Handwritten Cyclic Compound Classification with EfficientNetB0 and Explainable AI Methods

Md. Ismiel Hossen Abir*[1], Sanjana Islam Kasfia[2] and Abir Mahmud Shahariar[3]

[1, 2, 3] Department of Computer Science & Engineering,
International Standard University, Dhaka, Bangladesh
{ismielabir286, sanjanaislamkasfia2648,
abirmahmudshahariar}@gmail.com

Abstract. Handwritten chemical structures are commonly used in chemistry education and research, but recognizing these drawings automatically remains a challenging task. Variations in writing styles, line thickness, and incomplete patterns make accurate classification difficult for computer systems. This study presents a deep learning-based framework for the automatic identification of handwritten cyclic compounds. We created a dataset of handwritten chemical compounds that contains 599 images. Then, data augmentation was applied, resulting in a total of 15945 images across 15 chemical compounds in this study. Five pre-trained Convolutional Neural Network (CNN) architectures, such as VGG16, VGG19, ResNet101, MobileNetV2, and EfficientNetB0, were fine-tuned and used in this research. Among them, EfficientNetB0 achieved the best performance with an accuracy of 97.87%, showing strong generalization and minimal overfitting. To enhance interpretability, Explainable AI (XAI) techniques such as SHAP, Grad-CAM, and Saliency Maps were applied. Each of the XAI highlights the molecular regions that most influenced the model's decisions. The visualization results confirmed that the model focused on key chemical structures such as ring systems and functional groups. Overall, this research provides an interpretable and accurate approach for recognizing handwritten cyclic compounds.

Keywords: Handwritten Chemical Structure Recognition, Explainable Artificial Intelligence, Grad-CAM, SHAP, Cyclic Compound.

1 Introduction

Cyclic compounds are known as molecules where atoms connect to form closed rings [1]. This chain usually constructed from atoms of carbon whereas sometimes it contains other elements such as nitrogen and oxygen [1]. These chemical compounds have an important role in organic chemistry and can be found in a wide range of significant natural and manufactured substances, such as sugars, aromatic hydrocarbons, and numerous medications [2]. Chemical molecular structure has a regular place in educational interactions for instructions regarding chemistry [3]. It is crucial for chemical research and education to learn about their ring structures as they have a major influence on both reactivity and chemical properties [4]. Students and researchers usually sketch cyclic compounds by hand in labs and classrooms to illustrate molecular structures and analyze reactions [2]. Although the availability of automated

M. S. Arefin et al. (eds.), *Proceedings of the International Conference on Intelligent Data Analysis and Applications (IDAA 2025)*, Advances in Intelligent Systems Research 206,
https://doi.org/10.2991/978-94-6239-664-7_48

tools, hand drawn diagrams remain widely used due to their quick creation and ease of usage [3]. However, handwritten cyclic structures are difficult for technological devices to understand since they vary so much in both appearance and quality and it requires time along with domain expertise [2]. Understanding the specific cyclic compounds is difficult for non-chemistry students and beginners. Most of the time, students face difficulties when searching because of their complex names and patterns. Proper identification of these patterns is crucial for digital lab notes, automated student work grading and accelerating chemical research. Because of these factors, research into the automatic identification and decoding of handwritten cyclic compounds has grown rapidly in an effort to further improve scientific procedures and education.

In this study, we aimed to analyze the performance of numerous pre-trained Convolutional Neural Network (CNN) architectures for the classification of hand writing cyclic compounds. We created a dataset having 15 different classes of cyclic compounds which are 1,2-dimethylbenzene, Aniline, Anthracene, Benzene, Cyclobutane, Cyclohexane, Cyclohexanone, Cyclohexene, Ethylcyclopentane, Naphthalene, Phenol, Purine, Pyrimidine, Thymine, and Toluene [9]. Several pre-trained CNN models such as VGG16, VGG19, ResNet101, MobileNetV2, and EfficientNetB0 were used in order to preserve an impartial comparison. To ensure assessment reliability, all models were trained and evaluated using the same dataset and same fine-tuned function. While compared with other CNN architectures, EfficientNetB0 presented better predictions. We applied Explainable AI techniques specifically SHAP (SHapley Additive exPlanations), Grad-CAM (Gradient-weighted Class Activation Mapping), and Saliency Maps in order to ensure the clarity of EfficientNetB0 predictions. The crucial areas of an image that have the biggest impact on the model's predictions have been identified through Explainable AI techniques.

There has been some previous research on handwritten chemical compounds. Wang et al. implemented a component detection method for predicting offline handwritten chemical cyclic compound patterns [5]. By using SSD together with proposed NMAS (Non-Maximum Area Suppression) algorithm to identify text and graphical objects and then analyze their spatial connections their proposed method achieved an accuracy of 89.5% [5]. Keyrouz et al. developed a chemical structure recognition system in order to identify hand drawn chemical diagrams, which incorporates two trainable Bayesian networks with morphological image processing, LDA and PCA to manage noisy sketches and provide real time assistance [6]. Zheng et al. presented a CNN classifier based on VGGNet for offline chemical organic ring structure (ORS) symbol identification [7]. They applied data augmentation to analyze the rotational invariance of CNNs and acquired 92.4% accuracy with VGGNet-19 [7]. Weir et al. was invented an offline tool named ChemPix that maps hand drawn hydrocarbon structure images to SMILES representations using an LSTM decoder and CNN encoder [2]. Keyrouz et al. proposed a method for handling inconsistent sketches that combines two Bayesian based image recognizers having both feature and image related methods [4]. This technique performs faster and more accurately than neural networks and provides real time feedback and candidate suggestions [4]. Musazade et al. reviewed the development of chemical formula extraction from images which demon-

strates the transition from rule centered techniques to natural language processing methods because of the intricate and unpredictable nature of molecular representations [8]. Ouyang et al. was developed ChemReco to identify hand drawn chemical molecule structures of C,H, and O by using synthetic image formation and a hybrid EfficientNet + Transformer encoder decoder framework [3].

The motivation for this study comes forward from the limitations identified in recent research documents on handwritten chemical structure recognition. Although a number of previous studies have shown that machine learning and deep learning approaches can be used to identify chemical structures, they often suffer from serious drawbacks. Some methods focus on heuristic approaches, which limit their reliability as they are unable to manage the inherent noise and variability found in real world handwritten images. Since Explainable AI techniques were not incorporated into their frameworks, some researchers gave higher priority on increasing identification accuracy while paying less attention to the model's clarity. Moreover, a number of studies only concentrate on identifying certain components, for example hydrocarbon skeletons or basic ring patterns, and do not provide a complete assessment of various novel deep learning models in a single experimental environment. These limitations indicate the necessity of a study which includes the use of more diverse dataset of handwritten cyclic compounds, analyzes different modern CNN architectures to determine which method performs very well, and uses Explainable AI techniques to improve reliability and clarity. This research makes three main contributions:

(i) Created a novel handwritten cyclic compound dataset that consists of a wide range of augmented molecular compound images.
(ii) Presents a comparative analysis of five pre-trained CNN architectures which are VGG16, VGG19, ResNet101, MobileNetV2, and EfficientNetB0 on the same dataset and fine-tuned parameters to ensure consistent assessment.
(iii) Incorporates Explainable AI techniques, particularly SHAP, Grad-CAM, and Saliency Map to demonstrate and clarify model's prediction.

The remaining part of our study is organized into different sections, where Section 2 describes the methodologies, Section 3 demonstrates the experimental findings, Section 4 addresses the comparative analysis, and Section 5 concludes the study with future work directions, followed by the references.

2 Methodologies

In this study, various deep learning models were evaluated to classify handwritten cyclic compounds into 15 distinct categories such as 1,2-dimethylbenzene, Aniline, Anthracene, Benzene, Cyclobutane, Cyclohexane, Cyclohexanone, Cyclohexene, Ethylcyclopentane, Naphthalene, Phenol, Purine, Pyrimidine, Thymine, and Toluene. The working procedure is shown in the Figure 1.

2.1 Dataset Creation and Pre-processing

We have created this dataset consist of 599 handwritten images of 15 different cyclic compounds. In the dataset creation process used handwritten cyclic compounds where several students participants to write the compounds. After the writing, we collected each of the handwritten images by using camera and then used scanner for quality update each of the image. After that, we augmented the primary dataset and build another dataset contains total 15945 images. In the data augmentation technique used rotation, width shift, height shift, shear, zoom, horizontal and vertical flip. Then, the augmented dataset was used and split into 80% for training, 10% for testing, and 10% for validation. Figure 2 presents the sample images of this dataset.

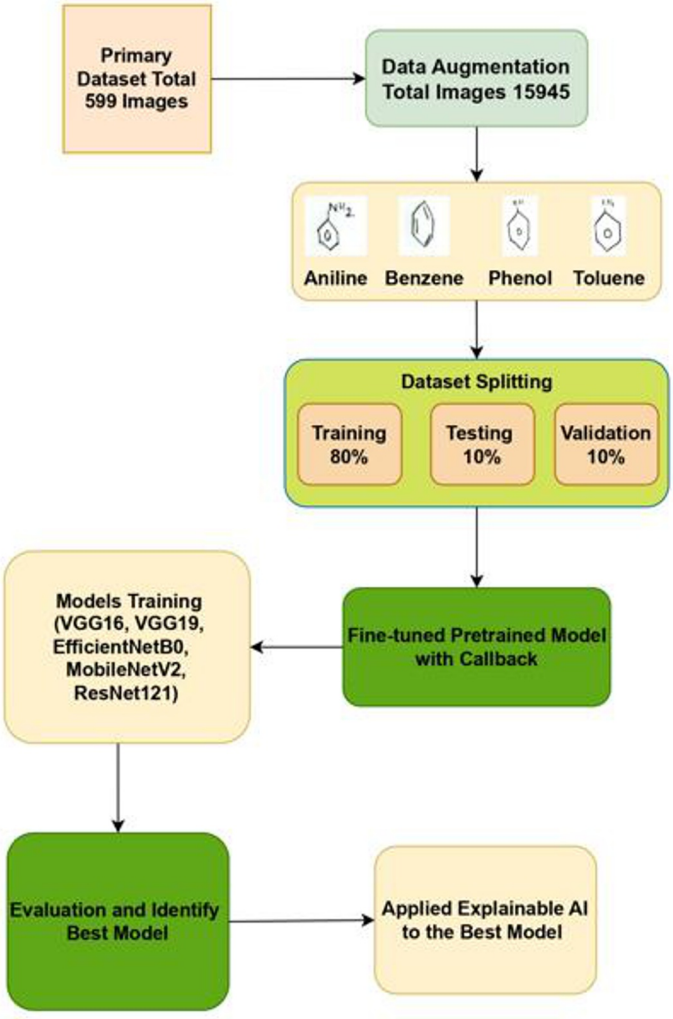

Fig. 1. Working Procedure

Total 12756 images were used for training purposes. The images were organized into a 50×50 pixel resolution with RGB channels. The batch size used 32. Model experiment environment was Kaggle notebook using GPU P100. Also used several python library such as Tensorflow, matplotlib, seaborn, scikit-learn, etc.

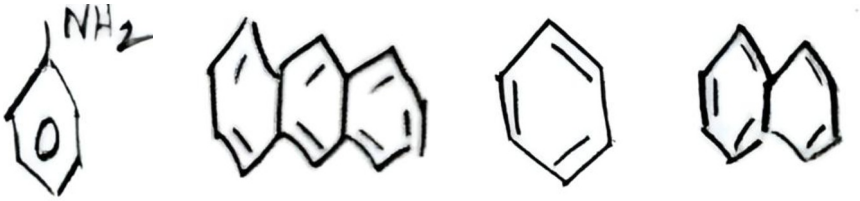

Fig. 2. Sample Images

2.2 Transfer Learning Models

Five Transfer Learning Models were fine-tuned for this classification task. These models are the VGG16, VGG19, ResNet101, MobileNetV2, and EfficientNetB0. The VGG16 model using small 3x3 filters which increase the depth of convolutional network where used 16-19 depth weight layers [10]. With all that VGG19 is a more profound variant of VGG16. In the ResNet101 used residual learning framework with very deep network [11]. This keeps the network's performance solid while providing a more extensive network. The MobileNetV2 model is designed for efficiency. It uses depth wise separable convolutions to reduce computation [12]. This makes it fast for mobile and embedded devices. The EfficientNetB0 model is also very efficient. It balances the depth, width, and input resolution of the network [13]. This balanced scaling leads to great performance while using fewer resources. Its success can be attributed to compound scaling, optimized network depth, and better feature reuse across layers.

In this study, all five pre-trained models were utilized and fine-tuned. We build a function that unfreezes the last few layers and adds new fully connected layers for classification. We applied Global Average Pooling, Batch Normalization, Dense, and Dropout layers to improve performance and reduce overfitting. This allows it to accurately capture complex patterns in handwritten cyclic compound structures. To ensure the training reliability of all models, the Adam optimizer was used with categorical cross-entropy loss. The models were trained for 30 epochs and included callbacks such as EarlyStopping and ReduceLROnPlateau. After that, each of the models was trained separately based on the fine-tuned function. Among all the models EfficientNetB0 gave better performance, and minimized overfitting. Therefore, for further analysis and explainable AI perspective, we used EfficientNetB0 model.

2.3 Explainable AI and Evaluation Metrics

The aim of our research is to increase model transparency and clarity. We applied three different Explainable AI techniques to accomplish this such as saliency map, Grad-CAM, and SHAP.

To interpret the model's predictions, we generated image-specific class saliency maps by computing the gradient [14]. Saliency maps highlight the image regions influential for the network's classification decision [14]. Grad-CAM (Gradient-weighted Class Activation Mapping) provides class discriminative heatmap feature [15]. This feature indicates the large complex areas that the model utilized to make a particular class decision. SHAP (SHapley Additive exPlanations) provides interpreting model predictions [16]. It assigns a distinct, mathematically derived value to each input feature.

There is several number of evaluation methods were used such as accuracy score, precision, recall, f1-score, confusion matrix, and explainable AI.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

$$F1\ Score = 2 * \frac{Precision\ x\ Recall}{Precision+Recall} \quad (4)$$

Here, TP means true positive, TN means true negative, FP means false positive, and FN means false negative.

These evaluation metrics provided a quantitative understanding of model performance. On the other hand, the XAI visualizations are qualitative insights into model decision-making. The combination of both quantitative and qualitative evaluation ensured that the model not only achieved high accuracy but also reliable and interpretable predictions.

3 Experimental Results

The experimental results section has two parts: analysis of the five fine-tuned pretrained models, and the use of explainable AI techniques on the best model.

3.1 Model Performance Evaluation

According to the table 1 shows classification report of EfficientNetB0 model. It clear showed that this model correctly predicts the fifteen cyclic compound classes. The model had an overall accuracy of 97.87%. Most classes, such as Ethylcyclopentane, Purine, Pyrimidine, and Thymine, achieved perfect precision, recall, and F1-scores. But some classes such as Phenol and Toluene shows slightly decreased performance compared to others classes. These results indicate that the EfficientNetB0 model effectively captures the structural patterns of different chemical compounds.

Table 1. Classification Report of EfficientNetB0 Model

Class Name	Precision	Recall	F1-Score
1,2-dimethylbenzene	0.97	0.94	0.95
Aniline	0.98	0.95	0.96
Anthracene	0.99	0.99	0.99
Benzene	1.00	0.99	0.99
Cyclobutane	1.00	0.99	0.99
Cyclohexane	0.93	0.97	0.95
Cyclohexanone	1.00	0.96	0.98
Cyclohexene	0.97	0.98	0.97
Ethylcyclopentane	1.00	1.00	1.00
Naphthalene	0.99	0.99	0.99
Phenol	0.89	0.97	0.93
Purine	1.00	1.00	1.00
Pyrimidine	1.00	1.00	1.00
Thymine	1.00	1.00	1.00
Toluene	0.95	0.93	0.94
Accuracy			0.97

The figure 3 shows the EfficientNetB0 model confusion matrix. It shows that most of the predictions lie along the diagonal, indicating that the model correctly classified the majority of samples for each cyclic compound class. But there is also a few misclassification 1,2-dimethylbenzene, Aniline, Cyclohexane, and Cyclohexanone classes.

The other four models showed similar performance, but VGG16 showed overfitting. Although VGG19 achieved a higher accuracy of 98.56% compared to EfficientNetB0, it also showed signs of overfitting. ResNet101 achieved an accuracy of 94.17%, and MobileNetV2 achieved 92.03%.

3.2 Explainable AI and EfficientNetB0

Figure 4 shows the SHAP visualization where original image from Aniline class which was output 1. The SHAP values highlight which image regions contributed most to the model's prediction across different output classes. Here, red regions indicate the positive influence and blue regions indicate the negative influence on the class decision. For Output 1, Aniline class, red regions indicate pixels that positively influenced the prediction, while blue areas show negative influence. This suggests that the model correctly identified the molecule as Aniline, with proper attention to its key structural components. The weak activations in other outputs indicate low confusion with other classes.

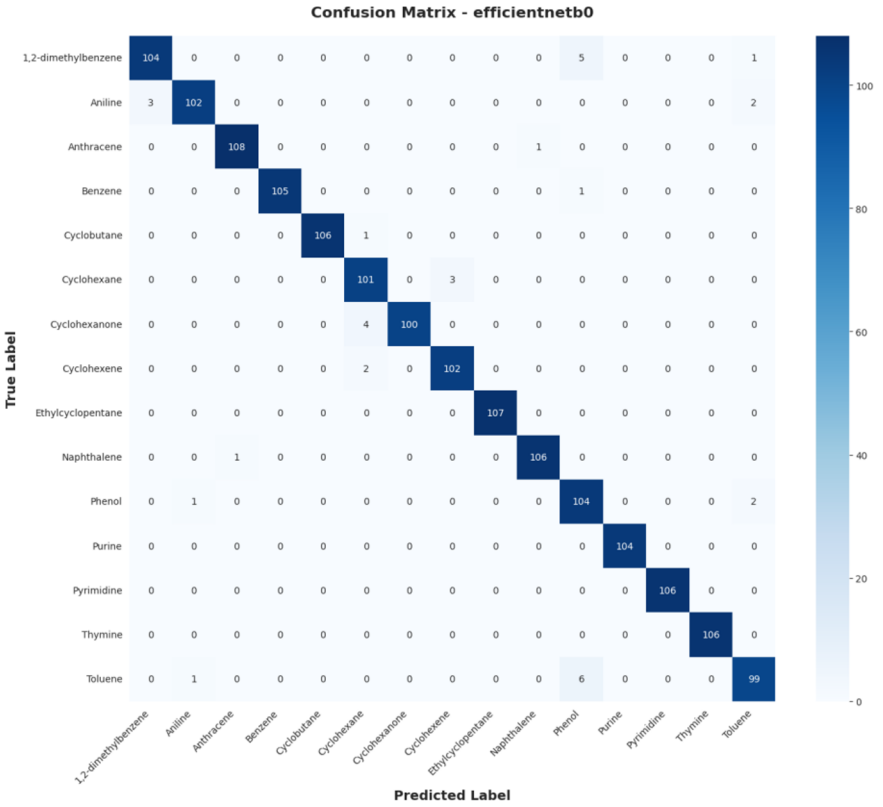

Fig. 3. Confusion Matrix

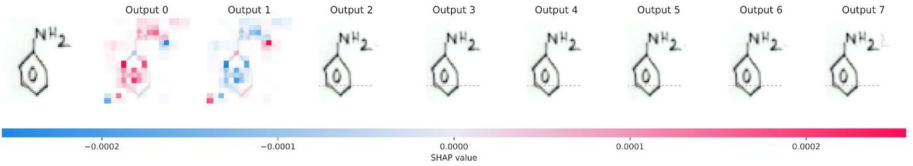

Fig. 4. SHAP for EfficientNetB0 Model

Figures 5 and 6 present the Grad-CAM visualizations. For the Purine molecule, the activation heatmap highlights the core ring structure, indicating that the model correctly identified the molecular framework responsible for classification. In the case of Toluene, the highlighted area covers the benzene ring and side chain, which are key structural features of the compound. So, the Grad-CAM confirms that the model's visual attention aligns with the true chemical structures.

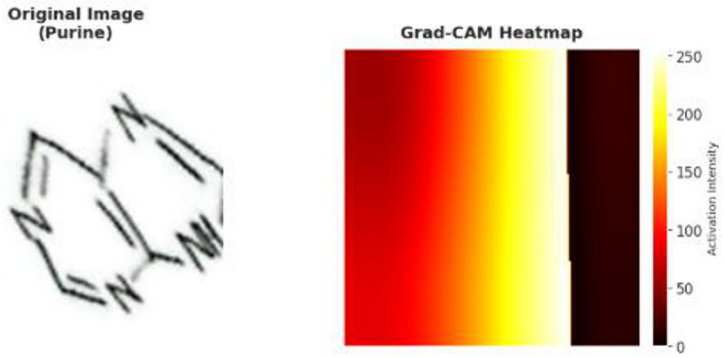

Fig. 5. Grad-CAM for EfficientNetB0 Model

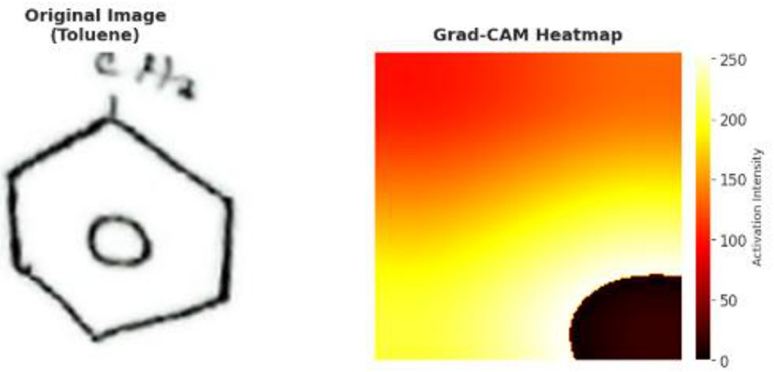

Fig. 6. Grad-CAM for EfficientNetB0 Model

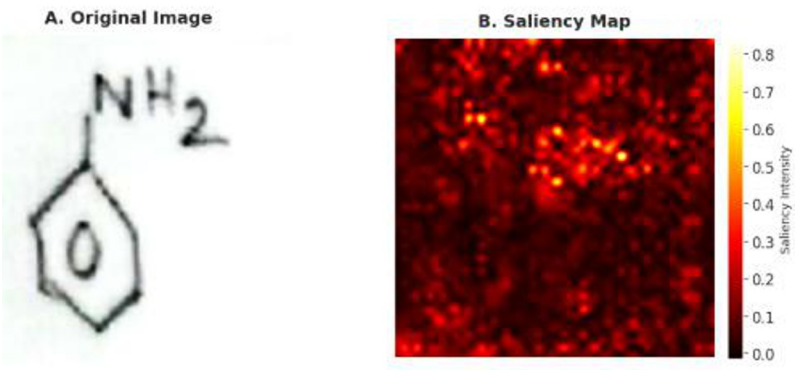

Fig. 7. Saliency Map for EfficientNetB0 Model

Figure 7 reflects the saliency map where it highlights the important regions of the model prediction for the specific class. The first image displays a hand-drawn chemi-

cal structure. This saliency map highlights the chemical bonds and the NH2 group as the most visually significant areas for analysis.

4 Comparative Analysis

The performance of several pre-trained CNN models for handwritten cyclic compound classification is compared in this section of our study. As demonstrated in the Table 2, MobileNetV2 possessed the lowest accuracy of 92.03% which means that it had insufficient capacity to identify complex patterns in the dataset. ResNet101 performed slightly better having and accuracy of 94.17% that implies more complex networks can improve learning. VGG16 performed well with an accuracy of 97.37%. Despite having the best accuracy, VGG16 and VGG19 went through overfitting during training, that reduces its reliability for real world applications. EfficientNetB0 maintained an ideal balance between accuracy and generalization with 97.87% accuracy. Therefore, we consider EfficientNetB0 as an ideal model for this research.

Table 2. Performance Comparison of Different Transfer Learning Models

Models	Accuracy Score (%)
MobileNetV2	92.03
ResNet101	94.17
VGG16	97.37
VGG19	98.56
EfficientNetB0	97.87

Table 3. Comparative Analysis

References	Methodology	Accuracy Score (%)	Explainable AI	Dataset
[2]	CNN (Encoder) + LSTM (Decoder) + Ensemble Learning (SMILES representation)	76%	No	~600 hand drawn (crowdsourced) +Synthetic (RDKit) images
[3]	Synthetic Dataset Generation + EfficientNet (Encoder) + Transformer (Decoder)	96.90%	No	~2598 Synthetic + Custom dataset of C–H–O structures
[5]	SSD + NMAS Algorithm	89.5%	No	~2100 images + Custom dataset
[7]	VGG19 + Rotation Invariant Data Augmentation	92.4%	No	~3600 images + Custom dataset
This Study	Fine-tuned EfficientNetB0	97.87%	Yes (SHAP, Grad-CAM, Saliency Map)	15945 images + Custom dataset

According to the comparative study shown in Table 3, majority of earlier studies had decent accuracy but lacked explainable AI approaches. Though they performed well, models such as CNN + LSTM and EfficientNet + Transformer suffered from clarity. The Deep Learning (SSD) + NMAS and VGG19 framework also showed acceptable outcomes but they were unable to clarify the results they predicted. The majority of the previous research applied small datasets, which reduced their ability to develop good predictions. However, our research applied a large custom dataset which contained 15,945 images of handwritten cyclic compounds. This helped the framework train better as well as function more accurately. Also, we incorporated Explainable AI techniques particularly SHAP, Grad-CAM, and Saliency Map to the EfficientNetB0 model. Because of this, our model is not only accurate but also transparent and reliable, which makes it easier for users to figure out its mechanism for identifies and categorizes cyclic molecules.

5 Conclusion

This study adopted a number of pre-trained deep learning models for classifying handwritten cyclic compounds. VGG16, VGG19, ResNet101, MobileNetV2, and EfficientNetB0 fine-tuned models were used in this study. The results demonstrated that EfficientNetB0 performed well than the other models with an accuracy of 97.87%. There was a decent balance between prediction and accuracy which was maintained properly by the EfficientNetB0. A large custom dataset having 15,945 images of handwritten cyclic compounds was used, which helped the models train well than usual. We applied Explainable AI techniques (SHAP, Grad-CAM, Saliency Map) on the EfficientNetB0. These methods helped to identify the most important areas in an image that influenced the model's prediction. The use of Explainable AI made the model more transparent and reliable. In future studies, a real time identification system can be developed based on the proposed model. This system can be used in classrooms and research lab to automatically identify and evaluate handwritten chemical structures.

References

1. S. Mondal, UNIT-I: Classification of Organic Compounds, Lecture Notes, Pharmaceutical Organic Chemistry-I (BP 202T), B.Pharm-II Sem., 2022. doi:10.13140/RG.2.2.19146.03525
2. Weir, H., Thompson, K., Woodward, A., Choi, B., Braun, A., Martínez, T.J.: ChemPix: Automated recognition of hand-drawn hydrocarbon structures using deep learning. Chemical Science 12(31), 10622–10633 (2021). https://doi.org/10.1039/D1SC02957F
3. Ouyang, H., et al.: ChemReco: Automated recognition of hand-drawn carbon–hydrogen–oxygen structures using deep learning. Scientific Reports 14, 17126 (2024). https://doi.org/10.1038/s41598-024-67496-7
4. Keyrouz, F., Tauk, L., Feghali, E.: Enhanced chemical structure recognition and prediction using Bayesian fusion. In: 2019 IEEE 9th Annual Computing and Communication Work-

shop and Conference (CCWC), pp. 611–616. IEEE, Las Vegas, USA (2019). https://doi.org/10.1109/CCWC.2019.8666586
5. Wang, Y., Zhang, T., Yu, X.: A component-detection-based approach for interpreting offline handwritten chemical cyclic compound structures. In: 2021 IEEE International Conference on Engineering, Technology & Education (TALE), pp. 785–791. IEEE, Wuhan, China (2021). https://doi.org/10.1109/TALE52509.2021.9678874
6. Keyrouz, F., Tauk, L., Feghali, E.: Chemical structure recognition and prediction: A machine learning technique. In: 2018 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB), pp. 1–7. IEEE, St. Louis, USA (2018). https://doi.org/10.1109/CIBCB.2018.8404964
7. Zheng, L., Zhang, T., Yu, X.: Recognition of handwritten chemical organic ring structure symbols using convolutional neural networks. In: 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW), pp. 165–168. IEEE, Sydney, Australia (2019). https://doi.org/10.1109/ICDARW.2019.40099
8. Musazade, F., Jamalova, N., Hasanov, J.: Review of techniques and models used in optical chemical structure recognition in images and scanned documents. Journal of Cheminformatics 14, 61 (2022). https://doi.org/10.1186/s13321-022-00642-3
9. Abir, M.I.H., Chomon, R.A.: Handwritten Cyclic Compounds Dataset. IEEE Dataport (2025). https://doi.org/10.21227/nvvq-jj53
10. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2015). https://doi.org/10.48550/arXiv.1409.1556
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. arXiv preprint arXiv:1512.03385 (2015). https://doi.org/10.48550/arXiv.1512.03385
12. Howard, A.G., et al.: MobileNets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017). https://doi.org/10.48550/arXiv.1704.04861
13. Tan, M., Le, Q.V.: EfficientNet: Rethinking model scaling for convolutional neural networks. arXiv preprint arXiv:1905.11946 (2020). https://arxiv.org/abs/1905.11946
14. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034 (2013).
15. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 618–626 (2017).
16. Lundberg, S.M., Lee, S.-I.: A unified approach to interpreting model predictions. In: Advances in Neural Information Processing Systems (NeurIPS) 30 (2017). https://arxiv.org/pdf/1705.07874