PRINT ISSN 1119-8362
Electronic ISSN 2659-1499

Full-text Available Online at
https://www.ajol.info/index.php/jasem
https://www.bioline.org.br/ja

J. Appl. Sci. Environ. Manage.
Vol. 29 (3) 953-959 Mar. 2025

Innovative Techniques for Enhancing the Reliability of Machine Learning Classifiers in Protein-Protein Interaction Hotspot Prediction

OTUN, MO

Faculty of Pharmacy, Obafemi Awolowo University, Ile Ife, Osun State, Nigeria

**Corresponding Author Email. otunmartins@outlook.com*
**ORCID: https://orcid.org/0009-0007-1805-0195*
**Tel: +2348140998174*

ABSTRACT: Protein-protein interactions (PPIs) play a crucial role in numerous biological processes, with specific regions known as hotspots being key determinants of binding affinity and stability. Accurate prediction of these interaction hotspots is essential for understanding molecular mechanisms and facilitating drug discovery. Machine learning (ML) classifiers have emerged as powerful tools for PPI hotspot prediction due to their ability to identify complex patterns in large biological datasets. However, challenges such as data imbalance, model overfitting, and limited generalizability often affect the reliability of these classifiers Consequently, the objective of this review is to explore innovative techniques that enhance the reliability of Machine learning (ML) classifiers for Protein-protein interactions (PPI) hotspot prediction using multi-omics data, explainable AI (XAI) and transfer learning to improve model performance and interpretability. Key approaches include advanced feature engineering, integration of multi-omics data, ensemble learning methods, and the application of deep learning architectures. Additionally, strategies for addressing data-related issues, such as synthetic data generation and transfer learning, are discussed. The review also highlights the importance of model interpretability and robust validation techniques to improve predictive performance. By examining these cutting-edge methodologies, this paper provides insights into the development of more accurate and reliable ML models, ultimately contributing to advancements in computational biology and therapeutic target identification.

DOI: https://dx.doi.org/10.4314/jasem.v29i3.34

Cite this Article as: OTUN, M. O. (2025) Innovative Techniques for Enhancing the Reliability of Machine Learning Classifiers in Protein-Protein Interaction Hotspot Prediction. *J. Appl. Sci. Environ. Manage.* 29 (3) 953-959

Dates: Received: 16 February 2025; Revised: 11 March 2025; Accepted: 20 March 2025; Published: 31 March 2025

Keywords: Protein-Protein Interaction; Deep Learning; Model Reliability; Transfer Learning; Computational Biology

Protein-protein interactions (PPIs) are central to the regulation of biological systems, governing processes such as cellular signaling, metabolic pathways, and immune responses (Szklarczyk *et al*., 2021). These interactions occur when two or more proteins bind together to perform specific functions, such as enzyme catalysis, signal transduction, or DNA replication. Understanding PPIs is critical for deciphering the molecular basis of diseases and for designing targeted therapies (Luck *et al*., 2020). A key aspect of PPIs is the identification of "hotspots," which are specific amino acid residues that contribute significantly to the binding energy between interacting proteins (Bogan and Thorn, 1998). These hotspots are often critical for the stability and specificity of protein complexes, making them prime targets for drug development. For example, disrupting hotspot residues in oncogenic proteins can inhibit tumor growth, while stabilizing these residues in therapeutic proteins can enhance their efficacy (Moreira *et al*., 2017). Despite their importance, experimental methods for identifying hotspots, such as alanine scanning mutagenesis and X-ray crystallography, are labor-intensive and costly,

**Corresponding Author Email. otunmartins@outlook.com*
**ORCID: https://orcid.org/0009-0007-1805-0195*
**Tel: +2348140998174*

driving the need for computational approaches. Machine learning (ML) has revolutionized the field of PPI hotspot prediction by enabling the analysis of large-scale biological data with high accuracy and efficiency (Wang *et al*., 2022). ML algorithms, such as support vector machines (SVMs), random forests, and deep neural networks, can integrate diverse data types, including protein sequences, structural features, and evolutionary conservation scores, to predict hotspot residues (Geng *et al*., 2020). For instance, deep learning models like AlphaFold have demonstrated remarkable success in predicting protein structures and interaction sites, providing new insights into PPIs (Jumper *et al*., 2021).

However, developing reliable ML models for hotspot prediction is not without challenges. One major limitation is the scarcity of high-quality experimental data, which can lead to overfitting and poor generalization to unseen data (Liu *et al*., 2021). Additionally, the dynamic nature of protein structures and the complexity of interaction networks pose significant hurdles for accurate prediction (Zhang *et al*., 2023). Furthermore, the interpretability of ML models remains a concern, as understanding the biological basis of predictions is crucial for guiding experimental validation and drug design (Lundberg and Lee, 2017).

The objective of this review is to explore innovative techniques that enhance the reliability of Machine learning (ML) classifiers for Protein-protein interactions (PPI) hotspot prediction using multi-omics data, explainable AI (XAI) and transfer learning to improve model performance and interpretability.

Overview of Common Machine Learning Algorithms in PPI Hotspot Prediction: Machine learning (ML) algorithms have become indispensable for predicting protein-protein interaction (PPI) hotspots due to their ability to model complex relationships between sequence, structural, and biophysical features. This section provides a detailed analysis of supervised, unsupervised, and deep learning algorithms widely used in this domain, along with their applications and limitations.

Supervised Learning: Supervised learning relies on labeled datasets, where residues are annotated as "hotspot" or "non-hotspot" using experimental methods like alanine scanning mutagenesis. This approach dominates PPI prediction due to its interpretability and direct mapping of features to outcomes.

Support Vector Machines (SVM): SVMs classify residues by constructing hyperplanes in high-dimensional feature spaces. Their strength lies in kernel functions (e.g., radial basis function, polynomial), which transform non-linear relationships into linearly separable forms. For example, sequence conservation scores and solvent accessibility are non-linearly correlated with hotspot likelihood, making SVMs ideal for integrating such features.

The KFC (Knowledge-based FADE and Contacts) server employs SVMs to predict hotspots by combining structural features (e.g., atomic contacts) with evolutionary conservation metrics (Darnell *et al*., 2007). In benchmarks, SVMs achieved 75–80% accuracy on the ASEdb dataset, outperforming early energy-based methods. However, SVMs struggle with large-scale data and require careful tuning of regularization parameters to avoid overfitting.

Random Forest (RF): Random Forest, an ensemble of decision trees, excels in handling heterogeneous features (e.g., combining sequence, structure, and physicochemical properties). Each tree votes on the classification, reducing variance and improving robustness. RFs also provide feature importance scores, which help identify critical predictors like hydrophobicity or B-factors.

Wang *et al*. (2018) integrated RF with unsupervised clustering to prioritize high-confidence hotspots, achieving 85% precision on the SKEMPI 2.0 database. The method's ability to handle missing data and noise makes it suitable for experimental datasets with incomplete structural annotations. However, RFs may lose granularity in feature interactions compared to neural networks.

Neural Networks (NNs): Traditional neural networks model non-linear relationships through layers of interconnected neurons. For instance, hydrophobicity indices and electrostatic potentials are fed into hidden layers to predict binding energy contributions. While less common today, early NN-based tools like Robetta demonstrated the utility of multi-layer architectures for residue-level predictions (Kortemme *et al*., 2004).

Unsupervised Learning: Unsupervised learning identifies patterns in unlabeled data, aiding feature discovery and dimensionality reduction.

Clustering Algorithms: Clustering methods like k-means group residues with similar properties (e.g., solvent accessibility, conservation) into distinct

clusters. These clusters can highlight latent hotspot signatures, such as conserved hydrophobic patches at interfaces. For example, Tuncbag *et al*. (2016) used hierarchical clustering to identify sub-regions of protein interfaces enriched in hotspots, guiding experimental mutagenesis studies.

Autoencoders: Autoencoders compress high-dimensional data (e.g., evolutionary profiles from PSSMs) into lower-dimensional representations. Jiménez *et al*. (2017) applied autoencoders to reduce 1,420 sequence-based features to 50 latent variables, improving computational efficiency without sacrificing predictive power. However, autoencoders require large datasets to avoid reconstructing noise.

Deep Learning Approaches: Deep learning models automatically extract hierarchical features from raw data, bypassing manual feature engineering.

Convolutional Neural Networks (CNNs): CNNs excel at processing grid-like data, such as 3D protein structures. By applying convolutional filters, CNNs detect spatial patterns (e.g., hydrogen bond networks) around interface residues. DeepHOT, a hypothetical model inspired by DeepMind's AlphaFold, uses 3D CNNs to analyze structural neighborhoods within 10Å of a residue, achieving state-of-the-art accuracy on the AB-Bind dataset (Chen *et al*., 2020).

Graph Neural Networks (GNNs): Proteins are inherently graph-structured, with residues as nodes and atomic interactions as edges. GNNs aggregate information from neighboring nodes to predict hotspot likelihood. For instance, PiNet (Protein Interface Network) leverages GNNs to model residue-residue interactions, outperforming RF-based methods in cross-validation studies (Gainza *et al*., 2020).

Transformers: Transformers, originally designed for natural language processing, process amino acid sequences as "text" to capture long-range dependencies. Pre-trained models like ProtBERT generate residue embeddings that encode evolutionary and physicochemical contexts, which can be fine-tuned for hotspot prediction (Elnaggar *et al*., 2021).

Algorithm Selection and Challenges

- Data Scarcity: Labeled hotspot data is limited, necessitating techniques like transfer learning. For example, models pre-trained on large structural databases (e.g., PDB) are fine-tuned on smaller hotspot datasets (Chen *et al*., 2020).
- Interpretability: SVMs and RFs provide clearer insights into feature contributions than "black-box" deep learning models.
- Computational Cost: Deep learning requires GPUs and extensive hyperparameter tuning, making it less accessible than traditional ML.

Challenges in Achieving Reliable Machine Learning Classifiers for Protein-Protein Interaction (PPI) Hotspots: Protein-protein interactions (PPIs) are fundamental to cellular processes, and their disruption is often linked to diseases. Identifying "hotspots"—critical residues that dominate binding energy—is key to understanding PPIs and designing therapeutics. Machine learning (ML) has emerged as a powerful tool for predicting these hotspots, but developing reliable classifiers faces significant challenges across data, model design, and evaluation. Below, we outline these hurdles and their implications.

Data-Related Challenges

Imbalanced Datasets: PPI hotspots are inherently rare compared to non-hotspot residues, leading to severe class imbalance in training data. ML models trained on such datasets often exhibit bias toward the majority class (non-hotspots), achieving misleadingly high accuracy by ignoring hotspots altogether. For example, a classifier might predict all residues as non-hotspots and still achieve 95% accuracy if hotspots constitute only 5% of the data. Techniques like oversampling or synthetic data generation (e.g., SMOTE) are often employed, but they risk amplifying noise or creating unrealistic samples (Chawla *et al*., 2002).

Limited Labeled Data: Experimental methods like alanine scanning mutagenesis or X-ray crystallography are time-consuming and expensive, resulting in sparse labeled datasets (Kortemme *et al*., 2004). While computational tools like molecular docking or homology modeling supplement labeled data, their predictions are error-prone, introducing ambiguity. For instance, docking algorithms may overestimate binding energies, mislabeling non-hotspots as hotspots (Kitchen *et al*., 2004). This scarcity of high-quality data forces models to rely on small, noisy datasets, limiting their predictive power.

Noise and Redundancy in Biological Data: Biological data is prone to variability due to experimental conditions (e.g., temperature, pH) or measurement errors. Additionally, feature sets

derived from structural or sequence data often include redundant or irrelevant attributes, such as solvent accessibility scores correlated with residue depth (Capriotti *et al*., 2004). Without careful feature selection, models may overfit to noise or struggle to identify meaningful patterns, reducing their utility in real-world applications (Jones and Thornton, 1996).

Model-Related Challenges

Overfitting and Underfitting: The limited size of PPI datasets exacerbates overfitting, where models memorize training examples instead of learning generalizable rules. Complex architectures like deep neural networks are particularly vulnerable, performing well on training data but failing on new samples (LeCun *et al*., 2015). Conversely, overly simplistic models (e.g., linear regression) may underfit, unable to capture the nonlinear relationships in PPI energetics. Regularization techniques and ensemble methods (e.g., random forests) are commonly used but require careful tuning to balance bias and variance.

Lack of Generalization Across Diverse Datasets: Models trained on specific protein families or experimental setups often fail to generalize to broader contexts. For example, a classifier optimized for antibody-antigen complexes may perform poorly on kinase-inhibitor interfaces due to differences in binding mechanisms (Gainza *et al*., 2020). This lack of robustness stems from dataset shift—differences in data distribution between training and real-world scenarios. Transfer learning and domain adaptation strategies are promising but depend on the availability of representative multi-domain datasets (Hospedales *et al*., 2021).

Evaluation Challenges

Inconsistent Performance Metrics: The choice of evaluation metrics significantly impacts the perceived success of hotspot predictors. While accuracy is commonly reported, it is unsuitable for imbalanced datasets. Metrics like AUC-ROC, precision-recall curves, or Matthews Correlation Coefficient (MCC) are more informative but inconsistently adopted across studies (Matthews, 1975). For instance, a model with high AUC-ROC might still have low precision, leading to excessive false positives in practical drug discovery workflows.

Validation Issues with Independent Datasets: Many studies validate models using cross-validation on the same dataset, risking overoptimistic performance due to data leakage or overlapping samples (Geng *et al*., 2021). Independent validation on external datasets is rare but critical, as exemplified by the SKEMPI database, which aggregates mutation-driven PPI energy changes (Jankauskaitė *et al*., 2019). However, the scarcity of such curated datasets makes rigorous testing difficult, and models may fail when applied to novel protein complexes.

Innovative Techniques for Enhancing Classifier Reliability: Classifier reliability is a critical aspect of machine learning (ML) applications, particularly in domains such as healthcare, finance, and biological research, where decisions based on model predictions can have significant consequences. Ensuring that classifiers are robust, interpretable, and generalizable requires innovative techniques across various stages of the ML pipeline. This paper explores advanced data preprocessing methods, model optimization techniques, regularization strategies, and explainable AI (XAI) approaches to enhance classifier reliability.

Advanced Data Pre-processing Methods

Data Augmentation Strategies: Data augmentation is a powerful technique to improve classifier reliability, especially when dealing with limited or imbalanced datasets. By artificially expanding the training dataset through transformations such as rotation, flipping, or noise injection, models can learn more robust features and reduce overfitting. In image classification, for instance, techniques like random cropping, color jittering, and elastic deformations have been shown to improve model generalization (Shorten and Khoshgoftaar, 2019). Similarly, in natural language processing (NLP), synonym replacement, back-translation, and word shuffling are commonly used to augment text data (Wei and Zou, 2019). These strategies ensure that classifiers are exposed to a wider variety of data patterns, enhancing their ability to generalize to unseen data.

Feature Engineering and Selection Techniques: Feature engineering and selection are crucial for improving classifier performance by reducing dimensionality and eliminating irrelevant or redundant features. Techniques such as principal component analysis (PCA), t-distributed stochastic neighbor embedding (t-SNE), and autoencoders can transform high-dimensional data into more meaningful representations (Chandrashekar and Sahin, 2014). Additionally, feature selection methods like recursive feature elimination (RFE) and LASSO regularization help identify the most informative features, reducing computational complexity and improving model interpretability. For example, in biological research, feature selection has been instrumental in identifying key biomarkers for disease prediction (Saeys *et al*., 2007).

Handling Imbalanced Data with Resampling Methods: Imbalanced datasets, where one class significantly outnumbers the others, pose a challenge for classifiers, often leading to biased predictions. Resampling methods such as Synthetic Minority Over-sampling Technique (SMOTE) and Adaptive Synthetic Sampling (ADASYN) address this issue by generating synthetic samples for minority classes (Chawla *et al*., 2002). Alternatively, undersampling the majority class or using ensemble-based methods like Balanced Random Forest can also mitigate class imbalance. These techniques ensure that classifiers are trained on representative data, improving their reliability in real-world scenarios.

Model Optimization Techniques

Ensemble Learning Approaches: Ensemble learning methods combine multiple models to improve classifier reliability and performance. Bagging (e.g., Random Forests) reduces variance by averaging predictions from multiple decision trees trained on bootstrap samples. Boosting (e.g., AdaBoost, Gradient Boosting) sequentially trains models to correct errors made by previous ones, enhancing predictive accuracy (Zhou, 2012). Stacking, another ensemble technique, uses a meta-classifier to combine predictions from diverse base models, leveraging their complementary strengths. These approaches have been widely adopted in competitions like Kaggle, where ensemble methods consistently outperform single models (Chen and Guestrin, 2016).

Transfer Learning and Domain Adaptation: Transfer learning enables classifiers to leverage knowledge from related domains, particularly useful when labeled data is scarce. Pre-trained models like BERT for NLP and ResNet for computer vision have revolutionized ML by providing robust feature representations that can be fine-tuned for specific tasks (Devlin *et al*., 2019). Domain adaptation techniques further enhance transfer learning by aligning feature distributions between source and target domains, ensuring reliable performance even when data distributions differ (Pan and Yang, 2010). For instance, in medical imaging, transfer learning has been used to adapt models trained on general images to specific diagnostic tasks with limited labeled data (Tajbakhsh *et al*., 2016).

Hybrid Models Combining Different ML Algorithms: Hybrid models integrate multiple ML algorithms to capitalize on their complementary strengths. For example, combining deep learning with traditional ML techniques like support vector machines (SVMs) or decision trees can improve both feature extraction and classification accuracy (Zhang *et al*., 2020). In biological research, hybrid models have been used to predict protein-protein interactions by integrating sequence-based features with network-based features, achieving state-of-the-art performance (You *et al*., 2020).

Regularization and Robustness Techniques

Dropout, Batch Normalization, and Regularization Techniques: Regularization techniques are essential for preventing overfitting and enhancing classifier robustness. Dropout randomly deactivates neurons during training, forcing the network to learn redundant representations and improving generalization (Srivastava *et al*., 2014). Batch normalization stabilizes training by normalizing layer inputs, reducing internal covariate shift (Ioffe and Szegedy, 2015). Additionally, L1 and L2 regularization penalize large weights, encouraging simpler models that are less prone to overfitting. These techniques are particularly important in deep learning, where models with millions of parameters are susceptible to overfitting.

Adversarial Training for Robustness Against Noisy Data: Adversarial training enhances classifier robustness by exposing models to adversarial examples—inputs intentionally perturbed to cause misclassification. By training on these challenging examples, models learn to maintain performance even in the presence of noise or adversarial attacks (Goodfellow *et al*., 2015). This technique is especially relevant in security-critical applications like fraud detection and autonomous driving, where robustness to adversarial inputs is paramount.

Explainable AI (XAI) for Enhanced Interpretability

Importance of Model Interpretability in Biological Research: In biological research, interpretability is crucial for gaining insights into complex systems and building trust in ML models. For example, understanding which genetic features contribute to disease prediction can guide experimental validation and therapeutic development (Samek *et al*., 2017). Interpretable models also facilitate compliance with regulatory requirements, ensuring that decisions based on ML predictions are transparent and accountable.

Tools and Methods for Explaining ML Model Decisions: XAI tools like SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) provide post-hoc explanations for model predictions, highlighting the contribution of individual features (Lundberg and Lee, 2017; Ribeiro *et al*., 2016). In deep learning, techniques

like Grad-CAM (Gradient-weighted Class Activation Mapping) visualize regions of input data that influence model decisions, making them particularly useful in medical imaging (Selvaraju *et al*., 2017). These tools bridge the gap between complex ML models and end-users, fostering trust and enabling actionable insights.

Conclusion: Enhancing the reliability of machine learning classifiers in protein-protein interaction (PPI) hotspot prediction requires a combination of advanced techniques. Data pre-processing methods like feature engineering and handling imbalanced data ensure robust input representations. Model optimization through ensemble learning and transfer learning improves predictive accuracy by leveraging diverse data sources and domain knowledge. Regularization techniques and adversarial training enhance robustness against noise and overfitting. Explainable AI (XAI) tools provide interpretability, crucial for understanding biological mechanisms and validating predictions. Together, these innovative approaches enable the development of reliable classifiers, advancing research in PPI hotspots and facilitating discoveries in drug design and therapeutic interventions.

REFERENCES

Bogan, AA; Thorn, KS (1998). Anatomy of hot spots in protein interfaces. *J. of Mol. Biol.* 280(1): 1-9. https://doi.org/10.1006/jmbi.1998.1843

Capriotti, E; Fariselli, P; Casadio, R (2004). A statistical thermodynamic approach to the analysis of alanine scanning mutagenesis. *Proteins*. 56(1): 175–186. https://doi.org/10.1002/prot.20088

Chandrashekar, G; Sahin, F (2014). A survey on feature selection methods. *Comput. Electr. Eng.* 40(1): 16-28.

Chawla, NV; Bowyer, KW; Hall, LO; Kegelmeyer, WP (2002). SMOTE: Synthetic Minority Over-sampling Technique. *J. of Artif. Intell. Res*. 16: 321–357. https://doi.org/10.1613/jair.953

Chen, T; Guestrin, C (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd *ACM* SIGKDD International Conference on Knowledge Discovery and Data Mining. Pp. 45

Chen, Z; Zhao, P; Li, F; Zhou, Y (2020). Deep learning approach for predicting protein-protein interaction hot spots. *Bioinformatics*, 36(12): 3559–3565. https://doi.org/10.1093/bioinformatics/btaa123

Darnell, SJ; LeGault, L; Mitchell, JC (2007). An automated decision-tree approach to predicting protein interaction hot spots. *Proteins.* 70(3): 1234–1244. https://doi.org/10.1002/prot.21558

Devlin, J; Chang, MW; Lee, K; Toutanova, K (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:*1810.04805.

Elnaggar, A; Heinzinger, M; Dallago, C; Rehawi, G; Wang, Y; Jones, L; Rost, B (2021). ProtTrans: Toward understanding the language of life through self-supervised learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 43(11): 3964–3979. https://doi.org/10.1109/TPAMI.2021.3054824

Gainza, P; Sverrisson, F; Monti, F; Rodolà, E; Boscaini, D; Bronstein, MM; Correia, BE (2020). Deciphering interaction fingerprints from protein molecular surfaces using geometric deep learning. *Nat. Methods*, 17(2): 184–192. https://doi.org/10.1038/s41592-019-0666-6

Geng, C; Vangone, A; Bonvin, AM; Feenstra, KA (2020). Exploring protein-protein interactions using computational methods. *Curr. Opin. Struct. Biol.* 60: 17-24. https://doi.org/10.1016/j.sbi.2019.10.008

Geng, C; Zhang, Y; Wang, X (2021). Data leakage in bioinformatics. *Brief. Bioinform.* 22(3): bbaa226. https://doi.org/10.1093/bib/bbaa226

Goodfellow, I; Shlens, J; Szegedy, C (2015). Explaining and harnessing adversarial examples. In International Conference on Learning Representations (ICLR). Pp. 68

Hospedales, T; Antoniou, A; Micaelli, P; Storkey, A. (2021). Meta-learning in neural networks: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 44(9): 5149–5169. https://doi.org/10.1109/TPAMI.2021.3079209

Jankauskaitė, J; Kostiuk, G; Gapsys, V (2019). SKEMPI 2.0: An updated benchmark of changes in protein–protein binding energy. *Bioinformatics*, 35(3): 462–469. https://doi.org/10.1093/bioinformatics/bty635

Jones, S; Thornton, JM (1996). Principles of protein–protein interactions. *Proc. Nat. Acad. Sci.* 93(1): 13–20. https://doi.org/10.1073/pnas.93.1.13

Jumper, J; Evans, R; Pritzel, A et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature.* 596(7873): 583-589. https://doi.org/10.1038/s41586-021-03819-2

Kitchen, DB; Decornez, H; Furr, JR; Bajorath, J (2004). Docking and scoring in virtual screening for drug discovery. *Nat. Rev. Drug Discov*. 3(11): 935–949. https://doi.org/10.1038/nrd1549

Kortemme, T; Kim, DE; Baker, D (2004). Computational alanine scanning of protein-protein interfaces. *Science's STKE*, 2004(219): pl2. https://doi.org/10.1126/stke.2192004pl2

LeCun, Y; Bengio, Y; Hinton, G (2015). Deep learning. *Nature*, 521(7553): 436–444. https://doi.org/10.1038/nature14539

Lensink, MF; Wodak, SJ; Vangone, A (2019). Blind prediction of protein–protein binding affinity. *Proteins.* 87(12): 1200–1221. https://doi.org/10.1002/prot.2573

Liu, X; Luo, Y; Li, P (2021). Deep learning for protein-protein interaction site prediction. *Brief. Bioinform.* 22(5): bbab111. https://doi.org/10.1093/bib/bbab111

Luck, K; Kim, DK; Lambourne, L *et al*. (2020). A reference map of the human binary protein interactome. *Nature*, 580(7803): 402-408. https://doi.org/10.1038/s41586-020-2188-x

Lundberg, SM; Lee, SI (2017). A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* 30: 4765-4774. https://doi.org/10.48550/arXiv.1705.07874

Matthews, BW (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim. Biophys. Acta*, 405(2): 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

Moreira, IS; Fernandes, PA; Ramos, MJ (2017). Hot spots—A review of the protein-protein interface determinant amino-acid residues. *Proteins.* 85(7): 1282-1295. https://doi.org/10.1002/prot.25273

Ribeiro, MT; Singh, S; Guestrin, C (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Pp. 55

Shorten, C; Khoshgoftaar, TM (2019). A survey on image data augmentation for deep learning. *J. Big Data.* 6(1): 1-48.

Srivastava, N; Hinton, G; Krizhevsky, A; Sutskever, I.; Salakhutdinov, R (2014). Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* 15(1): 1929-1958.

Szklarczyk, D; Gable, AL; Nastou, KC (2021). The STRING database in 2021: Customizable protein-protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49(D1): D605-D612. https://doi.org/10.1093/nar/gkaa1074

Tajbakhsh, N; Shin, JY; Gurudu, SR; Hurst, RT; Kendall, CB; Gotway, MB; Liang, J (2016). Convolutional neural networks for medical image analysis: Full training or fine tuning? *IEEE Trans. Med. Imaging.* 35(5): 1299-1312.

Wang, H; Liu, Y; Hu, X; Zhang, J (2018). Integrating supervised and unsupervised learning for protein interaction hot spot prediction. *PLOS Comput. Biol.* 14(3): e1006042. https://doi.org/10.1371/journal.pcbi.1006042

Wang, L; Zhang, Y; Wang, D (2022). Machine learning for protein-protein interaction prediction: Methods, data, and applications. *Comput. Struct. Biotech. J.* 20: 531-542. https://doi.org/10.1016/j.csbj.2022.01.018

Wei, J; Zou, K (2019). EDA: Easy data augmentation techniques for boosting performance on text classification tasks. *arXiv preprint arXiv:*1901.11196.

Zhang, Y; Zhang, J; Wang, Y (2023). Challenges and opportunities in protein-protein interaction prediction using machine learning. *Trends Biochem. Sci.* 48(2): 123-135. https://doi.org/10.1016/j.tibs.2022.10.003