Supplementary Information (SI) for Molecular Systems Design & Engineering.

Supporting Information for :

PepMNet: A Hybrid Deep Learning Model for Predicting Peptide Properties Using Hierarchical Graph Representations.

Authors: Daniel Garzon[1], Omid Akbari[1], Camille Bilodeau[1,†]

[1]University of Virginia, Chemical Engineering Department, 385 McCormick Road, Charlottesville, VA 22903

*Corresponding Author: cur5wz@virginia.edu

†Permanent address: 385 McCormick Road, Charlottesville, VA 22903

1. Hyperparameter Tunning

For both models, we set the NNConv[2,3] layer so that we could incorporate atom and bond information at the atomic level, using sum as the aggregation function for neighborhood information and ReLU as the activation function. After hyperparameter tuning, we set 3 layers for the retention time (RT) model and 2 layers for the antimicrobial peptide (AMP) classification model. An ARMA[4] layer was chosen for both models at the residue level. This layer provided us the best model performance by setting NNConv at the atomic level, as shown **Figure 5** and **Supplementary Figure Table 6.** The tuning was performed manually and was not exhaustive, allowing us to explore several key hyperparameters.

For hyperparameter tuning on the retention time model, we selected a subset of 10,000 sequences randomly drawn from the misc dataset to streamline the process. The best hyperparameters were selected based on performance on the training set, while the testing set was reserved for evaluating final model performance. We experimented incorporating 1, 2, and 3 layers at the atomic level, using layer sizes of 500, 250, and 100, respectively. At the amino acid level, we tested 1 and 2 layers but observed significant overfitting with 2 layers. Once the ARMA layer was fixed at the amino acid level, we explored different configurations, including 7 and 3 stacks, with 3, 5, 7, and 10 layers each. Additionally, we experimented with dropout rates of 0, 0.1, and 0.5 in the ARMA layer, and examined the effects of sharing or not sharing weights between layers. The layer sizes tested included 100, 25, 10, and 15. The final ARMA layer configuration consisted of a layer size of 15, 3 stacks, 7 layers, no dropout, and no weight sharing between layers. At the peptide level, we explored various configurations of four linear layers with size combinations of (90, 50, 5, 1), (150, 50, 5, 1), (50, 50, 5, 1), and (100, 50, 5, 1), selecting the latter for the final model. We also evaluated different learning rates, including 1e-4, 1e-2, and 1e-3, along with batch sizes of 30, 20, and 25, ultimately settling on a learning rate of 1e-3 and a batch size of 25. Furthermore, multiple train/test splits were evaluated—specifically 6:4, 7:3, 8:2, and 9:1 selecting this last one with match we the split in from the previous state of the art work[1]. The final hyperparameter configuration that yielded the best results is summarized in Supplementary Table 1.

For AMP predictions, we split the dataset, using 80% for training and 20% for validation. The best hyperparameters were selected based on performance on the validation set. We experimented with binary cross-entropy combined with a final sigmoid layer, as well as binary cross-entropy with logits, ultimately selecting the latter. Additionally, we explored models with 1, 2, and 3 layers at the atomic level, using layer sizes of 1000, 250, 100, 20, 15, and 10. From these, we selected a two-layer configuration at the atomic level with sizes of 20 and 10. At the amino acid level, we tested one ARMA layer with different configurations, including 7 and 3 stacks, and layers with sizes of 3, 5, and 7. We also experimented with dropout rates of 0, 0.1, 0.2, and 0.3 in the ARMA layer, evaluating the impact of sharing or not sharing weights between layers. The final ARMA configuration included a layer size of 50, 3 stacks, 7 layers, a dropout rate of 0.3, and no weight sharing between layers. At the peptide level, we tested several configurations of four linear layers, including size combinations of (5, 0, 0, 1), (50, 100, 0, 1), (200, 150, 100, 1), (300, 150, 100, 1), and (200, 100, 10, 1), selecting the latter for the final model. We also explored different learning rates (1e-4, 5e-4, 5e-3, 1e-3) and batch sizes (25, 70, 100, 200, 300, 500), ultimately choosing a learning rate of 1e-3 and a batch size of 25. Finally, we tested weight decay values for regularization (1e-6, 1e-5, 1e-4), selecting 1e-5 as the final value. The optimal hyperparameter configuration that produced the best results for AMP classification is summarized in Supplementary Table 1.

Table S1. Final Hyperparameter Configuration

Hyperparameter	RT	AMP
Optimizer	Adam	Adam
Learning Rate	1e-3	1e-3
Loss Function	Mean square error	Binary cross entropy with logits
Batch Size	25	100
Epochs	100	500
Training Size	90%	80%
Atomic Layer 1	500	20
Atomic Layer 2	250	10
Atomic Layer 3	100	-
Amino Acid Layer 1	15	50
Number Stacks	3	3
Number Layers	7	7
Shared Weights	False	False
Dropout-ARMA Layer	0	0.3
Peptide Layer 1	100	200
Peptide Layer 2	50	100
Peptide Layer 3	10	10
Peptide Layer 4	1	1
Dropout- Linear Layers	0	0

The non-hierarchical models were designed with the same number of layers and parameters as PepMNet to ensure a fair comparison. Each graph convolutional layer had its own set of hyperparameters, and after tuning, the final configurations are provided in Supplementary Table 2.

Table S2. Hyperparameter Configuration Nonhierarchical Layers

Layer	Hyperparameters
ARMA	Stack = 3 Layers = 7 Dropout = 0.1
EGConv	Aggregation = Mean Heads = 5 Bases = 2
GAT	Heads = 7 Concatenated = True Dropout = 0
SAGEConv	Aggregation = Mean Normalize = True Project = False
TransformerConv	Heads = 3 Concatenated = True Beta = False
GCNConv	Normalize = True Improved = False Cached = False

2. Regression Task- Retention Time Prediction:

Table S3. R^2 Comparison for RT Prediction Using Different Graph Convolutional Layers at The Amino Acid Level for HeLa and misc Dataset.

	Amino Acid Feature - No Concatenated					**Amino Acid Feature - Concatenated**				
Dataset - Layer	1	2	3	Avg	Std	1	2	3	Avg	Std
HeLa - ARMA	0.9597	0.9444	0.9566	0.9536	0.0081	**0.9625**	**0.9521**	**0.9553**	**0.9566**	0.0053
HeLa - EGConv	0.8743	0.8245	0.8911	0.8633	0.0346	0.9123	0.8529	0.9151	0.8934	0.0351
HeLa - GAT	0.9231	0.9335	0.8923	0.9163	0.0214	0.9473	0.9425	0.9335	0.9411	0.0070
HeLa - GCNConv	0.9559	0.9628	0.9389	0.9525	0.0123	0.9468	0.9429	0.9295	0.9397	0.0091
HeLa - SAGEConv	0.8789	0.8631	0.8434	0.8618	0.0178	0.8971	0.8664	0.8929	0.8854	0.0166
HeLa - TransformerConv	0.9449	0.9353	0.9601	0.9468	0.0125	0.8586	0.8334	0.8442	0.8454	0.0127
misc – ARMA	0.9848	0.9872	0.9886	0.9869	0.0019	**0.9880**	**0.9898**	**0.9892**	**0.9890**	0.0010
misc - EGConv	0.9513	0.9461	0.9542	0.9505	0.0041	0.9542	0.9536	0.9516	0.9531	0.0014
misc - GAT	0.9506	0.9496	0.9507	0.9503	0.0006	0.9439	0.9457	0.9490	0.9462	0.0026
misc - GCNConv	0.9516	0.9529	0.9528	0.9524	0.0007	0.9547	0.9540	0.9543	0.9543	0.0004
misc - SAGEConv	0.9562	0.9563	0.9564	0.9563	0.0001	0.9568	0.9571	0.9563	0.9567	0.0004
misc - TransformerConv	0.9550	0.9545	0.9561	0.9552	0.0008	0.9543	0.9558	0.9548	0.9550	0.0008

Table S4. R^2 Performance of PepMNet

PepMNet	1	2	3	Avg	Std
Atlantis Silica	0.9809	0.9800	0.9816	**0.9809**	0.0008
HeLa	0.9376	0.9463	0.9441	**0.9427**	0.0045
Luna HILIC	0.9833	0.9829	0.9860	**0.9841**	0.0017
Luna Silica	0.9773	0.9856	0.9857	**0.9829**	0.0048
misc	0.9890	0.9879	0.9886	**0.9885**	0.0006
SCX	0.9928	0.9950	0.9949	**0.9942**	0.0012
Xbridge	0.9892	0.9886	0.9849	**0.9876**	0.0023
yeast	0.9777	0.9837	0.9859	**0.9825**	0.0043

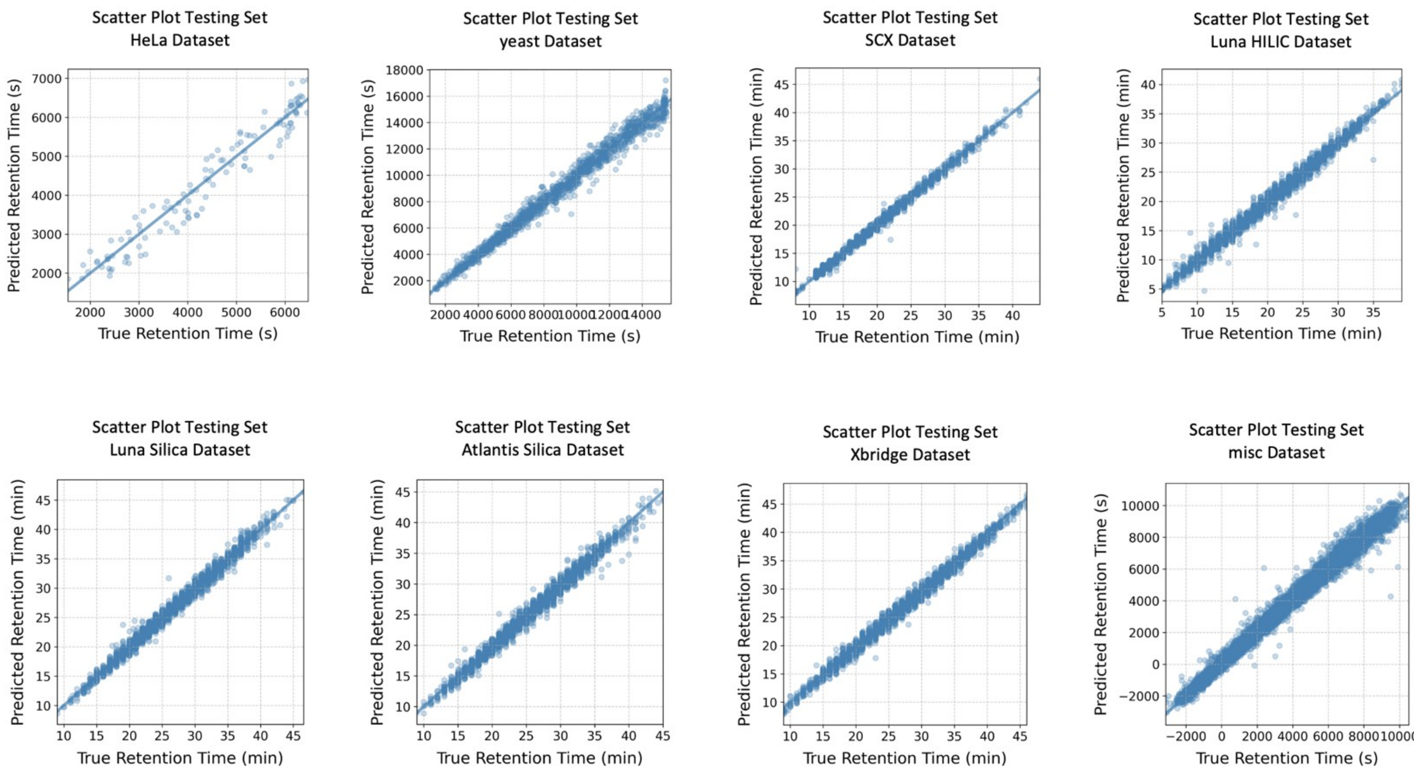

Figure S1. Scatter plots of retention time prediction on the test set across the eight different datasets.

Table S5. R^2 Performance of Nonhierarchical Models with Atomic Level

Dataset and Type of Layers **Nonhierarchical Atomic Level**	1	2	3	Avg	Std
Atlantis Silica-R2 Testing Using ARMA Layer	0.9792	0.9725	0.9791	**0.9769**	0.0039
Atlantis Silica-R2 Testing Using EGConv Layer	0.9728	0.9745	0.9736	**0.9736**	0.0009
Atlantis Silica-R2 Testing Using GAT Layer	0.9704	0.9704	0.9712	**0.9707**	0.0005
Atlantis Silica-R2 Testing Using GCNConv Layer	0.9720	0.9745	0.9745	**0.9737**	0.0014
Atlantis Silica-R2 Testing Using NNConv Layer	0.9728	0.9697	0.9730	**0.9718**	0.0019
Atlantis Silica-R2 Testing Using SAGEConv Layer	0.9752	0.9749	0.9750	**0.9750**	0.0002
Atlantis Silica-R2 Testing Using TransformerConv Layer	0.9739	0.9719	0.9744	**0.9734**	0.0013
HeLa-R2 Testing Using ARMA Layer	0.7623	0.8456	0.8784	**0.8288**	0.0599
HeLa-R2 Testing Using EGConv Layer	0.8266	0.8264	0.8264	**0.8265**	0.0001
HeLa-R2 Testing Using GAT Layer	0.8348	0.8235	0.8302	**0.8295**	0.0057
HeLa-R2 Testing Using GCNConv Layer	0.8181	0.8169	0.8185	**0.8178**	0.0009
HeLa-R2 Testing Using NNConv Layer	0.9497	0.9500	0.9537	**0.9511**	0.0022
HeLa-R2 Testing Using SAGEConv Layer	0.8649	0.8391	0.8382	**0.8474**	0.0152
HeLa-R2 Testing Using TransformerConv Layer	0.8556	0.8569	0.8571	**0.8565**	0.0008
Luna Hilic-R2 Testing Using ARMA Layer	0.9797	0.9797	0.9801	**0.9798**	0.0002
Luna Hilic-R2 Testing Using EGConv Layer	0.9728	0.9723	0.9715	**0.9722**	0.0006

Luna Hilic-R2 Testing Using GAT Layer	0.9706	0.9714	0.9719	**0.9713**	0.0006
Luna Hilic-R2 Testing Using GCNConv Layer	0.9730	0.9727	0.9723	**0.9727**	0.0003
Luna Hilic-R2 Testing Using NNConv Layer	0.9708	0.9724	0.9705	**0.9712**	0.0010
Luna Hilic-R2 Testing Using SAGEConv Layer	0.9740	0.9739	0.9740	**0.9739**	0.0001
Luna Hilic-R2 Testing Using TransformerConv Layer	0.9734	0.9731	0.9722	**0.9729**	0.0006
Luna Silica-R2 Testing Using ARMA Layer	0.9724	0.9686	0.9787	**0.9732**	0.0051
Luna Silica-R2 Testing Using EGConv Layer	0.9695	0.9713	0.9693	**0.9700**	0.0011
Luna Silica-R2 Testing Using GAT Layer	0.9708	0.9715	0.9707	**0.9710**	0.0004
Luna Silica-R2 Testing Using GCNConv Layer	0.9684	0.9663	0.9673	**0.9673**	0.0011
Luna Silica-R2 Testing Using NNConv Layer	0.9710	0.9709	0.9707	**0.9709**	0.0002
Luna Silica-R2 Testing Using SAGEConv Layer	0.9711	0.9724	0.9718	**0.9718**	0.0006
Luna Silica-R2 Testing Using TransformerConv Layer	0.9723	0.9719	0.9723	**0.9722**	0.0002
misc-R2 Testing Using ARMA Layer	0.9584	0.9569	0.9471	**0.9541**	0.0061
misc-R2 Testing Using EGConv Layer	0.9414	0.9413	0.9413	**0.9414**	0.0001
misc-R2 Testing Using GAT Layer	0.9414	0.9416	0.9420	**0.9417**	0.0003
misc-R2 Testing Using GCNConv Layer	0.9401	0.9412	0.9407	**0.9407**	0.0005
misc-R2 Testing Using NNConv Layer	0.9412	0.9403	0.9414	**0.9410**	0.0006
misc-R2 Testing Using SAGEConv Layer	0.9428	0.9427	0.9426	**0.9427**	0.0001
misc-R2 Testing Using TransformerConv Layer	0.9419	0.9419	0.9423	**0.9420**	0.0003
SCX-R2 Testing Using ARMA Layer	0.9931	0.9942	0.9942	**0.9938**	0.0006
SCX-R2 Testing Using EGConv Layer	0.9877	0.9873	0.9868	**0.9872**	0.0005
SCX-R2 Testing Using GAT Layer	0.9879	0.9868	0.9873	**0.9873**	0.0006
SCX-R2 Testing Using GCNConv Layer	0.9880	0.9846	0.9881	**0.9869**	0.0020
SCX-R2 Testing Using NNConv Layer	0.9869	0.9860	0.9865	**0.9865**	0.0005
SCX-R2 Testing Using SAGEConv Layer	0.9884	0.9885	0.9888	**0.9886**	0.0002
SCX-R2 Testing Using TransformerConv Layer	0.9881	0.9825	0.9885	**0.9864**	0.0033
Xbridge Amide-R2 Testing Using ARMA Layer	0.9827	0.9826	0.9820	**0.9824**	0.0004
Xbridge Amide-R2 Testing Using EGConv Layer	0.9774	0.9776	0.9758	**0.9769**	0.0010
Xbridge Amide-R2 Testing Using GAT Layer	0.9744	0.9749	0.9765	**0.9752**	0.0011
Xbridge Amide-R2 Testing Using GCNConv Layer	0.9726	0.9743	0.9776	**0.9748**	0.0026
Xbridge Amide-R2 Testing Using NNConv Layer	0.9778	0.9776	0.9753	**0.9769**	0.0014
Xbridge Amide-R2 Testing Using SAGEConv Layer	0.9785	0.9729	0.9759	**0.9758**	0.0028
Xbridge Amide-R2 Testing Using TransformerConv Layer	0.9783	0.9776	0.9786	**0.9782**	0.0005
yeast-R2 Testing Using ARMA Layer	0.9346	0.9440	0.9435	**0.9407**	0.0053
yeast-R2 Testing Using EGConv Layer	0.9474	0.9444	0.9461	**0.9460**	0.0015
yeast-R2 Testing Using GAT Layer	0.9433	0.9477	0.9463	**0.9457**	0.0023
yeast-R2 Testing Using GCNConv Layer	0.8767	0.8803	0.8753	**0.8774**	0.0026
yeast-R2 Testing Using NNConv Layer	0.9429	0.9386	0.9400	**0.9405**	0.0022
yeast-R2 Testing Using SAGEConv Layer	0.8972	0.9004	0.8979	**0.8985**	0.0017
yeast-R2 Testing Using TransformerConv Layer	0.9439	0.9428	0.9419	**0.9429**	0.0010

Table S6. Performance of Nonhierarchical Models with Amino Acid Level

Dataset and Type of Layers Nonhierarchical Amino Acid Level	1	2	3	Avg	Std
Atlantis Silica-R2 Testing Using ARMA Layer	0.8876	0.8905	0.8550	**0.8777**	0.0197
Atlantis Silica-R2 Testing Using EGConv Layer	0.9526	0.9523	0.9444	**0.9498**	0.0046
Atlantis Silica-R2 Testing Using GAT Layer	0.9506	0.9557	0.9480	**0.9515**	0.0039
Atlantis Silica-R2 Testing Using GCNConv Layer	0.9281	0.9302	0.9258	**0.9281**	0.0022
Atlantis Silica-R2 Testing Using SAGEConv Layer	0.9760	0.9689	0.9669	**0.9706**	0.0048
Atlantis Silica-R2 Testing Using TransformerConv Layer	0.9560	0.9470	0.9269	**0.9433**	0.0149
HeLa-R2 Testing Using ARMA Layer	0.6957	0.7995	0.7138	**0.7363**	0.0555

HeLa-R2 Testing Using EGConv Layer	0.5696	0.5753	0.5459	**0.5636**	0.0156
HeLa-R2 Testing Using GAT Layer	0.5636	0.5619	0.5629	**0.5628**	0.0009
HeLa-R2 Testing Using GCNConv Layer	0.4768	0.4768	0.4767	**0.4768**	0.0001
HeLa-R2 Testing Using SAGEConv Layer	0.7912	0.7906	0.7911	**0.7910**	0.0003
HeLa-R2 Testing Using TransformerConv Layer	0.8143	0.8160	0.8146	**0.8150**	0.0009
Luna Hilic-R2 Testing Using ARMA Layer	0.8782	0.9064	0.8709	**0.8852**	0.0188
Luna Hilic-R2 Testing Using EGConv Layer	0.9683	0.9677	0.9678	**0.9679**	0.0003
Luna Hilic-R2 Testing Using GAT Layer	0.9409	0.9491	0.9533	**0.9478**	0.0063
Luna Hilic-R2 Testing Using GCNConv Layer	0.9338	0.9322	0.9305	**0.9322**	0.0016
Luna Hilic-R2 Testing Using SAGEConv Layer	0.9776	0.9754	0.9752	**0.9760**	0.0013
Luna Hilic-R2 Testing Using TransformerConv Layer	0.9708	0.9717	0.9714	**0.9713**	0.0004
Luna Silica-R2 Testing Using ARMA Layer	0.9101	0.7773	0.8411	**0.8428**	0.0664
Luna Silica-R2 Testing Using EGConv Layer	0.9467	0.9478	0.9456	**0.9467**	0.0011
Luna Silica-R2 Testing Using GAT Layer	0.9493	0.9500	0.9485	**0.9493**	0.0008
Luna Silica-R2 Testing Using GCNConv Layer	0.8703	0.8763	0.8498	**0.8654**	0.0139
Luna Silica-R2 Testing Using SAGEConv Layer	0.9701	0.9676	0.9729	**0.9702**	0.0027
Luna Silica-R2 Testing Using TransformerConv Layer	0.9555	0.9545	0.9545	**0.9548**	0.0006
misc-R2 Testing Using ARMA Layer	0.9565	0.9718	0.9476	**0.9586**	0.0122
misc-R2 Testing Using EGConv Layer	0.9419	0.9415	0.9371	**0.9402**	0.0027
misc-R2 Testing Using GAT Layer	0.9319	0.9357	0.9345	**0.9341**	0.0019
misc-R2 Testing Using GCNConv Layer	0.9198	0.9191	0.9163	**0.9184**	0.0019
misc-R2 Testing Using SAGEConv Layer	0.9487	0.9467	0.9478	**0.9477**	0.0010
misc-R2 Testing Using TransformerConv Layer	0.9454	0.9448	0.9455	**0.9452**	0.0004
SCX-R2 Testing Using ARMA Layer	0.9349	0.9323	0.9303	**0.9325**	0.0023
SCX-R2 Testing Using EGConv Layer	0.9552	0.9740	0.9685	**0.9659**	0.0097
SCX-R2 Testing Using GAT Layer	0.9873	0.9853	0.9763	**0.9830**	0.0059
SCX-R2 Testing Using GCNConv Layer	0.9636	0.9656	0.9658	**0.9650**	0.0012
SCX-R2 Testing Using SAGEConv Layer	0.9921	0.9916	0.9916	**0.9918**	0.0003
SCX-R2 Testing Using TransformerConv Layer	0.9860	0.9893	0.9909	**0.9888**	0.0025
Xbridge-R2 Testing Using ARMA Layer	0.9265	0.9322	0.9260	**0.9282**	0.0034
Xbridge-R2 Testing Using EGConv Layer	0.9730	0.9653	0.9666	**0.9683**	0.0041
Xbridge-R2 Testing Using GAT Layer	0.9635	0.9557	0.9516	**0.9569**	0.0061
Xbridge-R2 Testing Using GCNConv Layer	0.9273	0.9317	0.9159	**0.9250**	0.0081
Xbridge-R2 Testing Using SAGEConv Layer	0.9784	0.9794	0.9801	**0.9793**	0.0008
Xbridge-R2 Testing Using TransformerConv Layer	0.9469	0.9559	0.9470	**0.9499**	0.0052
yeast-R2 Testing Using ARMA Layer	0.9180	0.9294	0.9238	**0.9237**	0.0057
yeast-R2 Testing Using EGConv Layer	0.8845	0.8806	0.8805	**0.8819**	0.0023
yeast-R2 Testing Using GAT Layer	0.8511	0.8646	0.8632	**0.8597**	0.0074
yeast-R2 Testing Using GCNConv Layer	0.7370	0.7287	0.7442	**0.7367**	0.0077
yeast-R2 Testing Using SAGEConv Layer	0.9451	0.9327	0.9409	**0.9396**	0.0063
yeast-R2 Testing Using TransformerConv Layer	0.9144	0.9184	0.9186	**0.9171**	0.0024

3. Classification Task- Antimicrobial Peptide Prediction:

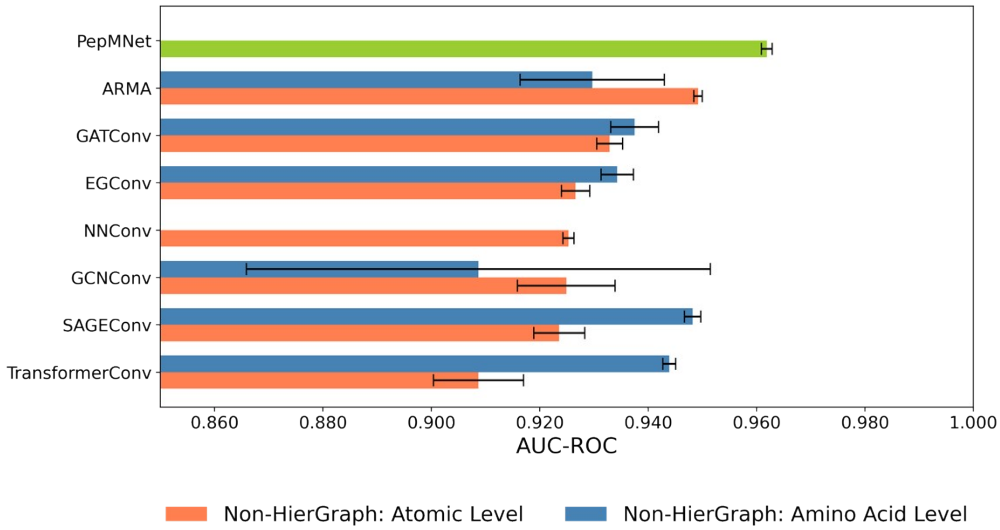

Figure S2. AUC Comparison Between PepMNet and Non-Hierarchical Models. The error bars are the standard deviation for the triplicate training implemented

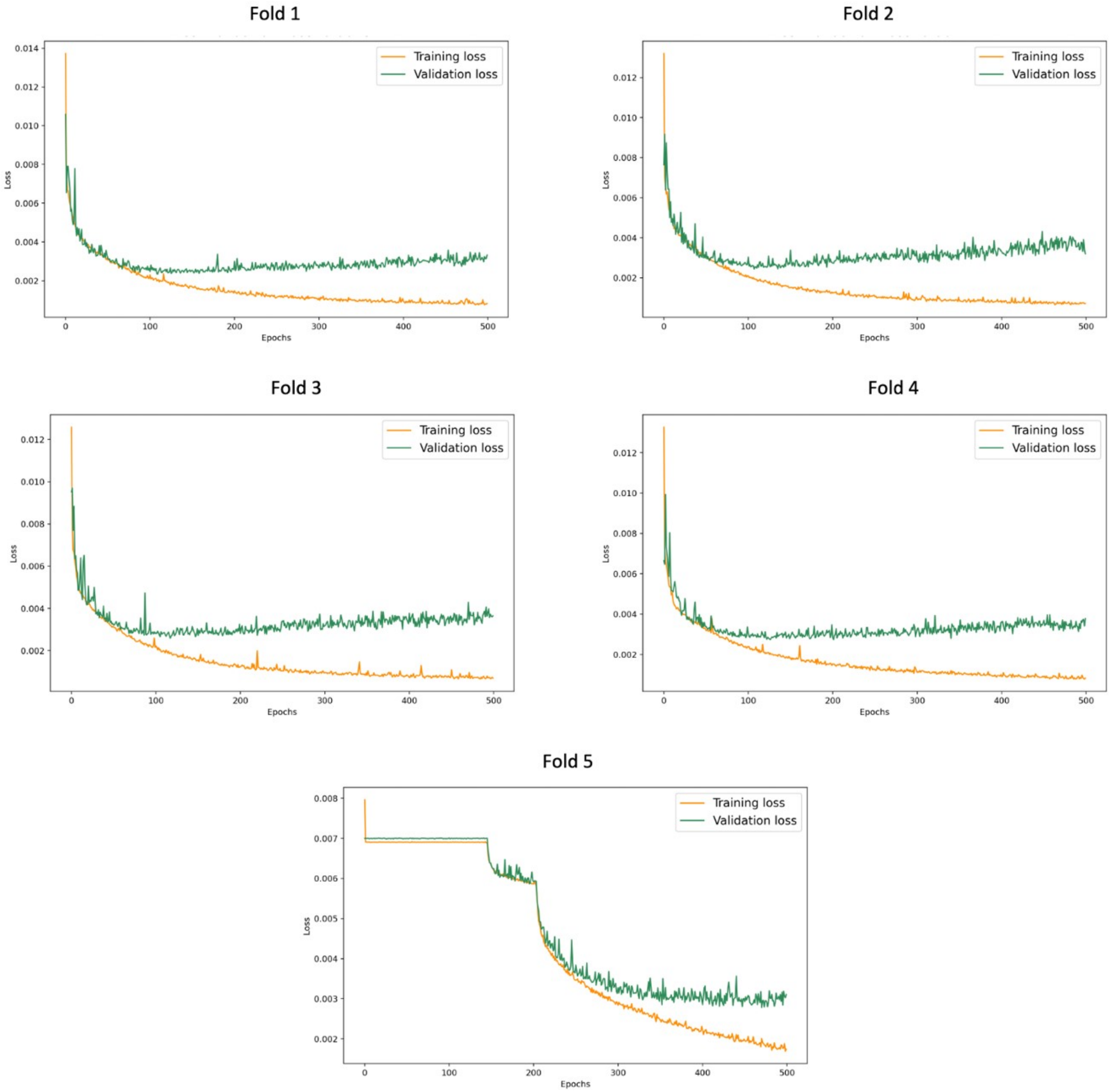

Figure S3. Loss Curves for Each Fold: Training and Validation Sets for Antimicrobial Classification

(a) (b)

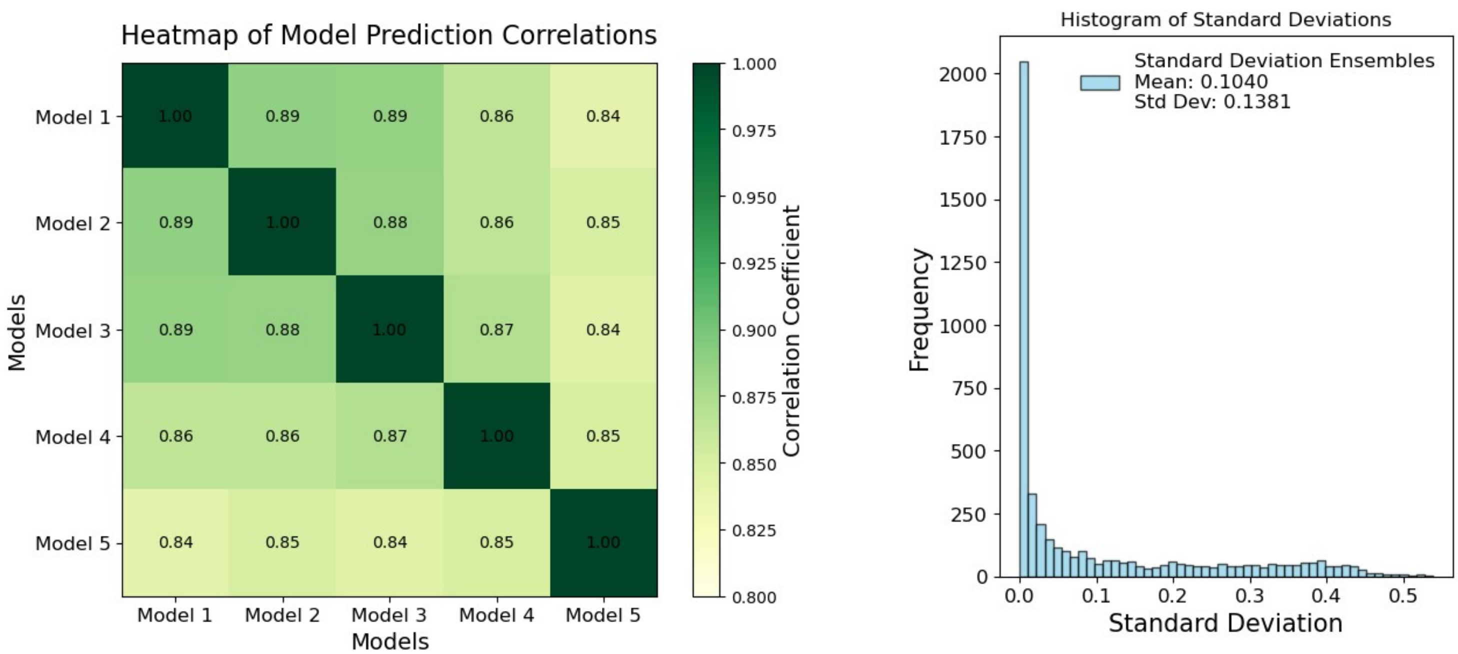

Figure S4. (a) Correlation Coefficient Between Model's Predictions for Each Fold (b) Distribution of the Standard Deviation of Predictions Across Folds