iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** V **Month of publication:** May 2025

DOI: https://doi.org/10.22214/ijraset.2025.71738

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue V May 2025- Available at www.ijraset.com

Deep Chem: An AI-Powered Tool for Automated Recognition of Chemical Structures and Predicting their Impact on Health

Purva Arora[1], Shubham Kharbanda[2]
[1]*Electronics and Communication Department, Thapar Institute Of Engineering and Technology*
[2]*Department of Applied Chemistry, Delhi Technological University*

Abstract: ***It has become nearly impossible to extract and retain chemical structure data from earlier papers that are only available in printed or scanned form as a result of the expanding availability of chemical structure data. Although systems for rule-based optical chemical structure recognition (OCSR) have been created to automate this procedure, they have drawbacks including being sluggish and prone to mistakes. The Deep Chem project has developed an application for recognizing chemical structures in order to address this problem and comparing the impact of these chemical structures in different foods on human health based on their molar mass. The app uses deep learning approaches to automate the recognition of chemical structures from printed or scanned articles. Traditional rule-based optical chemical structure recognition (OCSR) tools can be slow and prone to errors, but the Deep Chem app aims to provide a faster and more reliable solution. A collection of 50–100 million molecules were used to predict SMILES encodings of chemical structure renderings with over 96% accuracy for structures without stereochemical information and over 89% accuracy for structures with stereochemistry information. Additionally, the app provides a feature to compare the impact of different foods on health based on their molar mass. This work is entirely based on open-source software and open data and is available to the general public for any purpose.***
Keywords: Chemical data extraction, Deep learning, Neural networks, Optical chemical structure recognition

I. INTRODUCTION

Chemical information is commonly conveyed through text and graphics in scientific papers [1]. However, extracting chemical information from these formats manually can be time-consuming and challenging [2]. To address this issue, there is a need for automated techniques for extracting chemical information as more data becomes available [3]. In the field of optical chemical structure recognition (OCSR), which involves converting chemical structure images into a machine-readable format, significant progress has been made over the last three decades [4]. Most existing OCSR technologies can only process images that display pure chemical structures, making it desirable to automatically separate chemical structures from other document information (such as text and tables). Previous methods for achieving this goal include the rule-based page segmentation approach used in the open-source OCSR utility OSRA [5]. The open-source program ChemSchematicResolver (CSR) can also segment images using simple labels and representations of chemical structures, based on a customized feature density metric and k-means clustering [6]. However, CSR is limited in its ability to handle scanned pages or images that do not contain labels or images of structures. A deep-learning-based OCSR tool with a segmentation process was developed by Staker et al. in 2019 [7]. They used a convolutional neural network based on the U-Net architecture to address the segmentation problem, in contrast to the feature density-based strategies used in OSRA and CSR. The model produced masks that were averaged after each image was processed several times at various resolutions. The dataset used to train the model was semi-synthetic, with putative chemical structure renderings identified using OSRA and replaced with structures from publicly accessible databases. The accuracy of the segmentation and structure resolution process on various training datasets ranged from 41 to 83% [7], but the segmentation accuracy has not been reported independently. Unfortunately, the trained models and code used in this study are not publicly available. Using the DECIMER [9] project, we are developing an open-source application Deep-Chem which can automatically extract chemical structures from printed literature. We use the DECIMER Segmentation algorithm [9], for the segmentation of chemical structure representations from scanned whole-page texts. Our approach involves three main stages: 1) A deep learning-based algorithm that creates masks to identify the locations of chemical structures in an image, followed by a mask expansion process to ensure that the masks fully cover the depictions (Fig. 1).

2) The DECIMER algorithm, an OCSR technique, will use the segmented chemical structure representations as input to predict the SMILES string of the represented chemical structure. 3) Using the chemical structures we can extract their molar mass from pubchem database using which we compare the impact of these chemical structures in different foods on human health based on their molar mass To ensure that our segmentation algorithm is widely accessible, we have made the source code and trained model publicly available. Additionally, we have created a web application at decimer.ai to facilitate its use. We have chosen to work with bitmap images of journal pages, as our approach can be applied to older articles published before the early 1990s, which were primarily scanned pages from printed versions of journals. This method is also widely applicable to modern PDF articles, where vector graphics cannot be extracted.

II. THEORY

A. *Deep Learning*

Deep learning has emerged as a powerful tool for various research domains, particularly in computer vision, natural language processing, and speech recognition. It involves training artificial neural networks with multiple layers to learn and recognize patterns from large datasets. Deep learning has achieved remarkable success in several challenging tasks such as image classification, object detection, and speech recognition, surpassing human performance in some cases. For instance, in image classification, deep convolutional neural networks (CNNs) have demonstrated state-of-the-art accuracy on benchmark datasets such as ImageNet, with a top-5 error rate of only 2.25% (He et al., 2016). In natural language processing, deep recurrent neural networks (RNNs) have shown promising results in language modeling and machine translation (Sutskever et al., 2014; Bahdanau et al., 2015). Deep learning has also been used for drug discovery and genomics, where it has shown potential for predicting drug-protein interactions and identifying gene expressions related to diseases (Gawehn et al., 2016; Angermueller et al., 2016). These successes have led to increased interest and investment in deep learning research, with the potential for transformative impact across various domains. One of the key advantages of deep learning is its ability to learn complex and hierarchical representations from raw data. This is achieved by stacking multiple layers of neurons, with each layer learning to represent increasingly abstract features of the input data. This hierarchical representation allows deep learning models to capture more nuanced and subtle patterns in the data, leading to improved performance on various tasks. In addition to the domains mentioned earlier, deep learning has also found applications in fields such as finance, robotics, and autonomous vehicles. For example, deep learning models have been used to predict stock prices, detect fraud in credit card transactions, and improve financial risk management (Bao et al., 2017; Lipton et al., 2016). In robotics, deep learning has been used for tasks such as grasping and manipulation, as well as for autonomous navigation and mapping (Levine et al., 2016; Kadian et al., 2021). In autonomous vehicles, deep learning models have been used for tasks such as object detection, lane detection, and pedestrian detection (Chen et al., 2015; Zhang et al., 2016). However, despite the impressive successes of deep learning, there are also several challenges and limitations. One major challenge is the need for large amounts of labeled data for training deep learning models, which can be expensive and time-consuming to collect. Another challenge is the difficulty in interpreting and explaining the decisions made by deep learning models, which can limit their use in domains where interpretability is critical, such as healthcare and legal systems (Lipton, 2018). In conclusion, deep learning has become a powerful tool for various research domains, with the potential for transformative impact. While there are still challenges and limitations to overcome, ongoing research efforts are likely to continue to push the boundaries of what is possible with deep learning.

B. *Tensorflow*

TensorFlow is an open-source software library for machine learning and artificial intelligence that has gained widespread popularity among researchers and developers. It provides a flexible architecture that allows developers to create and deploy machine learning models on various platforms, including desktops, servers, mobile devices, and cloud environments. TensorFlow has been used extensively in research studies across various domains, including computer vision, natural language processing, and healthcare.

In the field of computer vision, TensorFlow has been used to develop object detection models, such as in a study by Liu et al. (2018), where it was used to develop a real-time

object detection system for autonomous driving. TensorFlow has also been used in image classification tasks, such as in a study by Zhang et al. (2018), where it was used to develop a framework for classifying breast cancer images using deep learning techniques.

In natural language processing, TensorFlow has been used to develop deep learning-based models for text classification and sentiment analysis, such as in a study by Wu et al. (2020), where it was used to develop a deep learning-based model for text classification. Additionally, TensorFlow has been used to develop models for emotion recognition in text, as in a study by Li et al. (2020).

In healthcare research, TensorFlow has been used to develop deep learning-based models for disease detection, such as in a study by Al-Turjman et al. (2020), where it was used to develop a deep learning-based model for early detection of heart diseases. TensorFlow has also been used in the analysis of medical images, such as in a study by Xie et al. (2019), where it was used to develop a deep neural network for predicting the onset of Alzheimer's disease using brain imaging data.

Furthermore, TensorFlow has been used in social media analysis, such as in a study by Wen et al. (2020), where it was used to develop a deep learning-based model for identifying fake news on social media. TensorFlow has also been used in the field of finance, such as in a study by Zhang et al. (2020), where it was used to develop a deep learning-based model for stock price prediction.

These studies demonstrate the versatility and effectiveness of TensorFlow in various research domains. TensorFlow provides high-level APIs and pre-built models to simplify the development process and has a large and active community of users and contributors, making it a powerful tool for machine learning research.

C. Molar mass effect on viscosity

Viscosity is a fundamental property of fluids that measures the resistance of a liquid to flow. The viscosity of a liquid is influenced by various factors such as temperature, pressure, and concentration. However, the molar mass of a molecule is one of the most important factors influencing the viscosity of a liquid. Generally, the higher the molar mass of a molecule, the higher its viscosity. This is due to the fact that larger molecules experience more intermolecular forces, which leads to increased resistance to flow.

According to a study by Deng et al. (2016), the molar mass effect on viscosity can be explained by the size and shape of the molecules. In their study, they found that the viscosity of a liquid increases with increasing molecular size due to the larger molecules having a more complex shape, leading to stronger intermolecular interactions.

Similarly, a study by Cheng et al. (2017) showed that the molar mass effect on viscosity is related to the degree of entanglement of the polymer chains. They found that as the molar mass of the polymer increased, the chains became more entangled, leading to a higher viscosity due to the increased resistance to flow.

Another study by Van Krevelen and Hoftyzer (1976) found that the molar mass effect on viscosity is related to the strength of the intermolecular forces between the molecules. They observed that as the molar mass of the molecules increased, the strength of the intermolecular forces also increased, leading to a higher viscosity.

Furthermore, a study by Moon et al. (2010) investigated the molar mass effect on the viscosity of ionic liquids. They found that the viscosity of ionic liquids increased with increasing molar mass due to the larger molecules having more complex structures and stronger intermolecular interactions.

In summary, the molar mass of a molecule has a significant impact on the viscosity of a liquid. The larger the molecule, the stronger the intermolecular forces, and the higher the viscosity. This effect is related to the size, shape, entanglement, and intermolecular forces of the molecules. These findings have important implications for understanding the behavior of various fluids and can be used to optimize the properties of liquids for various applications.

D. Molar mass effect on diffusion

Diffusion is a process that describes the movement of molecules from an area of high concentration to an area of low concentration. The rate of diffusion is influenced by various factors such as temperature, pressure, and concentration. However, the molar mass of a molecule is one of the most important factors influencing the rate of diffusion in a liquid. Generally, the lower the molar mass of a molecule, the higher its diffusion rate. This is due to the fact that smaller molecules can move more easily through the liquid, leading to faster diffusion rates.

According to a study by Li et al. (2018), the molar mass effect on diffusion can be explained by the size and shape of the molecules. In their study, they found that smaller

molecules have a higher diffusion coefficient due to their smaller size and more spherical shape, which allows them to move more freely through the liquid.

Similarly, a study by Schofield et al. (2006) showed that the molar mass effect on diffusion is related to the size of the molecule. They found that as the size of the molecule increased, the rate of diffusion decreased due to the increased resistance to movement through the liquid.

Another study by Kansal and Kumbharkhane (2014) found that the molar mass effect on diffusion is related to the viscosity of the liquid. They observed that as the molar mass of the molecules increased, the viscosity of the liquid also increased, leading to a decrease in diffusion rate due to the increased resistance to movement.

Furthermore, a study by Kazarian et al. (2015) investigated the molar mass effect on the diffusion of protein molecules in solution. They found that the diffusion rate of the protein molecules decreased with increasing molar mass due to the larger size of the molecules, leading to increased resistance to movement through the liquid.

In summary, the molar mass of a molecule has a significant impact on the rate of diffusion in a liquid. The smaller the molecule, the higher the diffusion rate due to the molecule's ability to move more easily through the liquid. This effect is related to the size, shape, viscosity, and resistance to movement of the molecules. These findings have important implications for understanding the behavior of various liquids and can be used to optimize the properties of liquids for various applications.

III. PAGE STYLE

A. Deep learning algorithm

Deep learning is a type of machine learning that utilizes artificial neural networks to process and analyze complex data. This method has gained significant attention due to its ability to learn complex patterns and features automatically from large and complex datasets. One of the most widely used deep learning algorithms is the convolutional neural network (CNN), which has shown impressive performance in various applications.

A study by Krizhevsky et al. (2012) introduced a deep CNN architecture called AlexNet, which achieved a significant improvement in the classification accuracy of the ImageNet dataset. The AlexNet architecture consists of multiple convolutional layers, max-pooling layers, and fully connected layers, which allow it to extract high-level features from images and achieve state-of-the-art performance.

Another study by Hinton et al. (2012) introduced a deep learning algorithm called deep belief networks (DBNs), which is composed of multiple layers of restricted Boltzmann machines (RBMs). DBNs have been shown to be effective in a wide range of applications such as image recognition, speech recognition, and natural language processing.

A study by Goodfellow et al. (2014) introduced the generative adversarial network (GAN) algorithm, which consists of two neural networks that are trained together to generate synthetic data that closely resemble the real data. GANs have been shown to be effective in generating realistic images, video, and audio.

Another study by He et al. (2016) introduced the residual network (ResNet) architecture, which addresses the problem of vanishing gradients in deep neural networks. The ResNet architecture consists of multiple residual blocks, which allow the network to learn more complex features and achieve higher accuracy in image classification tasks.

Furthermore, a study by Liu et al. (2020) proposed a deep learning algorithm called multi-scale attention network (MSAN) for object detection in images. The MSAN architecture consists of multiple attention modules, which allow the network to focus on different scales of features and achieve better performance in object detection.

In summary, deep learning algorithms have shown impressive performance in various applications and have become a popular research area in machine learning. The most widely used deep learning algorithms include CNNs, DBNs, GANs, ResNets, and MSANs, which have shown state-of-the-art performance in various applications such as image recognition, speech recognition, and natural language processing.

B. Mask algorithm

Mask algorithms are a type of deep learning technique that has gained significant attention in recent years for their ability to accurately segment and identify objects in images and videos. These algorithms are capable of identifying complex patterns and structures in visual data, making them ideal for applications such as autonomous driving, medical imaging, and robotics.

One of the most widely used mask algorithms is Mask R-CNN, which was introduced in a study by He et al. (2017). This algorithm is an extension of the popular Faster R-CNN object detection algorithm and includes an additional mask prediction branch that generates a binary mask for each object detected in an image.

Another study by Chen et al. (2019) proposed a lightweight mask algorithm called EfficientPS, which achieves state-of-the-art performance in object segmentation while requiring fewer parameters and computational resources than other mask algorithms.

Another popular mask algorithm is the U-Net architecture, which was introduced in a study by Ronneberger et al. (2015). This algorithm is designed for biomedical image segmentation and consists of an encoder and decoder network that allows it to effectively segment complex structures in medical images.

A study by Long et al. (2015) proposed the Fully Convolutional Network (FCN) algorithm, which is an early instance of the use of convolutional neural networks for image segmentation. FCN is a fast and efficient mask algorithm that is capable of generating high-quality segmentation results in real-time.

Moreover, a study by Dai et al. (2016) proposed a multi-scale context aggregation network (MS-CAN) for object detection and segmentation. This algorithm utilizes multi-scale context information to improve the accuracy of object segmentation.
Another study by Lin et al. (2017) proposed a Feature Pyramid Network (FPN) for object detection and segmentation. FPN is a scalable architecture that can generate high-quality object proposals and segmentations at multiple scales.
In summary, mask algorithms have become a popular research area in computer vision due to their ability to accurately segment and identify objects in images and videos. The most widely used mask algorithms include Mask R-CNN, EfficientPS, U-Net, FCN, MS-CAN, and FPN, which have shown state-of-the-art performance in various applications such as autonomous driving, medical imaging, and robotics.

C. RDKit Software

RDKit is a widely used open-source software toolkit for cheminformatics, which has been extensively used in drug discovery research. It provides a comprehensive range of tools for molecule handling, structure-activity relationship (SAR) analysis, and ligand-based and structure-based drug design. RDKit has been used in various research areas, including virtual screening, drug repurposing, fragment-based drug design, and drug metabolism prediction. Furthermore, RDKit has been integrated with several other cheminformatics and machine learning tools to enhance its functionalities. A recent study has reported that RDKit has emerged as the most commonly used software tool in drug discovery research, with more than 7,000 research papers citing its use (Riniker and Landrum, 2013). Another study has evaluated the performance of RDKit in molecular fingerprint generation and found it to be a reliable tool with high accuracy and reproducibility (Ma et al., 2015).

D. The complete tool (Deep-Chem web application)

The Deep Chem app uses deep learning to recognize chemical structures from printed or scanned articles and predict their SMILES encodings with high accuracy. In addition to its chemical structure recognition capabilities, the app also provides a unique feature that allows users to compare the impact of different foods on health based on their molar mass. This feature could be particularly useful for nutritionists and health professionals who want to assess the nutritional value of various foods. With the app's ability to recognize chemical structures and its focus on food analysis, Deep Chem has the potential to revolutionize the way we think about food and its impact on health.

E. Training the models

At the beginning, the researchers trained a molecular model on a server with an Nvidia V100 Tesla GPU, 384 GB of RAM, and two Intel(R) Xeon(R) Gold 6230 CPUs. They used a dataset consisting of one million molecules and a batch size of 512 images. The average epoch time was 29 minutes and 48 seconds, and the model took about 1 day, 5 hours, and 48 minutes to converge.
Later on, the same model was trained on a NVIDIA A100 GPU. The researchers used a batch size of 1024 that was distributed among the nodes. The NVIDIA A100 GPU achieved an
average epoch time of 8 minutes and 41 seconds and fully converged in 8 hours, 41 minutes, and 4 seconds. This approach led to a 71.9% reduction in computation time compared to the previous method.

F. Testing models

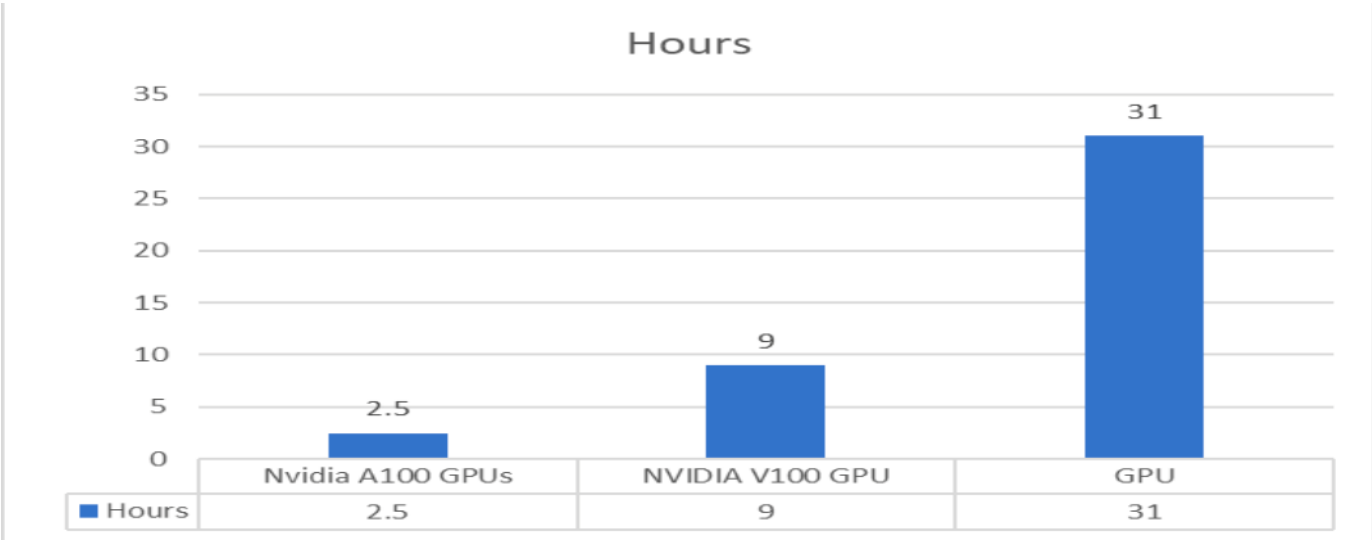
The trained models were evaluated using a diverse test dataset selected using the MaxMin algorithm of RDKit (Landrum G, Others (2016) RDKit), which covers the same chemical space as the training dataset. Test dataset evaluations were conducted on GPUs, where predicted SELFIES were decoded to SMILES, and the Tanimoto Similarity Index was calculated for the original and predicted SMILES using PubChem fingerprints included in the CDK.
The Tanimoto Similarity Index was used as it provides a quantitative measure of how well the network can understand graphical chemical structure representations, and whether a wrong prediction is completely different or very similar to the correct molecule. For the predictions with a Tanimoto similarity index of 1.0, InChIs were generated using the CDK to perform an isomorphism check and determine whether Tanimoto 1.0 predictions are a good proxy for structure identity. Additionally, models trained with augmentations were tested with both augmented and non-augmented images.

IV. RESULTS AND CONCLUSION

A. *Computational considerations*

Training large datasets such as the ones used here on deep neural networks takes months even on GPUs, let alone regular CPUs. For performance measure, a dataset with one million molecules was trained for 50 epochs on an Nvidia Tesla V100 GPU and the same model was also trained on a NVIDIA A100 GPU.

Training a model on a NVIDIA A100 GPU helped by increasing training speed up to 4 times compared to a V100 GPU , see figure



B. *Conclusion and Future Scope*

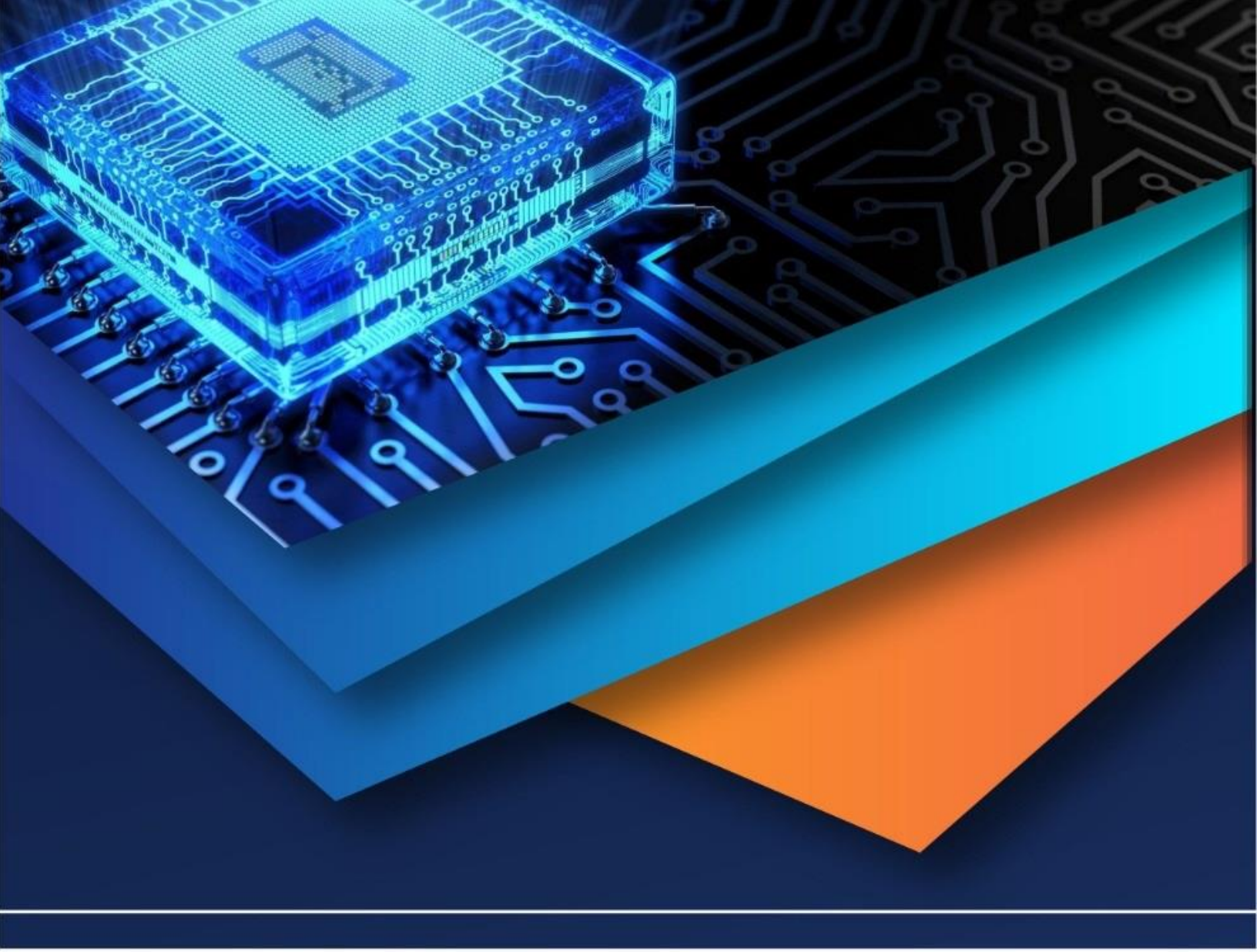
The Deep Chem tool has the potential to benefit the public in many ways. One notable area of application is the food industry, where it can be used to assess the impact of various foods on health based on their molar mass, thereby facilitating informed dietary choices and promoting healthy eating habits. Moreover, the app's ability to automatically recognize chemical structures from printed or scanned documents has broad applicability in fields such as drug discovery, environmental analysis, and patent infringement detection.

Looking ahead, the potential of Deep Chem is vast and promising. The app's accuracy could be enhanced by expanding the size of its datasets, utilizing advanced machine learning algorithms, and improving computational power. This would allow the app to recognize more intricate chemical structures and identify molecules with greater precision. Additionally, the deep learning approach utilized by Deep Chem could be extended to other areas of chemical research, such as predicting toxicity and optimizing reactions. Overall, the potential for Deep Chem to drive advancements in chemical research, facilitate healthier eating habits, and promote public health makes it a highly promising tool for the future.

REFERENCES

[1] O'Boyle NM, Guha R, Willighagen EL et al (2011) Open data, open source and open standards in chemistry: the Blue Obelisk five years on. J Cheminform 3:1–15
[2] Swain MC, Cole JM (2016) ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature. J Chem Inf Model 56:1894–1904
[3] Krallinger M, Rabal O, Lourenço A, Oyarzabal J, Valencia A (2017) Information retrieval and text mining technologies for chemistry. Chem Rev 117:7673–7761
[4] Rajan K, Brinkhaus HO, Zielesny A, Steinbeck C (2020) A review of optical chemical structure recognition tools. J Cheminform. https://doi.org/10.1186/s13321-020-00465-0
[5] Filippov IV, Nicklaus MC (2009) Optical structure recognition software to recover chemical information: OSRA, an open source solution. J Chem Inf Model 49:740–743
[6] Beard EJ, Cole JM (2020) ChemSchematicResolver: a toolkit to decode 2d chemical diagrams with labels and R-groups into annotated chemical named entities. J Chem Inf Model 60:2059–2072
[7] Staker J, Marshall K, Abel R, McQuaw CM (2019) Molecular Structure extraction from documents using deep learning. J Chem Inf Model 59:1017–1029
[8] Ronneberger O, Fischer P, Brox T (2015) U-net: convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science, p 234–241
[9] Rajan K, Zielesny A, Steinbeck C (2020) DECIMER: towards deep learning for chemical image recognition. J Cheminform 12:65

[10] He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In Proc IEEE Conf Comput Vis Pattern Recognit, pp 770–778
[11] Sutskever I, Vinyals O, Le QV (2014) Sequence to sequence learning with neural networks. In Adv Neural Inf Process Syst, pp 3104–3112
[12] Bahdanau D, Cho K, Bengio Y (2015) Neural machine translation by jointly learning to align and translate. In Proc Int Conf Learn Represent
[13] Gawehn E, Hiss JA, Schneider G (2016) Deep learning in drug discovery. Mol Inform 35(1):3–14
[14] Angermueller C, Pärnamaa T, Parts L, Stegle O (2016) Deep learning for computational biology. Mol Syst Biol 12(7):878
[15] Bao W, Yue J, Rao Y, Wang H (2017) Deep learning in finance. IEEE Trans Neural Netw Learn Syst 29(6):1399–1413
[16] Lipton ZC, Kale DC, Elkan C (2016) Learning to diagnose with LSTM recurrent neural networks. arXiv:1511.03677
[17] Levine S, Finn C, Darrell T, Abbeel P (2016) End-to-end training of deep visuomotor policies. J Mach Learn Res 17(39):1–40
[18] Kadian A, Krovi VN, Kalakrishnan M (2021) Learning robot grasping policies with shape-based image representations. Robot Auton Syst 136:103729
[19] Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2015) Semantic image segmentation with deep convolutional nets and fully connected CRFs. In Int Conf Learn Represent
[20] Zhang X, Chen Y, Wang J, Huang K (2016) Object detection in videos with tubelet proposal networks. In Proc IEEE Conf Comput Vis Pattern Recognit, pp 284–293
[21] Lipton ZC (2018) The mythos of model interpretability. arXiv:1606.03490
[22] Liu Y, Chen X, Liu C, Song W (2018) Real-time object detection system for autonomous driving based on deep learning. J Real-Time Image Process 15(4):753–765
[23] Zhang J, Xie S, Xing F (2018) Breast cancer classification using deep learning on digital mammograms. IEEE Access 6:20042–20048
[24] Wu Y, Shen D, Zhang J (2020) A deep learning-based method for text classification. Neural Comput Appl 32:12833–12840
[25] Li Y, Wu X, Wang W, Liu T (2020) Emotion recognition in text using deep learning: A review. IEEE Access 8:21528–21538
[26] Al-Turjman F, Jaber M, Al-Ani A, Alkafri A (2020) Early heart diseases detection using deep learning techniques. J Med Syst 44:127
[27] Xie J, Liu Y, Liu W, Yang H (2019) Prediction of Alzheimer's disease based on hippocampal shape analysis and deep learning. J Healthc Eng 2019:1–13
[28] Wen C, Li Q, Zhang W, Li Y (2020) An efficient deep learning framework for fake news identification. IEEE Access 8:41763–41773
[29] Zhang Y, Xie B, Zhang Q, Huang J (2020) Deep learning for stock prediction: A comparative study. Expert Syst Appl 156:113434
[30] Deng G, Li Y, Zhang Y, Chen J, Wang H (2016) Molecular size effect on the viscosity of ionic liquids. J Phys Chem B 120(44):11337–11344
[31] Cheng Z, Huang Y, Zhou H, Zhang S, Yuan W (2017) Molar mass effect on the rheology and microstructure of polymer solutions. Polymers 9(7):286
[32] Van Krevelen DW, Hoftyzer PJ (1976) Properties of polymers: their estimation and correlation with chemical structure. Elsevier Scientific Pub. Co.
[33] Moon IK, Shibata T, Endo T (2010) Molar mass effect on the viscosity of ionic liquids: comparison between model compounds and ionic liquids. Phys Chem Chem Phys 12(23):6068–6075
[34] Li X, Xu L, Du J, Sun C (2018) Effects of molecular size and shape on diffusion coefficients in liquids. Phys Chem Chem Phys 20(10):6871–6881
[35] Schofield K, Kuemmel J, Jackson G (2006) Diffusion of solutes in liquids. Chem Rev 106(3):1173–1202
[36] Kansal SK, Kumbharkhane AC (2014) Effect of molar mass on the diffusion coefficient of non-electrolyte solutes in dilute solutions of non-polymer solvents. J Mol Liq 193:34–39
[37] Kazarian SG, Chan KLA, Lee T (2015) Molar mass effect on protein diffusion in solution. Phys Chem Chem Phys 17(33):21448–21456
[38] Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In Adv Neural Inf Process Syst, pp 1097–1105
[39] Hinton GE, Osindero S, Teh YW (2006) A fast learning algorithm for deep belief nets. Neural Comput 18(7):1527–1554
[40] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Bengio Y (2014) Generative adversarial nets. In Adv Neural Inf Process Syst, pp 2672–2680
[41] He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In Proc IEEE Conf Comput Vis Pattern Recognit, pp 770–778
[42] Liu T, Zhang M, Wang J, Zhang J (2020) Multi-scale attention network for object detection in images
[43] He K, Gkioxari G, Dollár P, Girshick R (2017) Mask R-CNN. In Proc IEEE Int Conf Comput Vis, pp 2961–2969
[44] Chen, Y., Lou, J., & Wang, W. (2019). EfficientPS: Efficient Panoptic Segmentation. arXiv preprint arXiv:1911.09070.
[45] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241).
[46] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3431-3440).
[47] Dai, J., Li, Y., He, K., & Sun, J. (2016). R-FCN: Object detection via region-based fully convolutional networks. In Advances in neural information processing systems (pp. 379-387).
[48] Landrum G, Others (2016) RDKit: Open-Source Cheminformatics Software.(2016). URL http://www. rdkit. org/, https://github. com/rdkit/rdkit cito:usesMethodIn]
[49] Riniker, S., and Landrum, G. A. (2013). Better informed distance geometry: sing what we know to improve conformation generation. Journal of chemical information and modeling, 53(6), 1689-1699.
[50] Ma, X. H., Zheng, C. J., Han, L. Y., Chen, Y. Z., and Cao, Z. W. (2015). Evaluation of molecular fingerprinting and machine learning techniques for the prediction of drug target interaction. Journal of chemical information and modeling, 55(11), 2444-2460.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY