Advancements in Hand-Drawn Chemical Structure Recognition through an Enhanced DECIMER Architecture

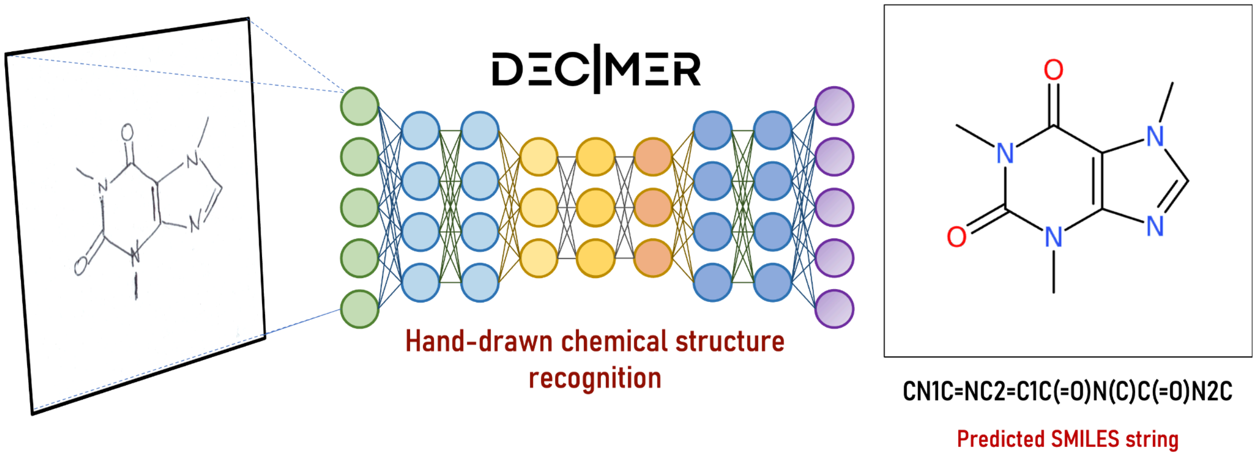
Kohulan Rajan[1], Henning Otto Brinkhaus[1], Achim Zielesny[2], and Christoph Steinbeck[1]*
Institute for Inorganic and Analytical Chemistry, Friedrich Schiller University Jena, Lessingstr. 8, 07743 Jena, Germany
Institute for Bioinformatics and Chemoinformatics, Westphalian University of Applied Sciences, August-Schmidt-Ring 10, 45665 Recklinghausen, Germany

*Corresponding author: christoph.steinbeck@uni-jena.de

Abstract

Accurate recognition of hand-drawn chemical structures is crucial for digitising hand-written chemical information found in traditional laboratory notebooks or for facilitating stylus-based structure entry on tablets or smartphones. However, the inherent variability in hand-drawn structures poses challenges for existing Optical Chemical Structure Recognition (OCSR) software. To address this, we present an enhanced Deep lEarning for Chemical ImagE Recognition (DECIMER) architecture that leverages a combination of Convolutional Neural Networks (CNNs) and Transformers to improve the recognition of hand-drawn chemical structures. The model incorporates an EfficientNetV2 CNN encoder that extracts features from hand-drawn images, followed by a Transformer decoder that converts the extracted features into Simplified Molecular Input Line Entry System (SMILES) strings. Our models were trained using synthetic hand-drawn images generated by RanDepict, a tool for depicting chemical structures with different style elements. To evaluate the model's performance, a benchmark was performed using a real-world dataset of hand-drawn chemical structures. The results indicate that our improved DECIMER architecture exhibits a significantly enhanced recognition accuracy compared to other approaches.

Scientific Contribution:
The new DECIMER model presented here represents a refinement of our previous research efforts and is currently the only open-source model tailored specifically for the recognition of hand-drawn chemical structures. The enhanced model performs better in handling variations in handwriting styles, line thicknesses, and background noise, making it suitable for real-world applications. The DECIMER hand-drawn structure recognition model and its source code have been made available as an open-source package under a permissive license.


DECIMER
Hand-drawn chemical structure recognition
O
N
N
O
N
N
CN1C=NC2=C1C(=O)N(C)C(=O)N2C
Predicted SMILES string

Graphical Abstract: Illustration of the DECIMER hand-drawn chemical structure recognition model.

Keywords

Hand-drawn chemical structures, Chemical Structure Recognition, OCSR, Optical Chemical Structure Recognition, DECIMER, Deep-Learning, Transformer

Introduction

For most of our cultural history, humans have used hand-drawing and hand-writing to create art and capture information. Digitising graphics is common, but capturing their deeper meaning is much more challenging. With the advent of so-called deep learning algorithms, the interpretation of images has seen considerable advances, ranging from the interpretation of medical images to the annotation of personal photo collections.

A key application of deep learning methods in chemistry is the mining of printed and hand-written documents for information on chemical compounds. Mining of past publications, for example, can augment present open-access databases [1]. While this information can often be found in printed literature, it is typically presented in unstructured, human-readable formats like text and images. Manually curating and organising this information to fill the database gaps is error-prone and time-consuming [2]. Therefore, automation is necessary to improve accuracy and efficiency [3]. A key task is detecting and interpreting chemical structure depictions to translate them into machine-readable formats, commonly referred to as Optical Chemical Structure Recognition (OCSR) [4].

Over the past few years, deep learning methods have been used extensively to conduct OCSR for detecting and converting chemical structure depictions from printed literature [4,5]. With improvements in computer vision and language models, the field has seen a lot of development

[6]. Molecular structures can be represented in images in various ways, using many different drawing styles. When representations of a variety of depiction styles are included in the training data, a data-driven deep-learning approach can be applied to reach a high degree of robustness and flexibility. Rule-based OCSR algorithms that are not based on deep learning have been shown to lack robustness and tend to fail when small distortions are added to the images in common benchmark datasets [7].

In addition to mining chemical information from printed literature, information can also be found in hand-written laboratory notebooks that were never before attempted to be digitised and mined for chemical structure information. In these notebooks, chemical structures are typically manually drawn, which means there is an even higher degree of diversity in the way molecular structures are depicted. Unless the chemists choose to publish their novel findings together with related information in a publication, these hand-drawn structures are never converted into machine-readable formats. Recognising and interpreting hand-drawn chemical structures is challenging due to the variety of drawing styles and the complexity of each individual's handwriting [8,9]. It is, therefore crucial to develop accurate tools for recognising hand-drawn chemical structures to digitise them. Digitising hand-written chemical structures enables high-quality data-driven research and preserves information for future use.

Like hand-written text recognition, hand-drawn chemical structure recognition can be categorised into online and offline recognition tasks [10]. Online chemical structure detection primarily denotes converting a chemical structure drawn on a digital medium, such as a tablet or personal computer, into a machine-readable format in real time. If the detection is inaccurate, the user can adjust their drawing style to make the system predict the molecule correctly. In contrast, offline chemical structure detection predominantly deals with previously drawn chemical structure images. These images exhibit a wide array of drawing styles, making it considerably more challenging to recognise them with high confidence [11].

Taking these considerations into account, we present an advanced deep-learning method for accurate hand-drawn chemical structure recognition. We introduce an encoder-decoder model that combines the EfficientNetV2 Convolutional Neural Network (CNN) with a Transformer Decoder-only model. This combination aims to identify and transform hand-drawn chemical structures into a machine-readable file format with higher confidence. Our approach builds upon the DECIMER image transformer [6][12], a deep learning-based OCSR method developed for extracting chemical structural data from printed literature. There is a growing interest in identifying hand-drawn chemical structure depictions, as this has the potential to streamline the automated digitisation of laboratory notebooks [13].

OCSR methods can be broadly categorized into two main groups: rule-based methods and deep learning-based methods [4]. Rule-based approaches typically involve a systematic sequence of processing steps, including vectorisation, atom detection, bond classification, Optical Character Recognition (OCR) [14], graph compilation, and post-processing. Various rule-based techniques, such as OSRA [15], Imago [16], and MolVec [17], follow a procedure along those lines. In 2021, Clévert et al. showed that the performance of the openly available rule-based systems on commonly used benchmark datasets decreases drastically when slight image distortions are introduced [7]. Apparently, the parameters in the rule-based procedures can be overfit to specific depiction styles and do not necessarily perform well on all types of chemical structure depictions.

In recent years, deep learning-based OCSR methods have become increasingly popular [5], driven by advancements in computer vision and powerful hardware for training complex models. Deep learning approaches excel in processing chemical structure depictions and can effectively process even distorted representations [7]. This capability provides a competitive edge when developing OCSR methods for hand-drawn chemical structures. Since deep learning algorithms can detect more complex patterns, they are an excellent choice for OCSR applications. Additionally, these methods can be trained with large amounts of diverse data, resulting in improved accuracy and reliability. Deep learning methods encompass a range of both closed-source approaches, such as MSE-DUDL [18], MICER [19], Image2SMILES [20], ABC-Net [21], Image-to-Graph Transformers [22], IMG2SMI [23], Molecular-InChI [24], and DeepOCSR [25]. On the other hand, several open-source deep learning algorithms have been published, including ChemGrapher [26], DECIMER Image Transformer [12], ChemPix [11], SwinOCSR [27], Img2Mol [7], MolScribe [28], and MolGrapher [29].

While deep learning methods were initially developed for broad applicability across various types of chemical structure depictions, ChemPix was explicitly designed to recognise hand-drawn chemical structure drawings. One notable constraint of ChemPix is its limited functionality, as it exclusively handles drawings of hydrocarbons and is unsuited for other classes of chemical structure representations. In our recently published study about the DECIMER Image Transformer [6], we provided evidence to show that even though our deep learning model was not explicitly trained on hand-drawn chemical structure representations, it exhibits a (limited) capability to interpret them. When compared with ChemPix, our model is capable of recognising various hand-drawn representations of small molecule structures that go beyond those of hydrocarbons. Furthermore, our findings suggest that the recognition performance of this model could be enhanced by training it on a dataset that contains a wide range of hand-drawn chemical structure images.

This work presents a working solution for translating hand-drawn chemical structures into SMILES representations of the depicted molecules [30]. It was specifically trained using artificial data generated by the open-source structure depiction toolkit RanDepict [31] with its synthetic hand-drawn feature capable of producing chemical structure representations that mimic hand-drawn chemical structure drawings [6]. The trained model has been benchmarked against the only available diverse hand-drawn chemical structure dataset, DECIMER hand-drawn images [32]. The approach followed here includes no hard-coded rules and is entirely data-driven. The model has been trained and tested only on openly available data sources.

Using this method, we can achieve recognition performance with high confidence in hand-drawn chemical structure depictions. Furthermore, we improved the recognition results' accuracy by enhancing the DECIMER Image Transformer model. To determine which encoder-decoder

model performs best on the same data set, three different models with different configurations of encoder-decoder architectures have been investigated in this study. Subsequently, the best-performing model was trained on datasets of hand-drawn-like chemical structure depictions of four different sizes generated using RanDepict. Finally, the best-trained model was benchmarked against other deep learning-based OCSR methods using a hand-drawn chemical structure dataset. As compared to other openly available OCSR applications, our approach produces better results, with an accuracy of 73.25% and a Tanimoto average of 0.94. This approach can be used to develop accurate and robust OCSR pipelines for real-world applications. Our hand-drawn chemical structure detection model, which we call the *DECIMER hand-drawn model,* has been incorporated into the DECIMER module and made publicly available. These resources are provided under permissive licenses and accompanied by comprehensive documentation.

Methods

Here, we introduce an improved version of the DECIMER model designed to recognise hand-drawn chemical structures. The model's architecture is illustrated in Figure 1. The final model consists of an EfficientNetV2-M encoder combined with a Transformer Decoder, specifically utilising only the decoder component of the transformer. The encoder processes the chemical structure images to generate a 2-dimensional feature vector, while the decoder then converts this feature vector into a SMILES string.

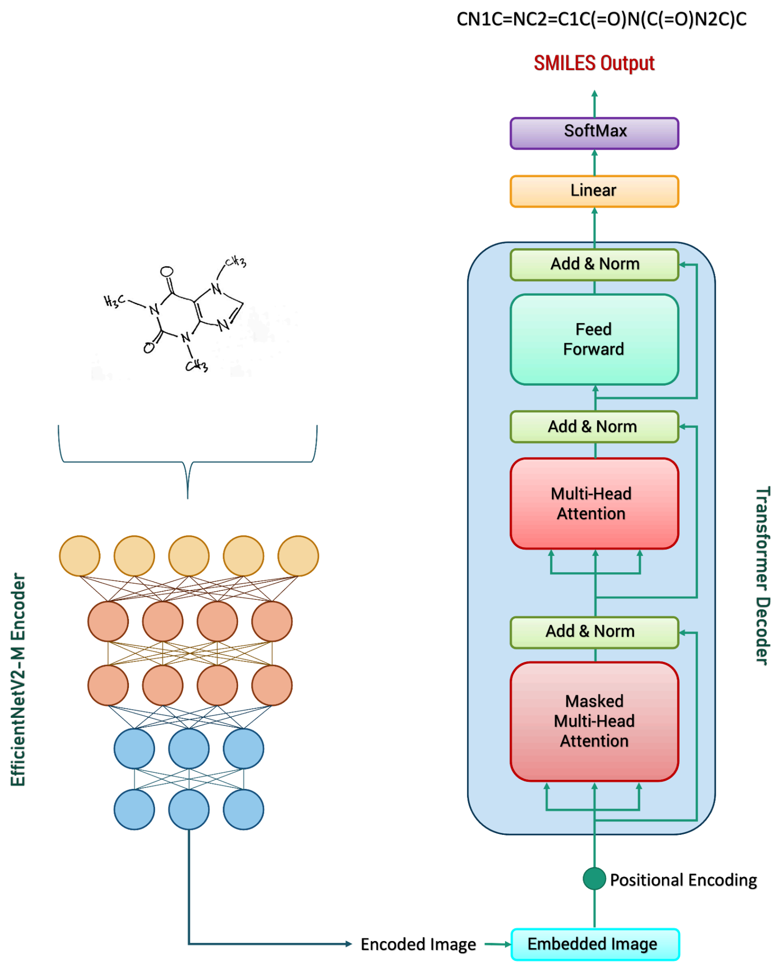


Figure 1: DECIMER hand-drawn chemical structure recognition OCSR model.

Model selection

An analysis of three different encoder-decoder models is presented in this work. All models feature a CNN encoder based on EfficientNet and a decoder based on the Transformer model [33]. The first model uses the original implementation from our recent publication [6]. It contains an EfficientNetV2-M [34] model as an encoder and a Transformer model as a decoder. The second model uses an EfficientNetV1-B7 [35] encoder and a Transformer decoder. For the third model, EfficientNetV2-M was used as the encoder. In models 2 and 3, only the decoder part of the Transformer model was utilised, while model 1 uses the complete Transformer model. The Transformer models used have six decoder layers, eight attention heads, and an embedding dimension of 512 parameters. A detailed summary of these models can be seen in Table 1. All three models were implemented using Python and TensorFlow. Among them, the best-performing model was selected as the final model (see Table 1).

Table 1: Configurations of the three tested DECIMER Image Transformer models.

Model ID	Encoder		Decoder		Batch size	Epochs	Average training time per epoch
	Type	Architecture	Type	Architecture			
1	EfficientNet-V2	M	Transformer	Encoder-Decoder	512	25	36 minutes
2	EfficientNet-V1	B7	Transformer	Decoder only	512	25	57 minutes
3	EfficientNet-V2	M	Transformer	Decoder only	512	25	34 minutes

Training the models

In this study, we trained all of our models on the Google Cloud Platform using the latest Tensor Processing Units (TPUs) - V4. TPUs were selected for this study based on our prior experience, which demonstrated significantly faster training times when compared to in-house Graphical Processing Units (GPUs). TensorFlow served as the backend framework, leveraging the TensorFlow distributed training Advanced Programming Interface (API). The TPU V4 has enabled us to train larger models with more extensive training datasets, yielding improved results. Moreover, TPUs are more energy-efficient than GPUs, facilitating more effective resource utilisation during training.

Testing the models

The initial models were tested using common OCSR benchmark datasets to determine which model performed best. It was then subjected to further testing later on (see below). The models were tested on these real-world datasets:

- JPO: a set of 450 chemical structure images from the Japanese Patent Office [36]
- CLEF: a set of 992 chemical structure images from the Conference and Labs of the Evaluation Forum test set [37,38]
- USPTO: a set of 5,719 chemical structure depictions from the US Patent Office [36]
- UOB: the dataset of 5,740 chemical structure depictions compiled by the University of Birmingham [39]

The models were primarily evaluated for their ability to recognise chemical structure depictions accurately. This evaluation was based on two key metrics. First, we conducted a one-to-one string comparison using Canonical SMILES for both the original and predicted SMILES representations. This analysis provided insight into how effectively each model predicts chemical structures from input images of chemical structure depictions, with even a single character mismatch in the predicted SMILES string considered as an incorrect prediction.

Additionally, a Tanimoto [40] similarity calculation was performed using PubChem fingerprints, employing the Chemistry Development Kit (CDK) [41] implementation, to compare the original and predicted molecular structures. This approach helped to assess the similarity between the predicted chemical structure and the original one, even when the model's SMILES prediction was inaccurate. This method is particularly valuable because not all predicted molecules precisely match the original, and a quantitative measure aids in understanding the model's performance in interpreting chemical structure depictions. As a result, this comprehensive evaluation approach enhances our understanding of the model's generalisation capabilities.

Datasets

This section discusses the data sources and the generation of images and textual molecular representations for the datasets used for training the models.

Selection of molecules for the datasets

For training and testing models 1 to 3, the latest ChEMBL-32 database was utilised. ChEMBL [42] database version 32 was acquired in the SDF (Structure-Data File) format. The dataset was processed using the CDK SMILES parser functionality to generate canonical SMILES representations preserving stereochemical information. These SMILES strings and their corresponding ChEMBL IDs were then stored in a text file. After analysing the frequency distribution of the length of the SMILES strings, those exceeding 300 characters were removed to eliminate rare, longer SMILES strings. The resulting dataset consisted of a total of 2,290,069 SMILES strings. To select the training and testing datasets, the RDKit [43] implementation of the MaxMin algorithm [44] was used to pick diverse data points for cross-validation. This resulted in training and a test dataset consisting of 2,187,669 and 102,400 molecules, respectively. From the resulting training dataset, a subset of 1,024,000 molecules were picked to be used for training the models in this experiment. These were used to train models 1 to 3 and later determine which model was suitable for further experiments.

Similarly, the whole PubChem [45] dataset was processed to select nearly 100 Million molecules for training and 100,000 data points for cross-validation. This dataset was later used to train and test the best-performing model for hand-written structure recognition.

Training Dataset Generation

Various chemical structure depictions of the selected SMILES strings were generated using the RanDepict toolkit [31]. The images were created with a resolution of 512 x 512 pixels per image. Each data point was represented by two 8-bit PNG images - one with and one without any image augmentations, excluding hand-drawn-like augmentations. The purpose of introducing augmentation on the images is to mimic real-world scanned pages and to add more complexity. The models were trained using a dataset consisting of 2,048,000 images. These generated images were used as the input for the encoder, and the SMILES strings were defined as the desired decoder output. The SMILES strings were split into meaningful tokens using the Keras tokenizer. The resulting tokenisation scheme splits the input after heavy atoms (such as "C" and "O"), open and closed brackets (such as "(" and ")"), bond symbols ("=" and "#"), special characters(“.”, “-”, ”+”, ”\”, ”/”, ”@”, ”%” and ”*”), as well as after every single-digit number. A start token "<start>" and an end token "<end>" were added to the beginning and end of each sequence, respectively. Additionally, each tokenised string was padded using "<pad>" tokens.

The generated images with their corresponding tokenised SMILES strings were then combined and converted into small chunks of TFRecord files of about 100 MB each. They were then moved to a Google Cloud bucket for training. Datasets were converted into TFRecord files primarily for training on Google Cloud using Tensor Processing Units (TPUs).

Similarly, the PubChem dataset was used to generate the training dataset for the final model. Using the selected SMILES strings, hand-drawn-like synthetic chemical structure depictions were generated using RanDepict (see Figure 3). Again, the image size was set to 512 x 512 and the generated data and the tokenised SMILES were saved into TFRecord files and moved to a Google Cloud bucket for training. Here, every molecule was depicted three times without augmentations and once with augmentations.

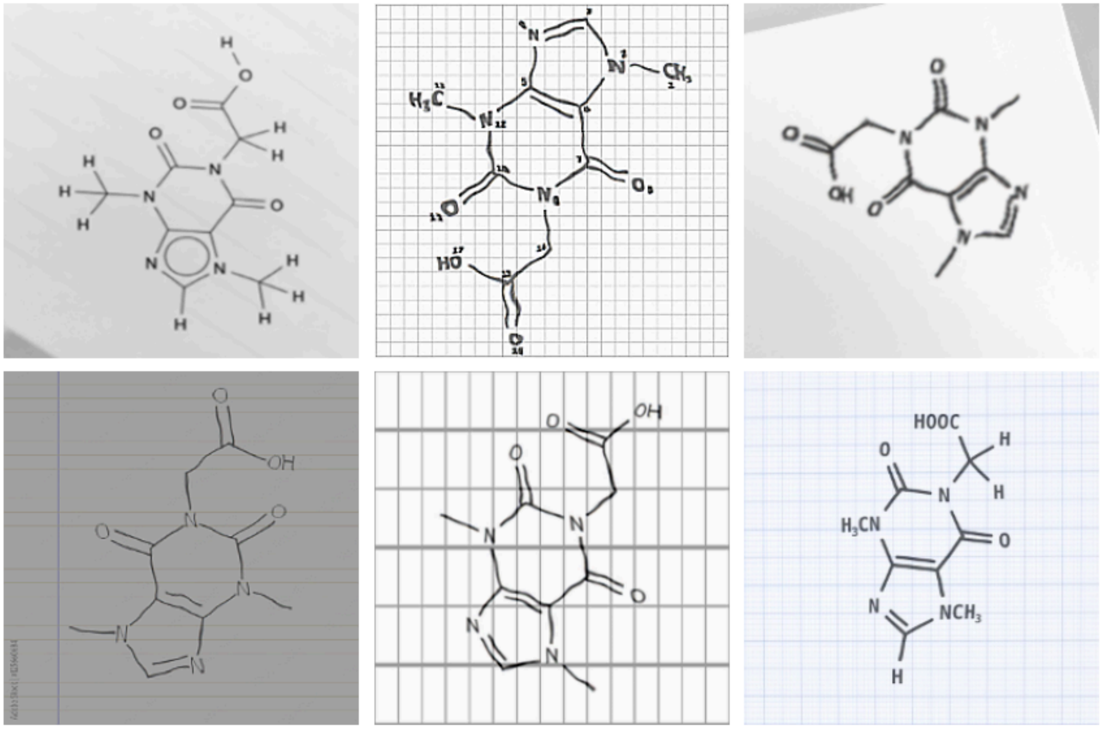


Figure 3: Examples of hand-drawn-like synthetic chemical structure depictions created for the Caffeine molecule through the use of RanDepict.

Results and Discussion

This section analyses the three models that we first selected to identify which model architecture yields the best results on all benchmark datasets. Subsequently, the best-performing model architecture was selected and used to carry out the next experiment to determine whether the model's accuracy could be improved with more training data.

Testing Different Model Architectures

The performance of the three models on real-world images was evaluated using the OCSR benchmark datasets listed under testing the models. The model performance is presented in Table 2, with '**P**' representing the percentage of identical predictions and '**T**' denoting the average Tanimoto similarity calculated across all structures in a dataset. This table serves as the basis for determining the best-performing model, which was considered a candidate for subsequent stages of the experiment.

Table 2: DECIMER Image Transformer model performance on OCSR benchmark datasets compared by identical predictions (P) and Tanimoto similarity (T).

	JPO		**CLEF**		**USPTO**		**UOB**		**Average**	
	P	**T**	**P**	**T**	**P**	**T**	**P**	**T**	**P**	**T**
Model 1	47.78%	0.86	62.00%	0.94	56.78%	0.95	78.55%	0.97	61.28%	0.93
Model 2	**64.00%**	**0.94**	60.58%	0.94	60.29%	0.97	86.17%	0.98	67.76%	0.96
Model 3	62.67%	**0.94**	**63.51%**	**0.95**	**64.01%**	**0.97**	**86.88%**	**0.99**	**69.27%**	**0.96**

Model 1's performance is poorer than that of Models 2 and 3: apparently, the usage of the entire Transformer model as a decoder leads to a reduction in performance compared to the usage of the decoder part of the Transformer architecture alone. By using only the Transformer decoder for decoding and removing the encoder part of the transformer, we achieved much better performance on all the OCSR benchmark datasets. Model 3 slightly outperforms Model 2. This is due to using EfficientNetV1 in Model 2, whereas Model 3 uses an updated architecture, EfficientNetV2. In general image recognition tasks, EfficientNetV2 outperforms EfficientNetV1 [34]. Additionally, due to the compact architecture of EfficientNet-V2, Model 3 could train approximately 2 times faster than Model 2 (see Table 1). After assessing the performance metrics and the training times, the model architecture of Model 3 was picked for further experiments.

Improvement in model prediction with increasing dataset size

Here, the improvement of the accuracy of the model predictions with an increase in the training dataset size and the introduction of hand-drawn-like images in the training data was assessed.

With the hand-drawn-like structure depictions, the complexity of the representations of the chemical structures was increased compared to the previously used clean depictions.

In this part of the experiment, we used the molecule datasets based on ChEMBL and PubChem that have been described under methods in datasets. All of the images that were used for training the models in this experiment were generated by RanDepict, which generated synthetic hand-drawn images for training the models. The models were then tested on a dataset of real-world images to assess their performance. The DECIMER - Hand-drawn images dataset [16], was used to evaluate the models' performance. The dataset consists of 5088 chemical structure drawings sketched by 23 volunteers. The drawings reflect a wide range of drawing styles. The dataset helps us to better understand how well the model that has been exclusively trained on artificially generated training data performs on real hand-drawn chemical structure images.

Training datasets

The ChEMBL and PubChem molecular structure datasets were each used to create two additional training datasets. Table 3 summarises the dataset sizes and the number of images with and without augmentations.

Table 3: Training dataset summary.

Dataset ID	Database	No of Molecules	No of images Without augmentations	No of images With augmentations	Total number of images
1	ChEMBL	2,187,669	2,187,669	2,187,669	4,375,338
2	ChEMBL	2,187,669	8,750,676	4,375,338	13,126,014
3	PubChem	9,510,000	28,530,000	9,510,000	38,040,000
4	PubChem	3,8040,000	114,120,000	38,040,000	152,160,000

There was no change in the number of molecules between datasets 1 and 2; however, there was a notable increase in the number of images depicted using each molecule. During the transition from Dataset 2 to Dataset 3, both the quantity of molecules and the number of depictions grew. Furthermore, as the number of molecules expanded from Dataset 3 to Dataset 4, there was a corresponding increase in the volume of depicted images.

Training implementation

The models were trained using TensorFlow version 2.13.0. The training process utilised the model 3 implementation, consisting of an encoder with an EfficientNetV2-M model employing default configurations and a transformer decoder with 6 layers (refer to Figure 1). These models underwent training for 25 epochs on a TPU V4-128 pod slice. Training employed focal loss and the Adam optimizer, complemented by a custom schedule for the learning rate, as specified in the original transformer paper [33]. A dropout rate of 0.1 was also used. To ensure compatibility with the encoder's settings, the images were preprocessed to attain a size of 512 x 512 before being fed into the encoder.

Performance on Hand-drawn Dataset

After training each model, it was tested against the DECIMER hand-drawn chemical structure images dataset for accuracy and similarity. The number of valid predictions, i.e. the returned SMILES string was syntactically valid and could be parsed into a molecular structure, is also measured. Table 4 provides the final average values for overall predictions by comparing each predicted structure with the original structure.

Table 4: Model performance with increasing dataset size against benchmark dataset.

Model ID	Training Dataset size	Percentage of Valid predictions	Model Accuracy	Average Tanimoto similarity
1	4,375,338	96.21%	5.09%	0.490
2	13,126,014	97.41%	26.08%	0.690
3	38,040,000	99.67%	70.34%	0.939
4	152,160,000	99.72%	73.25%	0.942

As expected, there is a significant improvement in performance by tripling the amount of training data from Model 1 via Model 2 to Model 3, reaching a high percentage of valid predictions above 99%, a substantial accuracy of about 70%, and an average Tanimoto similarity of 0.93, indicating similar input and output structures. However, the next quadrupling of the training data for Model 4 only leads to a slight improvement in performance compared to Model 3, suggesting that the potential of the selected training data has been exhausted and that in the future the diversity of the training data needs to be increased to address the weaknesses of the model specifically.

Performance comparison with other available methods

The performance of the final best model on the DECIMER Hand-Drawn Molecules dataset was compared with other available open-source OCSR methods. The tools were evaluated and compared by executing them on real-world hand-drawn images from the DECIMER Hand-Drawn dataset to provide valuable insights into the applicability of the available tools for processing real hand-drawn structure depictions. The summarised results of these comparisons are presented in Table 5. Our study incorporates both rule-based and deep-learning methods.

Table 5: DECIMER model performance compared with all available open-source methods.

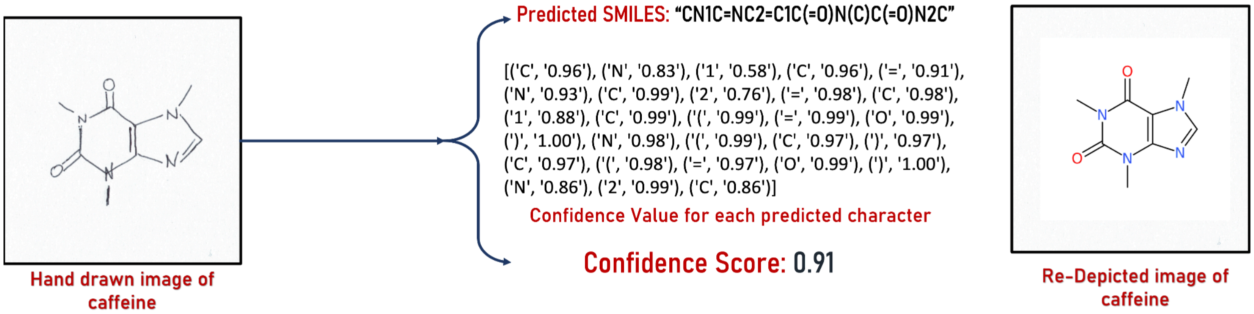
OCSR tool	Method	Percentage of Valid predictions	Model Accuracy	Average Tanimoto similarity
OSRA	Rule-based	54.66%	0.57%	0.17
Imago	Rule-based	43.14%	2.99%	0.22
MolVec	Rule-based	71.86%	1.30%	0.23
ChemGrapher	Deep Learning	69.56%	0.0%	0.09
Img2Mol	Deep Learning	98.96%	5.25%	0.52
SwinOCSR	Deep Learning	97.37%	5.11%	0.64
MolScribe	Deep Learning	95.66%	7.65%	0.59
MolGrapher	Deep Learning	99.94%	10.81%	0.51
DECIMER	Deep Learning	99.72%	73.25%	0.94

As can be seen from the above results, the DECIMER model overall performs much better than other deep learning models. According to the results, the rule-based methods perform significantly worse than all the currently available deep learning methods. It is primarily due to the handcrafted rules that were developed for chemical structure representations found in printed literature, as when we deployed them on a hand-drawn dataset, they were not able to function properly since they are not as flexible as the deep learning tools when it comes to processing hand-drawn chemical structures. While deep learning models tend to display a higher level of robustness on this dataset, the number of valid predictions generated by these models is significantly higher than those generated by rule-based methods since deep learning models are likely to pick up on patterns, contexts and subtleties in the hand-drawn structures since they are more robust to noise and variability because they learn the patterns directly from the training data rather than having hardcoded rules. As a result, they can take advantage of a lot more contextual data in the input to make predictions.

Confidence score and prediction analysis

It is difficult to estimate the quality of a result returned by the DECIMER Image Transformer in a real-world application without a manual assessment. Users have to re-depict the molecule represented by a predicted SMILES string and compare it to the molecule depicted in the original image. As this is a time-consuming and potentially error-prone procedure, an automated estimation of the quality of the generated SMILES string is highly desirable.

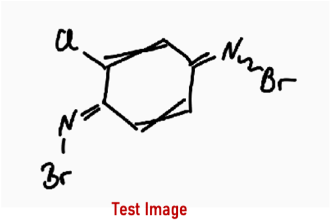
The DECIMER Image Transformer now includes a feature for extracting token-level confidence scores from the model's output. The final layer of the Transformer decoder generates a value for each token within the output vocabulary. These values are then normalised to sum up to 1 using a Softmax activation function. During each prediction step, the token associated with the highest value is selected, and the value itself is regarded as a measure of confidence. This process bears similarity to the methodology used for determining confidence scores in MolScribe, as previously described by Qian et al. [28]. The token-level scores can then be averaged to yield one score for the whole predicted SMILES string. This confidence score allows us to determine how well the model can interpret a given hand-drawn chemical structure image (see Figure 4).


Predicted SMILES: “CN1C=NC2=C1C(=O)N(C)C(=O)N2C”
Confidence Value for each predicted character
Confidence Score: 0.91
Hand drawn image of caffeine
Re-Depicted image of caffeine
O
N
N
N
N
O

```
[('C', '0.96'), ('N', '0.83'), ('1', '0.58'), ('C', '0.96'), ('=', '0.91'),
('N', '0.93'), ('C', '0.99'), ('2', '0.76'), ('=', '0.98'), ('C', '0.98'),
('1', '0.88'), ('C', '0.99'), ('(', '0.99'), ('=', '0.99'), ('O', '0.99'),
(')', '1.00'), ('N', '0.98'), ('(', '0.99'), ('C', '0.97'), (')', '0.97'),
('C', '0.97'), ('(', '0.98'), ('=', '0.97'), ('O', '0.99'), (')', '1.00'),
('N', '0.86'), ('2', '0.99'), ('C', '0.86')]
```

Figure 4: A hand-drawn representation of a Caffeine molecule was processed using the DECIMER hand-drawn model, and the resulting SMILES string, accompanied by associated confidence values, is presented. The predicted SMILES string serves as the basis for the re-depicted Caffeine structure.

The model's improvement was first assessed using the benchmark dataset. With each iteration of expanding the training dataset, we analysed the model's performance on the same benchmark data. To understand the extent to which the model improves with the increasing training dataset size, we examined the predicted SMILES and their associated confidence scores, as detailed above. Figure 5 shows that the model keeps improving its overall performance with the increasing training dataset size. Also, the confidence score increases.



Test Image

Model ID	I	II	III	IV
Predicted SMILES	C1=C/C(=N\[O-])/C=C(C1)Cl.[Zn+2]	C1=C/C(=N\[11CH3])/C(=CC1=[13Br])Cl	C1=C/C(=N\Br)/C(=CC1=NBr)Cl	C1=C/C(=N\Br)/C(=CC1=NBr)Cl
Tanimoto Similarity	0.77	0.74	1.0	1.0
Confidence Score	0.782	0.877	0.903	0.909

Figure 5: The DECIMER hand-drawn model performance on the same test image along with its Tanimoto Similarity value calculated using the original structure and the calculated confidence value. The columns represent increasing training set sizes used to train the presented model.

Conclusion

This study introduces an enhanced encoder-decoder model designed to recognise hand-drawn chemical structures. Leveraging recent advancements in computer vision and natural language processing, our model demonstrates significantly improved accuracy, particularly when trained on extensive datasets which contain synthetic hand-drawn images generated using RanDepict. Comparative analysis with already available open-source methods exhibits highly competitive performance when converting hand-drawn chemical structure depictions into computer-readable file format.

The DECIMER model for hand-drawn chemical structure recognition is now seamlessly integrated within the DECIMER modules and will also be available to use in the Decimer.ai platform soon. Our intent in providing both the model and its source code to the broader public is to make a substantial contribution to the field of chemical data mining. Furthermore, it will facilitate the development of innovative applications and tools for extracting valuable information from laboratory notebooks.

List of abbreviations

ABC-Net - Atom and Bond Center Network

CDK - Chemistry Development Kit

CLEF - Conference and Labs of the Evaluation Forum

CNN - Convolutional Neural Networks

DECIMER - Deep lEarning for Chemical ImagE Recognition

JPO - Japanese Patent Office

IMG2SMI - Image to SMILES

InChI - International Chemical Identifier

MICER - Molecular Image CaptionER

MSE-DUDL - Molecular Structure Extraction from Documents Using Deep Learning

OCR - Optical Character Recognition

OCSR - Optical Chemical Structure Recognition

OSRA - Optical Structure Recognition Application

PC - Personal Computer

PNG - Portable Network Graphics

SDF - Structure-Data File

SMILES - Simplified Molecular Input Line Entry System

TFRecord - TensorFlow Record

TPU - Tensor Processing Unit

UOB - University Of Birmingham

USPTO - United States Patent Office

V - Version

References

1. Brinkhaus, H.O.; Rajan, K.; Schaub, J.; Zielesny, A.; Steinbeck, C. Open Data and Algorithms for Open Science in AI-Driven Molecular Informatics. *Curr. Opin. Struct. Biol.* **2023**, *79*, 102542, doi:10.1016/j.sbi.2023.102542.
2. Swain, M.C.; Cole, J.M. ChemDataExtractor: A Toolkit for Automated Extraction of Chemical Information from the Scientific Literature. *J. Chem. Inf. Model.* **2016**, *56*, 1894–1904, doi:10.1021/acs.jcim.6b00207.
3. Rajan, K.; Zielesny, A.; Steinbeck, C. DECIMER: Towards Deep Learning for Chemical Image Recognition. *J. Cheminform.* **2020**, *12*, 65, doi:10.1186/s13321-020-00469-w.
4. Rajan, K.; Brinkhaus, H.O.; Zielesny, A.; Steinbeck, C. A Review of Optical Chemical Structure Recognition Tools. *J. Cheminform.* **2020**, *12*, 60, doi:10.1186/s13321-020-00465-0.
5. Musazade, F.; Jamalova, N.; Hasanov, J. Review of Techniques and Models Used in Optical Chemical Structure Recognition in Images and Scanned Documents. *J. Cheminform.* **2022**, *14*, 61, doi:10.1186/s13321-022-00642-3.
6. Rajan, K.; Brinkhaus, H.O.; Agea, M.I.; Zielesny, A.; Steinbeck, C. DECIMER.ai: An Open Platform for Automated Optical Chemical Structure Identification, Segmentation and Recognition in Scientific Publications. *Nat. Commun.* **2023**, *14*, 5045, doi:10.1038/s41467-023-40782-0.
7. Clevert, D.-A.; Le, T.; Winter, R.; Montanari, F. Img2Mol - Accurate SMILES Recognition from Molecular Graphical Depictions. *Chem. Sci.* **2021**, doi:10.1039/D1SC01839F.
8. Bluche, T.; Louradour, J.; Messina, R. Scan, Attend and Read: End-to-End Handwritten Paragraph Recognition with MDLSTM Attention. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR); IEEE, November 2017; Vol. 01, pp. 1050–1055.
9. Michael, J.; Labahn, R.; Grüning, T.; Zöllner, J. Evaluating Sequence-to-Sequence Models for Handwritten Text Recognition. In Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR); IEEE, September 2019; pp. 1286–1293.
10. Plamondon, R.; Srihari, S.N. Online and off-Line Handwriting Recognition: A Comprehensive Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 63–84, doi:10.1109/34.824821.
11. Weir, H.; Thompson, K.; Choi, B.; Woodward, A.; Braun, A.; Martínez, T.J. ChemPix: Automated Recognition of Hand-Drawn Hydrocarbon Structures Using Deep Learning. *ChemRxiv* 2021.
12. Rajan, K.; Zielesny, A.; Steinbeck, C. DECIMER 1.0: Deep Learning for Chemical Image Recognition Using Transformers. *J. Cheminform.* **2021**, *13*, 61, doi:10.1186/s13321-021-00538-8.
13. Andrews, D.M.; Broad, L.M.; Edwards, P.J.; Fox, D.N.A.; Gallagher, T.; Garland, S.L.; Kidd, R.; Sweeney, J.B. The Creation and Characterisation of a National Compound Collection: The Royal Society of Chemistry Pilot. *Chem. Sci.* **2016**, *7*, 3869–3878, doi:10.1039/c6sc00264a.
14. Casey, R.; Boyer, S.; Healey, P.; Miller, A.; Oudot, B.; Zilles, K. Optical Recognition of Chemical Graphics. In Proceedings of the Proceedings of 2nd International Conference on Document Analysis and Recognition (ICDAR '93); October 1993; pp. 627–631.
15. Filippov, I.V.; Nicklaus, M.C. Optical Structure Recognition Software to Recover Chemical

Information: OSRA, an Open Source Solution. *J. Chem. Inf. Model.* **2009**, *49*, 740–743, doi:10.1021/ci800067r.

16. Smolov, V.; Zentsev, F.; Rybalkin, M. Imago: Open-Source Toolkit for 2D Chemical Structure Image Recognition. In Proceedings of the TREC; Citeseer, 2011.
17. Peryea, T.; Katzel, D.; Zhao, T.; Southall, N.; Nguyen, D.-T. MOLVEC: Open Source Library for Chemical Structure Recognition. In Proceedings of the ABSTRACTS OF PAPERS OF THE AMERICAN CHEMICAL SOCIETY; AMER CHEMICAL SOC 1155 16TH ST, NW, WASHINGTON, DC 20036 USA, 2019; Vol. 258.
18. Staker, J.; Marshall, K.; Abel, R.; McQuaw, C.M. Molecular Structure Extraction from Documents Using Deep Learning. *J. Chem. Inf. Model.* **2019**, *59*, 1017–1029, doi:10.1021/acs.jcim.8b00669.
19. Yi, J.; Wu, C.; Zhang, X.; Xiao, X.; Qiu, Y.; Zhao, W.; Hou, T.; Cao, D. MICER: A Pre-Trained Encoder-Decoder Architecture for Molecular Image Captioning. *Bioinformatics* **2022**, *38*, 4562–4572, doi:10.1093/bioinformatics/btac545.
20. Khokhlov, I.; Krasnov, L.; Fedorov, M.V.; Sosnin, S. Image2SMILES: Transformer‐based Molecular Optical Recognition Engine. *Chemistry Methods* **2022**, *2*, doi:10.1002/cmtd.202100069.
21. Zhang, X.-C.; Yi, J.-C.; Yang, G.-P.; Wu, C.-K.; Hou, T.-J.; Cao, D.-S. ABC-Net: A Divide-and-Conquer Based Deep Learning Architecture for SMILES Recognition from Molecular Images. *Brief. Bioinform.* **2022**, *23*, doi:10.1093/bib/bbac033.
22. Yoo, S.; Kwon, O.; Lee, H. Image-to-Graph Transformers for Chemical Structure Recognition. In Proceedings of the ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); IEEE, May 23 2022.
23. Campos, D.; Ji, H. IMG2SMI: Translating Molecular Structure Images to Simplified Molecular-Input Line-Entry System. *arXiv [q-bio.QM]* 2021.
24. Kumar, N.; Rashmi, M.; Ramu, S.; Reddy Guddeti, R.M. Molecular-InChI: Automated Recognition of Optical Chemical Structure. In Proceedings of the 2022 IEEE Region 10 Symposium (TENSYMP); IEEE, July 1 2022.
25. Zhaopeng, Y.; Jianhua, L.I. DeepOCSR: A Deep Encoder-Decoder Network for Optical Chemical Structure Recognition. 华东理工大学学报 (自然科学版), doi:10.14135/j.cnki.1006-3080.20210916002.
26. Oldenhof, M.; Arany, A.; Moreau, Y.; Simm, J. ChemGrapher: Optical Graph Recognition of Chemical Compounds by Deep Learning. *J. Chem. Inf. Model.* **2020**, *60*, 4506–4517, doi:10.1021/acs.jcim.0c00459.
27. Xu, Z.; Li, J.; Yang, Z.; Li, S.; Li, H. SwinOCSR: End-to-End Optical Chemical Structure Recognition Using a Swin Transformer. *J. Cheminform.* **2022**, *14*, 41, doi:10.1186/s13321-022-00624-5.
28. Qian, Y.; Guo, J.; Tu, Z.; Li, Z.; Coley, C.W.; Barzilay, R. MolScribe: Robust Molecular Structure Recognition with Image-to-Graph Generation. *J. Chem. Inf. Model.* **2023**, *63*, 1925–1934, doi:10.1021/acs.jcim.2c01480.
29. Morin, L.; Danelljan, M.; Agea, M.I.; Nassar, A.; Weber, V.; Meijer, I.; Staar, P.; Yu, F. MolGrapher: Graph-Based Visual Recognition of Chemical Structures. *arXiv [cs.CV]* 2023, 19552–19561.
30. Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36, doi:10.1021/ci00057a005.
31. Brinkhaus, H.O.; Rajan, K.; Zielesny, A.; Steinbeck, C. RanDepict: Random Chemical Structure Depiction Generator. *J. Cheminform.* **2022**, *14*, 31, doi:10.1186/s13321-022-00609-4.
32. Brinkhaus, H.O.; Zielesny, A.; Steinbeck, C.; Rajan, K. DECIMER-Hand-Drawn Molecule Images Dataset. *J. Cheminform.* **2022**, *14*, 36, doi:10.1186/s13321-022-00620-9.

33. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. *arXiv [cs.CL]* 2017.
34. Tan, M.; Le, Q.V. EfficientNetV2: Smaller Models and Faster Training. *arXiv [cs.CV]* 2021.
35. Tan, M.; Le, Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. *arXiv [cs.LG]* 2019.
36. OSRA Validation Datasets Available online: https://sourceforge.net/p/osra/wiki/Validation/ (accessed on 24 June 2020).
37. CLEF-IP 2012 chemical image recognition task - qrels, 2012. (accessed on 14 November 2023).
38. Piroi, F. CLEF-IP 2012 2021.
39. Sadawi, N.M.; Sexton, A.P.; Sorge, V. Chemical Structure Recognition: A Rule-Based Approach. In Proceedings of the Document Recognition and Retrieval XIX; SPIE, January 23 2012; Vol. 8297, pp. 101–109.
40. Tanimoto, T.T. *An Elementary Mathematical Theory of Classification and Prediction*; International Business Machines Corporation, 1958;.
41. Steinbeck, C.; Han, Y.; Kuhn, S.; Horlacher, O.; Luttmann, E.; Willighagen, E. The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 493–500, doi:10.1021/ci025584y.
42. Mendez, D.; Gaulton, A.; Bento, A.P.; Chambers, J.; De Veij, M.; Félix, E.; Magariños, M.P.; Mosquera, J.F.; Mutowo, P.; Nowotka, M.; et al. ChEMBL: Towards Direct Deposition of Bioassay Data. *Nucleic Acids Res.* **2019**, *47*, D930–D940, doi:10.1093/nar/gky1075.
43. Landrum, G.; Tosco, P.; Kelley, B.; Ric; sriniker; gedeck; Vianello, R.; NadineSchneider; Kawashima, E.; Dalke, A.; et al. *Rdkit/rdkit: 2022_03_3 (Q1 2022) Release*; 2022;.
44. Ashton, M.; Barnard, J.; Casset, F.; Charlton, M.; Downs, G.; Gorse, D.; Holliday, J.; Lahana, R.; Willett, P. Identification of Diverse Database Subsets Using Property-Based and Fragment-Based Molecular Descriptions. *Quant. struct.-act. relatsh.* **2002**, *21*, 598–604, doi:10.1002/qsar.200290002.
45. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem in 2021: New Data Content and Improved Web Interfaces. *Nucleic Acids Res.* **2021**, *49*, D1388–D1395, doi:10.1093/nar/gkaa971.

Declarations

Availability of data and materials

DECIMER Image Transformer was developed using data obtained from ChEMBL and PubChem:

- PubChem: https://ftp.ncbi.nlm.nih.gov/pubchem/Compound/Extras/CID-SMILES.gz
- ChEMBL: https://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/releases/chembl_32/chembl_32.sdf.gz

Code availability: https://github.com/Kohulan/DECIMER-Image_Transformer

Model availability: https://doi.org/10.5281/zenodo.10781330
PyPi Package: https://pypi.org/project/decimer/

Competing interests

AZ is co-founder of GNWI - Gesellschaft für naturwissenschaftliche Informatik mbH, Dortmund, Germany. The remaining authors declare no financial and non-financial competing interests.

Funding

Open Access funding enabled and organised by Projekt DEAL. KR acknowledges the funding by the German Research Foundation under project number: 239748522- SFB 1127 ChemBioSys (Project INF). H.O.B., and C.S. acknowledge for this work by the Carl-Zeiss-Foundation.

Authors' contributions

KR initiated, designed, tested, applied and validated the software features. HOB implemented the confidence score calculation. KR, HOB, AZ, and CS wrote the manuscript. CS and AZ conceived the project and supervised the work. All authors contributed to and approved the manuscript.

Acknowledgements

We are grateful for the company Google making free computing time on their TPU Research Cloud infrastructure available to us, and KR acknowledges the research supported with Cloud TPUs from Google's TPU Research Cloud (TRC). Thanks to Dr. Isabel Agea for her contribution to the development of RanDepict's synthetic hand-drawn chemical structure generation and Dr. Jonas Schaub for his extensive revision of the paper.