REVIEW ON ROLE OF DEEP LEARNING TECHNIQUES IN PROTEIN STABILITY

A. Jasmine Sugil[1], Dr.K. Merriliance[2], S. Serina Hingis[3]

[1]Research Scholar, Reg.No:23111242282011, Department of Computer Applications & Research Centre, Sarah Tucker College (Autonomous), Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli,-627007, Tamilnadu, India, Email:findsugil@gmail.com

[2]Associate Professor, Department of Computer Applications &Research Centre, Sarah Tucker College (Autonomous), Tirunelveli,-627007, Tamilnadu, India, Email: merriliance@gmail.com

[3]Pharm. D (5th year), Department of Pharmacy, Annamalai University Chidambaram, India, Email:serinahingis@gmail.com

Abstract

In the modern world, technology plays a crucial role in data management. In this regard, bioinformatics is the use of computational and analytical tools to analyze and interpret biological data such as protein sequencing, molecular structure, and DNA sequences. Protein bio molecules are very important to any living organism for proper functioning. Estimate of protein stability is vital and presents challenges to bioinformatics engineers. This is made possible by the advancement of various modern technologies, such as deep learning and machine learning techniques. Deep learning methods have been widely used for protein stability prediction due to their ability to fully exploit the complex relationships between protein sequence, structure, and stability. The unparalleled potential collecting data and analytics has resulted in an increase in multi-disciplinary research within deep learning and biological data. We identify and discuss some of the key data issues, such as protein structure, function, folding, and stability. Further, this article concentrates on the impact of these data issues on various deep learning approaches within the context of protein sequence molecules. Finally, we identify some of the common challenges in deep learning and protein research, including the misapplication of deep learning evaluation techniques.

INDEX TERMS: Bioinformatics, Protein molecules, Protein structure, Protein stability, Deep learning.

1. Introduction

Proteins are important components of any living organism such as the cytoskeleton, mitochondria, and ribosomes, contributing to their structure and function. Protein thermodynamic stability depends on its function and their structure. Moreover, inner hydrophobicity and H-bonding among surface of protein and solvent are known to have major roles in protein stability. The effects of single point residue changes on protein stability can interpret the molecular biology mechanisms of human diseases and help in developing new drugs discovery [1-2]. Several methods have been devised to anticipate alterations in protein stability among variants, utilizing either protein sequence or structural characteristics. Stability refers to the ability of a protein to maintain its native structure and function under different conditions such as hydrophobic interactions, Hydrogen bonding, Disulfide bonds, Salt bridges, Chaperones, Ligand binding, Molecular crowding[Figure 1]. It is important to remember that outside variables such as temperature, pH, and the presence of denaturing chemicals can also have an impact on a protein's stability. Predict the change in the protein stability upon a single amino acid substitution based on sequence information, without the experimentally determined three-dimensional structure [3].

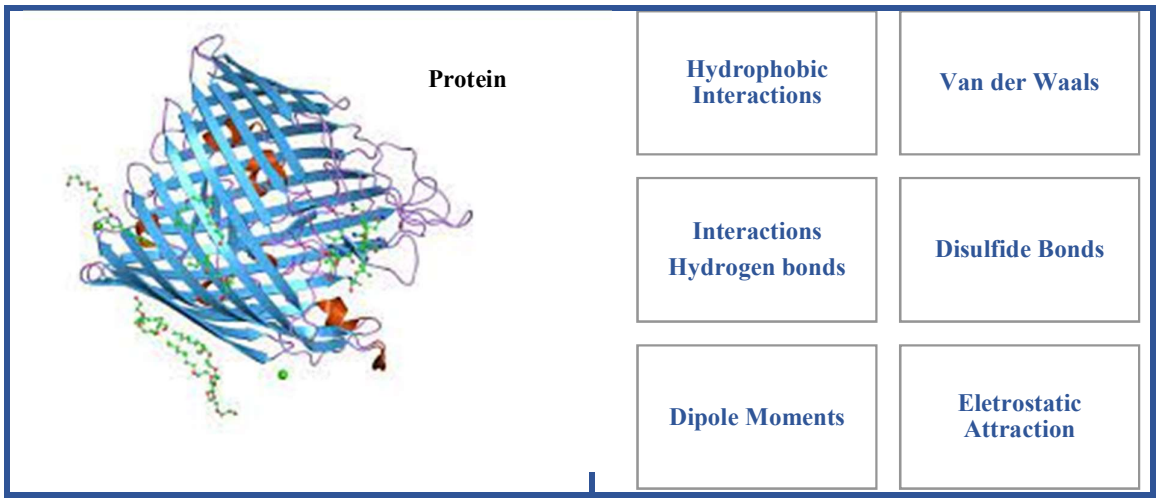

Figure 1. Forces Stabilizing Protein Structure

ELASPIC [4] Generated protein structures are produced by homology modelling using one of the computer prediction approaches for protein structures. The primary result of the web server is the expected adjustment in the Gibbs free energy (DDG) of binding and/or folding for every domain and interface that has been altered by the modification. According to the statement, those modifications will reduce EP300's stability and affinity for the SRC/p160 nuclear receptor coactivator family members NCOA1, NCOA2, and NCOA3, as well as the subsequent hypoxia-regulators HIF1A and EPAS1. The interaction models available on the ELASPIC website illustrate the structural alterations that ends in a decrease in affinity.

STRUM [5] predicts how single point mutations may affect stability with wild-type sequences are simulated using iterative threading assembly improvement (I-TASSER) to create three-dimensional models. It was assessed using 3421 experimentally found mutations from 150 different proteins utilising a 5-fold cross validation. Protein function annotation and human illness detection depend heavily on SNP [6] (Genetic Variants) mutation-induced stability alterations (DDG). In general, the characteristics of the wild-type amino acids are less than those of the mutant amino acids.

Predicting how genetic variations may affect protein stability is currently challenging. It is essential to comprehend the impact of amino acid sequence variations on human health and disease, mostly caused by non-synonymous (or missense) DNA mutations that alter or increase protein function [6–10].

A pathogenic missense variants association with protein stability alterations has already been shown [11] and it has been demonstrated that these perturbations help haplo insufficient genes lose function [12].

Protein customization to achieve desired distinctive features is the most difficult and developing area of biotechnology. Gaining insight into the mechanism whereby, proteins fold and unfold will help to improve the stability of proteins and further research on their function. The process by which a protein transforms from its linear form into a stable three-dimensional structure has long been the subject of discussion. The creation of sophisticated, comprehensible, and trustworthy deep learning-based techniques for forecasting the folding conformational stability of proteins in response to multipoint variations will result from filling up the research gap. These techniques have the potential to shed new light folding dynamics of proteins and advance our knowledge of disease processes, medication discovery, and precision medicine.

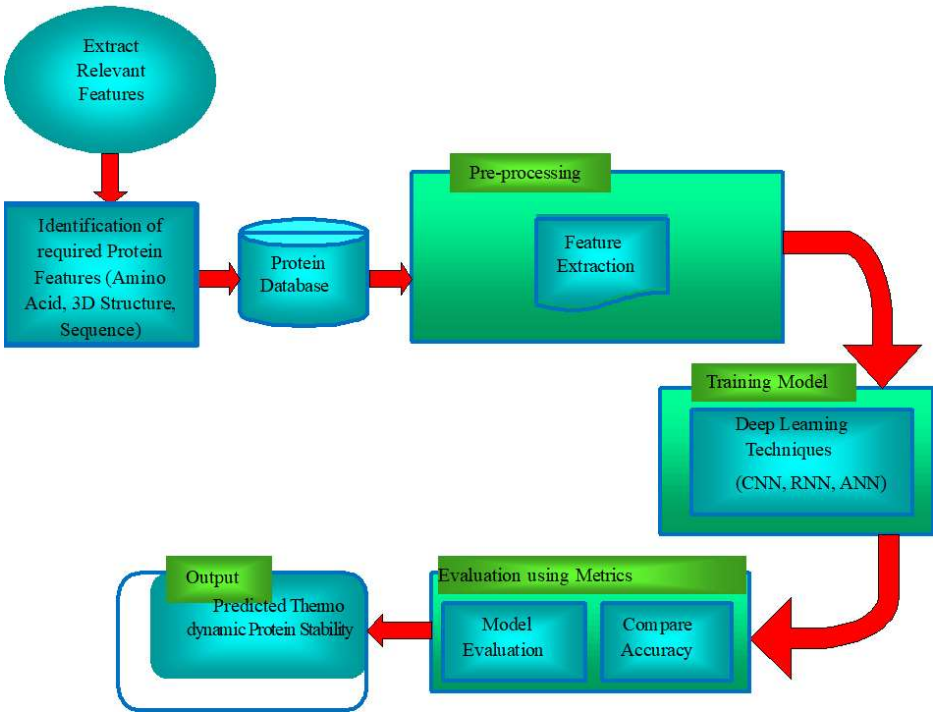

Figure 2: Activity Recognition

By systematically following ensemble methods [Figure 2], we can develop a robust deep learning-based framework for predicting the thermodynamic stability of proteins in response to single and multi-point variants, thereby advancing our knowledge of protein stability and its implications for various biological processes and disease mechanisms.

2. Analysis of Prediction Techniques

Mutations can affect protein stability by altering the sequence of amino acids, disrupting key interactions within the protein structure, changing the folding kinetics, or destabilizing secondary or tertiary structures. These alterations can lead to changes in Gibbs free energy, ultimately impacting the stability of the protein. Gibbs free energy quantifies the stability of proteins by assessing the balance between enthalpy and entropy changes during their folding or unfolding processes.

We have chosen the most well-known ΔΔG predictors in terms of speed and accuracy.

2.1 Below is a list of them along with a brief description of their traits.

STRUM [5]: integrates knowledge and physics-based energy functions acquired from protein structure models obtained by I-TASSER (Roy et al., 2010) using gradient boosting regression,

MAESTRO [31]: a multi-agent system that integrates statistical energy functions as primary characteristics with an artificial neural network, SVM, and linear regression.

ELASPIC [4] predicts the effects of mutations on protein-protein interactions and folding. ELASPIC website can be used to evaluate the effects of mutations on any protein, including modelled wild-type and mutant structures, that is included in the Uniprot database.

EASE-MM [32] consists of five specialised support vector machine (SVM) models, with the final prediction coming from a consensus of two models chosen based on the accessible surface area and predicted secondary structure of the mutated residue.

MM-TCNN [12] propose a convolutional neural network with a topological approach that is multi-tasking and multichannel. Topology Net predicts protein-ligand binding affinities, mutation-induced changes in the liberated energy of globular protein folding, and mutation-induced changes in the liberated energy of membrane protein folding better than the state-of-the-art techniques.

SDM [33] determines the difference in thermodynamic stability among the wild-type and altered proteins using conformationally controlled environment-specific replacement tables.

Rosetta [34] creates a three-dimensional (3D) structural paradigm of the mutant protein based on the wild-type structure. It then calculates the energy difference between the two structures using the sum of several empirical physics-based energy contributions, ([35] Coulomb electrostatic, Lennard-Jones atomic interactions, etc.)

PON-tstab[13] utilised a rectified dataset, they trained PON-tstab, a random forests-based technique that can handle variations. ProTherm data has been employed to train all machine learning methods currently available for variation stability. All variant stability predictors in machine learning have been developed on erroneous and biased data.

PoPMuSiCsym [14] Using black-box machine learning, they developed PoPMuSiCsym, an elegant solution to the bias problem by compelling physical symmetries under converse mutations on the model structure. This novel predictor represents an effective balance between bias-free performance and accuracy.

DeepDDG [16] developed a neural network-based technique to forecast how point mutations may affect the protein stability. The neural network outperformed and obtained a Pearson correlation coefficient of 0.48–0.56 for three independent test sets. A thorough examination of the input properties reveals the most significant feature, indicating that the buried hydrophobic region is the primary factor influencing protein stability.

BoostDDG [19] proposed a novel method Extreme Gradient Boosting, to predict stability changes upon point mutations from protein sequences using a sequential forward selection technique. Pearson correlation coefficient (PCC) of 0.535, which agrees with the 0.540 on an individualistic test, was caused by 14 traits from six groups. They emphasized that BoostDDG serves as a robust tool for forecasting stability alterations resulting from point mutations in protein sequences.

iStable 2.0[17] introduced an integrated predictor that integrates sequence information with forecasts from various approaches (I-Mutant, AUTOMUTE, MUPRO, PoPMuSiC, and CUPSAT) utilising an SVM algorithm. They Uses machine learning to integrate 11 tools for sequence- and structure-based prediction and to incorporate protein sequence data as features.

ACDC-NN [6] Exploiting the sequence data and addressed the antisymmetry property with **Single point** mutation.

DDGun [15] Generates basic anti-symmetric features using evolutionary information and sequence and structural data to estimate the ΔΔG for both single and multiple variations (DDGun3D). Without any training on the experimental datasets, our technique obtains impressive results, with Pearson correlation coefficients between the measured and predicted ΔΔG values of about 0.4 and ~0.5 for single and multiple site variations.

GMMA[36] uses a computer model called **Gobal Multi-Mutant Analysis** to analyse the effects of the majority of amino acid changes from a gene library that has been randomly altered. The method can identify mutations that improve the stability of even very stable proteins for which conventional selection algorithms have reached their limit by using a high mutation frequency. Global multi-mutant analysis (GMMA) can separate single changeover effects, which may be used to screen any protein for function.

MU3DSP [18] suggesting a novel computational framework, the 3D single point variants model surpasses current methods across multiple benchmarks, establishing itself as a dependable tool for evaluating both somatic and germline substitution variants, and aiding in protein design.

2.2 Current Practices in Deep Learning

Class imbalance [21]: It investigates the impact of data imbalance to deep learning model's performance, with a comparison of frequently used techniques to alleviate the problem.

Regulatory genomics, Cellular imaging [22]: A review of deep learning's applications in computational biology, with an emphasis on regulatory genomics and cellular imaging.

Acceleration and model compression [23]: An overview of techniques for lowering the computational demands of deep learning methods and compressing deep learning models. **Human illness and biomedicine** [24], gives an overview of the uses of deep learning in

biomedicine and goes into great detail on the challenges that come with implementing deep learning.

Machine learning, Genomic medicine [25] A review of machine learning tasks and related datasets in genomic medicine, not specific to deep learning.

Biomarker development, Transcriptomics [26] It explored the fundamental aspects of deep learning and numerous biomedical applications thereof.

Medical imaging, CNN [27] It focuses on the applications of CNN in biomedical image processing, as well as the challenges in that field.

Protein sequences [28] A tutorial-style review of using deep learning to handle sequence data, with three application examples.

Biomedicine, Model flexibility [29] It discusses the history and advantage of deep learning, arguing the potential of deep learning in biomedicine.

Overfitting, Regularization [30] A systematic review of regularization methods used to combat overfitting issue in deep learning.

Transfer learning, Data shortage [31] The review examines transfer learning methodologies within the realm of deep learning, addressing the challenge of limited data availability.

The research gap in the field of "Prediction of Protein Stability based on Mutations "lies in the limited exploration of accurate and interpretable means to foresee the intricate and context-dependent effects of multipoint genetic variations on protein folding stability. Despite substantial progress in foresee protein stability and deep learning applications several key gaps and challenges remain:

a. **Limited Coverage of Multipoint Variants:**
   Existing research predominantly focuses on single point mutations neglecting the extensive landscape of multipoint variants. The interactions and synergistic effects among multiple residues within the framework of folding conformations stability alterations have not been thoroughly investigated.
b. **Complexity of Multipoint Effects:**
   The simultaneous alteration of multiple amino acids introduces complex and non linear effects on protein folding stability. Current methodologies struggle to capture the nuanced interplay of various mutations and accurately predict their collective impact.
c. **Scarcity of Quality Experimental Data:**
   Reliable experimental data that quantifies the folding conformational stability changes upon multipoint variants is limited. This scarcity hampers the training and justification of prognostic models, hindering their reliability and generalization.
d. **Model Interpretability**:
   Deep learning models often lack interpretability, making it challenging to discern the molecular mechanisms and interactions underlying stability predictions. The ability to attribute predictions to specific structural features and residues remains an open challenge.
e. **Generalization across Proteins**: The generalization of Deep learning models across diverse protein families with varying sizes and structural characteristic is not well established. Ensuring the model's applicability beyond a specific set of protein is crucial for real word scenarios.
f. **Biological Insights**: While accurate prediction is important, gaining a deeper understanding of the biological insights and mechanisms underlying stability changes induced by multipoint variants remains an unexplored territory.
g. **Ethical Considerations:** As deep learning models gain prominence in biomedical research and decision-making ethical considerations related to transparency, bias and accountability in prediction protein stability alteration need to be addressed.

Addressing these research gap will lead to the development of advanced, interpretable, and reliable deep learning methods for forecasting the folding conformational stability of proteins in response to multipoint variants. Such methods hold the potential to unlock novel insights into

protein folding dynamics and contribute our understanding of disease mechanisms, drug development and precision medicine.

3. Research Methodology

Various biophysical techniques, such as circular dichroism spectroscopy, differential scanning calorimetry, and protein folding assays, are usually used to study and assess protein stability. The ability of deep learning methods to fully exploit the complex relationships between protein sequence, structure, and stability has led to their widespread usage for protein stability prediction. To assess protein stability by deep learning, some common techniques include Convolutional Neural Networks, Recurrent Neural Networks, Graph Neural Networks, Attention Mechanisms and Transformer Models.

ACDC-NN-Seq is a sequence-based tool capable of demonstrating comparable or superior performance compared to state-of-the-art methods, all while maintaining impeccable thermodynamic antisymmetry. This approach can be expanded to foresee the ΔΔGs of multiple site variations [6]. By taking advantage of a specially created loss function that reduces the firm deviation between the direct and reverse projections, ACDC-NN-Seq addresses the physical characteristics of anti-symmetry.

For ACDC-NN to forecast the alteration in the folding free energy (DDG) resulting from a point replacement in a protein sequence, it requires structural and evolutionary data regarding the protein. Data on the structure comes from a PDB file. The evolutionary data comes from a tab-parted table of the frequency of each residue at each place in homologous proteins found in a profile file. The DDG levels are becoming more stable.

This methodology aims to effectively capture the complex relationship between multi-point variants and protein folding stability.

The following steps outline the proposed methodology [Figure 3]:

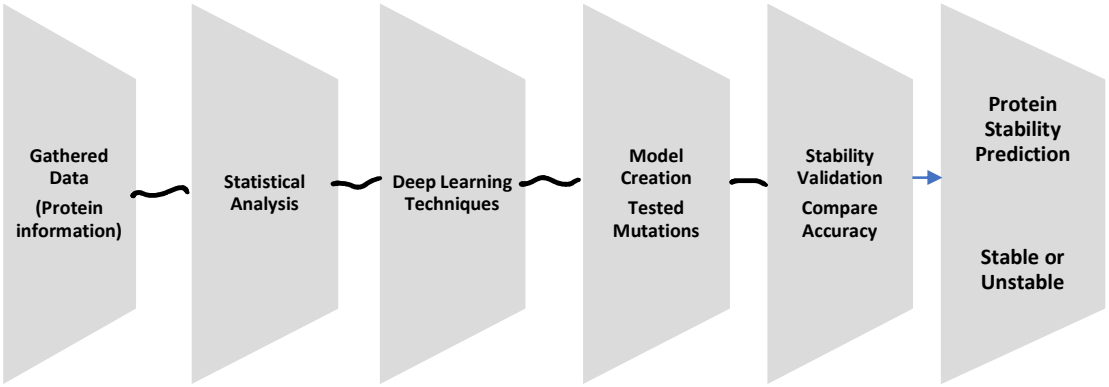

Figure 3. Processing Mechanisms for predicting protein stability.

1. **Data Collection and Pre-processing**:

   Compile a broad range of protein structures with multi-point variants that have experimentally measured changes in folding stability.

   Add pertinent annotations to the dataset, such as protein names, amino acid sequences, structural coordinates, and conformational stability values that go with them. Standardize the dataset's structures, fill in any gaps in the data, and normalize stability values.

2. **Feature Extraction and Representation**:

   Extract valuable information from protein structures, such as evolutionary conservation scores, secondary structure details, solvent accessibility, and amino acid physicochemical properties.

   Transform the retrieved features into appropriate numerical representations that reflect the protein's structural and sequence-based properties.

3. **Deep Learning Model Design**

Select a deep learning architecture that can accurately capture spatial dependencies and sequential linkages within protein structures, such as convolutional neural networks (CNNs), recurrent neural networks (RNNs), or graph neural networks (GNNs).

Adapt the model's architecture by defining layers, units, activation functions, and connectivity patterns based on the characteristics of the incoming data and the specific problem being addressed.

4. **Model Training and Validation**

To ensure model generalization and avoid over fitting, divide the dataset into training, validation, and test sets. Randomize or use pre-trained embeddings to initialize model parameters, then use back propagation algorithm and gradient descent-based optimization methods to optimize them.

To improve the model's stability and prevent over fitting, regularize it using methods like dropout or batch normalization. Keep track of how the model performs on the validation set and make any necessary hyper parameter adjustments to boost convergence and accuracy.

5. **Performance Evaluation**:

Use relevant evaluation measures, such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Pearson correlation coefficient, and others, to assess the trained model's performance on the test dataset. Perform a comparison analysis using current computational techniques for stability prediction to evaluate the model's generalizability and predictive capacity.

6. **Interpretability and Visualization**:

Create methods for interpreting the model's predictions and highlighting the structural elements, residues, and interactions that oversee the stability alterations given on by multiple-point variations. Produce visualizations that provide insights on the molecular mechanisms supporting the observed changes in folding conformational stability.

7. **Application to Disease Variants**:

Apply the trained model to predict the folding conformational stability changes of protein variants associated with specific hereditary disease or drug resistance situations. To verify the model's accuracy and applicability, compare its predictions to experimental findings and current biological information.

4. Summary and Discussion

Protein stability is crucial for various biological processes, including function, enzyme activity, protein folding and misfolding, cellular homeostasis, and drug targeting. Proteins need a stable three-dimensional structure to bind to molecules, catalyze reactions, and regulate cellular processes. Misfolding and aggregation of proteins can lead to diseases like Alzheimer's and Parkinson's. Protein stability also helps maintain cellular homeostasis, preventing cellular stress and dysfunction. Understanding protein stability helps design drugs that effectively interact with target proteins, leading to desirable therapeutic outcomes.

Moreover, protein stability plays a crucial role in maintaining cellular homeostasis, ensuring that cells function optimally. Stable proteins contribute to the proper functioning of cellular machinery, helping to prevent cellular stress and dysfunction. This balance is essential for overall health and vitality.

Understanding protein stability is also pivotal in drug discovery and development. Drugs often work by interacting with specific target proteins, modulating their activity to achieve therapeutic outcomes. By comprehensively understanding protein stability, researchers can design drugs that effectively bind to their targets, leading to desired therapeutic effects while minimizing side effects.

In summary, protein stability is a cornerstone of biological function, impacting processes essential for life. Its significance spans from basic cellular functions to the enhancement of novel therapeutics. By delving deeper into the mechanisms governing protein stability, researchers can unlock new insights into disease mechanisms and therapeutic strategies, ultimately improving human health and well-being.

5.Conclusions

In conclusion, protein stability is critical for maintaining the proper functioning of cells and organisms. It influences various cellular processes and has implications in numerous aspects of biology, medicine, and biotechnology. The study explored the application of deep learning techniques in analyzing the stability of proteins-based mutations. The researchers demonstrated that deep learning algorithms were effective in predicting the influence of mutations on protein stability, outperforming traditional methods. This finding has significant implications for protein engineering and drug discovery. Furthermore, the review discussed the various deep learning structures and approaches employed, showcasing the versatility and adaptability of these techniques in the field of protein stability analysis. This opens avenues for further research and development in the application of deep learning to tackle complex biological problems. The review supports the notion that deep learning techniques hold immense promise in analyzing protein stability based on mutations. The integration of artificial intelligence and deep learning methodologies in this domain has the potential to revolutionize protein engineering and contribute to advancements in biomedical research.

References:

1. Hung CL, Kuo YH, Lee SW, Chiang YW. Protein Stability Depends Critically on the Surface Hydrogen-Bonding Network: A Case Study of Bid Protein. J Phys Chem B. 2021 Aug 5;125(30):8373-8382. doi: 10.1021/acs.jpcb.1c03245. Epub 2021 Jul 27. PMID: 34314184.
2. Savojardo C, Fariselli P, Martelli PL, Casadio R. INPS-MD: a web server to predict stability of protein variants from sequence and structure. Bioinformatics. 2016 Aug 15;32(16):2542-4. doi: 10.1093/bioinformatics/btw192. Epub 2016 Apr 10. PMID: 27153629.
3. Folkman L, Stantic B, Sattar A, Zhou Y. EASE-MM: sequence-based prediction of mutation-induced stability changes with feature-based multiple models. J Mol Biol 2016;428:1394–405. https://doi.org/10.1016/j.jmb.2016.01.012
4. Witvliet DK, Strokach A, Giraldo-Forero AF, Teyra J, Colak R, Kim PM. ELASPIC web-server: proteome-wide structure-based prediction of mutation effects on protein stability and binding affinity. Bioinformatics 2016;32:1589–91. https://doi.org/10.1093/bioinformatics/btw031.
5. Lijun Quan, Qiang Lv, Yang Zhang, STRUM: structure-based prediction of protein stability changes upon single-point mutation, *Bioinformatics*, Volume 32, Issue 19, October 2016, Pages 2936–2946, https://doi.org/10.1093/bioinformatics/btw361
6. Pancotti, C., Benevenuta, S., Repetto, V., Birolo, G., Capriotti, E., Sanavia, T., & Fariselli, P. (2021). A deep-learning sequence-based method to predict protein stability changes upon genetic variations. *Genes*, *12*(6), 911.
7. Hartl, F. U. (2017). Protein misfolding diseases. *Annual review of biochemistry*, *86*, 21-26.
8. Martelli, P. L., Fariselli, P., Savojardo, C., Babbi, G., Aggazio, F., & Casadio, R. (2016). Large scale analysis of protein stability in OMIM disease related human protein variants. *BMC genomics*, *17*(2), 239-247.
9. Cheng, T. M., Lu, Y. E., Vendruscolo, M., Lio', P., & Blundell, T. L. (2008). Prediction by graph theoretic measures of structural effects in proteins arising from non-synonymous single nucleotide polymorphisms. *PLoS computational biology*, *4*(7), e1000135.
10. Compiani, M., & Capriotti, E. (2013). Computational and theoretical methods for protein folding. *Biochemistry*, *52*(48), 8601-8624.
11. Casadio, R., Vassura, M., Tiwari, S., Fariselli, P., & Luigi Martelli, P. (2011). Correlating disease-related mutations to their effect on protein stability: A large-scale analysis of the human proteome. *Human mutation*, *32*(10), 1161-1170.Birolo, G.; Benevenuta, S.; Fariselli, P.; Capriotti, E.; Giorgio, E.; Sanavia, T. Protein Stability Perturbation Contributes to the Loss of Function in Haploinsufficient Genes. Front. Mol. Biosci. 2021, 8, 10. [CrossRef]
12. Cang, Z., & Wei, G. W. (2017). TopologyNet: Topology based deep convolutional and multi-task neural networks for biomolecular property predictions. *PLoS computational biology*, *13*(7), e1005690.

13. Yang Y, Urolagin S, Niroula A, Ding X, Shen B, Vihinen M. PON-tstab: Protein Variant Stability Predictor. Importance of Training Data Quality. Int J Mol Sci. 2018 Mar 28;19(4):1009. doi: 10.3390/ijms19041009. PMID: 29597263; PMCID: PMC5979465.
14. Pucci F, Bernaerts KV, Kwasigroch JM, et al. Quantification of biases in predictions of protein stability changes upon mutations. Bioinformatics 2018;34(21):3659–65.
15. Montanucci L, Capriotti E, Frank Y, Ben-Tal N, Fariselli P. DDGun: an untrained method for the prediction of protein stability changes upon single and multiple point variations. BMC Bioinf 2019;20:335. https://doi.org/10.1186/s12859-019-2923-1. Received March 24, 2022; Revised April 15, 2022; Editorial Decision April 19, 2022; Accepted May 04, 2022
16. Cao H, Wang J, He L, Qi Y, Zhang JZ. DeepDDG: predicting the stability change of protein point mutations using neural networks. J Chem Inf Model 2019;59:1508–14. https://doi.org/10.1021/acs.jcim.8b00697.
17. Chen CW, Lin MH, Liao CC, Chang HP, Chu YW. iStable 2.0: Predicting protein thermal stability changes by integrating various characteristic modules. Comput Struct. Biotechnol J 2020;18:622–30. https://doi.org/10.1016/j.csbj.2020.02.021.
18. Gong, J., Wang, J., Zong, X., Ma, Z., & Xu, D. (2023). Prediction of protein stability changes upon single-point variant using 3D structure profile. *Computational and Structural Biotechnology Journal*, *21*, 354-364.
19. Xuan Lv, Jianwen Chen, Yutong Lu, Zhiguang Chen, Nong Xiao, and Yuedong Yang, Accurately Predicting Mutation-Caused Stability Changes from Protein Sequences Using Extreme Gradient Boosting,*Journal of Chemical Information and Modeling* **2020** *60* (4), 2388-2395.DOI: 10.1021/acs.jcim.0c00064
20. Mateusz Buda, Atsuto Maki, Maciej A Mazurowski, A systematic study of the class imbalance problem in convolutional neural networks, Neural Netw. 106 (2018) 249–259
21. C. Angermueller, T. Parnamaa, L. Parts, O. Stegle, Deep learning for computational biology, Mol. Syst. Biol. 12 (7) (2016) 878.
22. Yu Cheng, Duo Wang, Pan Zhou, Tao Zhang, A survey of model compression and acceleration for deep neural networks, 2017. arXiv:1710.09282
23. Travers Ching, Daniel S. Himmelstein, Brett K. Beaulieu-Jones, Alexandr A. Kalinin, Brian T. Do, Gregory P. Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M. Hoffman, et al., Opportunities and obstacles for deep learning in biology and medicine, J. R. Soc. Interface 15 (141) (2018) 20170387
24. Michael K.K. Leung, Andrew Delong, Babak Alipanahi, Brendan J. Frey, Machine learning in genomic medicine: a review of computational problems and data sets, Proc. IEEE 104 (1) (2016) 176–197.
25. Michael Wainberg, Daniele Merico, Andrew Delong, Brendan J. Frey, Deep learning in biomedicine, Nat. Biotechnol. 36 (9) (2018) 829.
26. Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram Van Ginneken, Clara I. Sánchez, A survey on deep learning in medical image analysis, Med. Image Anal. 42 (2017) 60–88.
27. Vanessa Isabell Jurtz, Alexander Rosenberg Johansen, Morten Nielsen, Jose Juan Almagro Armenteros, Henrik Nielsen, Casper Kaae Sønderby, Ole Winther, Søren Kaae Sønderby, An introduction to deep learning on biological sequence data: examples and solutions, *Bioinformatics*, Volume 33, Issue 22, November 2017, Pages 3685–3690, https://doi.org/10.1093/bioinformatics/btx531
28. Polina Mamoshina, Armando Vieira, Evgeny Putin, Alex Zhavoronkov, Applications of deep learning in biomedicine, Mol. Pharm. 13 (5) (2016) 1445–1454.
29. Jan Kukačka, Vladimir Golkov, Daniel Cremers, Regularization for deep learning: a taxonomy, 2017. arXiv:1710.10686.
30. 31.Chuanqi Tan, Fuchun Sun, Tao Kong, Wenchang Zhang, Chao Yang, Chunfang Liu, A survey on deep transfer learning, International Conference on Artificial Neural Networks, Springer, 2018, pp. 270–279.

31. Laimer, J., Hiebl-Flach, J., Lengauer, D., & Lackner, P. (2016). MAESTROweb: a web server for structure-based protein stability prediction. *Bioinformatics*, *32*(9), 1414-1416.
32. Folkman, L., Stantic, B., Sattar, A., & Zhou, Y. (2016). EASE-MM: sequence-based prediction of mutation-induced stability changes with feature-based multiple models. *Journal of molecular biology*, *428*(6), 1394-1405.
33. Pandurangan, A. P., Ochoa-Montano, B., Ascher, D. B., & Blundell, T. L. (2017). SDM: a server for predicting effects of mutations on protein stability. *Nucleic acids research*, *45*(W1), W229-W235.
34. Kellogg, E. H., Leaver-Fay, A., & Baker, D. (2011). Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins: Structure, Function, and Bioinformatics*, *79*(3), 830-838.
35. Alford, R. F., Leaver-Fay, A., Jeliazkov, J. R., O'Meara, M. J., DiMaio, F. P., Park, H., ... & Gray, J. J. (2017). The Rosetta all-atom energy function for macromolecular modeling and design. *Journal of chemical theory and computation*, *13*(6), 3031-3048.
36. Norrild, R. K., Johansson, K. E., O'Shea, C., Morth, J. P., Lindorff-Larsen, K., & Winther, J. R. (2022). Increasing protein stability by inferring substitution effects from high-throughput experiments. *Cell Reports Methods*, *2*(11).