Machine learning approach to improve drug discovery in natural product extracts

by
Robert T. Heussner

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Chemical Engineering
(Honors Associate)

Presented August 16, 2022
Commencement June 2023

AN ABSTRACT OF THE THESIS OF

Robert T. Heussner for the degree of <u>Honors Baccalaureate of Science in Chemical Engineering</u> presented on August 16, 2022. Title: <u>Machine learning approach to improve drug discovery in natural product extracts</u>.

Abstract approved:_____

Kevin Brown

Current natural product (NP) research is limited by its reliance on bioassay-guided fractionation to identify bioactive compounds in mixtures. Computational approaches may improve NP research by correlating mass spectra (MS) and nuclear magnetic resonance spectra (NMR) of bioactive mixtures to their bioactivity patterns. In this study, I use artificial neural networks (ANNs) to correlate both MS and NMR data of 40 fractions of hops (*Humulus lupulus*) extract to inhibition of iNOS-mediated formation of nitric oxide (provided by the Stevens Lab at Oregon State University). Xanthohumol and its derivatives, constituents of hops extract, are known to exhibit this anti-inflammatory activity. An MS-based model, NMR-based model, a model that concatenates MS and NMR as a single input, and model that treats them as separate inputs were investigated. The MS-based model predicted bioactivity with lowest error (MSE = 0.685) and identified xanthohumol as the top anti-inflammatory compound. The NMR-based model, concatenated model, and multichannel model predicted bioactivity with higher error: MSE = 9.48, 8.05, and 7.58, respectively, but they identified several known bioactive molecules and associated proton shifts as top predictors. In conclusion, ANNs have been shown to usefully predict bioactivity from MS/NMR data.

Keywords: Machine learning, artificial neural networks, mass spectra, nuclear magnetic resonance spectra, bioactivity, natural products, *Humulus lupulus*, xanthohumol

Corresponding e-mail address: robheussner@gmail.com

Machine learning approach to improve drug discovery in natural product extracts

by
Robert T. Heussner

A THESIS

submitted to

Oregon State University

Honors College

in partial fulfillment of
the requirements for the
degree of

Honors Baccalaureate of Science in Chemical Engineering
(Honors Associate)

Presented August 16, 2022
Commencement June 2023

<u>Honors Baccalaureate of Science in Chemical Engineering</u> project of Robert T. Heussner presented on August 16, 2022.

APPROVED:

_____

Kevin Brown, Mentor, representing Chemical, Biological, and Environmental Engineering and Pharmaceutical Sciences

_____

Adam Higgins, Committee Member, representing Chemical, Biological, and Environmental Engineering

_____

Paul Ha-Yeon Cheong, Committee Member, representing Chemistry

_____

Toni Doolen, Dean, Oregon State University Honors College

I understand that my project will become part of the permanent collection of Oregon State University, Honors College. My signature below authorizes release of my project to any reader upon request.

_____
Robert T. Heussner, Author

Acknowledgements

I would like to thank Kevin Brown for his invaluable mentorship of my thesis. I would not have been able to complete this work without his expertise, guidance, and patience over the last two years. I would like to thank Fred Stevens and Gisela Gonzalez-Montiel for the annotated MS and NMR data that made this work possible. I would like to thank Adam Higgins for helping me develop as a scientist. Finally, I would like to thank my close friends and family for their unwavering support throughout this process.

1. Introduction

Natural product (NP) research is a pillar of the pharmaceutical industry. Between 1981 and 2006, nearly 50% of new drugs were derived from NPs [1]. NPs are any compounds that have natural origins, but most often are small secondary metabolites of plants. Traditionally, NP drug discovery is carried out through bioassay-guided extract fractionation. First, NPs are extracted from the plant material with solvent and fractionated by mass. Each fraction is subjected to a bioassay that screens for the desired bioactivity. The active fractions are fractionated repeatedly until the bioactive compounds are isolated. The active fractions are also analyzed using UV, IR, MS, NMR, or other techniques to dereplicate the compounds [2]. However, this approach is expensive, time-consuming, and inefficient. Many fractionations, bioassays, and chemical analyses are required before bioactive compounds are discovered, and late dereplication causes high rediscovery rates.

Machine learning (ML) with artificial neural networks (ANNs) may improve NP drug discovery by correlating MS and NMR data of the 1st round extract fractions to their bioactivities. Feature importance methods, which quantify the importance of each feature to a model's prediction, can be used to determine the individual masses (features) in the NP mixture that are bioactive. This would negate the need for additional fractionation and provide early dereplication. Dereplication would further improve if the model leveraged NMR shifts to match local chemical environments that are present in the bioactive molecules. Partial least-squares analysis (PLS) has previously been used for making these predictions, with some success [3, 4]. However, a recent study showed that many MS/bioactivity studies (specifically hops extracts) have properties that invalidate this basic approach; a more complex modeling pipeline is needed [5].

In this paper, I use an ANN-based ML pipeline to correlate MS and NMR data from 40 lupulin extract fractions to anti-inflammatory bioactivity to predict which lupulin constituents and associated NMR shifts are responsible for the anti-inflammatory effects of the mixture. Lupulin is an extracellular resin of hops (*Humulus lupulus*), and contains xanthohumol, a prenylflavonoid that has antioxidant [6], anti-inflammatory [7, 8], antimicrobial [9], and immune modulatory activity [10]. Structural analogs of xanthohumol that are present in the mixture, such as xanthohumol B, have been shown to exert anti-inflammatory activity [11]. Lupulin also contains 8-prenylnaringenin, an estrogenic metabolite of xanthohumol, which has been hypothesized to contribute to xanthohumol's health benefits [12-13]. A novel ML pipeline with Elastic Net and Random Forests models was shown to identify xanthohumol and other previously studied compounds as responsible for anti-inflammatory activity in a lupulin extract from MS data, as well as identifying a previously unknown inhibitor of this activity [5]. However, this study used only MS data in the ML pipeline and neglected 1D proton NMR measurements.

2. Methods

2.1. Data

The data was provided by the Stevens lab at Oregon State University. LC-QTOF-MS/MS (Liquid chromatography quadrupole time-of-flight) and 1D proton NMR spectra were collected for 40 fractions of hops *(Humulus lupulus)* extract, and an anti-inflammatory cell assay was used to measure inhibition of iNOS-mediated formation of nitric oxide. After peak picking, the MS spectra had 39 masses and NMR spectra had 667 shifts, and were $log_2(1 + i)$ normalized before being fed into the ML models. Similarly, the bioactivity data was $log_2(1 + i)$ normalized before use.

2.2. Artificial neural networks

Three multilayer feed-forward neural networks and one multichannel network with a multilayer feed-forward branch and a 1D convolutional branch were investigated using Keras/Tensorflow in Python. The first model used MS data as the input, the second used NMR, the third used a concatenated array of the MS and NMR data, and the multichannel model used the MS and NMR data as inputs to separate channels. The architecture of each model is shown in Fig. 1.
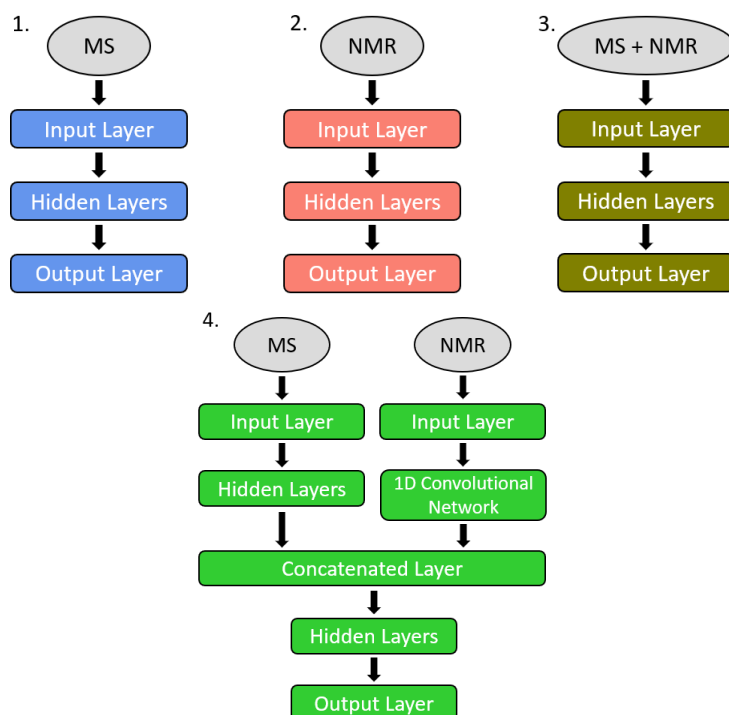
Figure 1. MS-based (1), NMR-based (2), concatenated MS/NMR-based (3), and multichannel-based model (4) diagrams.

In general, the models consisted of an input layer, 1-3 dense hidden layers, and a single node output layer. The hidden layer nodes had rectified linear activation functions and the models were compiled with the Adam optimizer and the mean squared error (MSE) loss function. For training, the batch size was 2 with 300 epochs and early stopping (patience = 4).

The models will be referenced by listing the input data type (MS, NMR, CONC for the concatenated input, or MC for the multichannel input) followed by the hidden layer sizes in sequential order, or for MC the number of filters in the convolutional layer. For example, an MS-based model with two hidden layers, the first with 10 nodes and the second with 5 nodes, would be referred to as MS-10-5. Likewise, MC-10 references the multichannel model with 10 filters.

The MS model was topologically optimized by changing the layer size and network depth, with a total of nine variations. The best model was chosen based on feature importance (see 2.3) and prediction error. The NMR and CONC models used the optimized MS topology, which was found to be a single hidden layer of 10 nodes (see Results). In the MC models, the MS channel was MS-10-5, and the NMR channel was a 1D convolutional neural network with one convolutional layer with 10 filters of size (1 x 3), a max pooling

layer with of pool size of 2, a flatten layer, and a 5-node dense layer. The outputs of the MS and NMR branches were concatenated, serving as the input to two dense hidden layers (10-5) and a single node output layer.

2.3. Feature importance

The connection weight approach (CWA) was used to determine feature importance for the MS, NMR, and CONC models, and permutation importance (PM) was used for the MC model. CWA calculates the dot product of the input-hidden layer weight matrix, hidden-hidden layer weight matrices, and hidden-output layer weight matrix [17, 18]. PM measures the difference in prediction error between a model with the feature of interest randomly shuffled and a model with the feature in its original order. A larger change in error corresponds to a greater importance.

2.4. Ensemble approach

An ensemble approach was used to mitigate overfitting due to the limited amount of data and its dimensional imbalance. Each model was evaluated 1000 times with random 2/3-1/3 training-testing splits, and the feature importance was computed using either CWA or PM. The feature importance for each iteration, $I_i^{(k)}$, was weighted by the inverse exponent of the iteration's *prediction* error (not training error), and the sum of the weighted importances was divided by the sum of the weights to estimate the model's overall feature importance $\langle \hat{I}^{(k)} \rangle$, as shown in Eqn. 1.

$$\langle \hat{I}^{(k)} \rangle = \frac{\sum_1^n I_i^{(k)} e^{-MSE_i}}{\sum_1^n e^{-MSE_i}} \tag{1}$$

Therefore, the results of models with lower holdout data prediction error are weighted exponentially better than those of higher error models. Previous work has established the ensemble approach as a viable solution to limited, imbalanced data, particularly when bioactivity is localized to only a handful of chromatographic fractions, as is the case with anti-inflammatory effects in hops extracts [5]. In simpler cases, when bioactivity is broadly shared across the fractions, model-to-model consistency in the ensemble is higher.

3. Results and Discussion

3.1. MS-based model:

The topology optimization study, shown in Fig. 2, revealed no systematic trend between network depth nor layer size and xanthohumol ranking ("1" is the top-ranked feature). In what follows, the listed MSEs are the ensemble average prediction errors, as opposed to the training errors. MS-10-5-3 had the highest MSE of 2.82 (SD = 3.42), and MS-40-20-10 had the lowest MSE of 0.461, (SD = 0.661). MS-10 ranked xanthohumol as the most important feature, with a MSE of 0.685, (SD = 0.800). MS-10-5 ranked xanthohumol least importantly of all models, at 34th, with a MSE 1.05 (SD = 2.04). It was expected that models with lower MSE would identify more known bioactive masses as top bioactivity predictors, but MS-10 had the 6th lowest MSE. However, the difference in MSE between MS-10 and MS-40-20-10 is relatively small, with ΔMSE=0.224, and all the models performed well at predicting bioactivity.

The top eight important masses from MS-10 are listed in Table 1, and Fig. 3 shows their intensities plotted across the 40 extract fractions. Although MS-10 did not find other known bioactive xanthohumol derivatives or metabolites in its top features, the 2nd ranked m/z, 355.2/3.2 (45), peaks in intensity near xanthohumol, around the 30th fraction, and may be worth further investigation. However, other important m/z like 745.4/3.3 (129), 331.2/3.2 (32), 399.3/3.3 (79) peak around the 10th fraction, and early fractions did not exhibit significant anti-inflammatory behavior in the bioassay. Interpreting the feature importance rankings of MS-based models is challenging because there are only 39 features, so in a random draw xanthohumol has a 1/39 probability of being ranked the most important. The high variability of the feature importance rankings with network topology emphasizes the necessity of topologically optimizing ANNs. In studies with extracts whose properties are less understood, comparison to previously identified compounds is not available. Models can still be ranked according to MSE, so it is encouraging that models with low MSE found previously identified anti-inflammatory compounds.
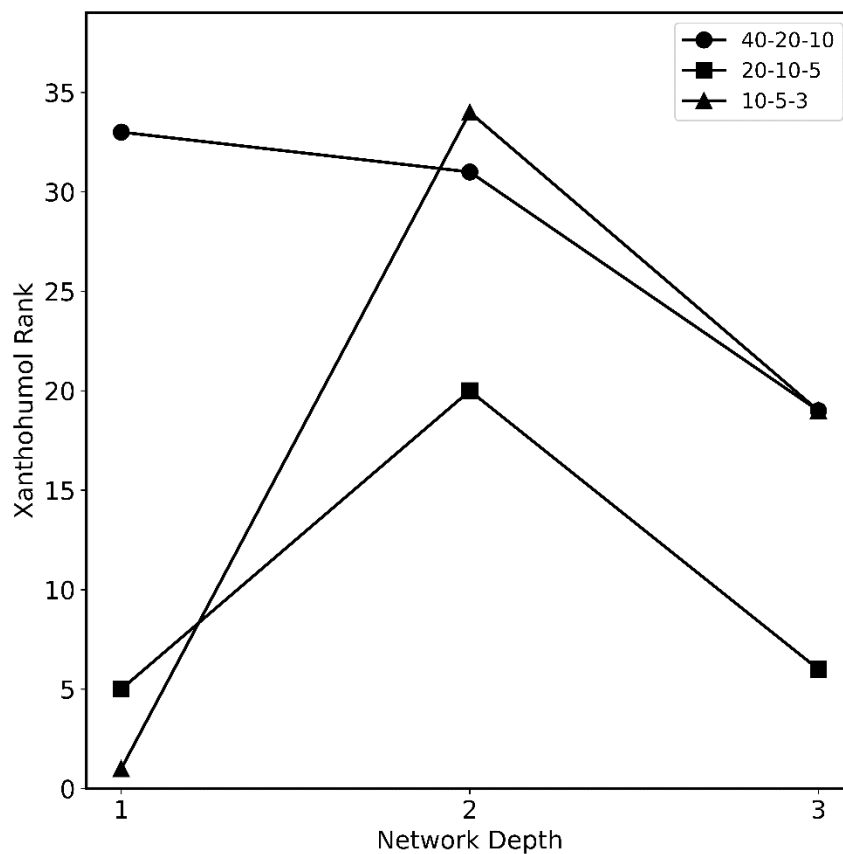
Figure 2. Network topology optimization based on xanthohumol ranking. The circles represent MS-40 (depth = 1), MS-40-20 (depth = 2), and MS-40-20-10 (depth = 3), the squares represent MS-20, MS-20-10, and MS-20-10-5, and the triangles represent MS-10, MS-10-5, and MS-10-5-3.

Table 1. The top eight masses based on CW from MS-10. Annotations provided by the Stevens Lab at Oregon State University.

Rank	m/Z	ID
1	353.1/3.2 (42)	xanthohumol
2	355.2/3.2 (45)	-
3	745.4/3.3 (129)	-
4	331.2/3.2 (32)	-
5	265.1/3.2 (11)	-
6	393.2/3.2 (76)	-
7	305.1/3.2 (20)	-
8	399.3/3.3 (79)	-
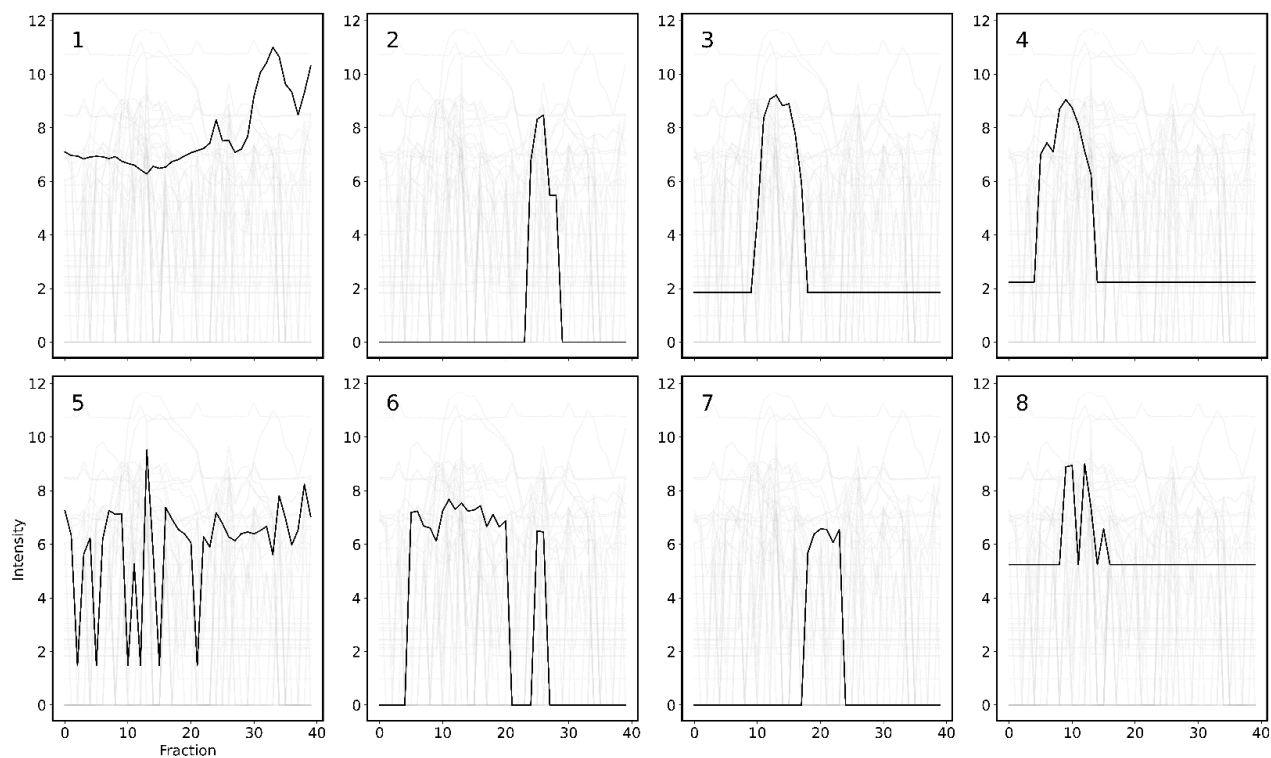
Figure 3. The $log_2(x+1)$ normalized m/z intensities plotted across the 40 extract fractions. The top-ranking masses are shown in black, and their rank is in the upper left of the plot with corresponding annotations in Table 1. The 31 remaining masses are plotted in grey.

3.2. NMR-based model:

NMR-10 predicted bioactivity worse than MS-10 with a MSE of 9.48 (SD = 14.0). NMR-10 models were capable of predicting with as low MSE as MS-10, but they occur less frequently in the 1000 model ensemble, as shown in Fig. 5. Table 2 shows that NMR-10 identified several annotated shifts in the top eight features (NMR annotations provided by Gisela Gonzalez-Montiel). The 2nd ranked shift, 0.86 ppm, is found in sitosterol-3-O-β-glucopyranoside, a sugar in lupuline. The 4th ranked shift, 2.65 ppm, is found in 6-prenylnaringenin, 8-prenylnaringenin, and isoxanthohumol, which are xanthohumol metabolites. The 8th ranked shift, 3.58 ppm, is found in xanthohumol G, a structural analog of xanthohumol. NMR-10 found more known bioactive shifts than MS-10 found masses because single NMR shifts reflect local chemical environments and thus can be found in many structurally similar molecules, whereas individual mass-to-charge ratios are unique molecular identifiers. The NMR-based models may have predicted bioactivity with higher error because the NMR-based models are less balanced ("wider" problems) [19] compared to the MS-based models, with 667 features versus 39.

Table 2. The annotated top eight shifts using CW from NMR-10. Annotations provided by Gisela Gonzalez-Montiel.

| Rank	δ	ID
1	7.89	-
2	0.86	sitosterol-3-O-β-glucopyranoside
3	1.46	-
4	2.65	6-prenylnaringenin, 8-prenylnaringenin, isoxanthohumol
5	7.03	-
6	0.24	-
7	1.43	-
8	3.58	xanthohumol G

3.3. Concatenated MS/NMR-based model:

CONC-10 predicted bioactivity with an average MSE of 8.05 (SD = 13.2). Table 2 shows the top eight masses and shifts from the model. NMR-10 identified 6.68 ppm as the 1st ranked shift, which is found in α, β-dihydroxanthohumol, and 5.51 ppm as the 6th ranked shift, which is found in 5″-hydroxyxanthohumol and xanthohumol C, all of which are xanthohumol derivatives that are lower in concentration than xanthohumol in the extract. The 3rd most important mass was xanthohumol, and the 4th was adhumulone, a lupulin constituent that has been associated with weak anti-inflammatory activity [5]. CONC-10 did not identify xanthohumol as most important feature like MS-10 did, but it did find an additional annotated mass (adhumulone) and two important shifts. Fig. 4a shows that shifts dominate the top- and bottom-ranked ~20 features, with the masses evenly distributed across the middle. The inclusion of MS data in

CONC-10 resulted in slightly lower MSE compared to NMR-10. CONC-10 may improve dereplication over NMR-10 and possibly over MS-10; The cost of worse average prediction may be worth the identification of important shifts.

Table 3. The annotated top eight masses and shifts from CONC-10.

Rank	δ	ID	m/z	ID
1	6.68	α, β-dihydroxanthohumol	331.2/3.2 (32)	-
2	2.91	-	447.3/3.4 (104)	noise
3	7.06	-	353.1/3.2 (42)	xanthohumol
4	4.33	-	361.2/3.3 (52)	adhumulone
5	1.59	-	375.2/3.2 (61)	-
6	5.51	5"-hydroxyxanthohumol, xanthohumol C	375.2/3.3 (62)	-
7	3.47	-	448.3/3.4 (105)	-
8	3.58	-	355.2/3.2 (45)	-
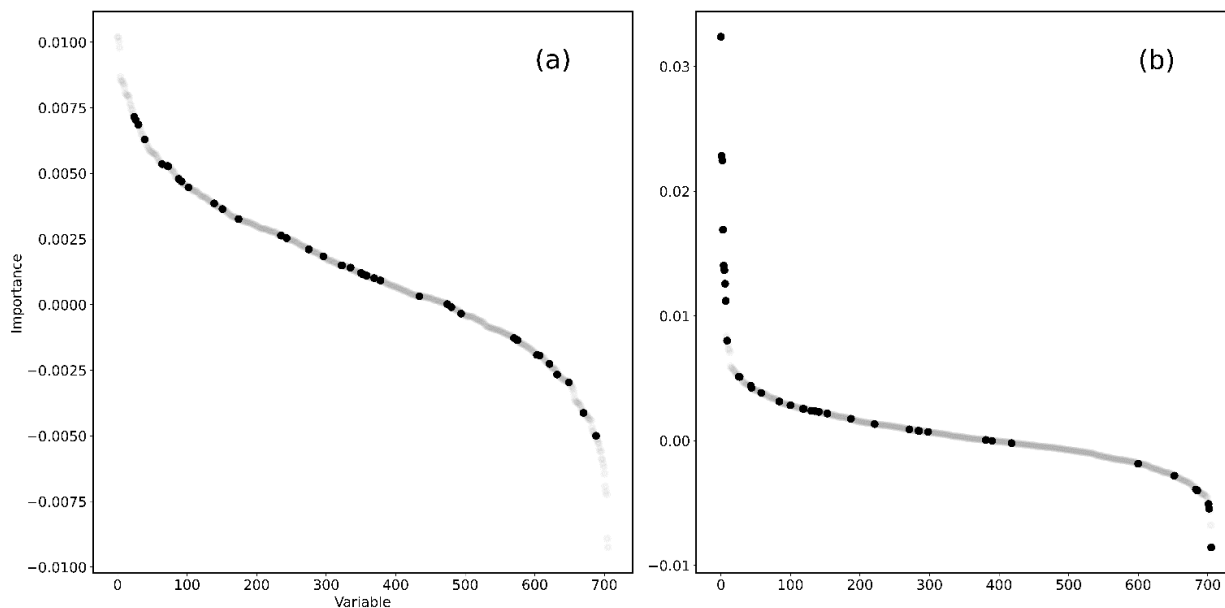
Figure 4 (a). The feature importance rankings of CONC-10. The black circles represent masses, and the grey circles represent shifts. **(b).** The feature importance rankings of MC-10. The black circles represent masses, and the grey circles represent shifts.

3.4. Multichannel MS/NMR-based model:

MC-10 predicted bioactivity with lower error compared to CONC-10 with an average MSE of 7.58 (SD = 11.5). Table 3 shows that MC-10 identified the most annotated masses of all models, ranking 4'-O-Methyl-xanthohumol as 1st and α,β-dihydroxanthohumol as 5th, both xanthohumol derivatives, xanthohumol B as 6th, a structural analog of xanthohumol, and cohumulinone as 7th, a low concentration humulone in the extract. MC-10 also identified two important shifts; 1st ranked 1.29 ppm, which is found in xanthohumol B, desmethylxanthohumol J, and xanthohumol I, all structural analogs of xanthohumol, and 7th ranked 2.66 ppm, which is found in isoxanthohumol, 6-prenylnaringenin, and 8-prenylnaringenin, all metabolites of xanthohumol. MC-10 may be better at dereplication than other models because it provides the most information about bioactive molecules in the mixtures. Unlike CONC-10, which found masses and shifts that are unconnected, MC-10 found both the mass for xanthohumol B and a shift associated with it. Fig. 4b shows that masses dominated the top-ranking features in MC-10, compared to shifts in CONC-10. As the MS-10 model predicted bioactivity the best of all models, this may be the reason that MC-10 had lower MSE than CONC-10, which placed more importance on the NMR data to make predictions.

Table 4. Top masses and shifts from MC-10.

Rank	δ	ID	m/z	ID
1	1.29	xanthohumol B, desmethylxanthohumol J, xanthohumol I	367.2/3.3 (54)	4'-O-Methyl-xanthohumol
2	0.44	-	237.1/3.1 (5)	-
3	1.75	-	251.1/3.1 (8)	-
4	0.45	-	305.1/3.2 (20)	-
5	0.24	-	355.1/3.2 (44)	α,β-dihydroxanthohumol
6	1.41	-	369.1/3.2 (55)	xanthohumol B
7	2.66	Isoxanthohumol, 6-prenylnaringenin, 8-prenylnaringenin	363.2/3.2 (53)	cohumulinone
8	7.89	-	409.3/3.5 (82)	-

The MSE probability density functions of the four models are shown in Fig. 5. MS-10 has the smallest range and lowest average MSE with one peak around 0.7 and a second, smaller peak around 3.6. NMR-10, CONC-10, and MC-10 have larger ranges, with one peak around 1 and a second around 40. The bimodality is likely due to the limited amount of data. Bioactivity is concentrated to a few fractions, meaning that some models will not train on bioactive data and therefore will predict bioactivity poorly. It may also be due to the stochastic nature of ANNs; at every model definition, the network weights are set to random values.

The gradient descent function also relies on randomness to adjust the network weights during training. If the loss function is nonconvex, this randomness results in the gradient descent finding different minima in different model iterations.
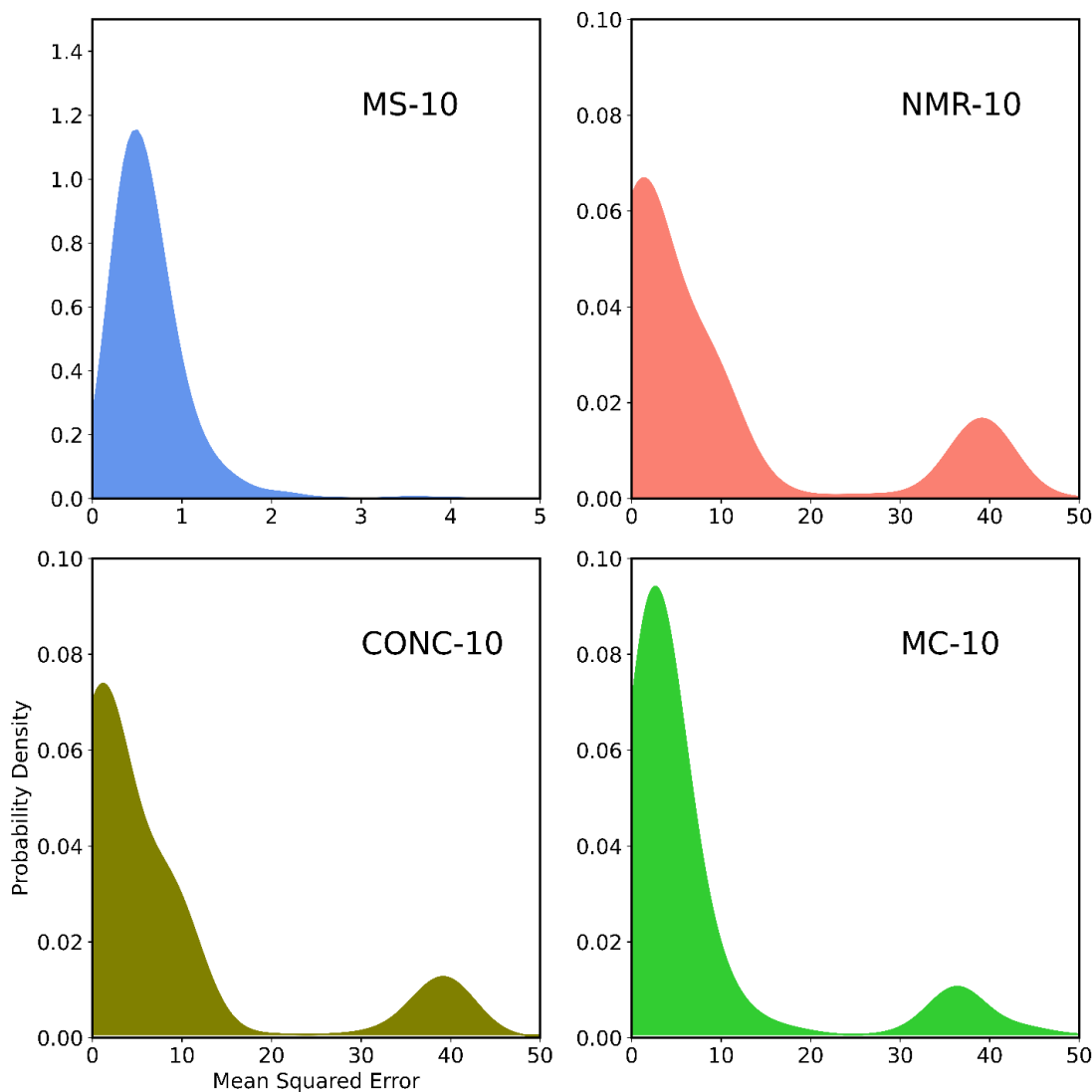
Figure 5. The truncated MSE probability density functions based on 1000 iterations for each model, computed using scipy.stats.gaussian_kde. MS-10 (mean = 0.685, SD = 0.800, min = 0.0841, max = 9.78) is cornflower blue, NMR-10 (mean = 9.48, SD = 14.0, min = 0.0784, max = 74.9) is salmon, CONC-10 (mean = 8.05, SD = 13.2, min = 0.122, max = 134) is olive, and MC-10 (mean = 7.58, SD = 11.5, min = 0.132, max = 83.4) is lime green.

3.6. Additional considerations and future work:

Moving forward, several improvements can be made to the ML pipeline. For a more consistent approach, it would be better to use PM to evaluate feature importance for all models rather than using CWA for MS, NMR, and CONC, and PM for MC. The downside to PM is that it has a much greater computational burden than CWA, but its upside is that it is compatible with any estimator. A version of PM where multiple

10

features are permuted simultaneously would be useful, especially to models that use NMR data because molecules are described by multiple shifts. However, even permuting just two shifts at a time would mean there are 222,111 shift combinations to test, which is unrealistic. This number could be lowered by only permuting shifts that have a positive importance according to CWA, but then it would be limited to the MS, NMR, and CONC models. Moreover, if only half the shifts were positive based on CWA, there would still be 55,278 combinations to test. A paired permutation test could also uncover synergistic compounds – those that have no effect alone but large effects when paired with other compounds, so the additional discovery power gained could be well worth the computational burden. Another caveat to this work is that only the MS-based model was topologically optimized, which showed great variability between xanthohumol ranking and layer size and network depth. Repeating topology optimization on the NMR, CONC, and MC models may yield a lower MSE and the identification of more known bioactive masses and associated shifts, making these models more competitive with the MS-based model.

4. Conclusion

In summary, the MS-based model MS-10 predicted bioactivity with the lowest holdout MSE of 0.685 and ranked xanthohumol, a known bioactive compound, as the top predictor. NMR- and MS/NMR-based models predicted bioactivity with higher MSEs; NMR-10 with 9.48, CONC-10 with 8.05, and MC-10 with 7.58. These models were less balanced because of the greater number of features, but identified more important masses than MS-10 and found several important NMR shifts. Shifts dominated the top important features in CONC-10, whereas masses dominated the top important features in MC-10, which may explain its lower relative MSE. Additionally, MC-10 found an important mass and shift that are associated with xanthohumol B, showing the most promise for improving dereplication despite worse prediction compared to MS-10. Further optimization of the NMR, CONC, and MC models is necessary, but this study has shown that ANNs can usefully predict bioactivity from limited MS/NMR data.

5. References

1. Kingston, D.G.I. Modern Natural Products Drug Discovery and its Relevance to Biodiversity Conservation. *J Nat Prod* **2011**; 74(3), pp. 496-511.
2. Sarker, S.D.; Latif, Z.; Gray, A.I. Natural Product Isolation. In Methods in Biotechnology. Humana Press **2006**; 20, pp. 1-25.
3. Caesar, L.K.; Nogo, S.; Naphen, C.N; Cech, N.B. Simplify: A Mass Spectrometry Metabolomics Approach to Identify Additives and Synergists from Complex Mixtures. *Anal Chem*, **2019**, *91(17)*, 11297-11305.
4. Kellogg, J.J.; Todd, D.A; Egan J.M.; Raja, H.A.; Oberlies, N.H.; Kvalheim, O.M; Cech, N.B. Biochemometrics for Natural Products Research: Comparison of Data Analysis Approaches and Application to Identification of Bioactive Compounds. J Nat Prod, **2016**, *79(2)*, 376-386.
5. Brown, K.S.; Jamieson, P.; Wu, W.; Vaswani, A.; Magana, A.A.; Choi, J.; Cheong, P.H.; Nelson, D.; Reardon, P.N.; Miranda, C.L.; Maier, C.S.; Stevens, J.F. Computation-assisted identification of bioactive compounds in botanical extracts: A cast study of anti-inflammatory natural products from lupulin (*Humulus lupulus*). *Antioxidants* **2022**; 11(7), 1400.
6. Miranda CL, Stevens JF, Ivanov V, McCall M, et al. Antioxidant and prooxidant actions of prenylated and nonprenylated chalcones and flavanones in vitro. *J Agric Food Chem.* **2000**; 48, pp. 3876–3884.
7. Lupinacci E, Meijerink J, Vincken JP, Gabriele B, et al. Xanthohumol from hop (Humulus lupulus L.) is an efficient inhibitor of monocyte chemoattractant protein-1 and tumor necrosis factor-alpha release in

LPS-stimulated RAW 264.7 mouse macrophages and U937 human monocytes. *J Agric Food Chem.* **2009**; 57, pp. 7274–7281.

8. Peluso MR, Miranda CL, Hobbs DJ, Proteau RR, Stevens JF. Xanthohumol and related prenylated flavonoids inhibit inflammatory cytokine production in LPS-activated THP-1 monocytes: structure-activity relationships and in silico binding to myeloid differentiation protein-2 (MD-2) *Planta Med.* **2010**; 76, pp. 1536–1543.

9. Gerhauser C. Broad spectrum anti-infective potential of xanthohumol from hop (Humulus lupulus L.) in comparison with activities of other hop constituents and xanthohumol metabolites. *Mol Nutr Food Res.* **2005**; 49, pp. 827–831.

10. Xuan NT, Shumilina E, Gulbins E, Gu S, et al. Triggering of dendritic cell apoptosis by xanthohumol. *Mol Nutr Food Res.* **2010**; 54 (Suppl 2), pp. S214–224.

11. Zhao, F.; Watanabe, Y.; Nozawa, H.; Daikonnya, A.; Kondo, K.; Kitanaka, S. Prenylflavonoids and Phloroglucinol Derivatives from Hops (*Humulus lupulus*). *J. Nat. Prod*. **2005**; 68, pp. 43-49.

12. Milligan S, Kalita J, Pocock V, Heyerick A, et al. Oestrogenic activity of the hop phyto-oestrogen, 8-prenylnaringenin. *Reproduction.* **2002**; 123, pp. 235–242.

13. Milligan SR, Kalita JC, Heyerick A, Rong H, et al. Identification of a potent phytoestrogen in hops (Humulus lupulus L.) and beer. *J Clin Endocrinol Metab.* **1999**; 84, pp. 2249–2252.

14. Milligan SR, Kalita JC, Pocock V, Van De Kauter V, et al. The endocrine activities of 8-prenylnaringenin and related hop (Humulus lupulus L.) flavonoids. *J Clin Endocrinol Metab.* **2000**; 85, pp. 4912–4915.

15. Stevens, J.F.; Revel, J. Chapter 15: Xanthohumol, what a delightful problem child! In *American Chemical Society Book: Chemistry and Biological Activities of Phenolic Compounds from Fruits and Vegetables*. ACS Symposium Series, "Advances in plant phenolics: From chemistry to human health" **2018**; 1286, pp. 283-304.

16. Brown, K.S.; Jamieson, P.; Wu, W.; Vaswani, A.; Magana, A.A.; Choi, J.; Cheong, P.H.; Nelson, D.; Reardon, P.N.; Miranda, C.L.; Maier, C.S.; Stevens, J.F. Computation-assisted identification of bioactive compounds in botanical extracts: A cast study of anti-inflammatory natural products from lupulin (*Humulus lupulus*). *Antioxidants* **2022**; 11(7), p. 1400.

17. Olden, J.D.; Joy, M.K.; Death, R.G. An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data. *Ecological Modelling* **2004**; 178, pp. 389-397.

18. Garson, D.G. Interpreting neural network connection weights. *AI Expert*, **1991**, 6, pp. 47-51.

19. Donoho, D.L. High-Dimensional Data Analysis: The Curses and Blessings of Dimensionality. In *AMS Conference on Math Challenges of the 21st Century* **2000**.