molecular informatics

DOI: 10.1002/minf.202200068

Identifying Chirality in Line Drawings of Molecules Using Imbalanced Dataset Sampler for a Multilabel Classification Task

Yong En Kok*[a] Simon Woodward*[b] Ender Özcan[c] and Mercedes Torres Torres[d]

Abstract: Chirality, the ability of some molecules to exist as two non-superimposable mirror images, profoundly influences both chemistry and biology. Advances in deep learning enable the automatic recognition of chemical structure diagrams, however, studies on discovering the molecule chirality are scarce and the machine-readable molecular representations are not always sufficient to fully support the encoding of this important property. Here, we pretrained networks on a ChEMBL+ dataset (79641 molecules) and fine-tuned them for the binary classification of chirality (achiral/chiral) or multilabel chirality type classifications (none/centre/axial/planar). To address the label combination imbalanced problem in the multilabel task, the study proposed a Formulated Imbalanced Dataset Sampler (FIDS) to sample a formulated amount of minority label combinations on top of the training set. On a 10-fold cross validation experiment using our CHIRAL dataset (1142 manually curated molecules), our models achieved up to an accuracy of 90% in the binary task. In the multilabel task incorporated with FIDS, the overall performance increases from 87% to 89% and the accuracy per label combination can attained up to a 50% increase. Through the study of heatmaps, our work also exemplified the potential of deep neural network to make predictions based on the actual location of chirality elements.

Keywords: Chirality, Deep Learning, Convolutional Neural Network, Image Recognition, Classification

1 Introduction

Molecules are chiral when their 3D structures exist as two non-superimposable mirror images forms of each other. If the two structures *can* be superimposed on each other (through any combination of translations, rotations and internal conformational (bond rotation) changes), then the molecule is *not* chiral. At its simplest level, the two mirror image chiral molecules are referred to as the "right-handed (*R*)" and "left-handed (*S*)" forms (which are also called their two enantiomers). Understanding molecular chirality is fundamentally important because different enantiomers of chiral molecules can have very different biological effects i.e. one may have a desired therapeutic effect, whilst the other one may produce detrimental side effects without any benefit. For example, left-handed (*S*)-Ibuprofen is a more effective pain killer, and causes less side-effects than its (*R*) mirror image form (enantiomer)[1] (see Figure 1). Since the Thalidomide disaster (one toxic enantiomer caused severe birth defects), it has become a legal requirement in drug development, that knowledge of the absolute stereochemistry(ies) present within any molecule must be accounted for.

Four common structural motifs are known to engender molecular chirality: (a) central/point, (b) axial, (c) planar and (d) helical chirality, as illustrated in Figure 2. The most common motif, central chirality occurs when an atom is connected to four different substituents (R^{1-4}). Axes of chirality can be found in molecules where a set of substituents is held in a non-planar arrangement that results in a non-superimposable mirror image forms[2] (it is common in conformationally restricted biaryls and allenes). Molecules with planar chirality exhibit a chiral plane containing different substituents that are arranged in a way such that the internal reflection symmetry is removed[2]. Finally, helical chirality[3] is a special case where the chirality element has a screw-liked shape, as seen in DNA. A complication in the study of chiral molecules is that the presence of more than one separate chiral motif can (but not always) lead to the total molecule becoming non-chiral (achiral) overall. For example, it is possible for a molecule containing two centres of chirality to become achiral due to the presence of a generated internal plane of symmetry or a S_n axis within the overall structure. These effects complicate the detection of chirality in molecules for both humans and expert systems.

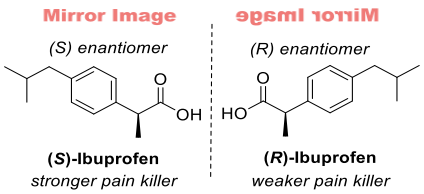


Figure 1. (*S*) and (*R*) forms of Ibuprofen.

[a] *Computer Vision Laboratory, School of Computer Science, University of Nottingham, Nottingham, NG81BB, UK.*
**e-mail: yong.kok@nottingham.ac.uk*

[b] *GlaxoSmithKline Carbon Neutral Laboratories for Sustainable Chemistry, University of Nottingham, Nottingham, NG81BB, UK.*
**e-mail: simon.woodward@nottingham.ac.uk*

[c] *Computational Optimisation and Learning Lab, School of Computer Science, University of Nottingham, Nottingham, NG81BB, UK*

[d] *B-hive Innovations Ltd, Unit 1.02, Boole Technology Centre, Beevor Street, Lincoln, LN6 7DJ, UK*

Chemists have used two-dimentional line drawings (such as those in Figures 1, 2 to represent three-dimentional chiral organic molecules for more than 150 years. Due to the increase of the diversity of chemical data and the presence of significant historical data backlogs in published scientific journals, there exists a need to systematically extract valuable information (including the presence of chiral elements) from classically drawn 2D representations 'line drawings' of organic molecules. Automation of such workflow would allow researchers access to larger amounts of data and identify potential correlations of structure to molecular effects. A particularly interesting aspect of 2D-to-3D structure analysis is to recover from 2D representations of organic molecules the presence (or absence) of 'chirality' in a given molecule. Chiral molecules are frequent high value components in modern pharmaceuticals and advanced materials, but chiral elements can be challenging to identify from 2D chemical line drawings using current image recognition approaches.

To interrogate the stereochemical information within a 2D molecule image, researchers usually follow a general pipeline of handcrafted rules or employ deep learning based algorithms to reconstruct the image into a machine-readable chemical structure or directly represent them in their molecular string representation[4–8], e.g. SMILES, InChI. However, these approaches are not always sufficient to represent the chirality in chemical compounds due to the loss of stereochemical information in the translation process. Additionally, the translated string molecular representations (SMILES, InChI) are not presently able to fully define molecular chirality in some cases. Presently, these nomenclatures can struggle to represent effectively axial, planar and helical chirality, thus limiting accurate chiral recognition/representation in many cases. As far as we are aware, there is only previous directly related study[9] where the automated detection and classification of chiral molecules is attempted, but the focus there is on supramolecular assemblies imaged under a scanning probe microscopy and the classification is only limited to a gross overall "left" or "right-handed" configuration.

In this work, we demonstrate a deep learning framework that can predict the overall molecule chirality and the chirality elements present in 2D line drawings typical in the chemical literature. Through the study of activation heatmaps, we show the potential of deep learning networks to make predictions that align with human understanding of chirality, thereby providing significant insights for both chemists and non-experts to quickly localise chirality element(s) in molecule images. Finally, our study proposes a new Formulated Imbalanced Dataset Sampler (FIDS) to deal with the label combination imbalance problem that is common in the classification of the molecules containing multiple types of chirality motif.

2 Methodology

We defined two separate classification tasks: 1) to identify if a given 2D line drawing is chiral, or not (binary classification), and 2) to identify the type of chirality elements (motifs) present in the molecules line drawing (multilabel classification).

2.1 Dataset

To the best of our knowledge, there is no readily publicly available dataset with unambigous chirality annotated to the associated molecule diagrams (line structures) from chemical literature across all chiral motifs. Hence, the dataset used in

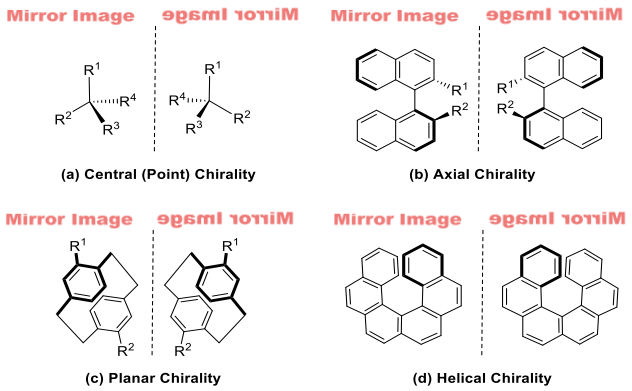


Figure 2. Different types of molecular chirality.

this study, called by us "CHIRAL", was manually curated from various chemical literature sources, chemical databases or those drawn using ChemDraw (the most common molecular line drawing software package used in the chemical sciences). CHIRAL consists of 1142 molecule diagrams labelled with their overall chirality and the type(s) of chirality elements present. Figure 3 shows the distribution of the dataset. To attain a better balanced dataset, we included 42 hypothetical molecules bearing underrepresented data types (Axial+Planar: 20, Centre+Axial+Planar: 22) which are presently unsynthesised. Such instances are very rare in native datasets which would not be sufficient for us to test out our methodology. Finally, molecules with helical chirality were excluded from this study as they account for less than 0.1% of the total chiral exemplars in the known chemical space. Even so, due to the imbalanced nature of the chirality elements present in these molecules, the CHIRAL dataset remained highly skewed, especially when label combinations are taken into account.

Given that Convolutional Neural Network (CNN) is a data-driven network, this limited amount of data present in the set of Figure 3 is simply not enough for us to properly train a neural network to examine the overall molecule chirality and chirality elements. Because of this we turned to ChEMBL, a database of bioactive molecules labelled with a chirality flag. The chirality flag indicates whether a molecule is a racemic mixture (a 1:1 mixture of "right" and "left" handed forms), single stereoisomer (enantiomer) or an achiral molecule. We reclassified all racemic mixtures and single stereoisomer under the chiral class. To maximise the generalisability of the network at the end of training, we referred ChEMBL as a source of input molecular string representations and used a range of structure diagram generators including RDKit[10], Indigo[11] and OpenEye[12] to generate a wider variety of molecular image representations that mimic the various styles portrayed in chemical literature. We also made use of UniChem, a chemical structure cross-referencing system to query for more molecular images from different databases. The resulting dataset is named ChEMBL+ whose inclusion flowchart is shown in Figure 4. The principal approach reported here is to attain a deep learning network using the large amount of dataset from ChEMBL+ to learn the weights leading to accurate binary classification of molecular chirality and then to further finesse the transfer learning in both the

binary and multilabel classification problem using our dataset, CHIRAL. To avoid overweighting large, inherently chiral, biomolecules of low symmetry, only molecules with less than or equal to 100 SMILES characters were included in the ChEMBL+ dataset. This is justified given that the majority of the molecules in the CHIRAL dataset are also small (<150 atoms, <100 SMILES characters) and this range encompasses the typical size of chiral building blocks used by chemists and thus these constitute the major portion (91%) of the ChEMBL+ database.

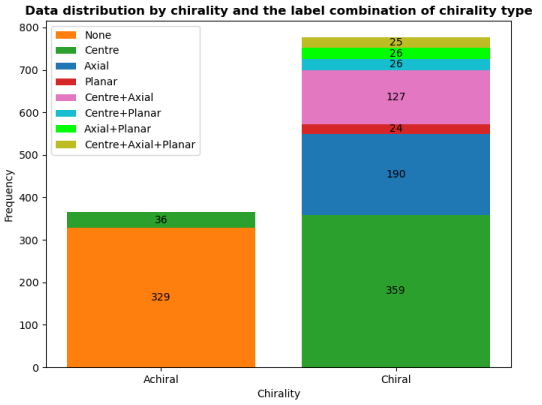


Figure 3. Data distribution of CHIRAL by chirality and the label combination of chirality type.

2.2 Preprocessing

Considering that both the ChEMBL+ and CHIRAL datasets were made up from a combination of primary data sources, some might contain an image background that is not desirable, for example a transparent background. To ensure consistency across the dataset, an automated inspection is performed to check for background transparency and, if necessary, a white background is added automatically. All images were then converted to grayscale since the colour highlights on atoms and bonds can vary across the different sources of primary dataset and they provide no useful information for chirality classification. Finally, all images were resized into a shape of 320x320 to match the model input size. All image data were then normalised based on the mean and standard deviation calculated on the ChEMBL+ dataset.

2.3 Network

Our final solution is a CNN pretrained on ChEMBL+ for a binary task of classifying chirality and then fine-tuned on the CHIRAL dataset for binary and multilabel classification. We compared both the EfficientNetV2-M[13] and ResNet50[14] models for our binary and multilabel classification problem. EfficientNetV2 was selected as it usually provides comparable accuracy to the state-of-the-art models. That is, it is built upon a mix of training-aware neural architecture search and supports scaling for the optimisation of training speed and parameter efficiency. Whereas ResNet is a deep residual network that uses skip connections to avoid the vanishing gradient problem and is considered as one of the most powerful backbone models for many computer vision tasks.

In the binary problem of classifying if a 2D structural representation is chiral or not, we simply replace the final classification layer and added a softmax layer at the end of the network to return the confidence score of the predictions.

In the case of the multilabel problem of classifying the chirality type, the classification layer is replaced by a combination of 2 blocks of Linear + ReLu + Batch Normalisation + Dropout(0.5) and a final Linear layer. We note that while using the deep architecture, EfficientNetV2, the batch normalisation layers do not improve the performance but lead to increased consumption of memory and time, thus they are excluded from the classification layer of EfficientNetV2. An additional sigmoid layer is then attached at the end of the network to return independent probabilities for each label

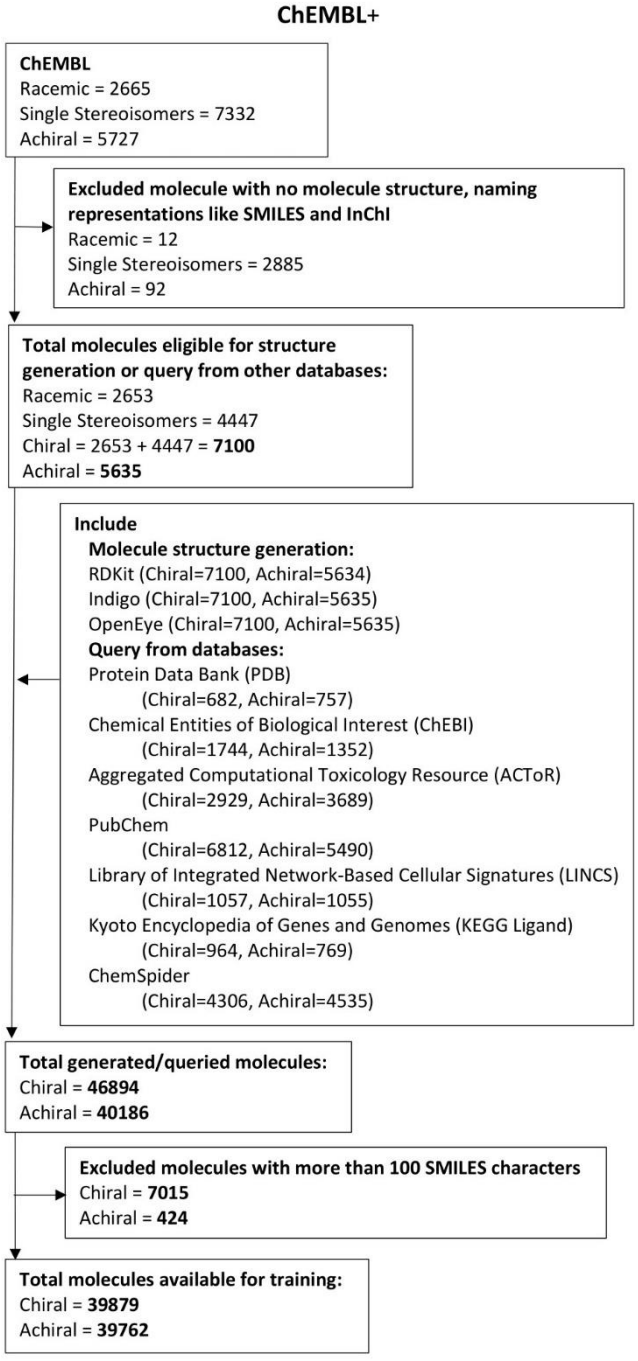


Figure 4. ChEMBL+ inclusion flowchart.

Table 1. Binary classification on CHIRAL data set: 10-fold cross-validation results.

Model	Accuracy	Precision	Recall	F1-score
ResNet50 (Partially Frozen)	**0.897±0.024**	**0.933±0.019**	**0.914±0.032**	**0.923±0.019**
EfficientNetV2-M (Partially Frozen)	0.773±0.032	0.811±0.034	0.873±0.040	0.840±0.022
ResNet50 (Fine-tune all)	0.865±0.020	0.902±0.020	0.900±0.022	0.901±0.015
EfficientNetV2-M (Fine-tune all)	0.877±0.034	0.909±0.035	0.911±0.028	0.910±0.025

because the outputs are non-exclusive. The molecule image is classified into labels in which the prediction probabilities are more than 50%.

Additionally, to visualise our network values when making predictions, we applied Grad-CAM++[15] to produce a heatmap that can provide highlights in the important area(s) that drive the network outputs.

2.4 Formulated Imbalanced Dataset Sampler (FIDS)

The Formulated Imbalanced Dataset Sampler (FIDS) is designed to address the extreme imbalance between the label combinations in the multilabel classification task. We introduce the FIDS approach starting from the Imbalanced Dataset Sampler[16]. The general idea of the Imbalanced Dataset Sampler is to rebalance the class distributions when sampling from the imbalanced dataset and this is often used in conjunction with data augmentation techniques to mitigate overfitting. While the method allows more minority class instances to be drawn to rebalance the class distribution, the instances of the majority class are less likely to be sampled and this can lead to the loss of information.

Thus, we proposed FIDS to sample a formulated amount of instances from the minority label combinations on top of the original dataset. This method intends to retain the information of the original dataset as well as dealing with the imbalanced problem through the addition of a formulated amount of samples from the minority label combinations. The method works by calculating a regulated number of images that can be sampled per instance (RNPI) (eq. 2) for the label combination that is considered as minority based on eq. 1. Using the RNPI calculation, we sampled the instances from the minority label combination accordingly and added them on top of the original dataset.

$$M = \{\, x \in labelCombinations \mid percentageInDataset(x) < minPercent \,\} \quad (1)$$

$$RNPI = \min_{m \in M}(NPI(m), maxI) \quad where: \quad (2)$$

$$NPI(m) = \frac{N \cdot addPercent}{n_m},$$

$maxI$ – maximum images that can be added per instance,

$addPercent$ – potential percentage of the minority label combination to be added

The minPercent, maxI and addPercent are tuneable parameters that need to be optimised based on the dataset used.

3 Experimental Design

3.1 Performance Measure

The evaluation metrics used for the binary task are accuracy, precision, recall and F1 score (eq. 3-6) while the multilabel task employs a strict metric, Exact Match Ratio (EMR). EMR is an extension of the accuracy metric for the single-label classification problem to a multilabel classification problem. EMR in eq. 7 can be defined as the percentage of instances that have all their labels classified correctly. This strict metric is particularly suitable for our problem context because we only have very few labels, i.e. (none/centre/axial/planar), and it is important that the partially correct labels are not taken into account.

$$Accuracy = \frac{TP}{\text{total number of instances}} \quad (3)$$

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (6)$$

where *TP* is true positives, *FP* is false positives and *FN* is false negatives.

$$EMR = \frac{1}{n}\sum_{i=1}^{n} I(g_i == p_i) \quad (7)$$

where *I* is the indicator function that returns 1 when condition is true and 0 if otherwise, g_i is the ground truth labels for ith training example and p_i is the predicted labels for ith training example.

3.2 Model Training

All the models are implemented using the Pytorch framework and the network backbones are originated from PyTorch Image Models[17].

Pretraining. We pretrained both the EfficientNetV2-M and ResNet50 models on the ChEMBL+ dataset with cross-entropy loss using Stochastic Gradient Descent optimizer at an initial learning rate of 0.5. The EfficientNetV2 and ResNet50 models were trained for 36 and 83 epochs respectively using the cosine annealing scheduler with a batch size of 32. The epochs for the pretraining were selected based on the model's performance on the CHIRAL dataset.

Fine-tuning. For both classification tasks, we fine-tuned the top few layers of the pretrained network on the CHIRAL dataset. To examine if we can further boost the accuracy, we fine-tuned the entire architecture by initialising the weights with that from the pretrained network.

Binary classification. In all settings, we performed a stratified 10-fold cross-validation and applied cross-entropy loss using Adam optimizer at a learning rate of 0.001 modulated by a cosine annealing scheduler. We trained both the EfficientNetV2-M and ResNet50 models for 16, 17, 18, 19 or 20 epochs with a batch size of 8, 16 or 32 (we were unable to train EfficientNetV2 with batch size 32 due to resource limitation). In the "partially frozen pretrained network", we fine-tuned the top few layers of both networks. Whereas in the "fine-tune all networks", we retrain all layers of the pretrained networks. The best results of the models are reported in Table 1.

Table 2. Comparison between 10-fold cross-validation results of EfficientNetV2-M and ResNet50 for multilabel classification on CHIRAL data set.

Model	Exact Match Ratio
ResNet50 (Partially Frozen)	**0.836±0.039**
EfficientNetV2-M (Partially Frozen)	0.466±0.047
ResNet50 (Fine-tune all)	0.832±0.029
EfficientNetV2-M (Fine-tune all)	0.658±0.036

Multilabel classification. We performed a multilabel stratified 10-fold cross-validation and implemented binary cross-entropy with logits loss using Adam Optimizer at a learning rate of 0.001 with a cosine annealing scheduler for all experiments. Data augmentation techniques such as rotation and translation are applied on the training set, but rotation was found to be non-improving, thus is omitted from our experiments. To select the appropriate network architecture, we start with the performance comparison between EfficientNetV2-M and ResNet50. We fine-tuned the top few layers of both networks and also retrain all layers for 20 epochs with a batch size of 16. Based on the results in Table 2, we decided to proceed to optimise ResNet50 for 16, 17, 18, 19, 20 epochs with batch size of 16 and 32. Using the best combination of epoch 20 and batch size 32, we applied FIDS to sample the dataset for training. ResNet50 was optimised for minPercent of 10% and 15%, addPercent of 10%, 15%, 20%, 25%, 30%, 35%, 40% and maxI of 16, 18, 20, 22. Note that when minPercent is 10%, Planar, Centre+Planar, Axial+Planar, Centre+Axial+Planar are considered as minority label combinations and when minPercent is 15%, Centre+Axial label combination is added to the minority list. The maxI value is limited to 22 because any value higher than that has the same result when used with the previous ranges of minPercent and addPercent. For comparison, the performances of the baseline models (ResNet50(Fine-tune all and Partially frozen) without the use of FIDS) were also evaluated.

4 Results and Discussion

4.1 Binary classification

Table 1 shows that the partially frozen ResNet50 has the best overall results in the 10-fold cross-validation of our binary problem. Even though this is an extremely imbalanced dataset, the model can achieve high precision and high recall, making a good separation between chiral and achiral molecules, The performance of EfficientNetV2

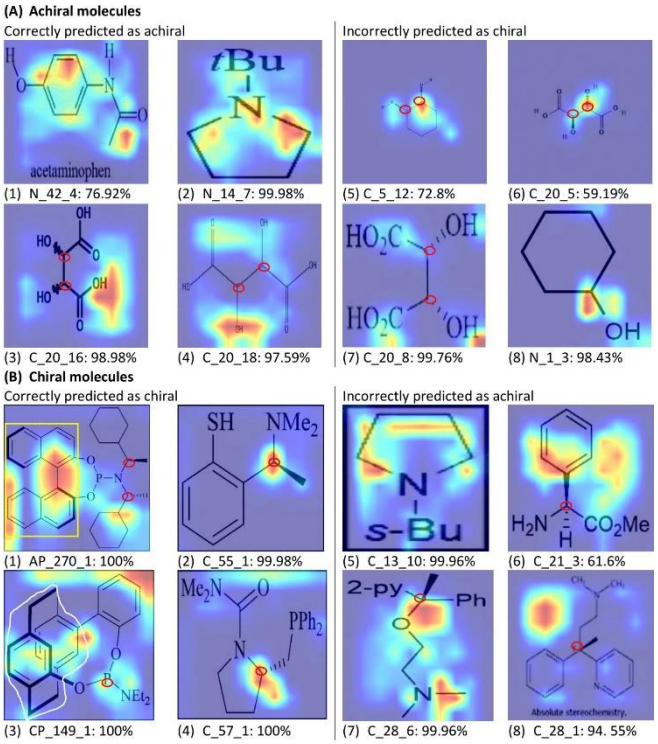


Figure 5. Binary classification using ResNet50 (Partially frozen): Heatmap visualisation when looking at predicted target. The percentage shown under each heatmap is the confidence score of the network prediction; The actual chiral centres are circled in red; The yellow rectangle shows the region of the actual axial chirality; The white line is drawn around the actual planar chirality.

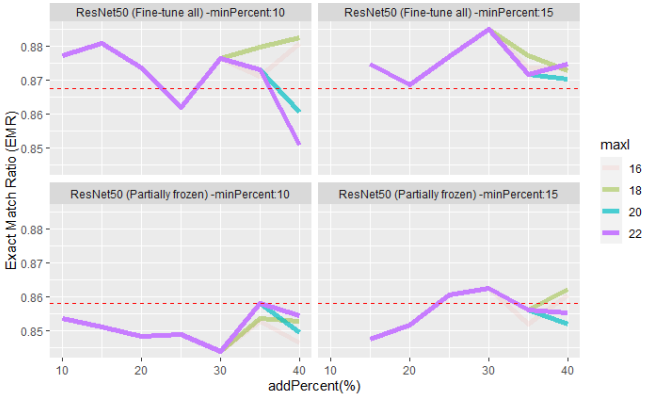


Figure 6. Multilabel classification on CHIRAL data set: 10-fold cross-validation results while changing minPercent, addPercent and maxI on FIDS. The red dashed line is the exact match ratio of the baseline i.e. without the use of FIDS.

can be further improved when we retrain all the layers, but it is still not better than the partially frozen ResNet50, implying that increasing complexity in architecture does not necessarily yield better performance in chirality classification. On the other hand, fine-tuning all layers of ResNet50 led to a slight drop in performance. A possible explanation is that fine-tuning all network layers caused the network to be more susceptible to overfitting, especially if the dataset is small.

We observed that our model, ResNet50 (Partially frozen) was able to detect potential chiral centres, *but also*, axial and planar chirality as regions of interests when predicting them as chiral (Figure 5.B.1-4). This is particularly interesting because the network was only trained with a simple chiral or achiral label without any

Table 3. Multilabel classification on CHIRAL data set: 10-fold cross-validation results of the baseline (ResNet50 (Fine-tune all)), FIDS(1st) and FIDS(2nd). Second row shows the performance of the algorithm as a whole. Third row onwards show the performance per label combination.

Chirality	Baseline	FIDS(1st), mean± SD (%increase[a])	FIDS(2nd), mean± SD (%increase[a])
Whole	0.868±0.023	0.885±0.031 **(+2.01)**	0.882±0.023 **(+1.33)**
None	0.909±0.062	0.912±0.052 **(+0.34)**	0.921±0.039 **(+2.94)**
Centre	0.883±0.061	0.899±0.063 **(+1.73)**	0.909±0.040 **(+2.94)**
Axial	0.858±0.102	0.846±0.090 (-1.46)	0.863±0.111 **(+0.52)**
Planar	0.609±0.293	0.913±0.127 **(+49.84)**	0.817±0.230 **(+34.04)**
Centre+Axial	0.852±0.133	0.898±0.071 **(+5.43)**	0.790±0.076 (-7.29)
Centre+Planar	0.867±0.176	0.817±0.200 (-5.77)	0.867±0.176 (0)
Axial+Planar	0.708±0.306	0.817±0.320 **(+15.30)**	0.892±0.175 **(+25.89)**
Center+Axial+Planar	0.674±0.336	0.696±0.301 **(+3.29)**	0.835±0.194 **(+23.90)**

[a] Percentage increase with respect to baseline model

additional annotation that shows the location of the chirality elements.

Our model encountered challenges with certain types of 2D image inputs. For example, when human chemists have indicated an ambiguous stereochemistry (one chemical might be 'up' on 'down' from the plane of the 2D line drawing) by a wavy line (〰). There were too few examples within our present dataset for accurate learning of this meaning. The present algorithm, although excellent at recognising individual chiral centres, is presently insufficiently trained to recognise that the presence of two chiral elements can still result in the overall molecule being achiral (e.g. meso compounds). These effects can be seen in heatmaps Figure 5.A.3,4 (correct prediction but the algorithm failed to recognise the potential for chirality in these molecules) and in heatmaps Figure 5.A.5-7 (where it was unable to realise that such molecules possess such internal reflection planes making the overall structure achiral).

4.2 Multilabel classification

Consistent with the binary classification results, the ResNet50 architecture performed significantly better than EfficientNetV2-M for both configurations in Table 2. This further confirms the observation made in CheXtransfer[18] and HistoTransfer[19] i.e. performance gain of network generated through neural architecture search optimised for ImageNet may not be transferable to dataset in other domains. That being so, we continued to optimise ResNet50 for FIDS. In contrast to our binary classification results, ResNet50 (Fine-tune all) has better overall results compared to ResNet50 (Partially frozen) based on Figure 6. We assume that this is due to the greater need to tune a pretrained network for a more complex problem. However, we also note that these results were heavily dependent on the amount of data and can vary. In this problem context, the best EMR (FIDS(1st)) was obtained when minPercent is 15, addPercent is 30 and maxI is 16, 18, 20, or 22 (see Figure 6: all has the same best result, maxI greater than 16 no longer has any effect). The next best EMR (FIDS(2nd)) was obtained when the minPercent is 10, addPercent is 40 and maxI is 18. Although FIDS(1st) had a slightly better EMR as a whole, but FIDS(2nd) had a more balanced performance increase across the label combinations (Table 3).

Our proposed method was able to significantly increase the EMR for most of the minority label combinations (Table 3). This implies that with a properly formulated amount of sampling from the minority label combinations, the performances of the minority label combinations can be improved without affecting the accuracies of the majority label combinations. In fact, our network was capable to view the actual chirality elements as regions of interest when predicting the corresponding chirality type (Figure 7). Again, no extra annotation about the chirality location was given when the model was trained.

Our network was more likely to fail to predict molecules with axial or planar chirality due to the lack of variation in these types of labels. Most molecules with planar chirality simply contain ferrocene (two pentagon rings) in an upright position as in Figures 7.D.1, F.1, F.2, G.1, G.2. As a result, the network was more susceptible to failure when a molecule with the rare shape of planar chirality showed up (Figure 7.D.2, H.2). On the other hand, molecules with axial chirality often appears containing 1,1'-binaphthyl units (4 closely connected hexagons, see Figure 2.b, 7.C.1, C.2, E.1). Hence, the network can find it difficult to recognise the allene arrangements of axial chirality as in Figure 7.G.2, H.1. To deal with these rare chirality appearances, a possible solution is that more molecules with such appearances can be purposely created to be included in the training set for further performance improvement.

5 Conclusion and Future work

In this paper, we conclude that the smaller model, ResNet50 tends to perform better and its performance is more transferable to the molecule dataset, unlike the EfficientNetV2 architecture. We studied the performances of retraining the top few layers and all layers of pretrained network for both binary and multilabel classification of the CHIRAL dataset, concluding that fine-tuning the top few layers of ResNet50 is decent enough for the binary problem whereas the more complex multilabel problem would require tuning for all layers to achieve the optimal result. Using our proposed sampling method--FIDS in the multilabel classification problem, we demonstrated a significant performance increase for most of the minority

label combinations and achieved slight overall performance improvement from 87% to 89%. Despite that, future work can focus on improving the FIDS algorithm as it is currently very dependent on the amount of data.

Furthermore, our models were also able to make predictions that were in line with the actual chiral regions i.e. locating the chirality elements within the 2D line drawing when predicting the corresponding chirality type. Now that we have evidence on the ability of the deep learning network to identify the correct location of chirality elements, future work can focus on using dataset annotated with the location of chirality in order for the network to predict the precise chirality element location. That said, we will continue collecting more molecule images with their chirality information through our demo website (https://chiral.cs.nott.ac.uk) and the Zooniverse platform (https://www.zooniverse.org/projects/shuxiang/chiral-molecules) to enrich the existing data. Our demo website also allows users to upload their molecule image reporting the recognised overall chirality and chiral elements in that molecule including a heatmap visualisation of the potential locations for those chiral elements.

Lastly, we acknowledge that our study has several limitations. We selected the pretrained models based only on their binary task performance on the CHIRAL dataset. An additional complication is that a few molecules in the CHIRAL dataset are also in the ChEMBL+ dataset, but we must stress that no identical molecule representations are used. These factors might lead to potential learning biases, however, we are confident that the resulting performances are still valid since the proposed network is shown to be able to make predictions based on the detected actual chiral regions rather than memorising the image for classification.

Acknowledgement

We thank Shuxiang Hu for setting up the “Chiral Molecules” project for collecting 2D molecule images via the Zooniverse platform. We are also grateful for the support from the University of Nottingham.

Data Availability Statement

The dataset and pretrained models are available on https://zenodo.org/record/5759416 and these will be maintained by Yong En Kok. The source code is accessible on https://github.com/janetkok/Chiral while the demo website is available on https://chiral.cs.nott.ac.uk.

Conflicts of interest

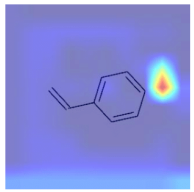
The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. M. Evans, *Clin Rheumatol* **2001**, *20*, 9–14.
[2] M. Karras, Synthesis of Enantiomerically Pure Helical Aromatics Such As NHC Ligands and Their Use in Asymmetric Catalysis, Charles University (Czech Republic), **2018**.
[3] M. Rickhaus, M. Mayor, M. Juríček, *Chemical Society Reviews* **2016**, *45*, 1542–1556.
[4] K. Rajan, H. O. Brinkhaus, A. Zielesny, C. Steinbeck, *Journal of Cheminformatics* **2020**, *12*, 60.
[5] K. Rajan, A. Zielesny, C. Steinbeck, *J Cheminform* **2020**, *12*, 65.
[6] H. Weir, K. Thompson, A. Woodward, B. Choi, A. Braun, T. J. Martínez, *Chemical Science* **2021**, *12*, 10622–10633.
[7] D.-A. Clevert, T. Le, R. Winter, F. Montanari, *Chemical Science* **2021**, *12*, 14174–14181.
[8] I. Khokhlov, L. Krasnov, M. V. Fedorov, S. Sosnin, *Chemistry–Methods* **2022**, *2*, e202100069.
[9] J. Li, M. Telychko, J. Yin, Y. Zhu, G. Li, S. Song, H. Yang, J. Li, J. Wu, J. Lu, X. Wang, *J. Am. Chem. Soc.* **2021**, *143*, 10177–10188.
[10] RDKit, **2021**, DOI 10.5281/zenodo.5242603.
[11] *Indigo Toolkit*, **2020**.
[12] *OpenEye Toolkits Dev Build OpenEye Scientific Software*, **2021**.
[13] M. Tan, Q. Le, in *Proceedings of the 38th International Conference on Machine Learning*, PMLR, **2021**, pp. 10096–10106.
[14] K. He, X. Zhang, S. Ren, J. Sun, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, **2016**, pp. 770–778.
[15] A. Chattopadhay, A. Sarkar, P. Howlader, V. N. Balasubramanian, in *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, **2018**, pp. 839–847.
[16] M. Yang, *Imbalanced Dataset Sampler*, **2021**.
[17] R. Wightman, *GitHub repository* **2019**, DOI 10.5281/zenodo.4414861.
[18] A. Ke, W. Ellsworth, O. Banerjee, A. Y. Ng, P. Rajpurkar, in *Proceedings of the Conference on Health, Inference, and Learning*, Association For Computing Machinery, New York, NY, USA, **2021**, pp. 116–124.
[19] Y. Sharmay, L. Ehsany, S. Syed, D. E. Brown, in *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI)*, **2021**, pp. 1–4.

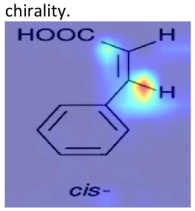
(A) Molecules with no chirality

(1) N_49_17
Correct prediction.



None: 99.96%
Centre: 0.03%; Axial: 0.01%; Planar: 0%

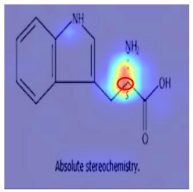
(2) N_24_5
Incorrectly predicted as of containing centre chirality.



Centre: 56.52%
None: 38.75%; Axial: 0.01%; Planar: 0.08%

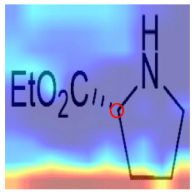
(B) Molecules with centre chirality only

(1) C_30_1
Correct prediction.



Centre: 99.89%
None: 0.04%; Axial: 0%; Planar: 0%

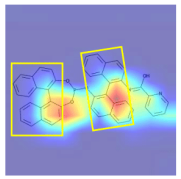
(2) C_17_2
Incorrectly predicted as of containing no chirality.



None: 96.26%
Centre: 3.9%; Axial: 0.01%; Planar: 0.02%

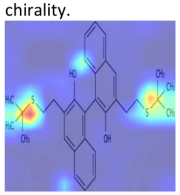
(C) Molecules with axial chirality only

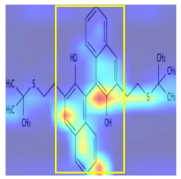
(1) A_137_12
Correct prediction.



Axial: 100%
None: 0 %; Centre: 0.42%; Planar: 0.03%

(2) A_159_13
Incorrectly predicted as of containing centre chirality.

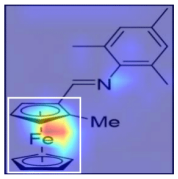


Centre: 99.47%


Axial: 57.39%
None: 0.14%; Planar: 0.27%

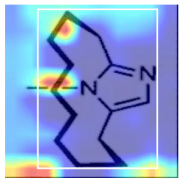
(D) Molecules with planar chirality only

(1) P_258_1
Correct prediction.



Planar: 100%
None: 0.01%; Centre: 0.04%; Axial: 0.04%

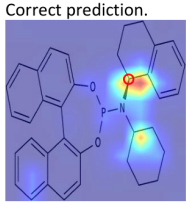
(2) P_226_1
Incorrectly predicted as of containing no chirality.

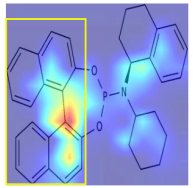


None: 75.12
Centre: 26.56%; Axial: 0.02%; Planar: 0.08%

(E) Molecules with Centre+Axial chirality

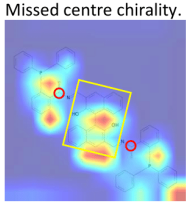
(1) CA_115_13
Correct prediction.





Centre: 95.20%
Axial: 99.96%
None: 0%; Planar: 0%

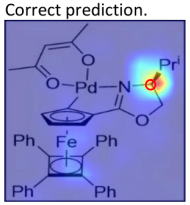
(2) CA_126_14
Missed centre chirality.

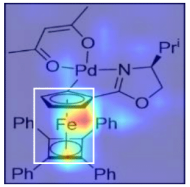


Axial: 100%
None: 0.03%; Centre: 3.24%; Planar: 0%

(F) Molecules with Centre+Planar chirality

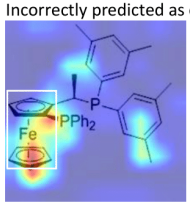
(1) CP_259_1
Correct prediction.

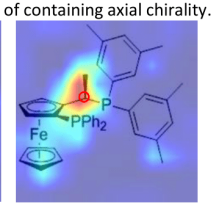


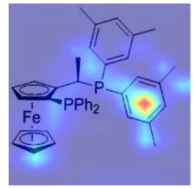


Centre: 100%
Planar: 100%
None: 0%; Axial: 31.56%

(2) CP_235_1
Incorrectly predicted as of containing axial chirality.



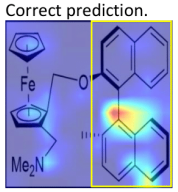


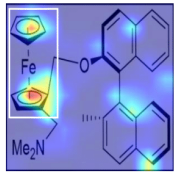


Planar: 100%
Centre: 100%
Axial: 99.99%
None: 0.07%

(G) Molecules with Axial+Planar chirality

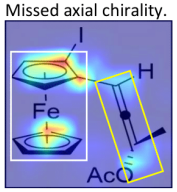
(1) AP_296_1 (hypothetical molecule)
Correct prediction.





Axial: 100%
Planar 100%
None: 0%; Centre: 0.07%

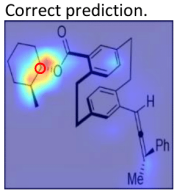
(2) AP_269_1 (hypothetical molecule)
Missed axial chirality.

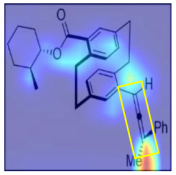


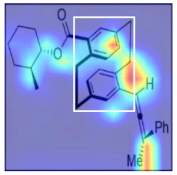
Planar: 100%
None: 0%; Centre: 0.02%; Axial: 32.22%

(H) Molecules with Centre+Axial+Planar chirality

(1) CAP_308_1 (hypothetical molecule)
Correct prediction.

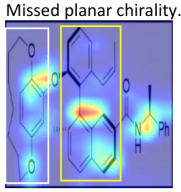


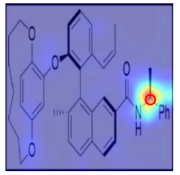




Centre: 100%
Axial: 99.99%
Planar: 99.70%
None: 0.01%

(2) CAP_306_1 (hypothetical molecule)
Missed planar chirality.





Axial: 99.99%
Centre: 99.27%
None: 0.04%; Planar: 24.58%

Figure 7. Multilabel classification using ResNet50 (Fine-tune all) with FIDS(2nd): Heatmap visualisation when looking at predicted chirality type. The percentage shown under each heatmap is the prediction probability of the corresponding chirality type; The actual chiral centres are circled in red; The yellow rectangle shows the region of the actual axial chirality; The white rectangle shows the actual planar chirality.

molecular informatics
Received: 30th March 2022
Accepted: 6th June 2022
Published online: 6th June 2022