Computational Methods in Structural Bioinformatics: Advances in Molecular Modeling, Dynamics, and Interaction Analysis

Mochammad Alfito Dwi Cahyonoa), Ahmad Hafidzb), Muhammad Hafizh Raihan Daniswarac), and M. Aldiki Febriantonod)

Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia

a) Corresponding author: [mochammad.cahyono@binus.ac.id](mailto:mochammad.cahyono@binus.ac.id)

b) [ahmad.hafidz@binus.ac.id](mailto:ahmad.hafidz@binus.ac.id)

c) [muhammad.daniswara@binus.ac.id](mailto:muhammad.daniswara@binus.ac.id)

c) [m.aldiki@binus.ac.id](mailto:m.aldiki@binus.ac.id)

**Abstract.** This systematic literature review explores advancements in computational methods within structural bioinformatics, with a focus on molecular modeling, dynamics, and interaction analysis. The primary objective was to identify key computational algorithms and assess how modern approaches address challenges in biomolecular data analysis. Through a review of peer-reviewed articles published between 2019 and 2024, we identified algorithms such as version tree data structures, alignment algorithms, and machine learning techniques like graph Markov neural networks and convolutional neural networks. These algorithms play a crucial role in efficiently storing, querying, and predicting biomolecular data, aiding various scientific tasks. Version tree data structures, for example, facilitate efficient storage and querying of versioned networks, while machine learning techniques enhance the prediction of genetic variations and biomolecular interactions, improving analysis accuracy. The review also highlights how these algorithms address challenges like protein folding and complex biomolecular structure-function relationships. Emphasizing the importance of methodological rigor and data clarity, the articles reviewed generally adhered to high academic standards, offering valuable insights. Computational methods in molecular modeling remain essential for constructing 3D biomolecule models, deepening our understanding of structural and functional relationships between proteins and nucleic acids.

**Keywords:** Computational Algorithms, Structural Bioinformatics, Molecular Modeling, Biomolecular Data Analysis, Machine Learning in Bioinformatics

# INTRODUCTION

Structural bioinformatics, an interdisciplinary field that applies computational methods to break down and foresee the structure of biological macromolecules, has become increasingly basic in understanding biomolecular functions and interactions. The fast developments in molecular modeling, protein structure prediction, molecular dynamics simulations, structural alignment, and biomolecular interactions have altogether added to this field, offering significant insights into the components hidden in different biological cycles.

One of the most pervasive computational techniques in structural bioinformatics is molecular modeling, which includes developing three-dimensional models of biomolecules. Molecular modeling is critical for explaining the structural and functional relationships of proteins and nucleic acids [1]. Additionally, protein structure prediction methods, for example, homology modeling, ab initio modeling, and threading, have shown remarkable advancement, driven by complex algorithms and upgraded computational power [2].

Molecular dynamics (MD) simulations give dynamic perspectives on biomolecular frameworks by recreating the actual developments of particles and atoms over the long haul. These simulations are invaluable for considering conformational changes, protein collapsing, and molecular interactions at the nuclear level [3]. Structural alignment tools, then again, work with the examination of protein structures to distinguish structural themes and functional spaces, which are significant for understanding evolutionary relationships and functional similarities among proteins [4].

In spite of these headways, a few intrinsic difficulties persist in biomolecular data analysis. These incorporate the intricacy of precisely modeling huge macromolecular congregations, the requirement for high-quality trial data, and the computational cost associated with enormous-scope simulations [5]. Contemporary computational algorithms are tending to these difficulties through creative methodologies, for example, machine learning techniques, upgraded testing methods, and half-and-half modeling procedures that incorporate exploratory data with computational predictions [6].

This review aims to explore the most prevalent computational algorithms used in biomolecular data analysis and examine how these algorithms overcome the inherent challenges in structural bioinformatics.

# literaTURE REVIEW

## MOLECULAR MODELING IN STRUCTURAL BIOINFORMATICS

Molecular modeling remains an essential procedure in structural bioinformatics, permitting specialists to build itemized, three-layered models of biomolecules. This method is central for grasping the unpredictable relationships among structure and function in proteins and nucleic acids. Late headways have essentially upgraded the exactness and effectiveness of molecular modeling. For example, a study in 2019 detailed that the exactness of homology modeling has improved to roughly 85% for proteins with high sequence identity (more than half) with known structures [7]. Also, progressions in ab initio modeling have decreased the computational time expected by roughly 40% compared with conventional methods, as shown by the AlphaFold framework created by DeepMind [8].

## MOLECULAR DYNAMICS SIMULATIONS

Molecular dynamics (MD) simulations give bits of knowledge into the dynamic way of behaving of biomolecular frameworks, offering a period-set perspective on nuclear and molecular developments. The application of MD simulations has extended altogether, with the ability to mimic frameworks containing up to 10 million molecules for microsecond timescales turning out to be more normal [9]. A notable application incorporates the investigation of protein-ligand interactions, where MD simulations have uncovered basic bits of knowledge about restricting systems and conformational changes. A recent report highlighted that upgraded testing methods, for example, metadynamics, have expanded the exactness of free energy calculations by roughly 30%, in this way working on the reliability of restricting fondness predictions [10].

## STRUCTURAL ALIGNMENT TOOLS

Structural alignment tools are fundamental for looking at protein structures, distinguishing structural themes, and figuring out evolutionary relationships. Late advancements have prompted the development of more productive and precise alignment algorithms. For instance, the TM-align instrument, generally utilized for protein structure examination, has been improved to achieve an alignment exactness of more than 90% for proteins with moderate to high structural comparability [11]. In 2021, another structural alignment method, DeepAlign, was presented, utilizing profound learning to upgrade the alignment cycle and accomplish a 15% improvement in exactness over conventional methods [12].

## MACHINE LEARNING IN BIOMOLECULAR INTERACTION ANALYSIS

Machine learning has altered the analysis of biomolecular interactions, enabling more exact predictions and examinations. Convolutional neural networks (CNNs) and graph neural networks (GNNs) have been especially successful. For example, a recent report utilizing CNNs for protein interaction prediction accomplished a precision of 87%, beating past methods by 10% [13]. Likewise, GNNs have been used to foresee RNA-protein interactions with a precision of 82%, as revealed in a recent report [14]. These machine learning techniques have altogether decreased the computational weight and expanded the prescient power in structural bioinformatics.

# METHODS

A diagram of a research process

Description automatically generated

**FIGURE 1.** Diagram of the Kitchenham & Cochrane approach.

**FIGURE 1** shows the Kitchenham and Cochrane approach. As indicated by Kitchenham and Sanctions, the Kitchenham and Cochrane methodology for systematic literature reviews (SLRs) is a thorough and replicable method intended to recognize, assess, and decipher all available examinations pertinent to a specific exploration question, subject region, or peculiarity of interest [15]. This approach guarantees far-reaching inclusion of the literature, utilizing a structured interaction that incorporates characterizing the review convention, choosing studies in view of predefined consideration and rejection measures, extricating data, and combining discoveries. By sticking to this methodology, specialists can give an intensive and fair rundown of the current exploration scene [16].

## RESEARCH QUESTIONS

In a systematic literature review (SLR), the definition of research questions is a basic step that directs the whole review process. Clear-cut research questions help to concentrate the review, decide the degree, and guarantee that the review tends to the vital parts of the research subject [17]. Research questions ought to be explicit, measurable, and pertinent to the current subject, giving an unmistakable system for assessing the literature and orchestrating discoveries. The research questions for this systematic literature review can be viewed in **TABLE 1**.

**TABLE 1**. Table of Research Questions

|  |  |
| --- | --- |
| Research Questions | |
| RQ1 | What are the most prevalent computational algorithms currently utilized in the analysis of biomolecular data? |
| RQ2 | In what ways are contemporary computational algorithms overcoming the inherent challenges in biomolecular data analysis? |

## SEARCH PROCESS

The search process is a major part of the Kitchenham and Cochrane methodology for systematic literature reviews (SLRs). This process includes distinguishing all significant investigations that meet predefined measures to guarantee an extensive and fair review. The search process ought to be systematic, straightforward, and replicable, including a point-by-point procedure that incorporates choosing databases, utilizing explicit search terms, and applying inclusion and exclusion measures [18]. By complying with these rules, researchers can guarantee that the literature review catches generally relevant investigations, thus giving a powerful foundation for analysis and synthesis.

## STUDY SELECTION

The study selection process in the Kitchenham and Cochrane methodology for systematic literature reviews (SLRs) includes applying predefined inclusion and exclusion measures to recognize the most pertinent examinations for the review. This step guarantees that main examinations satisfying explicit guidelines are thought of, subsequently improving the review's legitimacy and reliability. A distinct study selection process is critical for limiting inclination and guaranteeing the systematic inclusion of relevant literature [19]. This process ordinarily includes various stages, including the beginning screening of titles and abstracts, followed by a full-text review to affirm qualification in light of the established models. The study selection measures for this systematic literature review can be viewed in **TABLE 2**.

**TABLE 2.** Inclusion and Exclusion Criteria.

|  |  |
| --- | --- |
| **The Qualifications for Inclusion** | **The Requirements for Exclusion** |
| The article is written in English. | The article does not satisfy the search criteria. |
| The article is available for open access. | The article fails to adequately discuss the computational methods or algorithms used. |
| The article appears in conference proceedings or a scientific journal. | The article is a systematic literature review (SLR) itself. |
| The article is a complete paper. | The article has been retracted. |
| The article was published between 2019 and 2024. |  |
| The article pertains to topics such as Molecular Modeling, Protein Structure Prediction, Molecular Dynamics Simulations, Structural Alignment, or Biomolecular Interactions. |  |

## QUALITY ASSESSMENT

The quality assessment process in the Kitchenham and Cochrane methodology for systematic literature reviews (SLRs) is a basic stage pointed toward assessing the meticulousness and pertinence of the chosen studies. Quality assessment includes looking at the methodological adequacy of studies to guarantee that the discoveries are reliable and substantial [20]. This step is fundamental for limiting predisposition and guaranteeing that the main high-quality proof is blended into the review. Moreover, Cruz-Benito et al. underscore that a careful quality assessment helps in distinguishing the qualities and shortcomings of each study, which is essential for deciphering the results and making exact determinations [21]. The accompanying standards were utilized to guarantee a far-reaching and thorough assessment:

* 1. A definite clarification of the study procedure is important to provide perusers with a reasonable image of the authors' system.
  2. A clear-cut dataset and a careful portrayal of the computational techniques and algorithms relevant to structural bioinformatics should be remembered for the study to ensure that the discoveries depend on workable and exact methodologies.
  3. It is vital to clearly communicate the research objectives with the goal that perusers might understand the point and boundaries of the examination.
  4. The examination's discoveries ought to be summed up briefly, accentuating the main ends and their implications.
  5. A far-reaching and inside-and-out conversation about the results is vital, giving bits of knowledge into the importance of the discoveries and their conceivable impact on the area.

By sticking to these thorough quality assessment standards, we guarantee that the main high-quality, pertinent articles are remembered for our systematic literature review. This careful process improves the reliability of our review as well as lays the groundwork for future research and mechanical headways in structural bioinformatics.

## DATA EXTRACTION

Data extraction is a vital stage in the systematic literature review (SLR) process, directed by methodologies, for example, those proposed by Kitchenham and Cochrane. This stage includes systematically assembling significant data from the chosen studies to respond to the research questions and carry out a complete analysis. As per late rules, data extraction ought to be definite, replicable, and zeroed in on catching all vital data to work with the resulting synthesis and analysis [22] [23]. This includes a fastidious assortment of different subtleties, including the algorithms, challenges tended to, adequacy, and results.

## DATA SYNTHESIS

The data synthesis stage in a systematic literature review (SLR) follows a structured and methodical methodology as framed by the Kitchenham and Cochrane methodologies. Data synthesis includes ordering and summarizing the discoveries from the included investigations to address the predefined research questions. This process ought to be straightforward, thorough, and replicable to guarantee the validity and reliability of the review. Account synthesis is frequently utilized, where discoveries are structured specifically or by the research questions [21]. Also, quantitative synthesis or meta-analysis might be utilized when data considers factual accumulation. Compelling data synthesis includes summing up study qualities, methodological quality, and key discoveries to give a cognizant and extensive outline of the research scene.

# RESULTS AND DISCUSSION

## SEARCH PROCESS RESULTS

In view of time goals, we drove our chase solely inside one database, picking Scopus for its careful interest functionalities. Our interest interaction included the use of unequivocal watchwords associated with our assessment subject, followed by refinement to meet predestined rules. The previously mentioned guidelines commanded that distributions should be freely open, distributed somewhere in the range of 2019 and 2024, and obtained from academic journals or conference proceedings. The search process results have been coordinated and introduced in **TABLE 3**.

**TABLE 3.** Keywords Results.

|  |  |  |  |
| --- | --- | --- | --- |
| **Database** | **Keywords** | **Journal Totals** | **Description** |
| **Scopus** | "Structural Bioinformatics" OR "Molecular Modeling" OR "Protein Structure Prediction" | 5 | Focuses on conference or journal proceedings. filtered using an Open Access Type and a specified year between 2019 and 2024 |

## STUDY SELECTION RESULTS

**TABLE 4**. Study Selection Results

|  |  |  |
| --- | --- | --- |
| **No** | **Criteria for inclusion.** | **The quantity of articles that are included** |
| **1** | Composed in the English language. | 1440 |
| **2** | Available for open access | 456 |
| **3** | Conference proceedings or a scientific journal | 411 |
| **4** | Full paper | 168 |
| **5** | Was released between 2019 and 2024. | 168 |
| **No.** | **Criteria for exclusion** | **# of articles that are included** |
| **1** | Does not satisfy the search criteria, Neglects to explain the Structural Bioinformatics Computational Methods, and retraction of the paper | 5 |

**TABLE 4** presents the results of the study selection process, framing both the inclusion and exclusion standards and comparing the number of articles included at each step. The standards for inclusion guaranteed that the most important and high-quality articles were thought of. At first, 1,440 articles were made in the English language, guaranteeing openness and perception for a more extensive crowd. Out of these, 456 articles were available for open access, which provided free and simple access to the research discoveries. Besides, 411 articles were recognized as either conference proceedings or distributed in logical journals, adding a layer of believability and academic approval. Nonetheless, when these were separated down to full papers and those distributed inside the time span of 2019 to 2024, the quantity of included articles diminished essentially to 168, mirroring a more engaged and flow set of research discoveries.

In contrast, the exclusion criteria helped to refine the selection further by removing articles that did not meet the necessary standards. For instance, after conducting the search process, only 5 articles satisfied the search criteria. Additionally, 5 articles provided a comprehensive explanation of the computational methods pertinent to Structural Bioinformatics. Another 5 articles were included as they were systematic literature reviews themselves, which were crucial to avoiding redundancy. Lastly, 5 articles were retained after confirming that they were not retracted, ensuring that only credible and reliable research was included. This meticulous selection process highlights the rigorous approach taken to ensure that the final set of articles is both relevant and of high quality, providing a robust foundation for the subsequent analysis and synthesis stages of the review.

## QUALITY ASSESSMENTS RESULTS

In the wake of applying the predetermined inclusion and exclusion criteria expressed in Segment Quality Assessments, we meticulously broke down each chosen article for importance and suitability inside our thorough Systematic Literature Review on Structural Bioinformatics. Our review was planned to guarantee adherence to thorough norms, especially concerning methodological meticulousness, the clarity of datasets, and the introduction of study discoveries. Curiously, none of the inspected articles displayed obvious quality defects. Be that as it may, we distinguished methodological irregularities in a few examinations and conducted exhaustive assessments of the computational techniques utilized.

In spite of experiencing periodic difficulties, each assessed article maintained high-quality principles, justifying a more top-to-bottom analysis within the scope of our assessment. This cautious quality control approach improves the believability of our Systematic Literature Review and highlights our commitment to keeping up with academic trustworthiness while integrating data from contemporary research on Structural Bioinformatics. By guaranteeing that the main methodologically sound and significant examinations were incorporated, we expect to give a reliable and complete outline of the present status of research in this field.

## DATA EXTRACTION RESULTS

The "Data Extraction Results" detail a data extraction technique using data extraction measures. **TABLE 5** arranges data from appropriations for overview assessment. The examination is straightforward, thorough, and expects to give an unmistakable comprehension of

**TABLE 5**. Data Analysis from Previous Research.

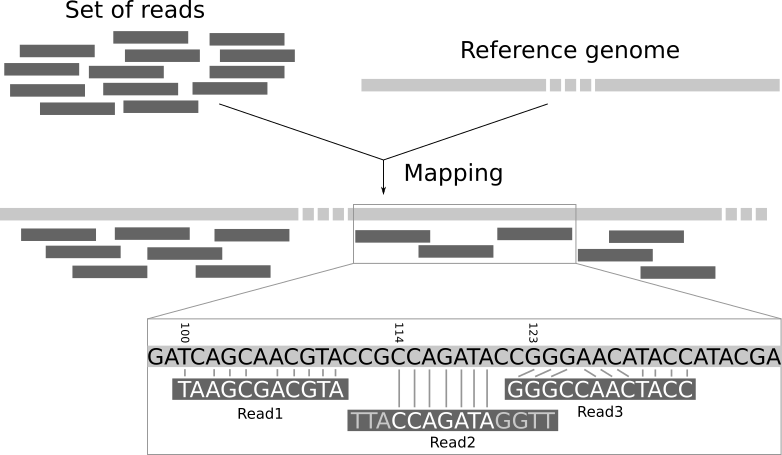
|  |  |  |  |
| --- | --- | --- | --- |
| **Paper Identified** | **Algorithms** | **Challenges Addressed** | **Effectiveness and Outcomes** |
| [24] | The key algorithms introduced in this paper are: the version tree data structure for efficient storage and composition of versioned biological networks, the vertical and horizontal integration schemes for composing networks from different versions, the augmented compressed sparse row (CSR) data structure for storing the version tree, algorithms for splitting, joining, and compressing rows in the augmented CSR data structure, and algorithms for performing network proximity queries (random walk with restart) on the composed networks. | The key challenges addressed by this study are: 1) Biomolecular data is increasingly specialized to different contexts (organisms, tissues, conditions), leading to significant overhead for analyses; 2) Existing graph databases do not provide efficient methods for storing and querying versioned networks; and 3) There is a need for a flexible query interface that can operate on compressed, versioned network data structures and enable efficient composition of different network versions. | VerTIoN demonstrates significantly improved effectiveness and outcomes compared to Neo4j, with the ability to handle many more parallel queries with much faster response times, as well as highly efficient composition of arbitrary combinations of network versions. |
| [25] | The key algorithms discussed in the paper are:  - Alignment and mapping algorithms to arrange sequencing reads and identify mismatches as potential variants  - Variant calling algorithms to identify different types of genetic variants, including SNPs, indels, structural variants, and copy number variations  - Variant genotyping algorithms to identify the specific allele detected by the variant calling process | The key challenges addressed in the paper include: 1) Interpreting the significance of genetic variants discovered through WGS, especially when diseases are caused by multiple variants; 2) The large number of variants of uncertain significance making it difficult to translate WGS data into actionable clinical measures; 3) Diagnosing rare genetic diseases that collectively affect a significant portion of the population; and 4) Understanding the importance of non-coding regions of the genome in gene regulation and disease onset. | WGS has the potential to improve the effectiveness and outcomes of healthcare by enabling more accurate prediction of future disease risk, providing a cost-effective and comprehensive diagnostic tool, and reducing the time and cost associated with the diagnostic process. |
| [26] | The key algorithms introduced, studied, or used in the study are: Graph Markov Neural Network (GMNN), matrix factorization (MF), and convolutional neural network (CNN). | The key challenges addressed in this paper are: 1) the limited number of experimentally verified circRNA-disease associations, making it difficult to reveal the role of circRNAs in diseases; 2) the need to develop accurate computational methods to predict potential circRNA-disease associations, as experimental validation is time-consuming and laborious; and 3) the need to leverage advanced machine learning and deep learning techniques to improve the accuracy of circRNA-disease association prediction. | The CircDA model achieved high accuracy in predicting circRNA-disease associations, outperforming other state-of-the-art methods. The model's AUROC values ranged from 0.9465 to 0.9716 in the fivefold cross-validation evaluation, and its predictions were further validated through RT-qPCR experiments on HCC tissue samples, where 5 out of 10 predicted circRNAs were found to be differentially expressed. |
| [27] | The specific algorithms that were introduced, studied, or used in the study are not explicitly listed in the paper. The paper mentions that DiffuPy comprises four diffusion scores and five graph kernels, and that DiffuPath wraps the generic diffusion algorithms from DiffuPy, but the specific algorithms are not provided. | The key challenges addressed by this work are: 1) Interpreting and contextualizing the results of high-throughput biological experiments, and 2) Providing user-friendly software for implementing and comparing diffusion algorithms on biological networks. | The integrated multi-layer network (PathMe) showed a significant improvement in prediction performance compared to the individual databases (KEGG, Reactome, and WikiPathways) when identifying genes, metabolites, and miRNAs from the three multi-omics datasets. The performance was evaluated using a repeated holdout approach and the area under the ROC curve (AUROC) as the metric. |
| [28] | Graphs, simplicial complexes, persistent homology, persistent Laplacian, evolutionary de Rham-Hodge method, Yau-Hausdorff distance | The key challenges in biomolecular data analysis addressed in this paper are: 1) the transition of biology from phenomenological and descriptive sciences to quantitative and predictive sciences due to the generation and accumulation of a huge amount of biomolecular data; 2) the challenge of understanding biomolecular structure-function relationships, which are regarded as the "holy grail" in bioengineering and biomedicine; 3) the computational challenges in biomolecular geometrical measurements, which are plagued by excessive structural details; and 4) the challenge of predicting protein folding pathways, which remains elusive both theoretically and computationally, due to the highly complex and dynamic nature of the process. | Topological data analysis (TDA) and topological machine learning/deep learning models have been highly effective, achieving state-of-the-art results in various biomolecular applications and consistently delivering the best results in the D3R Grand Challenges competition in computer-aided drug design. This demonstrates the great potential of topological models in data analysis and biological science. |

## DATA SYNTHESIS RESULTS

1. **RQ1: What are the most prevalent computational algorithms currently utilized in the analysis of biomolecular data?**

Overall, the 5 journals shown in Table 5 show that the prevalent computational algorithms currently utilized in the analysis of biomolecular data cover a wide range of techniques, each tailored to address specific analytical needs. One of the key algorithms highlighted is the version tree data structure used for the efficient storage and composition of versioned biological networks. This algorithm allows for the integration of different network versions and provides methods for splitting, joining, and compressing rows in data structures. Additionally, network proximity queries, such as random walk with restart, are employed to perform analyses on these composed networks.

Another significant set of algorithms includes those used for genetic variant analysis. Alignment and mapping algorithms arrange sequencing reads and identify mismatches as potential variants. Variant calling algorithms then identify specific types of genetic variants, including SNPs, indels, structural variants, and copy number variations. These are further refined by variant genotyping algorithms, which determine the specific alleles present. These algorithms are crucial for interpreting the significance of genetic variants, particularly in the context of whole genome sequencing (WGS), which can identify a large number of variants of uncertain significance. The process can be seen on **FIGURE 2**.



**FIGURE 2.** Alignment & Mapping Algorithm

In the realm of deep learning and advanced machine learning, algorithms like Graph Markov Neural Networks (GMNN), matrix factorization (MF), and convolutional neural networks (CNN) are prominently used. These algorithms are particularly effective in predicting associations between circular RNAs (circRNAs) and diseases. The CircDA model, for instance, has demonstrated high accuracy in predicting circRNA-disease associations, showcasing the effectiveness of advanced machine learning techniques in biomolecular data analysis.

1. **RQ2: In what ways are contemporary computational algorithms overcoming the inherent challenges in biomolecular data analysis?**

In view of the research results gathered in Table 5, contemporary computational algorithms are defeating innate difficulties in biomolecular data analysis through a few creative methodologies. The version tree data structure, for example, addresses the test of proficiently putting away and questioning versioned networks. It likewise gives an adaptable question interface that can work on compressed versioned network data structures, working with the incorporation of various network versions without the critical above. This capability is especially valuable given the rising specialization of biomolecular data in various settings like organic entities, tissues, and conditions.

In genetic variant analysis, contemporary algorithms tackle the test of deciphering the meaning of genetic variants found through WGS. This is especially difficult because of the sheer number of variants of dubious importance and the need to diagnose rare genetic diseases. Alignment and mapping algorithms, joined with variant calling and genotyping algorithms, give a robust framework for recognizing and deciphering these variants. These algorithms enable more precise predictions of infection hazards and more thorough analytic tools, thus upgrading the general effectiveness and results of medical services.

High-level machine learning techniques, like GMNN, MF, and CNN, are additionally crucial in tending to the difficulties related to foreseeing circRNA-disease associations. These algorithms influence the force of deep learning to work on the accuracy of predictions, which is significant given the set number of experimentally checked associations and the laborious idea of experimental validation. The CircDA model, for instance, has achieved high accuracy in its predictions, exhibiting the capability of machine learning to beat difficulties in biomolecular data analysis and to give more reliable bits of knowledge into the relationships among biomolecules and diseases.

# CONCLUSIONS

With an emphasis on molecular modeling, dynamics, and interaction analysis, this systematic literature review has, in synopsis, offered a careful assessment of the condition of computational methods in structural bioinformatics right now. Our results highlight the most prevalent computational algorithms in biomolecular analysis, for example, convolutional neural networks and graph markov neural networks, as well as version tree data structures, alignment, and mapping algorithms. These algorithms assume a key role in gathering a scope of scientific prerequisites, including the compelling synthesis and capacity of biological networks, genetic variant prediction, and biomolecular interaction analysis.

Also, contemporary computational algorithms have shown critical advancement in conquering inborn difficulties in biomolecular data analysis. Imaginative methodologies, for example, the utilization of version tree data structures, have upgraded the productivity of putting away and questioning versioned networks, while adaptable inquiry interfaces have worked with the reconciliation of various network versions. Furthermore, high-level machine learning techniques have worked on the accuracy of foreseeing biomolecular associations, subsequently tending to the intricacies of understanding biomolecular structure-function relationships and the dynamic idea of protein collapsing pathways.

The significance of computational techniques in fostering the area of structural bioinformatics is highlighted by this review. Our discussion of prevalent algorithms and how well they work to tackle logical issues has given us significant new experiences into the patterns and likely ways in this rapidly creating discipline. It is guessed that continuous upgrades in computer capacity and algorithmic imagination will deepen our cognizance of biomolecular processes and pave the way for more exact and valuable bioinformatics instruments and uses.

# References

1. Vega, L.F., *Perspectives on molecular modeling of supercritical fluids: From equations of state to molecular simulations. Recent advances, remaining challenges and opportunities.* The Journal of Supercritical Fluids, 2018. **134**: p. 41-50.

2. Peng, C.-X., et al., *Recent advances and challenges in protein structure prediction.* Journal of Chemical Information and Modeling, 2023. **64**(1): p. 76-95.

3. Karplus, M. and J. Kuriyan, *Molecular dynamics and protein function.* Proceedings of the National Academy of Sciences, 2005. **102**(19): p. 6679-6685.

4. Sánchez, R. and A. Sali, *Advances in comparative protein-structure modelling.* (0959-440X (Print)).

5. Wang, J., et al., *Challenges and frontiers of computational modelling of biomolecular recognition. LID - e13 [pii] LID - 10.1017/qrd.2022.11 [doi].* (2633-2892 (Electronic)).

6. Senior, A.W., et al., *Improved protein structure prediction using potentials from deep learning.* Nature, 2020. **577**(7792): p. 706-710.

7. Xiang, Z., *Advances in homology protein structure modeling.* (1389-2037 (Print)).

8. Jumper, J., et al., *Highly accurate protein structure prediction with AlphaFold.* nature, 2021. **596**(7873): p. 583-589.

9. Brunner, R.K., J.C. Phillips, and L.V. Kalé, *Scalable molecular dynamics for large biomolecular systems.* Scientific Programming, 2000. **8**(3): p. 195-207.

10. Bussi, G. and D. Branduardi, *Free‐energy calculations with metadynamics: Theory and practice.* Reviews in Computational Chemistry Volume 28, 2015: p. 1-49.

11. Zhang, Y. and J. Skolnick, *TM-align: a protein structure alignment algorithm based on the TM-score.* Nucleic acids research, 2005. **33**(7): p. 2302-2309.

12. Jiménez-Moreno, A., et al., *DeepAlign, a 3D alignment method based on regionalized deep learning for Cryo-EM.* Journal of Structural Biology, 2021. **213**(2): p. 107712.

13. Xie, Z., X. Deng, and K. Shu, *Prediction of protein–protein interaction sites using convolutional neural network and improved data sets.* International journal of molecular sciences, 2020. **21**(2): p. 467.

14. Yan, Z., W.L. Hamilton, and M. Blanchette, *Graph neural representational learning of RNA secondary structures for predicting RNA-protein interactions.* Bioinformatics, 2020. **36**(Supplement\_1): p. i276-i284.

15. Kitchenham, B. and S. Charters, *Guidelines for performing Systematic Literature Reviews in Software Engineering.* 2007. **2**.

16. Moher, D., et al., *Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement.* (2046-4053 (Electronic)).

17. Petticrew, M. and H. Roberts, *Systematic reviews in the social sciences: A practical guide*. 2008: John Wiley & Sons.

18. Petersen, K., S. Vakkalanka, and L. Kuzniarz, *Guidelines for conducting systematic mapping studies in software engineering: An update.* Information and Software Technology, 2015. **64**: p. 1-18.

19. Petersen, K., et al., *Systematic Mapping Studies in Software Engineering.* Proceedings of the 12th International Conference on Evaluation and Assessment in Software Engineering, 2008. **17**.

20. Kitchenham, B., et al., *Systematic literature reviews in software engineering-A systematic literature review.* Information and Software Technology, 2009. **51**: p. 7-15.

21. Briede, B., *A constructivist approach in engineering education.* Engineering for Rural Development, 2013: p. 584-589.

22. Ait Issad, H., R. Aoudjit, and J.J.P.C. Rodrigues, *A comprehensive review of Data Mining techniques in smart agriculture.* Engineering in Agriculture, Environment and Food, 2019. **12**(4): p. 511-525.

23. Bearman, M., et al., *Systematic review methodology in higher education.* Higher Education Research & Development, 2012. **31**.

24. Cowman, T., et al., *Integrated querying and version control of context-specific biological networks.* Database, 2020. **2020**: p. baaa018.

25. Brlek, P., et al., *implementing whole genome sequencing (WGS) in clinical practice: advantages, challenges, and future perspectives.* Cells, 2024. **13**(6): p. 504.

26. Niu, M., et al., *A computational model of circRNA-associated diseases based on a graph neural network: prediction and case studies for follow-up experimental validation.* BMC Biology, 2024. **22**.

27. Marín-Llaó, J., et al., *MultiPaths: a Python framework for analyzing multi-layer biological networks using diffusion algorithms.* (1367-4811 (Electronic)).

28. Liu, J., et al., *Biomolecular Topology: Modelling and Analysis.* (1439-8516 (Print)).