Structure-to-Shift Prediction for ^{19}F NMR: Comparison of Graph Neural Networks, HOSE Code, and Large Language Models

Nathan O'Neill[1], Adam P Jones[1], Katharina T Huber[1], E Kate Kemsley[1,3], Carlos Cobas[2,3], David Williamson[3], Gary Sharman[3], John Hollerton[3]

1.University of East Anglia (UEA), Norwich Research Park, NR4 7TJ, Norwich, UK
2.Centro de Investigación Mestrelab (CIM), Av. de Barcelona, 7, 15706, Santiago de Compostela, Spain
3.SciY-Mestrelab, Av. de Barcelona, 7, 15706, Santiago de Compostela, Spain

Motivation and Methods

Fluorine (^{19}F) NMR is useful for structural verification of fluorinated compounds, important in the pharmaceutical and agrichemical industries. This study compared different approaches for predicting ^{19}F chemical shifts directly from molecular structure:

Graph Convolutional Neural Networks (GCNNs)

Trained on a proprietary database of approximately 15,000 structures annotated with groundtruth ^{19}F chemical shifts; uses a 'message-passing' network architecture.

Mnova's ModGraph-NMRPredict (SciY-Mestrelab)

Commercial tool that uses the same underlying database as the GCNN approach but applies empirical HOSE-code prediction.

Large Language Models (LLMs) from Anthropic and OpenAI

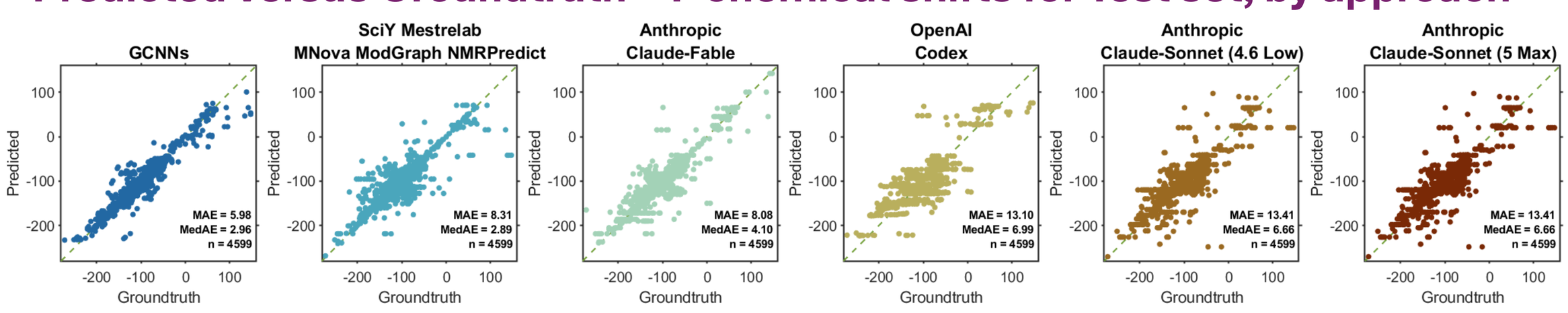
Prompted to design, develop and execute their own zero-shot prediction workflows, based on molecular environment analysis and chemical reasoning.

Test Dataset

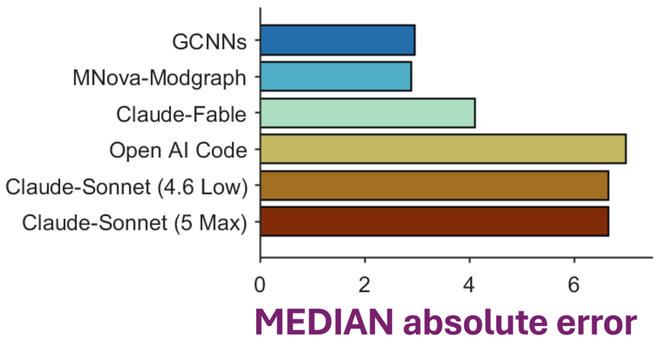
A test collection was sourced from the public *nmrshiftdb2* database, comprising 795 diverse fluorinated structures annotated with ^{19}F chemical shifts.

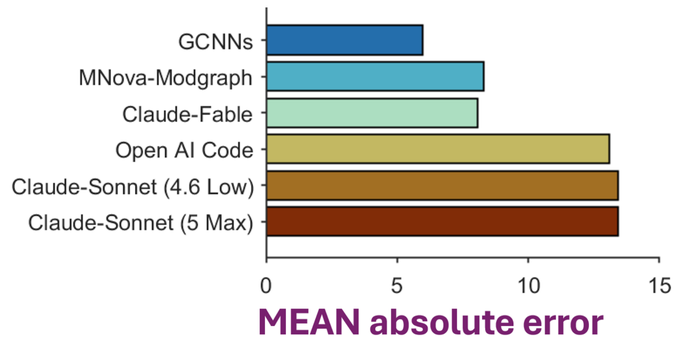
None of these were present in the GCNN/ModGraph training set. The LLMs were specifically instructed to refrain from search and retrieval. Thus, all methods were benchmarked exclusively on unseen compounds.

Predicted versus Groundtruth ^{19}F chemical shifts for Test set, by approach

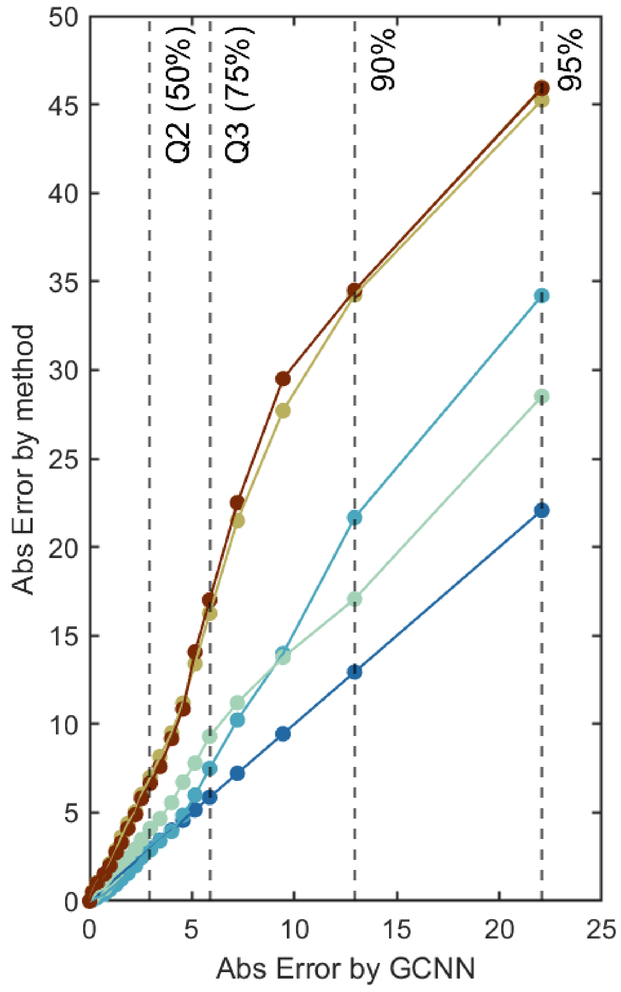


- GCNNs and ModGraph achieved similar **MEDIAN Absolute Errors** (2.96 and 2.89 ppm respectively), both outperforming the LLMs.
- GCNNs provided the best overall accuracy and generalisation to unseen compounds and achieved the lowest **MEAN Absolute Error** (5.98 ppm), >25% lower than any other approach.





Q-Q plots



Shows percentiles of each absolute error distribution plotted against that of the GCNNs.

This highlights a key advantage of the GCNN approach: it produces far fewer large prediction errors than the competing methods.

Conclusion

Although modern LLMs demonstrate impressive chemical reasoning capabilities, for quantitative ^{19}F chemical shift prediction they currently do not match the performance of dedicated NMR models trained on high-quality experimental data.

Acknowledgments

C. Cobas thanks the creation of the Mestrelab Research Center (CIM) subsidized by the Axencia Galega de Innovación. Operation funded by the Xunta de Galicia, through the business aid program for the creation and integration of new business research centers 001_IN853D_2022.

A. Jones & UEA co-authors thank SciY Mestrelab for supporting the ongoing Ph.D. project "A.I. and Graph Theoretical Approaches to Small Molecule Structure Elucidation"