Journal of Communicable Diseases
Volume 57, Issue 1 - 2025, Pg. No. 91-95
Peer Reviewed & Open Access Journal

Research Article

Efficiently Predicting HIV-1 Protease Cleavage Sites by Using Deep CNN-Assisted Hybridised Approach

Navneet Kaur Bawa[1], Satish Saini[2], Gagandeep Kaur[3]

[1]Research Scholar, Department of Computer science & Engineering, Rimt University, Mandi Gobindgarh, Punjab, India
[2]Department of Electrical, Electronics and Communication Engineering,RIMT University, Mandi Gobindgarh, Punjab, India
[3]Department of Electronics and Communication Engineering, Chandigarh Group of College Jhanjeri, Mohali, Punjab, India
DOI: https://doi.org/10.24321/0019.5138.202511

I N F O

Corresponding Author:
Satish Saini, Department of Electrical, Electronics and Communication Engineering,RIMT University, Mandi Gobindgarh, Punjab, India
E-mail Id:
satishsainiece@gmail.com
Orcid Id:
https://orcid.org/0000-0002-9194-3068
How to cite this article:
Saini S, Kaur G. Efficiently Predicting HIV-1 Protease Cleavage Sites by Using Deep CNN-Assisted Hybridised Approach. J Commun Dis. 2025;57(1):91-95.

Date of Submission: 2024-10-25
Date of Acceptance: 2025-02-12

A B S T R A C T

Acquired Immuno Deficiency Syndrome (AIDS) is considered a vital danger over the feasible development and on account of its pestilence effect and nonattendance of reparable medicines. HIV-1 AIDS can be constrained by utilising protease inhibitors. Different procedures for predicting sites are utilised to comprehend the highlighted sites of those proteases. Arrangement-based sites like physiochemical elements and construction-based sites are separated from HIV-1 proteases. In this article, a procedure for choosing those sites using a deep CNN-assisted hybridised approach will be used for effectively predicting cleavage sites. The proposed methodology was evaluated based on various Type-1 and Type-2 parameters. The proposed approach gives superior results on Type-1 and Type-2 parameters. Data746_setset provides an accuracy of 0.924, data 1625_set provides an accuracy of 0.946, Data_schilling_set provides an accuracy of 0.9389, Data_impens_set provides an accuracy of 0.911 and average dataset provides an accuracy of 0.921 for Type-1 parameters. Data746_setset provides an accuracy of 0.84664, data 1625_set provides an accuracy of 0.34179, data_schilling_set provides an accuracy of 0.69529 and data_impens_set provides an accuracy of 0.59511 for Type-2 parameters.

Keywords: HIV-1, Protease Inhibitor, Deep CNN

Introduction

From the beyond twenty years, AIDS a tainted infection ordinarily known as HIV-1 AIDS was accounted for by the US community for infectious prevention. In an exceptionally brief timeframe, the sickness spreads like a pestilence extent in the entire world. There were an estimated 39.9 million people living with HIV at the end of 2023, 65% of whom are in the WHO African Region. In 2023, an estimated 630 000 people died from HIV-related causes and an estimated 1.3 million people acquired HIV.[1] Though many studies have been conducted, no technique has been developed that can fix AIDS to date. HIV-1 focuses on the safe framework and debilitates an individual's resistant framework, contaminated individuals gradually become immune deficient, as it reduces CD4 cells in the blood.[2,3] Weakened immunity can lead to an extensive variety of sickness, diseases and malignant growths. Notwithstanding, there are strategies to fix AIDS up to some degree. Since the disclosure of HIV, numerous enemies of HIV substrates have been recognised and recommended by the US Food

and Drug Administration (FDA). Some of them are inhibitors of HIV protease. The combination of inhibitors of HIV-1 PR, invert transcriptase inhibitor and integrase inhibitor generally known as exceptionally dynamic antiretroviral treatment (HAART) is the most dynamic treatment.[4–7] HIV-1 protease is an enzyme which assumes a significant part in the replication cycle in the human body. HIV-1 protease inhibitor is a tiny particle that typically ties to HIV-1 protease at the dynamic cleavage site, to such an extent that the substrate should not tie to the protease.[8–12] It has been observed experimentally that protease attaches to the protein in an octapeptide way and cleaves its scissile bond. A scissile bond is a covalent synthetic bond that can be broken by an enzyme. Models would be the cleaved bond in oneself severing hammerhead ribozyme. There are 20 amino acids that are naturally available, so the number of octapeptides can be 208, which is very large in number and difficult to process using experimental methods and prediction of protease cleavage using computerised methods are cost-effective and simple solution.[13–17]

Review of Literature

Gok and Özcerit predicted HIV-PR cleavage sites using a combination of a biased SVM Classifier with an asymmetric bagging strategy.[18] Feature extraction techniques used different coding schemes, including AAI, chemical properties and VLCP (predicted coronavirus 3CL protease cleavage on the viral poly-peptides). Cleavage prediction was done using random forest model. This machine learning method achieved an accuracy of 0.96 in cross-validation. 3CLP named web server was created to anticipate the cleavage locales of 3CL proteases. Brik and Wong developed a hybrid octapeptide sequence information, AA binary profiles and physicochemical properties acting as mathematical descriptors.[19] In this research, 10-fold cross-validation techniques were used which included LR and DT classifiers. K-nearest neighbor and perceptron classifier showed lower performance with regard to AUC and F-Score and B.Acc, whereas logistic regression and multilayer perceptron classifier had the most accurate performance on the test set with an AUC of 0.97, f.score of 0.89, and B.Acc > 90%.

Wang developed a method in which sites of substrate are treated as non-labelled samples in place of negative ones. This research used a positive non-labelled learning method called PU for effective prediction extracted from substrate, feature sequences, AAI, co-evolutionary sequence and chemical properties. By tuning weighted errors from unlabelled positive samples, a biased SVM classifier was built.[17]

Rahman and Rashid has proposed a new technique to combine the multiple WC optimally by means of integrating the information extracted from several encoding approaches of amino acids.[6] As a consequence, various encoding models that produce the particular heterogeneous information were paired with every classifier and this locally trained classifier gets aggregated to form the concluding prediction. The GA algorithm has been applied to achieve the ideally consolidated feature and classifier sequence. The general simulation has made sense of the huge increment of the proposed model over the prior models considering normal accuracy, area under the curve and standard deviation. Rognvaldsson et al. have introduced a novel technique to encode and characterise the amino acids sequence and a new technique to predict the cleavage of AA sequence using HIV protease.[16] The implemented encoding structure has used the amalgamation of spatial and structural characteristics of amino acids along with an amino acid sequence count of 20 to ensure the sequencing and physicochemical features. The implemented HIV-1 amino acid cleavage prediction method has been introduced along with the genetic programming and SVM. The performance of the proposed work was compared which resulted in better prediction. Feng et al. has deployed two new notions for the identification of HIV Cleavage sites.[9] Initially, an optimal ensemble formation approach has been proposed for search space optimisation of 228 using four SVM kernels (7x4) and seven encoding approaches with the utilisation of GA. Secondly, the AUC has been used as a measure of fitness for evaluating the optimal ensemble technique. The interrelationship among the encoding methodology pair in an ensemble has been decided by the evolutionary algorithm that encoded with binary string pattern. By using this, the implemented model has achieved better prediction of HIV sites. A past study by Rahman and Rashid has introduced the cognitive framework to cope with the significant changes among nonlinear and linear separable data.[6] This framework has used the evolutionary algorithm's dexterity for formulating the encoding approaches and best adaptive settings in the classifier. This, in turn, enhances the predictive capacity of the approach. In this, the framework has chosen the three decisive factors, SVM kernel selection for linear or non-linear data, physicochemical property for peptide encoding, and parameter tuning of the selected kernel. The performance of the implemented model has been validated on the basis of simulation outcomes and thus validates its efficiency. Wang has introduced a novel feature amino acid encoding approach for anticipating the HIV-1 protease cleavage sites.[17] The implemented model was the combination of Taylor's Venn diagram and orthonormal encoding. For the testing, the linear SVM has been used.[19–21] The analysis has been evaluated by comparing the proposed model in the view of some feature encoding methods. The evaluation of the test was made on PR-3261 and PR-1625 datasets. The simulation outcome has thus shown better performance on classification over other encoding approaches. Nanni and Lumini have developed

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202511

a generative method on the basis of a generalisation of variable order Markov chains (VOMC) for peptide sequences and further assigned this method to predict the cleavability of particular proteases.[22] Moreover, the variable context Markov chains (VCMC) model has intended to identify the patterns equivalent on the basis of evolutionary similarities among every amino acid. Then, it was adjusted for HIV-1 protease cleavage site expectation issues and has demonstrated betterment over the conventional models in view of prediction accuracy on a normal dataset. Mafarja et al. have implemented the prediction model that integrates the structural, sequence, and physicochemical features within several ML (Machine learning) algorithms.[12] After that, in feature selection, a bidirectional stepwise selection approach was integrated to determine the discriminative features. Moreover, several encoding structures were used to calculate the selected features and these were given as input for LR, ANN and DT. The objective presentation of cleavage site expectation has been assessed by applying a more thorough three-way information split technique. Fathi and Sadeghi has presented a novel feature subset selection approach called FS-MLP, which has chosen the appropriate features by means of multi-layered perceptron (MLP) learning.[8] The model was incorporated with the training dataset and after that, the feature subset selection by means of the decomposition technique was applied to analyse the trained MLP. The simulation outcome has explained that the verified performance of FS-MLP over the seven feature selection models has shown advancement in multi-variate, non-straight and high-dimensional areas.

Objectives of the Study

Cleavage Site Prediction in HIV-AIDS has become a challenging task in the past few years. Many researchers have applied various biological prediction techniques that are very time-consuming. On reviewing the work of researchers, a need was felt to explore new methodologies to predict the cleavage sites as well as to conduct a comparative study to find an efficient way of evaluating Type-1 and Type-2 parameters.

The main objectives of this study are:

1. To find the optimised solution for HIV protein cleavage site in HIV prediction using Deep CNN in an efficient way
2. To compare the results of the proposed work with existing techniques for Type-1 and Type-2 parameters

Material and Methods

Certain research works have been already done on 'HIV-1 protease cleavage site prediction' that suffer from inaccurate prediction results due to the lack of sustained logic or diplomacy. With this in mind, this proposal intends to propose a new HIV-1 protease cleavage site forecast under two significant stages: (i) Feature Extraction and (ii) Classification that involves different logic apart from the conventional works. Wavelet decomposition-based feature extraction will be carried out in the first phase. Subsequently, the extracted features will be input to the classification of DCNN will be used. In fact, in order to make the prediction more accurate, this proposal aims to insist on the logic of optimisation under two processes: (i) Deep CNN training and (ii) Optimal tuning of the activation function. To solve these optimisation problems, a new hybrid algorithm will be introduced that hybridises the concept of Moth Search (MS) and Dragonfly (DA) algorithm. Both the MS and DA are renowned optimisation algorithms that have the ability to solve complex optimisation problems under various applications. The workflow of the study is shown in Figure 1.

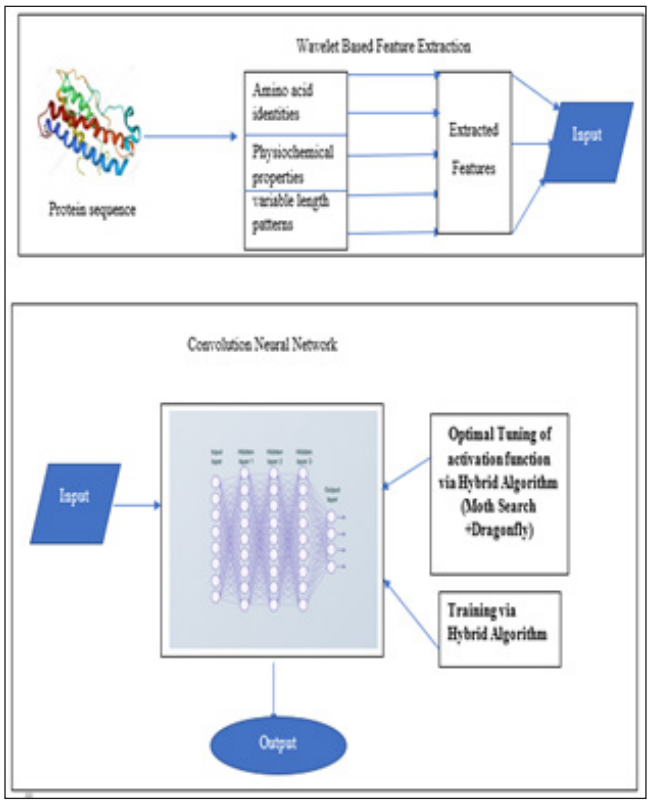

Figure 1.Steps in the Protease Cleavage Site Prediction

Results and Discussion

There are various techniques and methodologies devised in the due course of time to predict the PR Cleavage sites for HIV-1 AIDS, the performance parameters discussed by researchers are categorised into two main types i.e. Type-1 and Type-2. Accuracy is one such parameter which it measures how often it correctly predicts the outcome. Table 1 shows the accuracy of the training dataset; all four datasets are standard datasets. The table clearly shows the proposed model outflanks from past methods.

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202511

Table 1.Experimental Results of Accuracy (DA_MSO) (Type-1 Parameters)

Training Dataset	Nanni and Lumini[22]	Shen and Chou[23]	Gok and Özcerit[18]	Rögnvaldsson et al.[16]	Faithi and Sadeghi[8]	Proposed Method (DA_MSO)
Data746_setset	0.892	0.929	0.894	0.885	0.910	0.924
Data 1625_set	0.886	0.891	0.870	0.856	0.856	0.946
Data_schilling_set	0.926	0.753	0.919	0.925	0.925	0.9389
Data_Impens_set	0.910	0.687	0.898	0.908	0.908	0.911
Average	0.903	0.815	0.895	0.893	0.911	0.921

The F1 score in Table 2 shows higher reliability of the classification model in this research paper as compared to the Ensemble-based Classification method predicted by Hu et al.[3] The F1 score defines the harmonic mean of precision and recall. A score greater than 0.7 or higher is considered good for any classification method.

Table 2.Experimental Result F1 Score (Type-1)

Dataset	DCNN	EM-HIV [2022]	(Proposed Method) DA_MSO
Data746_setset	0.81481	-	0.91457
Data 1625_set	0.9667	0.86	0.97216
Data_schilling_set	0.9637	0.62	0.96556
Data_Impens_set	0.93.496	0.68	0.9499

Mathew Correlation Coefficient draws the summary of Confusion matrix and its value is between +1 and -1. +1 value depicts the best performance between the actual and predicted values. As shown in Table 3, dataset_746 shows the best performance in the proposed method, i.e. 0.84664 as compared to DCNN, MSO and DA individually.

Table 3.Experimental Result of MCC (Type-2)

Dataset	DCNN	MSO	DA	Proposed Method (DA_MSO)
Data746_ setset	0.6814	0.78537	0.80977	0.84664
Data 1625_ set	0.2688	0.2332	0.3137	0.34179
Data_ schilling_ set	0.65096	0.69193	0.6783	0.69529
Data_ Impens_set	0.5142	0.60309	0.52726	0.59511

These results have significant ramifications for the creation of successful HIV-1 protease inhibitors and need further research on the use of machine learning methods in this field. The capacity of the suggested approach to precisely forecast PR cleavage sites will help to build new inhibitors, thereby improving the therapy results for HIV-1 patients. Future studies may look at using this approach to different viral proteases and look at its possibility to find new binding sites.

The future of AI holds immense potential for revolutionising numerous industries and aspects of life, with advancements in machine learning, natural language processing, and computer vision expected to drive significant innovation and transformation.[24–26]

Conclusion

Although there are varied machine learning techniques available to predict the PR-Cleavage sites in HIV-1 AIDS, a well-designed prediction model with accuracy better than the cutting-edge model was not achieved by any presently applied ML model. Since extracting features was also a concern, the proposed model uses wavelet-based decomposition to extract features from the available benchmark datasets. After the feature extraction phase, the deep convolution NN model is applied to the input data to optimise the tuning of the activation function. Tuning is done using a hybrid approach. The hybrid approach uses two optimisation algorithms, moth search and dragonfly. In this research work, the parameters achieved like accuracy, F1 score and MCC are better than the other competing models. The proposed methodology was evaluated based on various Type-1 and Type-2 parameters. The proposed approach gives superior results on Type-1 and Type-2 parameters. Data746_set provides an accuracy of 0.924, Data 1625_set provides an accuracy of 0.946, Data_schilling_set provides an accuracy of 0.9389, Data_Impens_set provides an accuracy of 0.911 and average dataset provides an accuracy of 0.921 for Type-1 parameters. Data746_setset provides an accuracy of 0.84664, Data 1625_set provides an accuracy of 0.34179, Data_schilling_set provides an accuracy of 0.69529 and Data_Impens_set provides an accuracy of 0.59511 for Type-2 parameters.

Conflict of Interest: None

Source of Funding: None

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202511

Declaration of Generative AI and AI-Assisted Technologies in the Writing Process: None

References

1. World Health Organization [Internet]. HIV and AIDS; [cited 2023 Oct 23]. Available from: https://www.who.int/news-room/fact-sheets/detail/hiv-aids
2. Gokuldhev M, Singaravel G. Local pollination-based moth search algorithm for task-scheduling heterogeneous cloud environment. Comput J. 2022;65(2):382-95. [Google Scholar]
3. Hu L, Li Z, Tang Z, Zhao C, Zhou X, Hu P. Effectively predicting HIV-1 protease cleavage sites by using an ensemble learning approach. BMC Bioinformatics. 2022;23(1):447. [PubMed] [Google Scholar]
4. Onah E, Uzor PF, Ugwoke IC, Eze JU, Ugwuanyi ST, Chukwudi IR, Ibezim A. Prediction of HIV-1 protease cleavage site from octapeptide sequence information using selected classifiers and hybrid descriptors. BMC Bioinformatics. 2022;23(1):466. [PubMed] [Google Scholar]
5. Chen H, Zhu Z, Qiu Y, Ge X, Zheng H, Peng Y. Prediction of coronavirus 3C-like protease cleavage sites using machine-learning algorithms. Virol Sin. 2022;37(3):437-44. [PubMed] [Google Scholar]
6. Rahman CM, Rashid TA. Dragonfly algorithm and its applications in applied science survey. Comput Intell Neurosci. 2019;2019:9293617. [PubMed] [Google Scholar]
7. Singh D, Singh P, Sisodia DS. Evolutionary based ensemble framework for realizing transfer learning in HIV-1 Protease cleavage sites prediction. Appl Intell. 2019;49:1260-82. [Google Scholar]
8. Fathi A, Sadeghi R. A genetic programming method for feature mapping to improve prediction of HIV-1 protease cleavage site. Appl Soft Comput. 2018;72:56-64. [Google Scholar]
9. Feng Y, An H, Gao X. The importance of transfer function in solving set-union knapsack problem based on discrete moth search algorithm. Mathematics. 2019;7(1):17. [Google Scholar]
10. Strumberger I, Tuba E, Bacanin N, Beko M, Tuba M. Hybridized moth search algorithm for constrained optimization problems. Proceedings of 2018 International Young Engineers Forum (YEF-ECE); 2018 May. p. 1-5. [Google Scholar]
11. Singh D, Singh P, Sisodia DS. Evolutionary based optimal ensemble classifiers for HIV-1 protease cleavage sites prediction. Expert Syst Appl. 2018;109:86-99. [Google Scholar]
12. Mafarja MM, Eleyan D, Jaber I, Hammouri A, Mirjalili S. Binary dragonfly algorithm for feature selection. Proceedings of 2017 International Conference on New Trends in Computing Sciences (ICTCS); 2017 Oct. p. 12-7. [Google Scholar]
13. Mirjalili S. Dragonfly algorithm: a new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems. Neural Comput Appl. 2016;27:1053-73. [Google Scholar]
14. Singh O, Su EC. Prediction of HIV-1 protease cleavage site using a combination of sequence, structural, and physicochemical features. BMC Bioinformatics. 2016;17:478. [Google Scholar]
15. Li Z, Zhou Y, Zhang S, Song J. Lévy-Flight Moth-Flame algorithm for function optimization and engineering design problems. Math Probl Eng. 2016;2016:1423930. [Google Scholar]
16. Rögnvaldsson T, You L, Garwicz D. State of the art prediction of HIV-1 protease cleavage sites. Bioinformatics. 2015;31(8):1204-10. [PubMed] [Google Scholar]
17. Wang GG. Moth search algorithm: a bio-inspired metaheuristic algorithm for global optimization problems. Memet Comput. 2018;10(2):151-64. [Google Scholar]
18. Gok M, Özcerit AT. A new feature encoding scheme for HIV-1 protease cleavage site prediction. Neural Comput Appl. 2013;22:1757-61. [Google Scholar]
19. Brik A, Wong CH. HIV-1 protease: mechanism and drug discovery. Org Biomol Chem. 2003 Jan 7;1(1):5-14. [PubMed] [Google Scholar]
20. Song J, Tan H, Perry AJ, Akutsu T, Webb GI, Whisstock JC, Pike RN. PROSPER: an integrated feature-based tool for predicting protease substrate cleavage sites. PloS One. 2012;7(11):e50300. [PubMed] [Google Scholar]
21. Kim G, Kim Y, Lim H, Kim H. An MLP-based feature subset selection for HIV-1 protease cleavage site analysis. Artif Intell Med. 2010;48(2-3):83-9. [PubMed] [Google Scholar]
22. Nanni L, Lumini A. Using ensemble of classifiers for predicting HIV protease cleavage sites in proteins. Amino Acids. 2009;36(3):409-16. [PubMed] [Google Scholar]
23. Shen HB, Chou KC. Using ensemble classifier to identify membrane protein types. Amino Acids. 2007;32(4):483-8. [PubMed] [Google Scholar]
24. Akhai S. The impact of artificial intelligence on healthcare: opportunities and challenges. J Adv Res Med Sci Technol. 2024;11(1&2):1-6. [Google Scholar]
25. Akhai S. Healthcare record management for healthcare 4.0 via blockchain: a review of current applications, opportunities, challenges, and future potential. In: Malviya R, Sundram S, editors. Blockchain for healthcare 4.0. Vol. 1. CRC Press; 2023. p. 211-23. [Google Scholar]
26. Akhai S, Khang A. Innovations in medical manufacturing: a review of 3D printing, robotics, and Internet of Things (IoT). In: The quantum evolution. Vol. 1. CRC Press; 2024. p. 226-41. [Google Scholar]

ISSN: 0019-5138
DOI: https://doi.org/10.24321/0019.5138.202511