pp. i522–i528, Sep. 2012, doi: 10.1093/bioinformatics/bts383.

[15] Y. Yamanishi, E. Pauwels, and M. Kotera, "Drug side-effect prediction based on the integration of chemical and biological spaces," *J. Chem. Inf. Model.*, vol. 52, no. 12, pp. 3284–3292, Dec. 2012, doi: 10.1021/CI2005548.

[16] Z. Wang, N. R. Clark, and A. Ma'ayan, "Drug-induced adverse events prediction with the LINCS L1000 data," *Bioinformatics*, vol. 32, no. 15, pp. 2338–2345, Aug. 2016, doi: 10.1093/bioinformatics/btw168.

[17] C. Moon, C. Jin, X. Dong, S. Abrar, W. Zheng, R. Y. Chirkova, and A. Tropsha, "Learning drug-disease-target embedding (DDTE) from knowledge graphs to inform drug repurposing hypotheses," *J. Biomed. Informat.*, vol. 119, Jul. 2021, Art. no. 103838, doi: 10.1016/j.jbi.2021.103838.

[18] Y. Ding, J. Tang, and F. Guo, "Identification of drug-side effect association via semisupervised model and multiple kernel learning," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 6, pp. 2619–2632, Nov. 2019, doi: 10.1109/JBHI.2018.2883834.

[19] C. Y. Lee and Y. P. Chen, "Descriptive prediction of drug side-effects using a hybrid deep learning model," *Int. J. Intell. Syst.*, vol. 36, no. 6, pp. 2491–2510, Jun. 2021, doi: 10.1002/int.22389.

[20] G. M. Dimitri and P. Lió, "DrugClust: A machine learning approach for drugs side effects prediction," *Comput. Biol. Chem.*, vol. 68, pp. 204–210, Jun. 2017, doi: 10.1016/j.compbiolchem.2017.03.008.

[21] M. Jiang, B. Zhou, and L. Chen, "Identification of drug side effects with a path-based method," *Math. Biosciences Eng.*, vol. 19, no. 6, pp. 5754–5771, 2022, doi: 10.3934/mbe.2022269.

[22] X. Liang, Y. Fu, L. Qu, P. Zhang, and Y. Chen, "Prediction of drug side effects with transductive matrix co-completion," *Bioinformatics*, vol. 39, no. 1, Jan. 2023, doi: 10.1093/bioinformatics/btad006.

[23] M. M. Hatmal, M. A. I. Al-Hatamleh, A. N. Olaimat, R. Mohamud, M. Fawaz, E. T. Kateeb, O. K. Alkhairy, R. Tayyem, M. Lounis, M. Al-Raeei, R. K. Dana, H. J. Al-Ameer, M. O. Taha, and K. M. Bindayna, "Reported adverse effects and attitudes among Arab populations following COVID-19 vaccination: A large-scale multinational study implementing machine learning tools in predicting post-vaccination adverse effects based on predisposing factors," *Vaccines*, vol. 10, no. 3, p. 366, Feb. 2022, doi: 10.3390/vaccines10030366.

[24] D. N. Swathi and U. Kumaran, "Predicting drug side-effects from open source health forums using supervised classifier approach," in *Proc. 5th Int. Conf. Commun. Electron. Syst. (ICCES)*, 2020, pp. 796–800, doi: 10.1109/ICCES48766.2020.9138065.

[25] K. O'Connor, P. Pimpalkhute, A. Nikfarjam, R. Ginn, K. L. Smith, and G. Gonzalez, "Pharmacovigilance on Twitter? Mining tweets for adverse drug reactions," in *Proc. AMIA Annu. Symp.*, 2014, p. 924.

[26] *SIDER Side Effect Resource*. Accessed: Aug. 16, 2022. [Online]. Available: http://sideeffects.embl.de/

[27] *PubChem*. Accessed: Aug. 16, 2022. [Online]. Available: https://pubchem.ncbi.nlm.nih.gov/

[28] *Fever: Symptoms, Causes, Care & Treatment*. Accessed: Sep. 24, 2022. [Online]. Available: https://my.clevelandclinic.org/health/symptoms/10880-fever

[29] *COVID-19 Dashboard | DrugBank Online*. Accessed: Feb. 9, 2023. [Online]. Available: https://go.drugbank.com/covid-19#drugs

[30] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019, doi: 10.1186/S40537-019-0197-0.

[31] J. Yang, X. Hu, G. Xiao, and Y. Shen, "A survey of knowledge enhanced pre-trained models," 2021, *arXiv:2110.00269*.

[32] M. A. Morid, A. Borjali, and G. Del Fiol, "A scoping review of transfer learning research on medical image analysis using ImageNet," *Comput. Biol. Med.*, vol. 128, Jan. 2021, Art. no. 104115, doi: 10.1016/j.compbiomed.2020.104115.

[33] *Transfer Learning and Fine Tuning of Neural Networks*. Accessed: Mar. 23, 2023. [Online]. Available: https://www.indusmic.com/post/transfer-learning-and-fine-tuning-of-neural-networks

[34] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," 2018, *arXiv:1801.04381*.

[35] *ReLu Definition | DeepAI*. Accessed: Mar. 23, 2023. [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/relu

[36] *Activation Functions in Neural Networks | By SAGAR SHARMA | Towards Data Science*. Accessed: Mar. 23, 2023. [Online]. Available: https://towardsdatascience.com/activation-functions-neural-networks-1cbd9f8d91d6

[37] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[38] *Understanding VGG16: Concepts, Architecture, and Performance*. Accessed: Feb. 13, 2023. [Online]. Available: https://datagen.tech/guides/computer-vision/vgg16/

[39] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger. (2017). *Densely Connected Convolutional Networks*. Accessed: Apr. 7, 2023. [Online]. Available: https://github.com/liuzhuang13/DenseNet.

[40] G. Varoquaux and O. Colliot, "Evaluating machine learning models and their diagnostic value," in *Machine Learning for Brain Disorders* (Neuromethods), vol. 197, O. Colliot, Ed. New York, NY, USA: Humana, 2023, doi: 10.1007/978-1-0716-3195-9_20.

[41] *Overfitting and Underfitting With Machine Learning Algorithms—MachineLearningMastery.com*. Accessed: Jan. 16, 2024. [Online]. Available: https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/

**MUHAMMAD ASAD ARSHED** received the master's degree in computer science from The Islamia University of Bahawalpur (IUB), Pakistan, where he is currently pursuing the Ph.D. degree. He has experience in bioinformatic research from his position as a Research Officer with KICS, University of Engineering and Technology, Lahore, Pakistan. His research interests include medical imaging, bioinformatics, machine learning, deep learning, and NLP.

**MUHAMMAD IBRAHIM** received the Ph.D. degree from the Department of Computer Science, The Islamia University of Bahawalpur. He is currently a Lecturer with the Department of Computer Science, The Islamia University of Bahawalpur. He has published his work in many international journals. His research interests include NLP, machine learning, and deep learning.

**SHAHZAD MUMTAZ** received the master's degree in computer science from The Islamia University of Bahawalpur, Pakistan, in 2005, and the Ph.D. degree in computer science from Aston University, U.K., in 2015. He was the Assistant Director (Computer) with the National Highway Authority, Pakistan, from December 2005 to October 2007. His research projects include Probabilistic Modeling of Blood Glucose Through Eye Parameters, An Analysis of the Protein Family of Major Histocompatibility Complex, Predictive Modeling of Accidents and Emergency Arrivals and Admissions, Patient-Specific Recommendation Systems for HIP Joint Patients, and Predictive Modeling of Extreme Content from Twitter in the Context of Afghanistan. His research interests include the areas of machine learning and data mining and their application to health informatics domains but recently he has worked in other areas, such as natural language analytics and high-performance computing.

**TENVIR ALI** received the Ph.D. degree from the Department of Information and Communication, Yeungnam University, South Korea. He is currently a Faculty Member of the Department of Computer Science, The Islamia University of Bahawalpur. His research interests include search engine optimization, computer vision, deep learning, and blockchain.

**GYU SANG CHOI** (Member, IEEE) received the Ph.D. degree from the Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA, USA, in 2005. He was a Research Staff Member with the Samsung Advanced Institute of Technology (SAIT), Samsung Electronics, from 2006 to 2009. Since 2009, he has been a Faculty Member with the Department of Information and Communication, Yeungnam University, South Korea. His research interests include non-volatile memory and storage systems.