



A comparative study of deep learning models and classification algorithms for chemical compound identification and Tox21 prediction

Yusuf Alaca^a, Berkay Emin^b, Akif Akgul^{a,*}

^a Department of Computer Engineering, Faculty of Engineering, Hittit University, Corum, 19030, Turkiye

^b Department of Electronics and Automation, Osmancik Omer Derindere Vocational School, Hittit University, Corum, 19500, Turkiye



ARTICLE INFO

Keywords:

Chemical property prediction
Image processing
Deep learning
Convolutional neural network
SMILES format

ABSTRACT

Chemical compound classification, toxicity prediction, and environmental risk assessments are critically important in various applications within the field of chemistry. Deep learning models provide highly effective tools for extracting features from complex large datasets and performing classification tasks. Four different deep learning models, namely ResNet50V2, VGG19, InceptionV3, and MobileNetV2, have been compared with the random forest (RF) and k-nearest neighbors (KNN) algorithms. The results obtained from experiments conducted using QR CODE images of the Tox21SMILES dataset demonstrate the effectiveness of deep learning models for classifying chemical compounds and showcase the performance of different classification algorithms. The findings of the study thoroughly evaluate the performance of deep learning models and classification algorithms in the task of chemical classification. While ResNet50V2 and VGG19 models achieve high accuracy and precision, InceptionV3 and MobileNetV2 models provide more balanced results. Additionally, in terms of classification algorithms, the k-nearest neighbors (KNN) algorithm generally outperforms the Random Forest (RF) algorithm. Although the RF algorithm achieves good accuracy, the KNN algorithm proves to be more effective in terms of sensitivity and F1 score. These results emphasize the factors to consider when choosing which deep learning model or classification algorithm to use in chemical classification tasks. In conclusion, this study presents a comprehensive analysis comparing the performance of deep learning models and classification algorithms in chemical classification tasks. The selection of the most suitable model and algorithm for a specific task supports achieving better results in the classification of chemical compounds and related applications.

1. Introduction

Chemical compounds play a vital role in the modern world, serving as foundational elements in various fields like drug development, the food industry, agriculture, and materials science. Toxicity refers to the extent to which a substance can cause harm or damage to living organisms. Toxicology investigates the adverse effects of chemicals on humans and other living beings (Smedley et al., 2013; Chapman, 2000). A compound's reactivity, solubility, toxicity, and numerous other chemical and physical properties offer critical insights into its potential uses and effects (Borosky and Laali, 2015; Sar et al., 2008; Sun et al., 2022). Hence, accurately predicting these properties is integral to studies in this domain.

In recent years, studies based on machine learning have aimed to predict chemical compound properties (Mayr et al., 2016; Xia et al., 2018; Yang et al., 2021). Machine learning methodologies leveraging

pattern recognition algorithms can predict mathematical relationships and properties based on empirical observations of small molecules (Lo et al., 2018). Additionally, machine learning techniques have been applied to predict compounds within metabolic pathways and their roles in metabolomics, offering significant insights (Liu et al., 2021).

Recent literature explores the use of deep learning models for toxicity prediction and molecular structure representation using SMILES notation (M. Hirohara et al., 2018; Hu et al., 2020; Saini, 2023). Insights are provided regarding the application of deep learning algorithms like convolutional neural networks for toxicity prediction based on molecular structures (Klambauer et al., 2017; Bagal et al., 2021; Sinha et al., 2023). The significance of substructure patterns and the random forest algorithm in computer-aided toxicity prediction is also highlighted (Cao et al., 2012).

This study's primary objective is to comprehensively evaluate deep learning models and traditional classification algorithms employed for

* Corresponding author.

E-mail address: akifakgul@hitit.edu.tr (A. Akgul).

effective chemical compound classification. Specifically, the aim is to comprehend and compare the performance of these models and algorithms in chemistry applications such as toxicity prediction and environmental risk assessments. Results from experiments conducted on QR images of the Tox21 dataset showcase the effectiveness of various deep learning models and classification algorithms in chemical classification tasks. Mathematical equations are utilized to forecast chemical properties, often derived through regression analysis (Bouarab-Chibane et al., 2019). QSAR finds extensive use in drug design and toxicology research (Wang et al., 2021). Physicochemical models are also employed to predict chemical compound properties (Sharma et al., 2022; Madani et al., 2021), leveraging principles from quantum chemistry to compute molecular structure's physical and chemical attributes (Sharma et al., 2022; Madani et al., 2021). Techniques such as density functional theory (DFT) or molecular dynamics simulation (MDS) are utilized (Patel, 2020; Rochlani et al., 2023; Sib Tul Hassan Shah and Naeem, 2023). Hansch Analysis (Kubinyi, 1988), a traditional approach, predicts the biological activities of chemical compounds using molecular descriptors and forms the basis of Quantitative Structure-Property Relationship (QSPR) studies (Platzer et al., 2022). Molecular Graph Theory (MGT) is another traditional method of assessing chemical compound topological structures (A Tudoran and V Putz, 2015; Rouvray, 1971), mathematically representing relationships between nodes and edges in molecules.

This study aims to comprehensively evaluate deep learning models and traditional classification algorithms used for effective chemical compound classification. Specifically, the goal is to understand and compare the performance of these models and algorithms in applications within the field of chemistry, such as toxicity prediction and environmental risk assessments. The results obtained from experiments conducted on QR images of the Tox21 dataset demonstrate the effectiveness of different deep learning models and classification algorithms in chemical classification tasks.

In this context, the specific objectives of the study are as follows:

- This study focuses on comparing the performance of four different deep learning models, namely ResNet50V2, VGG19, InceptionV3, and MobileNetV2, to determine which model is more effective in classifying chemical compounds.
- The performance of deep learning models is compared with traditional classification algorithms such as Random Forest (RF) and k-nearest neighbors (KNN), with feature selection performed using transfer learning. This helps identify which approach performs better in classifying chemical compounds.
- A unique approach is adopted, which involves converting the SMILES representations of chemical compounds into QR code images. This provides a more effective method for visualizing chemical structures.
- Large datasets of chemical compounds, such as the Tox21SMILES dataset, are used to evaluate the performance of deep learning models and classification algorithms. This dataset represents the diversity of real-world chemical compounds.
- A comprehensive analysis is provided to determine which model or algorithm is more suitable for a specific task, evaluating the performance of each model and algorithm in detail.
- The obtained results are elaborated in detail to understand the strengths and weaknesses of each model and algorithm.

Conducted with these objectives in mind, the study aims to comprehend the performance of models and algorithms used in classification tasks within the field of chemistry and to direct this knowledge towards relevant applications.

2. Material and methods

This study aims to develop an innovative approach for predicting the properties of chemical compounds. The study utilizes the Tox21 dataset

represented in SMILES notation. Initially, SMILES images were converted into QR codes. These QR code images were then used to train ResNet50V2, VGG19, InceptionV3, and MobileNetV2 convolutional neural network (CNN) architectures, categorizing the data into two distinct categories: toxic and nontoxic. Subsequently, the data obtained from CNN architectures underwent classification using Random Forest (RF) and K-Nearest Neighbor (KNN) algorithms, and their performances were compared based on the achieved results.

2.1. Dataset and data preprocessing

This study employs the Tox21 dataset, which plays a crucial role in understanding the properties and behaviors of chemical compounds. Each compound is represented by Simplified Molecular Input Line Entry System (SMILES) notations sourced from chemical databases or literature (Weininger et al., 1989). SMILES notation is a system designed to encode the structure of chemical compounds in a text-based format, explicitly expressing molecular structures, chemical bonds, and atoms.

Our study presents results obtained from experiments utilizing QR code images of the Tox21 SMILES dataset, evaluating the effectiveness of deep learning models and the performance of various classification algorithms in classifying chemical compounds. The Tox21 dataset, acquired from the MoleculeNet benchmark (Wu et al., 2018), is characterized in Table 1. It includes labels indicating the structure and toxicity status of chemical compounds, representing different tasks aimed at predicting various properties and activities of these compounds. Therefore, the proposed model is deemed suitable for assessing general chemical property prediction capabilities.

Tox21 is a database containing 12 targets and 12,707 small molecules, all of which are compounds tested for their toxicities. The classification of compounds follows a binary structure, with each compound categorized as either toxic or nontoxic. Five out of the 12 targets are associated with hormones, such as the estrogen receptor (ER) and its ligand-binding domain, the androgen receptor (AR) and its ligand-binding domain, as well as aromatase. Both ER and AR receptors regulate gene expression and play significant roles in sexual maturation and pregnancy. Aromatase catalyzes a reaction converting testosterone to the estradiol hormone. An aromatase deficiency can lead to delayed puberty in females, osteoporosis in males, and virilization in pregnant mothers. Conversely, an excess of aromatase can result in early puberty and breast development in males.

The remaining seven targets belong to stress response pathways (HSE, MMP, ATAD5, PPAR γ , ARE, AhR, and p53). Cells activate the heat shock response element (HSE) in response to stressful conditions. Mitochondrial membrane potential (MMP) decreases during apoptotic cell death. Expression of 5 containing the AAA domain (ATAD5) leads to an extended lifespan, resulting in increased ineffective replication factories and delayed progression in the S-phase.

Peroxisome Proliferator-Activated Receptor γ (PPAR γ) plays a crucial regulatory role in energy homeostasis. The antioxidant response

Table 1
The Tox21 dataset includes labels indicating the structure and toxicity status of chemical compounds.

Datasets	Task
Tox21	Nuclear Receptor Panel AhR AR AR-LBD Aromatase ER ER-LBD NR.PPAR.gamma ARE ATAD5
	Stress Response Panel HSE MMP p53

element (ARE) regulates cytoprotective genes critical for redox homeostasis. In addition to regulating metabolic enzymes through gene expression, the aryl hydrocarbon receptor (AhR) plays a role in immune regulation, stem cell maintenance, and cellular differentiation. The induction of metabolic enzymes can lead to the production of toxic metabolites.

When cells undergo DNA damage, the tumor suppressor protein p53 is expressed, causing a growth pause to balance effects, repairing DNA, or initiating the cell death process. However, it is also associated with drug resistance in cancer cells. Table 2 provides representations and formulas for some chemicals found in the Tox21 database.

2.2. CNN models

Convolutional Neural Networks (CNNs) are artificial neural networks specifically developed for visual data analysis (LeCun et al., 2015). They are utilized to achieve high performance in tasks such as image classification, object recognition, and face detection (Lu et al., 2021). CNNs have played a significant role in applications such as natural object classification, segmentation, handwriting recognition, and face recognition (Jasim et al., 2022). They can learn complex relationships between input and output without explicitly expressing these relationships through a mathematical equation (Aufar and Sitanggang, 2022). A basic CNN model consists of key components such as convolutional layers, activation functions, and pooling layers. CNN architectures have shown promising results in various computer vision and machine learning tasks (Wang et al., 2018).

In this study, CNNs have been customized for the analysis of molecular structures. The aim is to predict the features of chemical compounds by converting their SMILES representations into QR code images. The CNN sub-models used in this study include ResNet50V2, VGG19, InceptionV3, and MobileNetV2.

- ResNet50V2:** Residual Network 50 Version 2 (ResNet50V2) is a 50-layer CNN model developed by Microsoft Research (He et al., 2016). It has been employed to convert SMILES representations into QR code images. The model consists of four blocks with a total of 16 connected blocks, containing 48 convolutional layers. Fig. 1 illustrates the architecture of the ResNet50V2 model. In the study, the

Table 2
The structure and formulas of chemical compounds in Tox21.

Structure	Preferred Name	Mol. Formula
	Fusaric acid	C10H13NO2
	Ethelyn(triethoxy)silane	C8H18O3Si
	2-Ethylhexaldehyde	C8H16O
	tert-Butylamine	C4H11N
	Methyl trifluoromethanesulfonate	C2H3F3O3S

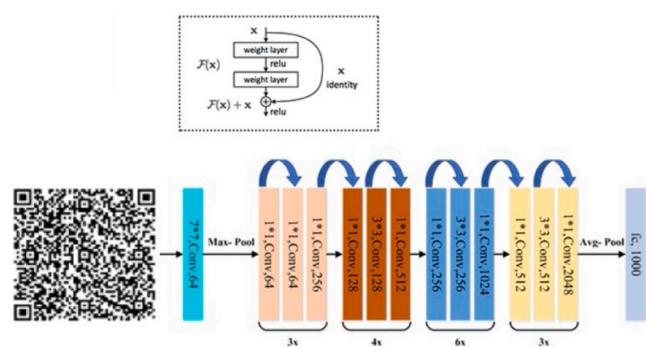


Fig. 1. ResNet50V2 architecture and its implementation in the proposed model.

ResNet50V2 deep learning model is utilized to demonstrate the classification performance of chemical compounds based on the Tox21 SMILES dataset, using QR code images and two different classification algorithms.

- VGG19:** VGG19 is a deep learning model developed by Simonyan and Zisserman (K. Simonyan and Zisserman, 2015). It has an input layer of size $224 \times 224 \times 3$ and is used to identify and classify features in images (K. Simonyan and Zisserman, 2015). Fig. 2 illustrates the architecture of the VGG19 model. In the study, the classification performance of chemical compounds based on the Tox21 SMILES dataset is demonstrated using QR code images and two different classification algorithms, with the VGG19 deep learning model.
- InceptionV3:** InceptionV3 is a deep learning model developed by Google, specifically designed for visual processing tasks like image classification and object recognition (Szegedy et al., 2016). It has a complex CNN architecture with specialized modules. Fig. 3 illustrates the architecture of the Inception V3 model. In the study, the performance of chemical compound classification based on the Tox21 SMILES dataset is presented using QR code images and two different classification algorithms with the Inception V3 deep learning model.
- MobileNetV2:** MobileNetV2 is a lightweight and efficient deep learning model developed by Google, specifically designed for resource-constrained platforms such as mobile and edge devices (Sandler et al., 2018). It incorporates convolutional layers, inverted residuals, global average pooling, and fully connected layer structures (Zhang et al., 2022). Fig. 4 illustrates the architecture of the MobileNetV2 model. The study explores the performance of chemical compound classification based on the Tox21 SMILES dataset using QR code images and two different classification algorithms with the MobileNetV2 deep learning model.

The study examines the performance of chemical compound classification based on the Tox21 SMILES dataset using QR code images and two different classification algorithms. These models demonstrate significant potential in the field of chemical data processing, particularly in the analysis and classification of molecular structures.

2.3. Classification algorithms

The primary goal of classification algorithms is to categorize or assign data into specific classes or categories. These algorithms find applications in various fields such as data mining (Alaoui et al., 2017), image recognition (Li, 2022), natural language processing (Sait and Ishak, 2023), and medicine (Xing and Bei, 2020). In this study, Random Forests and K-Nearest Neighbors classification algorithms have been employed.

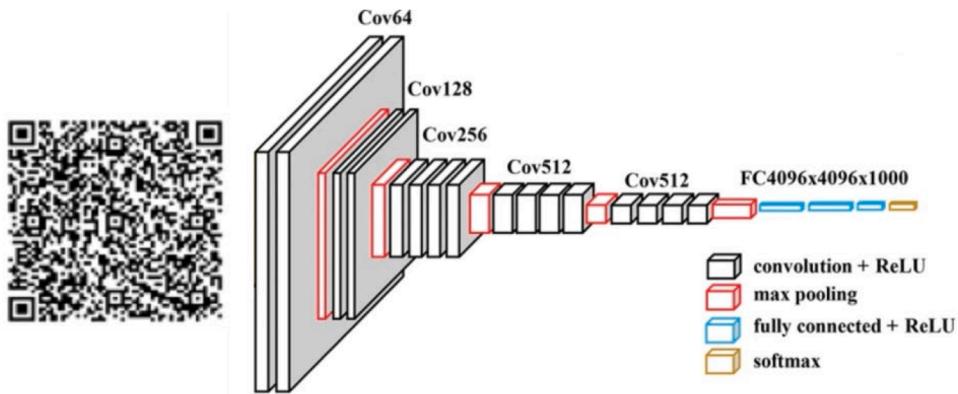


Fig. 2. VGG19 architecture and its implementation in the proposed model.

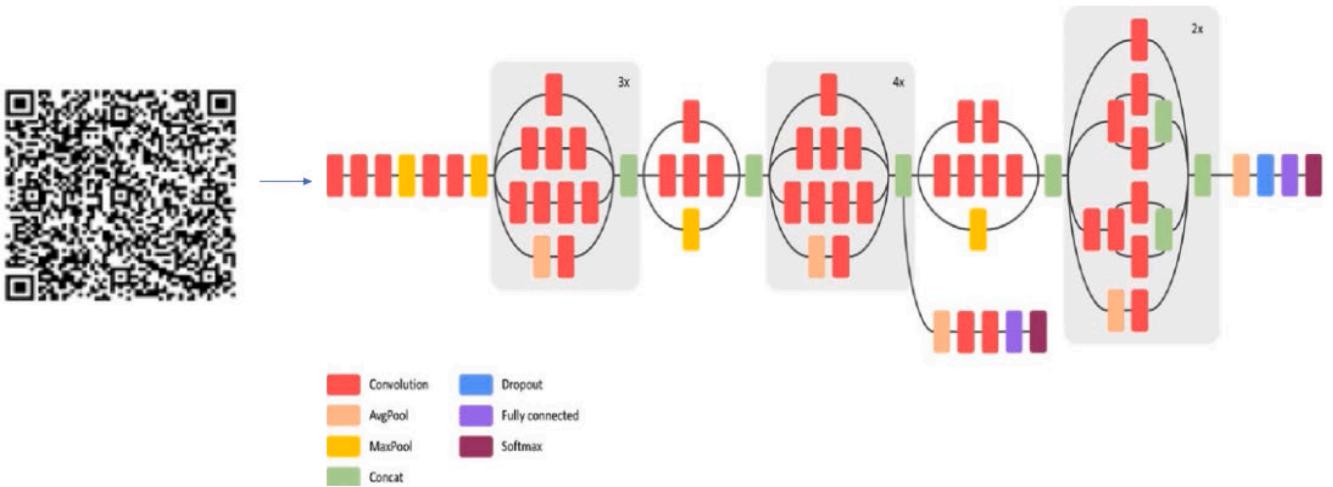


Fig. 3. InceptionV3 architecture and its implementation in the proposed model.

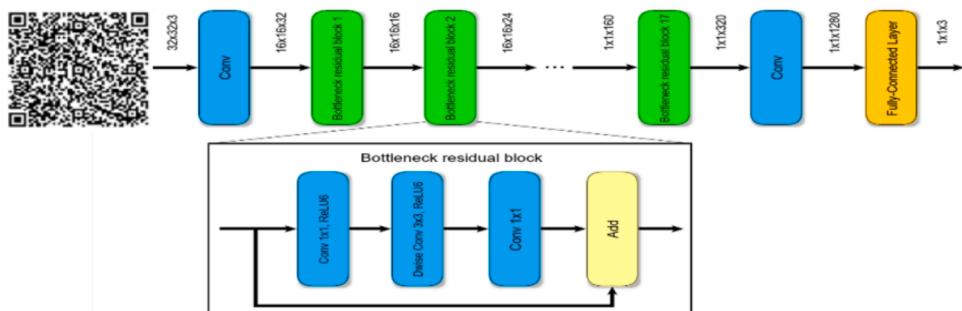


Fig. 4. MobileNetV2 architecture and its implementation in the proposed model.

2.3.1. Random forest (RF)

Random Forest is an ensemble method where multiple decision trees come together to form a collective approach. Each decision tree is trained with randomly selected samples and features. As a result, each tree focuses on different features and data points, thus evaluating the dataset from various perspectives. The Random Forest (RF) algorithm was proposed by Leo Breiman in 2001 (Breiman, 2001). Random Forest is a machine learning algorithm used as an ensemble learning method. RF constructs a model by combining many trees built based on decision trees. Each tree is trained using random samples and features from the dataset, and then these trees are aggregated to produce the result (Kilinc et al., 2020). The RF algorithm is commonly employed in various fields such as classification (Speiser et al., 2019), regression (Suha and Sanam,

2022), and anomaly detection (Aldweesh et al., 2020).

The algorithm starts by gathering random samples from the dataset and constructing decision trees based on these samples. At each node of the decision trees, a feature selection is made to determine the best split, ensuring robustness and accuracy. This process is repeated for different subsets of the dataset by each tree. After constructing all trees, the classification or regression process occurs by aggregating the predictions of all trees, typically through averaging or majority voting. RF can be expressed as Equation

$$\hat{Y} = \text{aggregate}(f_1(X), f_2(X), \dots, f_t(X)) \quad (1)$$

Here, Y denotes the final prediction, $f_t(X)$ represents the prediction of the t -th decision tree, and the aggregate function generally signifies

taking the majority vote in the case of classification.

This article uses the RF algorithm to predict chemical compound properties in the proposed models. The proposed model employs the RF algorithm to predict the properties of chemical compounds. Initially, the SMILES representations of chemical compounds are transformed into interpretable QR code images by the deep learning models used in this study. This transformation visually represents the information containing molecular structures, chemical bonds, and atoms from the SMILES notation in a QR code image format. Subsequently, these QR code image representations are processed using the RF algorithm to predict chemical properties. RF combines the results from all trained trees and performs the classification operation by aggregating them.

2.3.2. k-nearest neighbor (KNN)

The K-Nearest Neighbor (KNN) algorithm, initially developed in 1967, remains a fundamental method in the realm of machine learning. It operates as a non-parametric, instance-based approach, serving both classification and regression problems in a supervised manner (Cover and Hart, 1967). KNN, as its name suggests, determines the class of an item by examining the 'k' nearest neighbors. The Euclidean distance is typically used as a distance measure in the KNN algorithm. It is calculated as the square root of the sum of the squares of the differences between the respective coordinates along each dimension for two points. The Euclidean distance formula is represented by Eq. (2):

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} \quad (2)$$

Here, 'p' and 'q' denote two points, and 'n' represents the number of dimensions.

The steps involved in the K-Nearest Neighbor (KNN) algorithm are as follows:

1. Select K labeled training set points (neighbor points) with known labels.
2. Calculate the Euclidean distances between the test example and its K nearest neighbors.
3. Identify the K nearest neighbors based on the computed distance values.
4. Count the number of training examples in each category among the K neighbors.
5. Determine the category of the test example based on the categories of these examples, with the majority category prevailing.

In the context of this study, the focus is on employing the kNN algorithm to classify and predict the properties of chemical compounds. Initially, the SMILES representations of these compounds transform into QR code images, a process facilitated by the deep learning models utilized. This conversion visually encodes information regarding molecular structures, chemical bonds, and atoms from the SMILES notation into a QR code image format. Subsequently, the QR code image representations undergo processing using the kNN algorithm to predict various chemical properties. The kNN algorithm accomplishes this by identifying the nearest neighbors to the chemical compound under examination and leveraging their characteristics for classification purposes.

3. Proposed model

In this study, four different deep-learning models and two classification algorithms were employed for the classification of chemical compounds. The deep learning models include ResNet50V2, VGG19, InceptionV3, and MobileNetV2, while the classification algorithms used were Random Forest (RF) and K-Nearest Neighbors (KNN). Deep learning models are highly effective tools for extracting features from large datasets and performing classification tasks. ResNet50V2 is a deep learning model based on the residual learning framework, primarily used for image classification tasks. This model addresses the overfitting

problem that arises with deeper models, allowing it to learn more complex datasets and achieve higher accuracy.

VGG19, with its 19-layer architecture, is another deep learning model known for high performance in image classification tasks. It incorporates a deeper network structure to learn more complex datasets. InceptionV3, developed by Google, is designed for more complex and extensive datasets. Despite having numerous parameters, InceptionV3 ensures computational efficiency through weight sharing and modular design. MobileNetV2 is a lightweight and fast deep learning model specifically designed for mobile and edge devices, achieving high performance even with less computational resources.

The workflow of the proposed model is illustrated in Fig. 5 of this study. The study consists of two main stages: the selection of the most suitable deep learning model and the subsequent selection of the appropriate classification algorithm. Emphasis is placed on highlighting the factors to consider when choosing a deep learning model or classification algorithm for chemical classification tasks. Selecting the most suitable model and algorithm for a specific task supports achieving better results in the classification of chemical compounds and related applications. Therefore, this study provides a comprehensive analysis comparing the performance of deep learning models and classification algorithms in chemical classification tasks. The selection of the most suitable model and algorithm for a specific task supports achieving better results in the classification of chemical compounds and related applications.

4. Results

This study extensively analyzes the performance of four distinct deep-learning models utilized in the classification and toxicity prediction tasks of chemical compounds. The evaluation encompasses ResNet50V2, VGG19, InceptionV3, and MobileNetV2 models using QR code images from the Tox21SMILES dataset. The results obtained offer a comprehensive assessment of each model's performance and algorithmic efficacy. Performance metrics, including accuracy, sensitivity, specificity, precision, and F-score, are computed through a confusion matrix analysis. The mathematical formulations for these metrics are provided in Eqs. (3), 4, 5, 6, and 7 (M. Hirohara et al., 2018; Matsuzaka and Uesawa, 2019). Performance metrics such as accuracy, precision, specificity, sensitivity, and F1 score are utilized to evaluate different aspects of classification results. Accuracy measures the ratio of correctly classified examples, while precision gauges the ability to correctly classify toxic compounds. Specificity measures the ability to accurately classify nontoxic compounds. Sensitivity expresses the capacity to reduce the number of false positive classifications. The F1 score is used as a balance between sensitivity and specificity.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (3)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (5)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (6)$$

$$F - \text{Scor} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}} \quad (7)$$

The results presented in Table 3 demonstrate varying performances among different deep-learning models and classification algorithms in the chemical classification task. These results specifically highlight the performance of each model. Notably, the ResNet50V2 model achieved the highest accuracy rate at 0.9767, indicating its overall success. Additionally, this model showed high sensitivity (0.9950), specificity

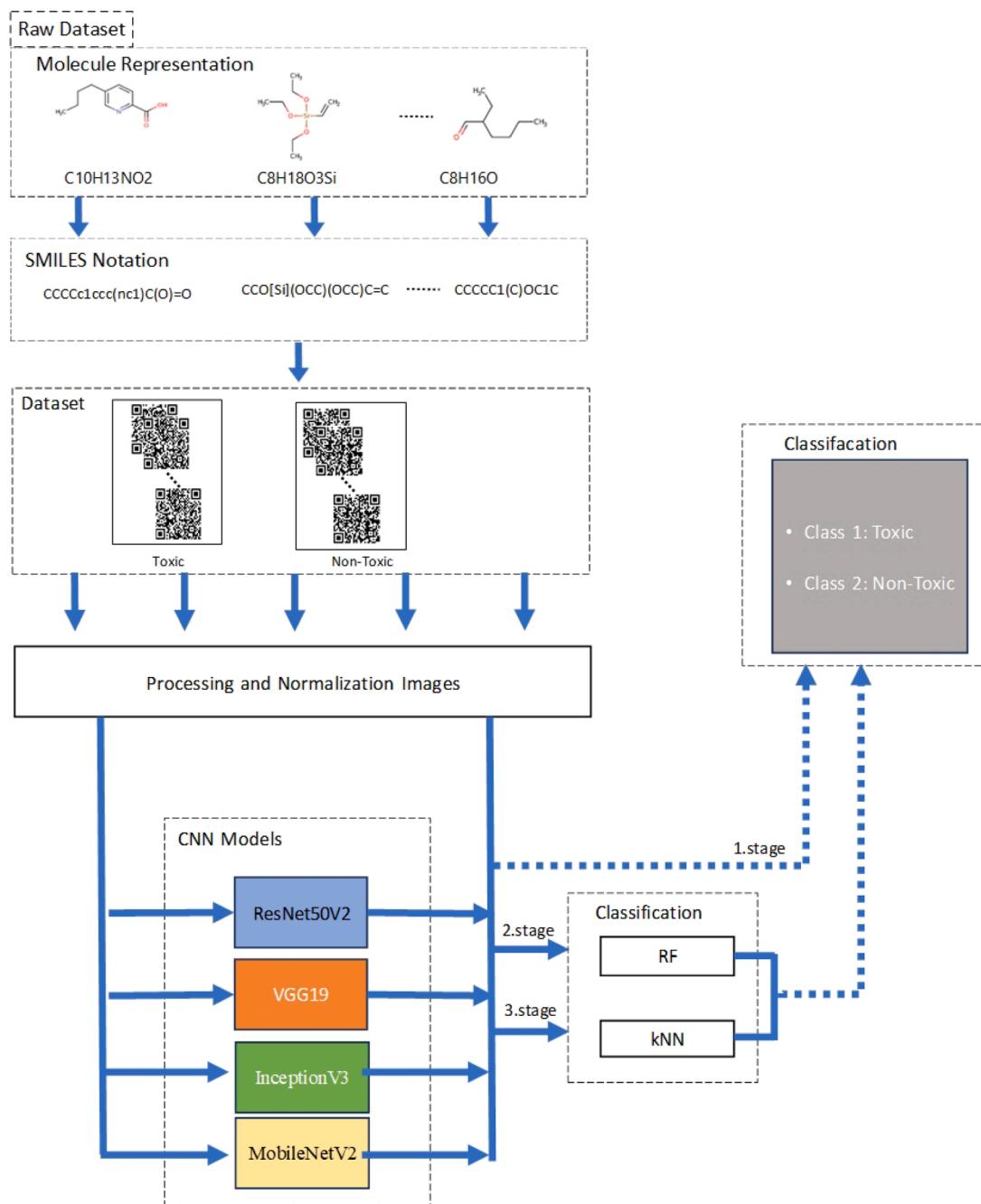


Fig. 5. Displaying the flowchart of the main model.

Table 3

Presents the performance results of CNN models on SMILES images.

CNN Models	Accuracy	Sensitivity	Specificity	Precision	F1_score
ResNet50V2	0.9767	0.9950	0.9572	0.9944	0.9754
VGG19	0.8372	0.9950	0.6684	0.9920	0.7987
InceptionV3	0.9741	0.9950	0.9197	0.9944	0.9556
MobileNetV2	0.9715	0.9950	0.9197	0.9888	0.9530

(0.9572), and precision (0.9944) values, with its F1 score (0.9754) further supporting its effectiveness. On the other hand, the VGG19 model stands out with a low accuracy rate (0.8372). However, this model still achieves high sensitivity (0.9950) while demonstrating weak performance in specificity (0.6684) and precision (0.9920) values. The F1 score (0.7987) reflects a lower level of success for the VGG19 model

compared to the accuracy value. The InceptionV3 model achieves high accuracy (0.9741) and sensitivity (0.9950) while delivering satisfactory performance in terms of specificity (0.9197) and precision (0.9944). The F1 score (0.9556) of this model indicates a balanced classification ability. Finally, the MobileNetV2 model draws attention with high accuracy (0.9715) and sensitivity (0.9950). It also obtains satisfactory results in terms of specificity (0.9197) and precision (0.9888). The F1 score (0.9530) reflects that MobileNetV2 is an effective model in the classification task. These results demonstrate that different deep learning models exhibit different performances in the chemical compound classification task.

When assessing the performance of deep learning models using confusion matrices and relevant metrics, as shown in Fig. 6, it becomes clear that each model possesses distinct advantages and disadvantages. In general, the InceptionV3 and MobileNetV2 models achieve a good

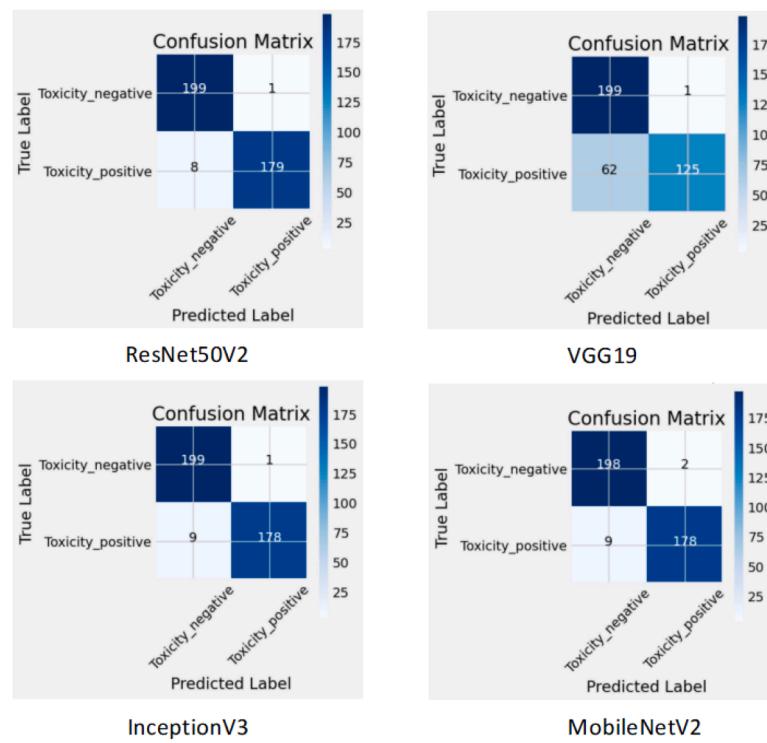


Fig. 6. Confusion matrices of four different deep learning models.

balance with high sensitivity and specificity. Particularly, InceptionV3 stands out in terms of the F1 score. The ResNet50V2 model showcases high sensitivity and specificity but appears to have a relatively higher number of false positives. The VGG19 model demonstrates good performance in terms of sensitivity but shows weak performance in specificity along with high false positives.

The comparison of obtained confusion matrices and metrics highlights different strengths and weaknesses for each model. ResNet50V2

attains high accuracy with both high sensitivity and specificity, albeit with a slightly higher number of false positives. The F1 score is also notably high, indicating a balanced classification ability. This model tends to successfully identify toxic compounds but tends to produce false alarms. VGG19 exhibits high sensitivity but a considerably high number of false positives. Specificity is low, and the F1 score is lower compared to other models. This model tends to correctly identify toxic compounds but tends to falsely classify nontoxic compounds as toxic. InceptionV3

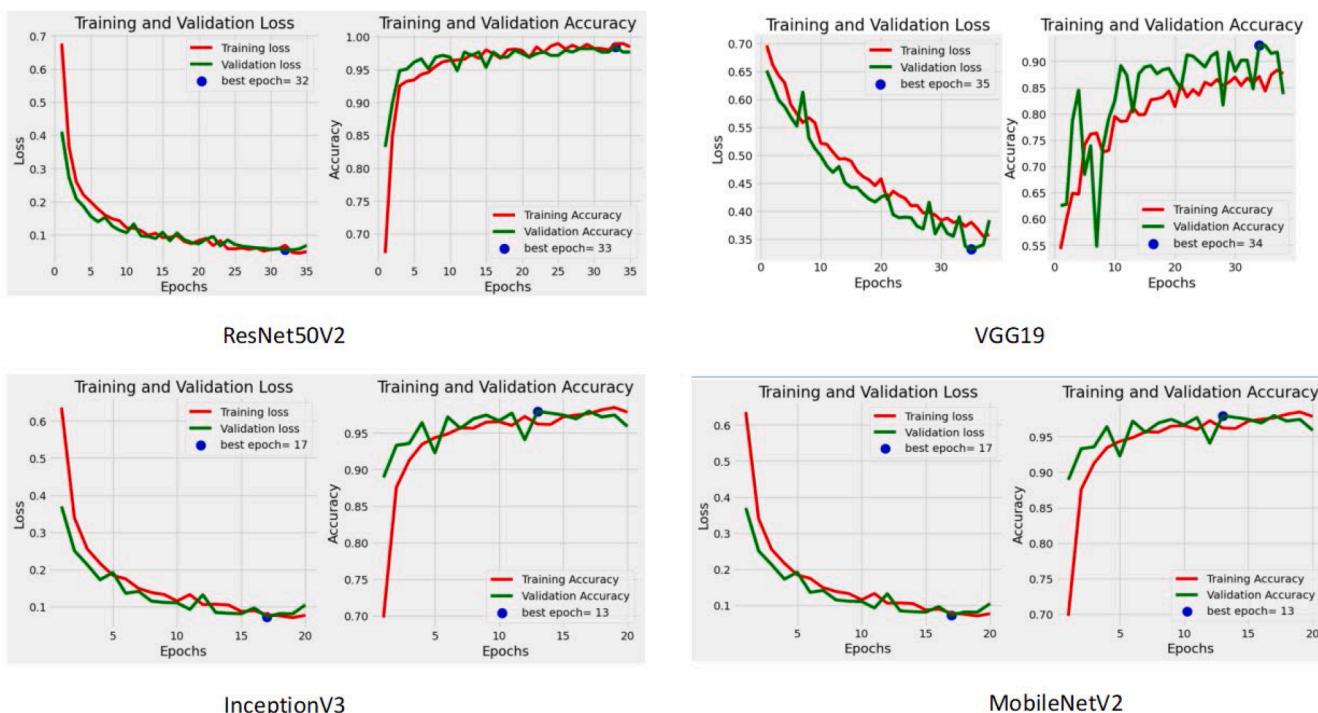


Fig. 7. The accuracy-loss graphs of four different models.

demonstrates successful performance with high sensitivity and specificity, and a high F1 score, showcasing balanced classification ability. It effectively recognizes both toxic and nontoxic compounds. MobileNetV2 achieves high sensitivity and specificity along with a high F1 score. This model successfully identifies both toxic and nontoxic compounds.

When selecting a model, it's crucial to consider the specific requirements of the application. Factors such as not missing toxic substances (high sensitivity) or avoiding incorrect classification of nontoxic substances as toxic (high specificity) should be prioritized. The F1 score provides a balanced measure considering both sensitivity and specificity, aiding in making informed decisions about model selection.

This study was conducted to evaluate the performance of four different deep-learning models in the task of toxicity classification of chemical compounds. The results were represented and interpreted visually using accuracy and loss graphs in Fig. 7. The accuracy graphs for each of the four deep learning models on the training data start from a low level and gradually increase over time. This indicates the process of the models adapting to the training data. In the graphs, an increasing trend of accuracy during training epochs is observed. The accuracy graph on the training data shows that each model achieves high accuracy rates. While ResNet50V2, InceptionV3, and MobileNetV2 models achieve high accuracy on the training data, the VGG19 model starts from a slightly lower point. The accuracy graph on the validation data shows that all four models maintain high accuracy rates on the validation data at the end of training epochs. The generalization abilities of the models indicate that they adapt well to new and unseen data. The loss graph on the training data shows that the training process of each of the four models starts, and the error rate decreases over time. The models learn the training data better and improve their predictions. The loss graph on the validation data reflects the generalization abilities of the models. The graphs show that even at the end of training epochs, the loss values are low, indicating that the models adapt well to new data without overfitting. The accuracy and loss graphs demonstrate that each of the four deep learning models effectively performs the task of chemical toxicity classification and exhibits high performance on both training and validation data. The training processes of the models are successful, and their generalization abilities are high.

Table 4 summarizes the results of experiments using four deep learning models—ResNet50V2, VGG19, InceptionV3, and MobileNetV2—along with two classification algorithms, RF (Random Forest) and KNN (K-Nearest Neighbors). These results are crucial for evaluating how well these models and algorithms classify chemical compounds.

Accuracy indicates the percentage of correctly classified examples, where high accuracy implies good overall classification performance of the model. **Table 3** provides accuracy values for each model and classification algorithm. Notably, the VGG19 model combined with the RF classification algorithm achieves high accuracy values, particularly with a score of 0.9965, indicating strong overall performance.

Sensitivity measures how many of the total positive examples are correctly classified, with high sensitivity indicating the model's tendency not to miss toxic examples. RF and KNN classification algorithms exhibit high sensitivity values for each model. Specificity measures how many of the total negative examples are correctly classified, and high specificity indicates the model's tendency not to misclassify nontoxic

examples as toxic. The VGG19 model demonstrates high performance in terms of specificity.

Precision measures the ratio of true positives to the total positive predictions, with high precision indicating the model's ability to keep false positives low. The VGG19 model combined with the RF classification algorithm achieves high precision values.

The F1 score represents the harmonic mean of precision and sensitivity, with a high F1 score indicating a balanced classification capability. This metric attains high values for each model and classification algorithm.

In conclusion, this table provides a comprehensive analysis, evaluating the performance of each deep learning model and classification algorithm in detail. Model and algorithm selection should be based on factors specific to a given application context. The high values of accuracy, sensitivity, specificity, precision, and F1 score suggest the effectiveness of these methods in the classification task of chemical compounds.

5. Discussion

This study was conducted to examine the performance of deep learning models used in the prediction of toxicity for chemical compounds. Various deep learning models, with different architectures and input data, were tested on QR code images of the Tox21 dataset. The results highlight some strengths and weaknesses of this study. The proposed approach stands out with a remarkably high success rate of 99.65%, indicating that deep learning models can be effective tools in toxicity prediction. Additionally, the study has diversified by employing different deep learning model architectures, showcasing that various approaches can be employed in predicting the toxicity of chemical compounds. Furthermore, adopting a different input method using QR code images from the Tox21 dataset helps in exploring new approaches to toxicity prediction. In comparison with other studies in **Table 5**, the work of Du et al. (Du et al., 2017) also achieves a high success rate of 91.00%. However, their input data only includes SMILES transformation fingerprints, which limits the diversity of input data. Studies by Hirohara et al. (M. Hirohara et al., 2018) and Matsuzaka et al. (Matsuzaka and Uesawa, 2019) also achieve high success rates, but when compared to the proposed study, they exhibit lower success rates. The uniqueness of the proposed approach lies in the use of QR code images as input data, offering a different visual representation of chemical structures and presenting a novel approach to toxicity prediction. To conclude, this study demonstrates that deep learning models can be effective tools in predicting the toxicity of chemical compounds. Various model architectures and input data types can be utilized in toxicity prediction, offering a new perspective for research in this field. This study contributes to exploring new approaches in toxicity prediction.

6. Conclusion

This study presents a comprehensive analysis of deep learning models and classification algorithms in the context of chemical compound classification. It evaluated the performance of various algorithms using the Tox21 dataset, focusing on identifying the most effective

Table 4
Experimental Test Results with CNN and Classification" can be the English translation of your phrase.

CNN Models	Classification	Accuracy	Sensitivity	Specificity	Precision	F1_score
ResNet50V2	RF	0.9484	0.9539	0.9424	0.9492	0.9458
	KNN	0.9914	0.9901	0.9928	0.9892	0.9910
VGG19	RF	0.9965	0.9967	0.9964	0.9964	0.9964
	KNN	0.9879	0.9901	0.9856	0.9891	0.9873
InceptionV3	RF	0.9450	0.9605	0.9280	0.9555	0.9416
	KNN	0.9501	0.9605	0.9388	0.9560	0.9473
MobileNetV2	RF	0.9329	0.9473	0.9172	0.9409	0.9289
	KNN	0.9725	0.9671	0.9784	0.9645	0.9714

Table 5
Comparing the recommended model with other studies.

Authors	Methods	Datasets	Input Types	Acc (%)
2020, Du et al. (Du et al., 2017)	MACCS Fingerprint and Combined Classifier (CC)	Tox21 10 K Compound Library (Tox21Dataset)	SMILES convolution fingerprint	91.00
2019, Yang et al. (Yang et al., 2019)	GCN	Tox21 (Toxicity) Datasets	SMILES	62.00
2018, Hirohara et al. (M. Hirohara et al., 2018)	CNN	Tox21 (Toxicity) Datasets	2D Images	81.30
2021, Matsuzaka et al. (Matsuzaka and Uesawa, 2019)	CNN	Tox21 (Toxicity) Datasets	2D Images	71.00
2022, This paper	CNN (ResNet50V2, VG19, InceptionV3, MobileNetV2)	Tox21 (Toxicity) Datasets	QR CODE Images	99.65

models and approaches for classifying chemical compounds. The results of this study emphasize the high accuracy and precision achieved by deep learning models in chemical compound classification tasks. Models such as ResNet50V2 and VGG19 exhibited high accuracy and precision, while InceptionV3 and MobileNetV2 achieved more balanced results. Additionally, in terms of classification algorithms, the KNN algorithm outperformed the RF algorithm, especially in terms of sensitivity and F1 score. These findings highlight the importance of considering various factors when selecting the most suitable model or algorithm for chemical compound classification tasks. Furthermore, this study draws attention to the significance of using different data representations, such as QR code images, as input types for deep learning models. The exploration of different input types guides innovative approaches in chemical compound classification. Overall, this research contributes to understanding the performance of deep learning models and classification algorithms in chemical compound classification tasks, particularly achieving significant results in toxicity prediction. Understanding the strengths and weaknesses of different models and algorithms in terms of accuracy, precision, and sensitivity helps researchers and practitioners make informed decisions when selecting the most appropriate approach for specific chemical compound classification needs. Moreover, this study encourages advancements in the field of chemical compound classification by guiding and exploring new input types and model architectures.

CRediT authorship contribution statement

Yusuf Alaca: Writing – original draft, Software, Data curation. **Berkay Emin:** Writing – original draft, Methodology, Investigation. **Akif Akgul:** Writing – review & editing, Methodology, Investigation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- Alaoui, S.S., Farhaoui, Y., Aksasse, B., 2017. Classification algorithms in data mining – a survey. *A Compar. Study Classific. Techn. Data Mining Algor.* 6, 1–6.
- Aldweesh, A., Derhab, A., Emam, A.Z., 2020. Deep learning approaches for anomaly-based intrusion detection systems: a survey, taxonomy, and open issues. *Knowl. Based. Syst.* 189, 105124.
- A Tudoran, M., V Putz, M., 2015. Molecular graph theory: from adjacency information to colored topology by chemical reactivity. *Curr. Org. Chem.* 19, 359–386.
- Aufar, Y., Sitanggang, I.S., 2022. Face recognition based on siamese convolutional neural network using Kivy framework. *Indon. J. Electrical Eng. Comp. Sci.* <https://doi.org/10.11591/jeeecs.v26.i2.pp764-772>. Epub ahead of print.
- Bagal, V., Aggarwal, R., Vinod, P.K., et al., 2021. MolGPT: molecular generation using a transformer-decoder model. *J. Chem. Inf. Model.* <https://doi.org/10.1021/acs.jcim.1c00600>. Epub ahead of print.
- Borosky, G.L., Laali, K.K., 2015. *In Silico* study on chemical properties and reactivity of enal derivatives. *Euro. J. Org. Chem.* <https://doi.org/10.1002/ejoc.201500853>. Epub ahead of print.
- Bouarab-Chibane, L., Forquet, V., Lantéri, P., et al., 2019. Antibacterial properties of polyphenols: characterization and QSAR (Quantitative structure–activity relationship) models. *Front. Microbiol.* 10, 829.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Cao, D., Yang, Y.-N., Zhao, J., et al., 2012. Computer-aided prediction of toxicity with substructure pattern and random forest. *J. Chemom.* <https://doi.org/10.1002/cem.1416>. Epub ahead of print.
- Chapman, P.M., 2000. Whole effluent toxicity testing—usefulness, level of protection, and risk assessment. *Environ. Toxicol. Chem.* <https://doi.org/10.1002/etc.5620190102>. Epub ahead of print.
- Cover, T., Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Trans. Inf. Theory.* 13, 21–27.
- Du, H., Cai, Y., Yang, H., et al., 2017. In silico prediction of chemicals binding to aromatase with machine learning methods. *Chem. Res. Toxicol.* 30, 1209–1218.
- He, K., Zhang, X., Ren, S., et al., 2016. Deep residual learning for image recognition. In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* 2016, pp. 770–778. -Decem.
- Hirohara, M., Saito, Y., Koda, Y., et al., 2018a. Convolutional neural network based on SMILES representation of compounds for detecting chemical motif. *BMC. Bioinform.* 19, 526.
- Hirohara, M., Saito, Y., Koda, Y., et al., 2018b. Convolutional neural network based on SMILES representation of compounds for detecting chemical motif. *BMC. Bioinform.* 19 <https://doi.org/10.1186/s12859-018-2523-5>. Epub ahead of print.
- Hu, S.S., Chen, P., Gu, P., et al., 2020. A deep learning-based chemical system for QSAR prediction. *IEEE J. Biomed. Health Inform.* 24, 3020–3028.
- Jasim, W.N., Almola, S.A.S., Alabiech, M.H.H., et al., 2022. Citrus diseases recognition by using CNN. *Informatica.* <https://doi.org/10.31449/inf.v46i7.4284>. Epub ahead of print.
- Kılınç, M., Tarhan, Ç., Aydin, C., 2020. Kitle fonlaması projelerinin karar ağacı ve rastgele orman algoritmalarıyla sınıflandırılması. *J. Info. Syst. Mgmt. Res.* 2 (2), 16–25.
- Klambauer, G., Unterthiner, T., Mayr, A., et al., 2017. DeepTox: toxicity prediction using deep learning. *Toxicol. Lett.* <https://doi.org/10.1016/j.toxlet.2017.07.175>. Epub ahead of print.
- Kubinyi, H., 1988. Free wilson analysis. Theory, applications and its relationship to Hansch analysis. *Quant. Struc.-Activity Relation.* 7, 121–133.
- LeCun, Y., Bengio, Y., Hinton, G.E., 2015. Deep learning. *Nature.* <https://doi.org/10.1038/nature14539>. Epub ahead of print.
- Li, Y., 2022. Research and application of deep learning in image recognition. In: *2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications, ICPECA 2022*, pp. 994–999.
- Liu J., Ali H., Yang Z., et al. Supervised learning techniques to predict compounds in pathway modules based on molecular properties. Epub ahead of print 2021. [10.2130/rs.3.rs.1140648/v1](https://doi.org/10.2130/rs.3.rs.1140648/v1).
- Lo, Y.C., Rensi, S., Tornig, W., et al., 2018. Machine learning in chemoinformatics and drug discovery. *Drug Discov. Today.* <https://doi.org/10.1016/j.drudis.2018.05.010>. Epub ahead of print.
- Lu, R., Li, Y., Yan, Y., et al., 2021. The object recognition research based on convolution neural network. In: *2021 International Conference on Intelligent Computing, Automation and Applications (ICA)*, pp. 275–278.
- Madani, A., Sibous, L., Hellal, A., et al., 2021. Synthesis, density functional theory study, molecular dynamics simulation and anti-corrosion performance of two benzidine Schiff bases. *J. Mol. Struct.* 1235, 130224.
- Matsuzaka, Y., Uesawa, Y., 2019. Prediction model with high-performance constitutive androstanone receptor (car) using deepsnap-deep learning approach from the Tox21 10K compound library. *Int. J. Mol. Sci.* 20, 4855.
- Mayr, A., Klambauer, G., Unterthiner, T., et al., 2016. DeepTox: toxicity prediction using deep learning. *Front. Environ. Sci.* <https://doi.org/10.3389/fenvs.2015.00080>. Epub ahead of print.
- Patel, A., 2020. *Silico Novel Identification of Anti-cancer Drugs Using Density Functional Theory and Molecular Dynamics Simulation* (Doctoral dissertation). Maharaja Sayajirao University of Baroda, India.
- Platzer, M., Kiese, S., Tybussek, T., et al., 2022. Radical scavenging mechanisms of phenolic compounds: a quantitative structure–property relationship (QSPR) study. *Front. Nutr.* 9, 882458.
- Rochlani, S., Bhatia, M., Rathod, S., et al., 2023. Exploration of limonoids for their broad spectrum antiviral potential via DFT, molecular docking and molecular dynamics simulation approach. *Nat. Prod. Res.* 1–6.

- Rouvray, D.H., 1971. Graph theory in chemistry. Royal Inst. Chem., Rev. 4, 173–195.
- Saini, V., 2023. Machine learning prediction of empirical polarity using SMILES encoding of organic solvents. Mol. Divers. 27, 2331–2343.
- Sait, A.R.W., Ishak, M.K., 2023. Deep learning with natural language processing enabled sentimental analysis on sarcasm classification. Comp. Syst. Sci. Eng. 44, 2553–2567.
- Sandler, M., Howard, A., Zhu, M., et al., 2018. Sandler_MobileNetV2_Inverted_Residuals_CVPR_2018_paper.pdf. ArXiv. 4510–4520.
- Sar, S.K., Mandavi, R., Rathore, N., et al., 2008. Effect of PEG-4000 & PEG-8000 on the reactivity of hydroxamate ion. J. Dispers. Sci. Technol. <https://doi.org/10.1080/01932690701866468>. Epub ahead of print.
- Sharma, G., Shukla, R., Singh, T.R., 2022. Identification of small molecules against the NMDAR: an insight from virtual screening, density functional theory, free energy landscape and molecular dynamics simulation-based findings. Net. Model. Anal. Health Inform. Bioinform. 11, 31.
- Sib Tul Hassan Shah, S., Naeem, I., 2023. In-silico targeting TMPK from monkey pox virus: molecular docking analysis, density functional theory studies and molecular dynamic simulation analysis. J. Biomol. Struct. Dyn. 1–13.
- Simonyan, K., Zisserman, A., 2015a. Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings, pp. 1–14.
- Simonyan, K., Zisserman, A., 2015b. Very deep convolutional networks for large-scale image recognition. In: 3rd International Conference on Learning Representations (ICLR 2015). Computational and Biological Learning Society, pp. 1–14.
- Sinha, K., Ghosh, N., Sil, P.C., 2023. A review on the recent applications of deep learning in predictive drug toxicological studies. Chem. Res. Toxicol. <https://doi.org/10.1021/acs.chemrestox.2c00375>. Epub ahead of print.
- Smedley J., Dick F., Sadhra S. Principles of toxicology. Epub ahead of print 2013. 10.1093/med/9780199651627.003.0033.
- Speiser, J.L., Miller, M.E., Tooze, J., et al., 2019. A comparison of random forest variable selection methods for classification prediction modeling. Expert. Syst. Appl. 134, 93–101.
- Suha, S.A., Sanam, T.F., 2022. A machine learning approach for predicting patient's length of hospital stay with random forest regression. In: 2022 IEEE Region 10 Symposium, TENSYMP 2022, pp. 1–6.
- Sun, C., Khalid, A., Usman, H.M., et al., 2022. On neighborhood degree-based topological analysis of polyphenylene network. Math. Probl. Eng. <https://doi.org/10.1155/2022/1951226>. Epub ahead of print.
- Szegedy, C., Vanhoucke, V., Ioffe, S., et al., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2016, pp. 2818–2826. Decem.
- Wang, X., Zhu, X., Wu, X., et al., 2018. Image encryption algorithm based on multiple mixed hash functions and cyclic shift. Opt. Lasers. Eng. 107, 370–379.
- Wang, Y.-L., Wang, F., Shi, X.-X., et al., 2021. Cloud 3D-QSAR: a web tool for the development of quantitative structure–activity relationship models in drug discovery. Brief. Bioinform. 22, bbaa276.
- Weininger, D., Weininger, A., Weininger, J.L., 1989. SMILES. 2. algorithm for generation of unique SMILES notation. J. Chem. Inf. Comput. Sci. 29, 97–101.
- Wu, Z., Ramsundar, B., Feinberg, E.N., et al., 2018. MoleculeNet: a benchmark for molecular machine learning. Chem. sci. 9 (2), 513–530.
- Xia, C., Fu, L., Zuo-yi, L., et al., 2018. Aquatic toxic analysis by monitoring fish behavior using computer vision: a recent progress. J. Toxicol. <https://doi.org/10.1155/2018/2591924>. Epub ahead of print.
- Xing, W., Bei, Y., 2020. Medical health big data classification based on KNN classification algorithm. IEE Access. 8, 28808–28819.
- Yang, K., Swanson, K., Jin, W., et al., 2019. Analyzing learned molecular representations for property prediction. J. Chem. Inf. Model. 59, 3370–3388.
- Yang Y., Yao K., Repasky M.P., et al. Efficient exploration of chemical space with docking and deep-learning. Epub ahead of print 2021. 10.26434/chemrxiv.14153819.v1.
- Zhang, J., Yu, X., Lei, X., et al., 2022. A novel CapsNet neural network based on MobileNetV2 structure for robot image classification. Front. Neurorobot. 16 <https://doi.org/10.3389/fnbot.2022.1007939>. Epub ahead of print.