WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Comparative Analysis of Optimization and Hybrid Deep Learning Models for Protein–Protein Interaction Prediction in Bioinformatics

Preeti Thareja [a*], Rajender Singh Chhillar [b]

[a] *Research Scholar, Department of Computer Science and Applications, Maharshi Dayanand University, Rohtak (124001), Haryana, India*
[b] *Professor, Department of Computer Science and Applications, Maharshi Dayanand University, Rohtak (124001), Haryana, India*

A B S T R A C T

This study compares two recent breakthroughs in bioinformatics optimization: the Integrated Bioinformatics Optimization Model (IBOM) and the Aquila Influenced Shark Smell Optimization (AISSO)-based hybrid deep learning model. While IBOM optimizes human protein interaction predictions using classic optimization techniques such as Whale Optimization Algorithm (WOA), Genetic Algorithm (GA), Ant Colony Optimization (ACO), Multi Verse Optimization (MVO), and 5-fold Cross Validation (CV), the AISSO-based approach uses hybrid neural networks and innovative optimization approaches to achieve accurate sequence-based PPI prediction. The analysis looks at performance measures, computational efficiency, adaptability to protein datasets, and the practical implications for bioinformatics applications. The results show that IBOM has higher precision and interpretability, whereas AISSO excels in generalization and advanced feature extraction via semantic enrichment.

Keywords: Protein–Protein Interaction, Bioinformatics, Optimization Algorithms, Deep Learning, AISSO, IBOM, Computational Efficiency

1. Introduction

Protein-protein interactions (PPIs) are the foundation of many biological processes, including cellular signaling, immunological responses, metabolic regulation, and gene expression. Studying these relationships is critical for advancing biomedical research, particularly in drug development, illness modeling, and therapeutic target identification. However, because to the complexity and dynamic nature of biological systems, accurately predicting PPIs is a substantial computational problem (Gündüz et al., 2024). The complexity of biological data, marked by high dimensionality, sparsity, and semantic heterogeneity, necessitates advanced mathematical models that go beyond typical statistical methods.

To solve these issues, the combination of optimization approaches with machine learning, particularly deep learning, is growing as a viable strategy in computational biology (Ahmed et al., 2024). Optimization methods allow for efficient search inside huge, complicated solution spaces, improving the accuracy and adaptability of PPI prediction algorithms (Hu & Ohue, 2024). At the same time, architectures for deep learning can extract significant representations from raw biological information, eliminating the requirement for human feature engineering. The combination of these methodologies brings up new possibilities for more accurate and computationally feasible PPI prediction..

This research compares the performance of two sophisticated models created for PPI prediction. The first is the IBOM, which uses a stacking ensemble of various optimization techniques to improve robustness and predictability. The second is an AISSO-based deep learning model that uses semantic analysis and gene ontology (GO)-based features in a deep architecture to improve sequence-based predictions. While the IBOM model focuses on optimization-driven integration and adaptation across datasets, the AISSO-based model excels at capturing semantic linkages and biological significance using GO characteristics.

The primary goal of this study is to assess the strengths and limits of these models from a variety of angles, including prediction performance, computational efficiency, interpretability, and adaptability to various bioinformatics datasets. By investigating their theoretical foundations, architectural designs, and empirical outcomes, we hope to gain a better understanding of their applicability for diverse PPI-related tasks in bioinformatics research. These comparative studies are critical for aiding academics and practitioners in picking models that are best suited for biological contexts and computing restrictions.

The rest of the paper is structured as follows: Section 2 discusses the technique, including the architecture, optimization strategies, and dataset properties of the IBOM and AISSO-based models. Section 3 summarizes the findings of the comparative analysis, including quantitative performance measurements and qualitative observations. Section 4 analyzes the consequences of the findings, emphasizing the trade-offs and application-specific concerns for each model. Finally, Section 5 summarizes the paper and suggests future research paths in PPI prediction utilizing optimization and hybrid deep learning algorithms.

2. Methodology

2.1 Model 1: IBOM Approach

The IBOM is a modular design that combines multiple nature-inspired metaheuristic algorithms to improve predicted performance for PPI categorization tasks. The IBOM technique uses a structured and modular workflow to optimize human protein interaction prediction, as shown in Fig. 1. The procedure starts with data gathering and preprocessing, in which raw protein interaction data from the Figshare repository is cleaned and standardized using min-max scaling methods. Following that, Principal Component Analysis (PCA) is used to extract features and reduce dimensionality, and Gaussian Mixture Models (GMM) are used to cluster the dataset and find essential distributions. The next phase entails integrating optimization algorithms to improve model performance. Four metaheuristic algorithms (WOA, GA, ACO, and MVO) are used independently to tune hyperparameters and choose the most relevant features. The model's adaptability is assessed using a 5-fold CV technique, which yielded remarkable results: accuracy of 98.61%, precision of 99.30%, and specificity of 96.59%. Finally, the system is implemented in MATLAB R2020a, which includes powerful built-in functions for PCA, GMM, and classification applications.

Fig. 1 – Pipeline of Model 1.

2.2 Model-2: AISSO Approach

DL-PPI and other contemporary deep learning frameworks have shown great promise in sequence-based prediction of PPIs (Wu et al., 2023), frequently relying on semantic-rich features and biological embeddings to improve generalization (Taha, 2025)(Soleymani et al., 2022). Deep architectures, such as RNNs and DBNs, have proven very useful in simulating complex biological sequences (Gonzalez-Lopez et al., 2018)(Dang & Vu, 2024). Furthermore, novel techniques such as AlphaFold have transformed structure prediction, enabling improved downstream tasks such as PPI analysis (Lee et al., 2023).

The AISSO model expands on this basis by combining deep sequence modeling with semantic enrichment via GO, drawing motivation from transformer-based fusion (Aksamit et al., 2024), and using optimization approaches like the Shark Smell Optimization (SSO) algorithm (Mohammad-Azari et al., 2018). Feature fusion procedures similar to those used in xCAPT5 and fusion-based models (Dang & Vu, 2024)(Tran et al., 2024) contribute to the AISSO architecture's effectiveness.

The hybrid deep learning model built for PPI prediction reflects findings observed in a survey of recent PPI approaches, which show that deep models outperform standard models when appropriately tuned (Kumar, 2017). These findings are consistent with the work of Thareja and Chhillar (Thareja et al., 2024), demonstrating the accuracy of semantic-informed hybrid models in PPI prediction tasks.To improve PPI prediction, the AISSO-based model uses a biologically inspired deep learning architecture that has been augmented with semantic and GO-based enrichment. The AISSO-based model is based on a hierarchical neural architecture designed to maximize the use of semantic and sequence-based data in PPI prediction, as can be seen in Fig. 2. The procedure starts with feature construction, which uses protein sequence data to derive three important categories of features: (1) physical features of amino acid sequences, (2) GO-based annotations of biological processes, molecular activities, and cellular components, and (3) semantic similarity scores computed using ontology-driven metrics that compare protein phrases.

The enriched feature set is analyzed using a dual neural network design. An Improved Recurrent Neural Network (IRNN) is used to recognize the sequential nature of protein sequences, while a Deep Belief Network (DBN) made up of stacked Restricted Boltzmann Machines recovers high-level hierarchical representations. Their results are integrated using a score-level fusion process, which improves the strength of the resultant feature representation.

To maximize the efficiency of this hybrid network, the model employs Aquila Influenced Shark Smell Optimization (AISSO), a metaheuristic that combines the exploratory capability of the Aquila optimizer coupled with the exploitative refining skills of the SSO technique. AISSO refines model weights, fusion scores, and important hyperparameters.

Ultimately, during the training and validation phase, the framework is trained using backpropagation, and validation accuracy and loss are monitored to evaluate performance. The results showed that the AISSO model had better generalization, particularly on semantically enriched datasets.

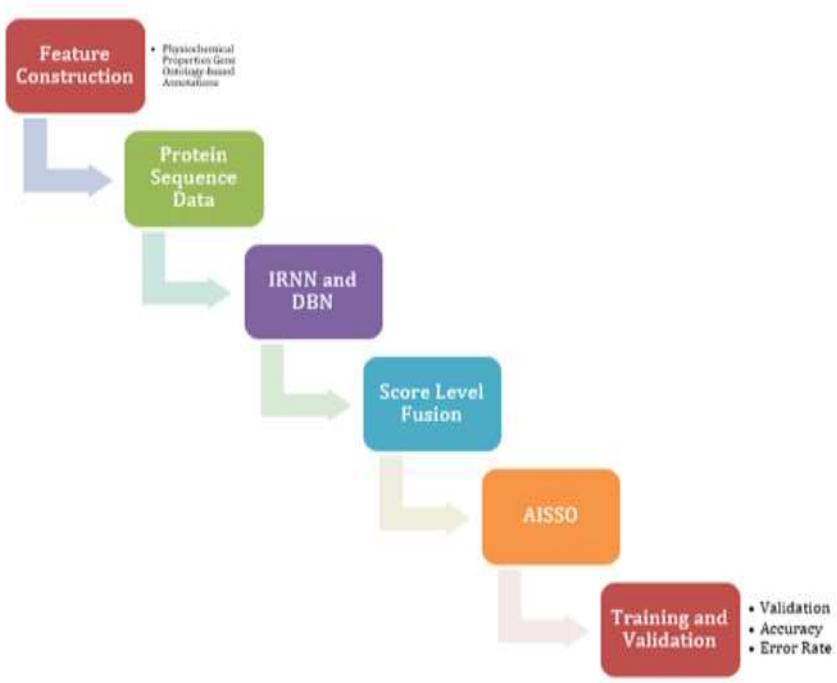

Fig. 2 – Pipeline of Model 2.

2.3 Dataset and Preprocessing

Both the IBOM and AISSO models rely on publically available datasets curated from credible bioinformatics repositories to provide superior pertinent data for PPI prediction tasks. The IBOM dataset, obtained from the Figshare repository, consists of binary protein interaction records, with each entry containing a number of attributes representing protein pair properties. These data are classified as positive or negative depending on whether or not the proteins interact. In contrast, the AISSO dataset includes protein sequences with Gene Ontology (GO) annotations. This dataset has been carefully filtered to include a varied range of interacting and non-interacting proteins. To preserve integrity of the data, redundancy is reduced using sequence similarity thresholds, and only high-confidence GO annotations are retained.

Both models' preprocessing phases include a number of significant stages. Missing value treatment is carried out using mean or mode substitution, based on the attribute type. Normalization strategies vary according on the nature of the features: min-max scaling is applied for IBOM, while AISSO utilizes a combination of Z-score normalization and min-max scaling. PCA is used to reduce dimensionality in the IBOM dataset, while the AISSO model reduces GO embeddings using feature ranking algorithms. Furthermore, the AISSO approach uses encoding and transformation techniques including one-hot encoding and embedding layers to efficiently accommodate categorical protein information, increasing the model's ability for learning from semantic and structural representations.

3. Results

3.1 Summary of Techniques

Table 1 illustrates the important distinctions within the IBOM and AISSO models on a variety of essential characteristics. The IBOM model's main technique is based on metaheuristic optimization, which includes methods such as GA, WOA, ACO, and MVO. Conversely, AISSO uses a hybrid deep learning technique, combining AISSO algorithm and deep learning algorithms such as IRNN and DBN.

In respect to dataset focus, IBOM emphasizes binary human protein interaction data, whereas AISSO is intended to operate with sequence-based PPIs enhanced with semantic characteristics. In terms of interpretation, IBOM delivers greater transparency because of its modular and rule-based optimization structure, whereas AISSO offers moderate interpretation due to the complex nature of its deep learning elements.

Finally, in the overall scheme of semantic enrichment, AISSO includes GO concepts to increase biological relevance, which is missing from the IBOM model. This semantic integration enables AISSO to more accurately detect operational and contextual links between proteins, providing an advantage in knowledge-based prediction scenarios.

Table 1 – Summary of Techniques.

Feature	IBOM	AISSO
Core Method	Metaheuristic Optimization	Hybrid Deep Learning
Algorithms Used	GA, WOA, ACO, MVO	AISSO (Aquila + SSO), IRNN, DBN
Dataset Focus	Human Proteins	Sequence-based PPIs
Interpretability	High	Moderate
Semantic Enrichment	No	Yes (GO terms)

3.2 Performance Metrics

Both models' performance parameters were measured using accuracy, precision, sensitivity, specificity, and the F1-score, as shown in table 2 along with the comparison bar chart in Fig. 3. These indicators offer an extensive viewpoint on each model's prediction abilities:

Accuracy: IBOM obtained 98.61% accuracy, demonstrating great overall consistency in classifying interactions. AISSO achieved approximately 88% accuracy, which, while lower than IBOM, is nevertheless regarded strong given the difficulty of sequence-based PPI predicting tasks.

Precision: The model's capability to find meaningful occurrences. IBOM achieved a near-perfect precision of 99.30%, which is especially useful in scenarios where false positives must be reduced. AISSO's precision is said to be high, however it is not quantitatively stated in its source, indicating a need for additional analysis.

Sensitivity (Recall): IBOM had a sensitivity of 88.64%, indicating it accurately identified a significant number of actual interactions. AISSO was moderately sensitive, indicating an equilibrium between detecting interactions while preventing misclassification.

Specificity: IBOM demonstrated 96.59% specificity in distinguishing between interacting and non-interacting proteins. AISSO also performed well in terms of specificity, because of the use of GO and semantic data to interpret biological roles.

F1-Score: IBOM had an F1-score of 92.31%, indicating a high precision and recall. AISSO did not provide an exact F1-score, but showed excellent results based on qualitative assessments of its semantic-driven feature selection.

Table 2 – Summary of Techniques.

Metric	IBOM	AISSO
Accuracy	98.61%	~88%
Precision	99.30%	High
Sensitivity	88.64%	Moderate
Specificity	96.59%	High
F1-Score	92.31%	Not stated
Error Rate	10.80%	Not stated

Fig. 3 – Comparison of Model 1 and 2 on the basis of performance metrics.

3.3 Computational Efficiency

The IBOM model is highly efficient since it is implemented with MATLAB's optimized libraries and relies on deterministic optimization procedures. These characteristics contribute to a shorter training duration—usually under 20 minutes—and a memory usage of less than 4 GB. This makes IBOM ideal for usage in contexts with low computing resources. In comparison, the AISSO model, while superior in semantic feature representation, is highly computational. It typically takes 4-6 hours of training time, GPU acceleration for neural components like IRNN and DBN, and memory use ranging from

12 to 16 GB. Despite the increased resource demand, AISSO's capacity to incorporate biologically rich, ontology-based information validates its use in high-stakes and research-intensive informatics applications. The comparison is shown in table 3.

Table 3 – Computational Efficiency Comparison.

Parameter	**IBOM**	**AISSO**
Implementation Tool	MATLAB R2020a	MATLAB R2020a
Memory Footprint	Low (≤ 4 GB)	High (12–16 GB)
Training Time	Short (≤ 20 minutes)	Long (4–6 hours)
Hardware Requirement	Standard CPU system	High-end GPU-supported system
Model Complexity	Shallow, rule-based optimization	Deep hybrid architecture with semantic feature fusion
Scalability	High for structured datasets	High for complex, high-dimensional semantic data
Best Use Case	Lightweight bioinformatics applications	Research-intensive, semantic-rich biomedical modeling

3.4 Adaptability

IBOM is adaptable to different structured biological datasets. Its optimization algorithms and modular pipeline allow for easy adaptation and extension with minimum retraining. In contrast, AISSO excels in semantically rich, unstructured, and complex data settings.

IBOM offers improved interpretability, rapid prototyping, and precise forecasts for clean, structured protein datasets.

AISSO excels in deep biological insight, semantic integration, and PPI prediction tasks with complicated annotations like gene ontology or functional pathways.

Together, the results show a clear trade-off: IBOM offers a realistic solution for rapid and accurate analysis, whereas AISSO focuses on depth, learning complexity, and generalization, although at the expense of computing demand.

4. Discussion

4.1 Strengths and Limitations

The IBOM model's main strength is its ease of use, accessibility, and outstanding accuracy on structured datasets. The use of widely recognized optimization methods such as WOA, GA, ACO, and MVO offer adaptability and versatility across a wide range of biological datasets. Its low computing requirement makes it ideal for locations with limited resources, such as academic labs and diagnostic centres. Furthermore, the inclusion of 5-fold cross-validation improves generalizability while minimizing overfitting. However, IBOM's reliance on structured and pre-engineered features may limit its usefulness when interacting with unstructured or semantically complicated data sets.

In contrast, AISSO combines powerful deep learning models with biologically rich data like gene ontology annotations and semantic similarity measurements. This enables AISSO to uncover deeper contextual associations between proteins, leading to robust generalisation powers even with noisy or high-dimensional datasets. Its score-level combination of deep belief networks and recurrent neural networks enables it to simulate both sequential dependencies and high-level abstractions. The trade-off includes increased processing costs and lower transparency, making AISSO less interpretable and more difficult to implement in resource-constrained contexts. Table 4 shows the strengths and limitations of both the models.

Table 4 – Strength and Limitations of Model 1 and 2.

Aspect	IBOM	AISSO
Strength	High accuracy with traditional methods	Advanced learning with semantic depth
Limitation	Less adaptable to unstructured inputs	-
Aspect	IBOM	AISSO
Strength	High accuracy with traditional methods	Advanced learning with semantic depth
Limitation	Less adaptable to unstructured inputs	-

4.2 Comparative Insights

The comparison of IBOM and AISSO indicates complementing benefits that are associated with diverse practical contexts in bioinformatics. IBOM excels in situations where fast, reliable predictions are required and structured datasets are easily available. It is suited for use cases such as high-throughput screening, biomarker identification, and circumstances that require interpretability.

On the other hand, AISSO is extremely useful in exploratory or research-driven applications that require a deeper understanding of the biological semantics of protein activity and interactions. Its architecture is ideal for integrative omics research, complicated PPI network design, and annotation-rich biomedical datasets.

From a deployment standpoint, IBOM's ease of adoption and maintenance enables routine and repeatable procedures. While AISSO necessitates more sophisticated computing settings and skills, it is best suited for cutting-edge research and applications where the depth of biological interpretation warrants the computational expenditure.

Conclusion and Future Work

This comparative investigation of the IBOM and AISSO models shows that both have great potential for improving PPI prediction in bioinformatics. IBOM provides a practical, effective approach with excellent interpretability and performance on well-structured datasets. Its dependence on traditional optimization approaches makes it simple to implement and fine-tune in a variety of biological applications.

In contrast, AISSO offers a novel paradigm for incorporating semantically rich biological information into a deep learning pipeline. Although it requires more computer power and has interpretability issues, it excels at capturing the intricate, nonlinear relationships seen in sequence-based and ontology-enriched datasets.

Future study could investigate the combination of these two techniques, incorporating AISSO's biologically inspired feature extraction into IBOM's optimization framework. This could result in a new class of interpretable, semantically powerful models. Furthermore, verifying our models against multi-species datasets, real-world experimental PPI networks, and future drug-target interactions would demonstrate their value and robustness in translational bioinformatics.

References

Ahmed, N. Y., Alsanousi, W. A., Hamid, E. M., Elbashir, M. K., Al-Aidarous, K. M., Mohammed, M., & Musa, M. E. M. (2024). An Efficient Deep Learning Approach for DNA-Binding Proteins Classification from Primary Sequences. International Journal of Computational Intelligence Systems, 17(1). https://doi.org/10.1007/s44196-024-00462-3

Aksamit, N., Hou, J., Li, Y., & Ombuki-Berman, B. (2024). Integrating transformers and many-objective optimization for drug design. BMC Bioinformatics, 25(1), 1–25. https://doi.org/10.1186/s12859-024-05822-6

Dang, T. H., & Vu, T. A. (2024). xCAPT5: protein–protein interaction prediction using deep and wide multi-kernel pooling convolutional neural networks with protein language model. BMC Bioinformatics, 25(1), 1–20. https://doi.org/10.1186/s12859-024-05725-6

Gonzalez-Lopez, F., Morales-Cordovilla, J. A., Villegas-Morcillo, A., Gomez, A. M., & Sanchez, V. (2018). End-to-end prediction of protein-protein interaction based on embedding and recurrent neural networks. 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2344–2350. https://doi.org/10.1109/BIBM.2018.8621328

Gündüz, H. A., Mreches, R., Moosbauer, J., Robertson, G., To, X. Y., Franzosa, E. A., Huttenhower, C., Rezaei, M., McHardy, A. C., Bischl, B., Münch, P. C., & Binder, M. (2024). Optimized model architectures for deep learning on genomic data. Communications Biology, 7(1), 1–10. https://doi.org/10.1038/s42003-024-06161-1

Hu, W., & Ohue, M. (2024). SpatialPPI: Three-dimensional space protein-protein interaction prediction with AlphaFold Multimer. Computational and Structural Biotechnology Journal, 23(November 2023), 1214–1225. https://doi.org/10.1016/j.csbj.2024.03.009

Kumar, A. (2017). Predictive Maintenance. Icces, 725–729.

Lee, J. W., Won, J. H., Jeon, S., Choo, Y., Yeon, Y., Oh, J. S., Kim, M., Kim, S., Joung, I., Jang, C., Lee, S. J., Kim, T. H., Jin, K. H., Song, G., Kim, E. S., Yoo, J., Paek, E., Noh, Y. K., & Joo, K. (2023). DeepFold: Enhancing protein structure prediction through optimized loss functions, improved template features, and re-optimized energy function. Bioinformatics, 39(12). https://doi.org/10.1093/bioinformatics/btad712

Mohammad-Azari, S., Bozorg-Haddad, O., & Chu, X. (2018). Shark Smell Optimization (SSO) Algorithm BT - Advanced Optimization by Nature-Inspired Algorithms (O. Bozorg-Haddad (ed.); pp. 93–103). Springer Singapore. https://doi.org/10.1007/978-981-10-5221-7_10

Soleymani, F., Paquet, E., Viktor, H., Michalowski, W., & Spinello, D. (2022). Protein–protein interaction prediction with deep learning: A comprehensive review. Computational and Structural Biotechnology Journal, 20, 5316–5341. https://doi.org/10.1016/j.csbj.2022.08.070

Taha, K. (2025). Protein-protein interaction detection using deep learning: A survey, comparative analysis, and experimental evaluation. Computers in Biology and Medicine, 185(August 2024), 109449. https://doi.org/10.1016/j.compbiomed.2024.109449

Thareja, P., Chhillar, R. S., Dalal, S., Simaiya, S., Lilhore, U. K., Alroobaea, R., Alsafyani, M., Baqasah, A. M., & Algarni, S. (2024). Intelligence model on sequence-based prediction of PPI using AISSO deep concept with hyperparameter tuning process. Scientific Reports, 14(1), 1–22. https://doi.org/10.1038/s41598-024-72558-x

Tran, H. N., Nguyen, P. X. Q., Guo, F., & Wang, J. (2024). Prediction of Protein–Protein Interactions Based on Integrating Deep Learning and Feature Fusion †. International Journal of Molecular Sciences, 25(11). https://doi.org/10.3390/ijms25115820

Wu, J., Liu, B., Zhang, J., Wang, Z., & Li, J. (2023). DL-PPI: a method on prediction of sequenced protein–protein interaction based on deep learning. BMC Bioinformatics, 24(1), 1–21. https://doi.org/10.1186/s12859-023-05594-5