Multilayer perceptron and evolutionary radial basis function neural network models for discrimination of HIV-1 genomes

Ashok Kumar Dwivedi* and Usha Chouhan

Mathematics and Computer Applications, Department of Bioinformatics, Maulana Azad National Institute of Technology, Bhopal 462 003, India

High rate of mutation and frequent recombination cause evolution of HIV-1 very diverse and adaptive. Revealing the recombination patterns in HIV-1 is a computationally intensive problem. Techniques based on phylogenetic analysis are not suitable for genome-level studies. Here we elucidate approaches based on multilayer perceptron and evolutionary radial basis function neural network for the analysis of 4130 HIV-1 genomes. These techniques show remarkable improvement over other machine learning techniques used for such classification. The models outperformed other machine learning models having 92% classification accuracy. Multilayer perceptron achieved sensitivity and specificity of 82% and 96%, whereas radial basis function neural network achieved sensitivity and specificity of 78% and 98% on tenfold cross-validation respectively.

Keywords: Artificial neural network, HIV-1 genome, machine learning, multilayer perceptron.

THE human immunodeficiency virus (HIV) causing acquired immunodeficiency syndrome (AIDS), has severely affected life expectancy and the world economy by infecting millions of people worldwide[1]. HIV has turned out to be ubiquitous causing worldwide health problems, since its first identification[2].

The major types of HIV are HIV type 1 (HIV-1) and HIV type 2 (HIV-2). The most prevalent and pathogenic strain of this virus is HIV-1; in contrast, HIV-2 is much less prevalent[3]. HIV-1 is described by a fast acclimating population with extraordinary diversity, resilient selection and recurrent recombination[4]. HIV-1 strains have been categorized into three key groups known as major (M), outlier (O), and non-M, non-O (N) group based on phylogenetic analysis. The leading group of HIV-1 strain is M, which is sub-divided into nine distinct subtypes (A–K)[5,6]. Some subtypes like A and F are further subdivided into subsubtypes like A (A1, A2) and F (F1, F2) based on phylogenetic clustering. Additionally, there are circulating recombinant forms (CRFs), which are the result of recombination between two or more HIV-1 subtypes[7]. This diversity poses challenges in the treatment for HIV/AIDS since characteristics of the virus deviate from strain to strain[8]. This exertion is further convoluted by the existence of various CRFs and further complicates the development of effective treatment of HIV/AIDS.

Nonetheless, drug amalgamations have to be designated for a particular strain of the virus. This is critical to recognize the strain that infects a patient in order to recommend a suitable course of medication. It is therefore indispensable to develop competent and effective classification techniques for the determination of HIV-1 subtypes. Most of the established techniques are targeted to classify HIV-1 and HIV-2 sequences or the subtypes[9]. Furthermore, these methods are based on finding pairwise distance amongst sequences or constructed on phylogenetic studies. It can be envisioned that phylogenetic studies based on whole genomes are much more reliable than those based on short segments of the HIV-1 genome. Nevertheless, such an analysis with phylogenetic techniques is not feasible due to inherent computational complexity of these techniques.

Heritable recombination and astonishing rate of mutations in HIV-1 genome increases diversity in HIV-1 genomes, which aids viruses to drip undeniably from host immune systems or encourage resistance for antiviral drugs. Identification of recombination in HIV-1 is important from the point of view of epidemiological issues and accurate detection of super infection in patients which can foster the strategy of prospective vaccines and treatments against HIV-1/AIDS (ref. 10). Previous work has used reference genomes for the identification of recombination[11]. Here we consider the problem of classification of CRFs from other basic subtypes. CRF genomes are composed of regions of several subtypes and hence it is computationally difficult to discriminate such genomes[12]. Therefore, it is indispensable to devise an proficient technique for the detection and classification of recombination in the HIV-1 strain.

There are several methods to detect recombination, and a few techniques for reconstruction of recombination events when it is putative. Nevertheless, further studies are essential in order to develop enhanced techniques for this problem. The foremost impediments in detecting recombination is that patterns in the data that are indicative

*For correspondence. (e-mail: dwivedi.ashok@gmail.com)

of recombination can exist due to other reasons like insufficient data, lineage sorting, inadequate analyses, etc., and it is problematic to distinguish among these different conditions.. Hence, contemporary techniques have not been completely efficacious. Two methods have been used here to deal with this problem.

Phylogenetic methods: Phylogenetic techniques help find the genes involved in the recombination process by examining contradictory phylogenetic trees. These techniques can be further divided into two groups. First, which equate phylogenetic trees unequivocally and second, which usages surrogate trees in position of phylogenetic trees[13]. The benefit of phylogenetic approaches is that they can use genetic material of many species and assimilate them with appriaite model of evolution. While limitations of phylogenetic techniques are their computational complexity and they cannot handle sequence at genome scale[14].

Parametric methods: Parametric techniques employ genomic characteristics (genomic signatures), specific to a particular genome, species or clades. If a portion of the genome strongly differs from the signature of that genome, this may be the sign of latent recombination occurrence. For example, GC% variation in different genomes which can be exploited as genomic abrasion. Different genomic signatures used include nucleotide composition[15], structural landscapes of the genome[16] and oligonucleotide occurrences[17]. Recombination is crucial aspect for HIV-1 diversity[18]. Discrimination of CRFs from basic subtypes provides an opportunity to the research community for better understanding the evolution of HIV-1, which may foster the development of effective vaccines for the treatment of HIV-1. Moreover, by discriminating CRFs from other subtypes, the correct treatment could be established. There are various techniques for predicting the genetic subtype of HIV-1 bases on sequence alignment[11,19,20]. However, such techniques are computationally intensive. The application of such techniques for classification at genome level is not feasible for a large number of genomes[21]. Furthermore, identifying CRFs from the HIV-1 genome is a challenging task[11]. Different algorithms have been applied in this endeavour, including the top string relative entropy method[19] and on NCBI sliding window method for creating BLAST similarity scores[11]. These methods have a classification accuracy of 87% and 77% respectively. High utility of NCBI tool for genotyping of recombinant has been emphasized sequences[19] with a prediction accuracy of 73.4% (ref. 11).

The results of Wu *et al.*[11] suggested high prediction accuracy. However, the prediction accuracy differs significantly while testing for complete pair subtype match (70%)[22]. Other claims and pitfalls of these techniques have been discussed by Eliuk *et al.*[22]. Our recent study has shown that machine learning technique can be applied for the recombination analysis using genome signatures[23].

The objective of this work is to analyse whole genome sequences of HIV-1 strains from Los Alamos National Laboratory[24], for classification of recombinant and basic sub type of HIV-1 genomes and to investigate the use of multilayer perceptron and evolutionary radial basis function neural network for this classification. We also compared their performance with other machine learning techniques.

Machine learning techniques explore the potential hypothesis to find the best mathematical representation of the data which best suit the data in hand and prior hypothesis about the data[23,25]. In this study, we have considered the problem of supervised learning. Neural networks have widespread applications like series forecasting, business modelling, in finance, medical applications, etc.[26–29]. Evolutionary algorithms have been used in the design of neural networks and in various other applications[30–32]. These techniques have also been applied in bioinformatics, for microarray analysis, gene identification, etc.[33–35]. Details of other applications of neural networks can be found in Dwivedi and Chouhan[35] and Yasdi[36] and references therein. Here we present multilayer perceptron and evolutionary radial basis function neural network models for the classification of recombinant and basic subtypes of HIV-1 genomes. Performance of these models was evaluated on the basis of different indices like classification accuracy, sensitivity and specificity. Moreover, the performance was compared with those of other machine learning models.

Materials and methods

Datasets

Genome sequences of all complete genomes of HIV-1 basic subtypes and HIV-1 CRFs were retrieved from the database at Los Almos Laboratory, USA[24]. Totally 4130 complete genomes of HIV-1 were processed after removal of duplicate genomes. These genomes were clustered in two sets: CRFs (1192) and basic type (NON-CRF; 2938). The codon composition of every genome were determined using the software package DAMBE[37]. This constructs the feature vector of 64 dimensions for each genome. Thus a total number of 4130 records with 64 features were used in this study. The implications of codon usage bias have been discussed by Jenkins and Holmes[38]. This dataset is similar to that used by us in a previous study[23] except that we have excluded genomes which have similar codon composition. This had a significant impact on the performance of different machine learning models, as presented later in the text. All records were labelled as CRF or NON-CRF according to the class they belonged to. Records were divided into

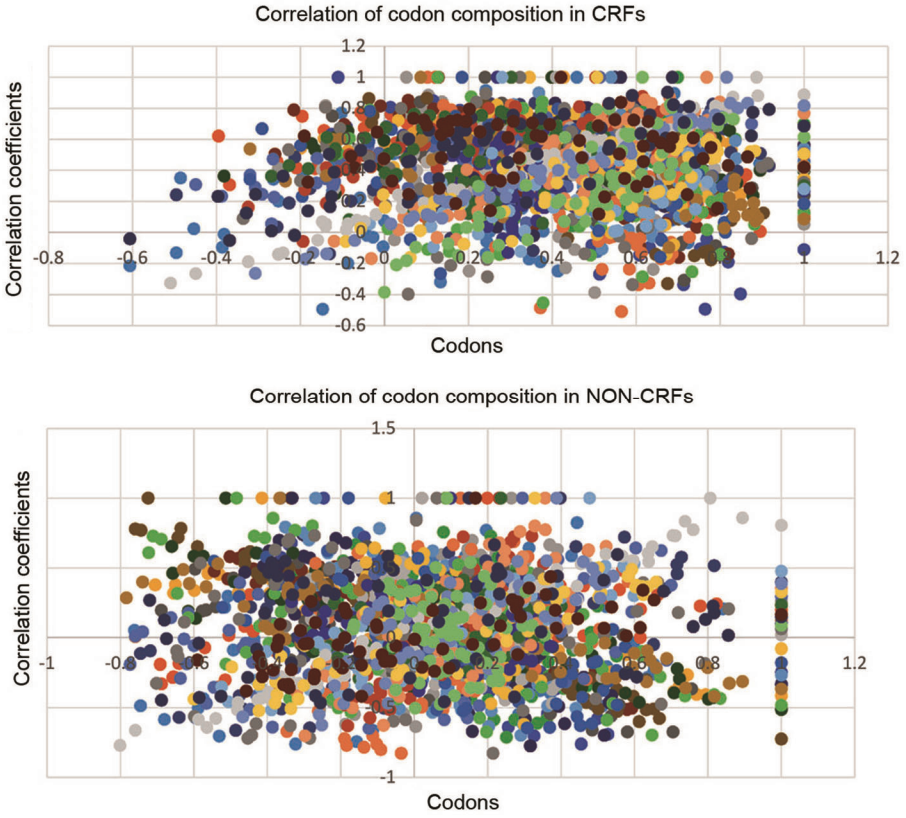


Figure 1. Scatter plot of correlation coefficients for codon composition of CRF and NON-CRF genomes indicating that there is positive correlation between genomes of the same type.

ten sets for tenfold cross-validation. We determined the correlation in codon composition between CRF and NON-CRF genomes, which indicates high correlation between CRF and NON-CRF (Figure 1). Furthermore, we have determined standard deviation and variance in the codon composition between CRF and NON-CRF genomes, which indicates that there is high variance and deviation among the genomes of CRF than basic subtypes (Figure 2).

Techniques used for classification

Various techniques of machine intelligence have been used in this study. These include artificial neural network, naive Bayes classifier, logistic regression classifier, support vector machine and classification trees. Detailed description of these techniques can be found in previous studies and references therein[27,29,39,40]. In the following section we will explore radial basis function neural network (RBFN) and evolutionary RBFN (EV-RBFN), which are of the focus of this study.

Radial basis function neural network: RBFN is based on biologically inspired concept of local and tuned receptive fields of neurons[41,42]. RBFNN has three levels of layers. The first layer known as input layer is used to feed data into the next layer which is known as the hidden or intermediate layer; this layer processes the received data and subsequently provides them to the output. In this type of network neurons of hidden layer exemplify a basis function of the output universe with respect to the response space. Intermediate layers of the network provide the means to project the input onto greater dimensional space where patterns could become linearly separable. Training samples are used to learn the weights of hidden layer and output layers weights are used for fitting model on the data. Here activation functions are known as radial functions having the characteristic of increasing or decreasing monotonically from the centre of the given data[43]. Gaussian kernel is a typical example of radial function

$$f(x) = \exp\left(-\frac{(x-c)^2}{r^2}\right),$$

where c is the epicentre and r is the radius of hyperplane.

Evolutionary RBFN: RBFN can approximate any multivariate continuous function when appropriate network settings are provided[44]. This is possible since once centre

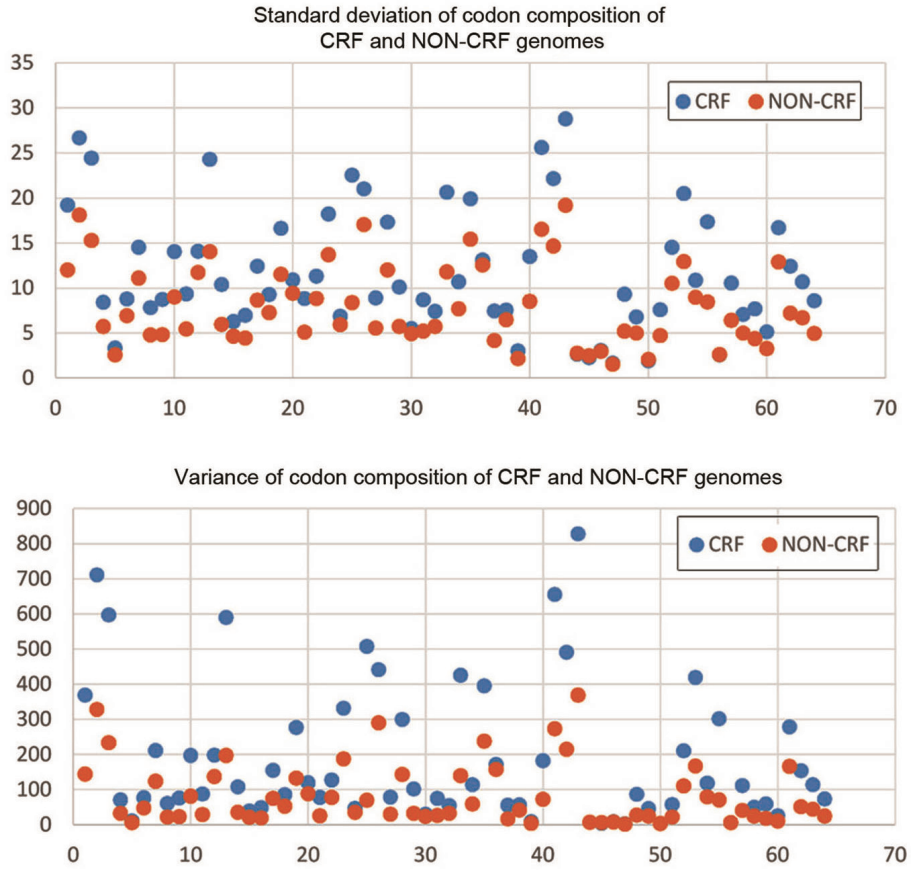


Figure 2. Standard deviation and variance in codon composition of the HIV-1 genomes.

and radii have been found, the weights of connections amongst intermediate and output layers can be tuned by means of appropriate analytical technique to resolve linear algebraic equalities as opposed to training algorithms used in others model like MLP[45]. The main challenges in RBF NN design are to find the number of neurons, their midpoints and radii which have been discussed by Broomhead and Wowe's[46]. We have used standard genetic algorithm-based approach for this, as discussed by Rivas *et al.*[46] and references therein.

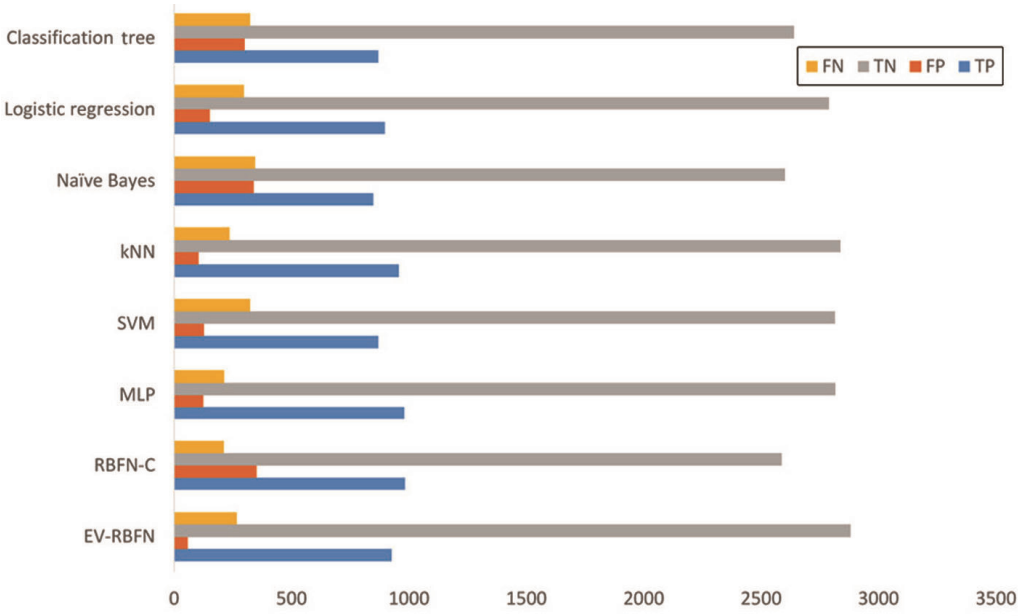
Measurement of classification performance

The results of classification were evaluated using tenfold cross-validation technique. The classification accuracy of models was evaluated using five different quality measures, which include classification accuracy, sensitivity, specificity, precision and F1 measure. These measures have been described in previous work and references therein[23,47,48]. We have examined these quality measures for classification accuracy analysis for classification of HIV-1 CRF and basic subtypes. CRFs were considered as positive class and basic subtypes (NON-CRF) were considered as negative class.

Results and discussion

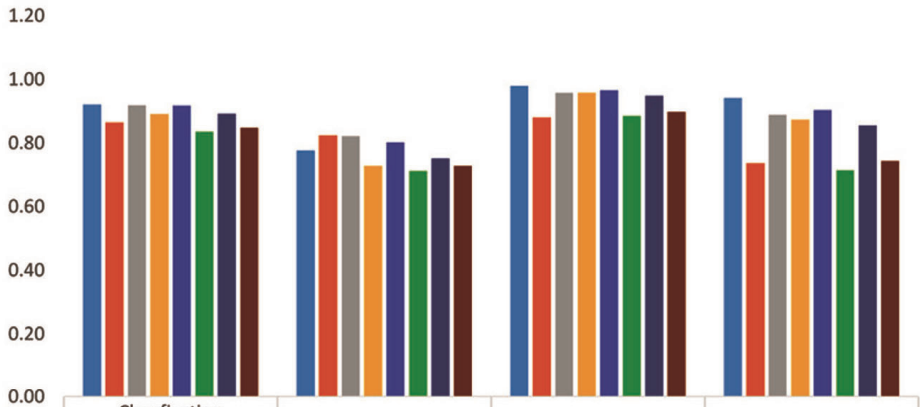
We have analysed the application of multilayer perceptron and evolutionary neural network models for genome classification of HIV-1 CRFs and basic subtypes. We have also compared the performance of these models with those of other machine learning techniques. Totally 4130 genomes of HIV-1, including 1192 CRF and 2938 NON-CRF genomes were used in the classification. Codon composition of each genome was used as sequence features, as done in our previous work[23]. Data samples were divided into tenfold; one fold was used for testing and the rest of were used for training. Classification results of all techniques are given in Figure 2 by multilayer perceptron (MLP), support vector machine (SVM), k nearest neighbor (kNN), naive Bayes, logistic regression, classification tree, EV-RBFN and RBFN.

Figure 3 plots the predictions of all the models. It clearly indicates that the neural network model predicts the highest number of true positives, i.e. CRF identified as CRF. RBFN predict the highest number of TP (982 out of 1192) and MLP have second highest TP (980). Figure 3 also shows that EV-RBVN has the highest number of TN (2880/2938). EV-RBFN has the lowest number of FP



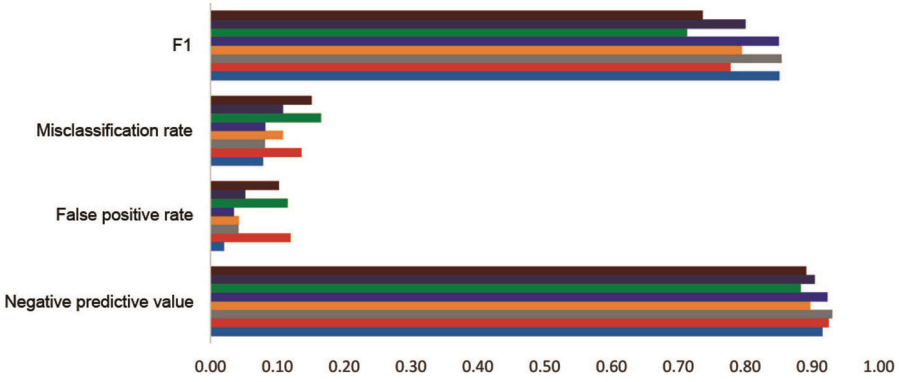
	EV-RBFN	RBFN-C	MLP	SVM	kNN	Naïve Bayes	Logistic regression	Classification tree
FN	266	210	212	323	236	344	296	323
TN	2880	2587	2815	2813	2836	2600	2787	2638
FP	58	351	123	125	102	338	151	300
TP	926	982	980	869	956	848	896	869

Figure 3. Classification results of machine learning techniques. True positive (TP), Number of actual CRFs predicted as CRFs. False positive (FP), Number of actual NON-CRFs predicted as CRFs. True negative (TN), Number of actual NON-CRFs predicted as NON-CRFs. False negative (FN), Number of actual CRFs predicted as NON-CRFs.



	Classfication accuracy	Sesnsitivity	Specificity	Precision
EV-RBFN	0.92	0.78	0.98	0.94
RBFN-C	0.86	0.82	0.88	0.74
MLP	0.92	0.82	0.96	0.89
SVM	0.89	0.73	0.96	0.87
kNN	0.92	0.80	0.97	0.90
Naïve Bayes	0.83	0.71	0.88	0.72
Logistic regression	0.89	0.75	0.95	0.86
Classification tree	0.85	0.73	0.90	0.74

Figure 4. Classification performance measurements of machine learning algorithm.



	Negative predictive value	False positive rate	Misclassification rate	F1
Classification tree	0.89	0.10	0.15	0.74
Logistic regression	0.90	0.05	0.11	0.80
Naïve Bayes	0.88	0.12	0.17	0.71
kNN	0.92	0.03	0.08	0.85
SVM	0.90	0.04	0.11	0.80
MLP	0.93	0.04	0.08	0.85
RBFN-C	0.92	0.12	0.14	0.78
EV-RBFN	0.92	0.02	0.08	0.85

Figure 5. Classification performance measurements of machine learning algorithm.

Table 1. Confusion matrix for tenfold cross-validation using EV and RBF

		Predicted class			
		EV-RBFNN		RBFN-C	
		CRF	NON-CRF	CRF	NON-CRF
Actual class	CRF 1192	926 (TP) 94.1%	266 (FN) 8.5%	982 (73.7%)	210 (7.5%)
	NON-CRF 2938	58 (FP) 5.9%	2880 (TN) 91.5%	351 (26.3%)	2587 (92.5%)
Total	4130	984	3146	1333	2797

Table 2. Classification performance measure indices for tenfold cross-validation using machine learning techniques

	CA	Sens.	Spec.	Pre.	NPV	FPR	RMC	F1
EV-RBFN	0.922	0.777	0.980	0.941	0.915	0.019	0.078	0.851
RBF	0.864	0.823	0.880	0.737	0.925	0.119	0.136	0.778
MLP	0.919	0.822	0.958	0.888	0.929	0.041	0.081	0.854
SVM	0.892	0.729	0.957	0.874	0.897	0.042	0.108	0.795
kNN	0.918	0.802	0.965	0.904	0.923	0.035	0.082	0.849
Naive bayes	0.835	0.711	0.884	0.715	0.883	0.115	0.165	0.713
Logistic regression	0.892	0.752	0.948	0.858	0.904	0.051	0.108	0.800
Classification tree	0.849	0.729	0.898	0.743	0.891	0.102	0.151	0.736

(58), whereas RBF has the lowest number of FN (210). Overall, multilayer perceptron, RBF and EV-RBF performed well in comparison to other machine learning models by having lower type-I and type-II errors.

Figure 4 plots four classification performance measurements, namely classification accuracy, sensitivity, specificity and precision. It shows that multilayer perceptron and EV-RBFN outperform all other machine learning techniques by having the highest classification accuracy of 92% (Tables 1 and 2). kNN also has the same classification accuracy, but MLP and EV-RBFN have the highest sensitivity and specificity respectively. Figure 4 also shows that RBF and MLP have the highest sensitivity of 82% compared to all techniques. Figure 5 compares misclassification rate, negative predictive value, and false positive rate and F1 measures of all machine learning techniques. The figure clearly illustrates that MLP and EV-RBFN have negligible false predictive rate of 0.02 and 0.04 respectively, whereas EV-RBFN, MLP and kNN have the lowest misclassification rate of 0.08, which is in

agreement with our previous result[23]. All machine learning techniques have negative predictive value of >90%, which indicates that these techniques are suitable in the identification of basic subtypes of HIV-1. MLP and EV-RBFN also outperformed other techniques on F1 measure (0.85). Thus, our experimental results indicate that MLP and EV-RBFN perform better than other classification techniques for the classification of CRFs and basic subtypes of HIV-1 genomes.

Conclusion

Genome wide analysis of HIV-1 is crucial for the development of effective treatments against AIDS. This process becomes complex by the manifestation of CRFs. Genome scale classification of HIV-1 using conventional phylogenetic techniques is not feasible. Herein, we have categorized all available complete genomes of HIV-1 into two classes of CRFs and basic subtypes using machine learning techniques. We assessed the ability of machine learning techniques for the effective classification of recombinant and no-recombinant sequences HIV-1 sequence on genome scale using simple codon composition feature. MLP and EV-RBFN models have been shown to provide good generalization proficiencies. Additionally, performance of these machine learning techniques was validated using tenfold cross-validation and evaluated using different classification performance measurements. A significant classification accuracy of 92% was achieved on tenfold cross-validation using MLP and EV-RBFN models. Our results are noteworthy since we have examined almost all available HIV-1 complete genomes. Furthermore, our work demonstrates competence of different machine learning techniques for classification of HIV-1 at genome scale. In future these technique can be used for genome wide subtyping of HIV-1.

1. Safrit, J. T., Fast, P. E., Gieber, L., Kuipers, H., Dean, H. J. and Koff, W. C., Status of vaccine research and development of vaccines for HIV-1. *Vaccine*, 2016.
2. Cihlar, T. and Fordyce, M., Current status and prospects of HIV treatment. *Curr. Opin. Virol*., 2016, **18**, 50–56.
3. Sharp, P. M. and Hahn, B. H., Origins of HIV and the AIDS pandemic. *Cold Spring Harbor Perspect. Med*., 2011, **1**, a006841.
4. Zanini, F., Brodin, J., Thebo, L., Lanz, C., Bratt, G., Albert, J. and Neher, R. A., Population genomics of intrapatient HIV-1 evolution. *eLife*, 2016, **4**, e11282.
5. Robertson, D. *et al.*, HIV-1 nomenclature proposal. *Science*, 2000, **288**, 55.
6. McCutchan, F. E., Global epidemiology of HIV. *J. Med. Virol.*, 2006, **78**, S7–S12.
7. Robertson, D. L., Hahn, B. H. and Sharp, P. M., Recombination in AIDS viruses. *J. Mol. Evol*., 1995, **40**, 249–259.
8. Palella Jr, F. J. *et al.*, Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. *N Engl. J. Med*., 1998, **338**, 853–860.
9. Rambaut, A., Robertson, D. L., Pybus, O. G., Peeters, M. and Holmes, E. C., Human immunodeficiency virus: phylogeny and the origin of HIV-1. *Nature*, 2001, **410**, 1047–1048.
10. Ntemgwa, M., Gill, M. J., Brenner, B. G., Moisi, D. and Wainberg, M. A., Discrepancies in assignment of subtype/recombinant forms by genotyping programs for HIV type 1 drug resistance testing may falsely predict superinfection. *AIDS Res. Hum. Retroviruses*, 2008, **24**, 995–1002.
11. Wu, X., Cai, Z., Wan, X.-F., Hoang, T., Goebel, R. and Lin, G., Nucleotide composition string selection in HIV-1 subtyping using whole genomes. *Bioinformatics*, 2007, **23**, 1744–1752.
12. Hoelscher, M. *et al.*, Detection of HIV-1 subtypes, recombinants, and dual infections in East Africa by a multi-region hybridization assay. *AIDS*, 2002, **16**, 2055–2064.
13. Dessimoz, C., Margadant, D. and Gonnet, G. H., DLIGHT–lateral gene transfer detection using pairwise evolutionary distances in a statistical framework. In *Annual International Conference on Research in Computational Molecular Biology*, Springer, Berlin, Heidelberg, 2008, pp. 315–330.
14. Truszkowski, J. and Brown, D. G., More accurate recombination prediction in HIV-1 using a robust decoding algorithm for HMMS. *BMC Bioinform*., 2011, **12**, 1.
15. Daubin, V., Lerat, E. and Perrière, G., The source of laterally transferred genes in bacterial genomes. *Genome Biol*., 2003, **4**, R57.
16. Worning, P., Jensen, L. J., Nelson, K. E., Brunak, S. and Ussery, D. W., Structural analysis of DNA sequence: evidence for lateral gene transfer in *Thermotoga maritima*. *Nucleic Acids Res*., 2000, **28**, 706–709.
17. Lawrence, J. G. and Ochman, H., Molecular archaeology of the *Escherichia coli* genome. *Proc. Natl. Acad. Sci. USA*, 1998, **95**, 9413–9417.
18. Jetzt, A. E., Yu, H., Klarmann, G. J., Ron, Y., Preston, B. D. and Dougherty, J. P., High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol*., 2000, **74**, 1234–1240.
19. Rozanov, M., Plikat, U., Chappey, C., Kochergin, A. and Tatusova, T., A web-based genotyping resource for viral sequences. *Nucleic Acids Res*., 2004, **32**, W654–W659.
20. Wu, X. *et al.*, Whole genome phyogeny construction via complete composition vectors. *Int. J. Bioinform. Res. Appl.*, 2006, **2**, 219–248.
21. Thompson, K. and Charnigo, R., Parallel computing in genome-wide association studies. *J. Biometr. Biostat*., 2015, **6**(1), 1.
22. Eliuk, A. S., Keith Ruiter, B. and Pierre Boulanger, C., Classifying HIV-1 circulating recombinant forms. In Proceedings of the International Conference on Bioinformatics and Computational Biology (BIOCOMP), The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (World Comp), 2011, p. 1.
23. Dwivedi, A. K. and Chouhan, U., Comparative study of machine learning techniques for genome scale discrimination of recombinant HIV-1 strains. *J. Med. Imaging Health Inf*., 2016, **6**, 425–430.
24. Briesmeister, J. F., Los Alamos National Laboratory. Oak Ridge National Laboratory, MCNP-4B Monte Carlo N-Particle Transport Code System, Manual La-12625-M, version B, 2000, **4**, 1997.
25. Mitchell, T. M., *Machine Learning*, McGraw Hill, Burr Ridge, IL, USA, 1997, **45**.
26. Hoptroff, R. G., The principles and practice of time series forecasting and business modelling using neural nets. *Neural Comput. Appl*., 1993, **1**, 59–66.
27. Dwivedi, A. K., Performance evaluation of different machine learning techniques for prediction of heart disease. *Neural Comput. Appl*., 2016, **29**(10), 685–693.
28. Dwivedi, A. K., Artificial neural network model for effective cancer classification using microarray gene expression data. *Neural Comput. Appl*., 2016, **29**(12), 1545–1554.
29. Dwivedi, A. K. and Chouhan, U., Comparative study of artificial neural network for classification of hot and cold recombination

regions in *Saccharomyces cerevisiae*. *Neural Comput. Appl*., 2016, **29**(2), 529–535.
30. Jones, A. J., Genetic algorithms and their applications to the design of neural networks. *Neural Comput. Appl*., 1993, **1**, 32–45.
31. Venkatesan, D., Kannan, K. and Saravanan, R., A genetic algorithm-based artificial neural network model for the optimization of machining processes. *Neural Comput. Appl*., 2009, **18**, 135–140.
32. Radcliffe, N. J., Genetic set recombination and its application to neural network topology optimisation. *Neural Comput. Appl*., 1993, **1**, 67–90.
33. Brown, M. P. *et al.*, Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci*., *USA*, 2000, **97**, 262–267.
34. Furey, T. S., Cristianini, N., Duffy, N., Bednarski, D. W., Schummer, M. and Haussler, D., Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics*, 2000, **16**, 906–914.
35. Dwivedi, A. K. and Chouhan, U., Genome-scale classification of recombinant and nonrecombinant HIV-1 sequences using artificial neural network ensembles. *Curr. Sci*., 2016, **111**, 853–860.
36. Yasdi, R., A literature survey on applications of neural networks for human–computer interaction. *Neural Comput. Appl*., 2000, **9**, 245–258.
37. Xia, X. and Xie, Z., DAMBE: software package for data analysis in molecular biology and evolution. *J. Hered.*, 2001, **92**, 371–373.
38. Jenkins, G. M. and Holmes, E. C., The extent of codon usage bias in human RNA viruses and its evolutionary origin. *Virus Res*., 2003, **92**, 1–7.
39. García-Pedrajas, N., Hervás-Martínez, C. and Ortiz-Boyer, D., Cooperative coevolution of artificial neural network ensembles for pattern classification. *IEEE Trans. Evol. Comput*., 2005, **9**, 271–302.
40. Yao, X. and Liu, Y., Making use of population information in evolutionary artificial neural networks. *IEEE Trans*. *Syst. Man Cybern., Part B*, 1998, **28**, 417–425.
41. Haykin, S., *Neural Networks: a Comprehensive Foundation*, 1994. McMillan, New Jersey, USA, 2010.
42. Saha, A., Wu, C.-L. and Tang, D.-S., Approximation, dimension reduction, and nonconvex optimization using linear superpositions of Gaussians. *IEEE Trans*. *Comput.*, 1993, **42**, 1222–1233.
43. Lowe, D. and Broomhead, D., Multivariable functional interpolation and adaptive networks. *Complex Syst*., 1988, **2**, 321–355.
44. Lee, C.-C., Chung, P.-C., Tsai, J.-R. and Chang, C.-I., Robust radial basis function neural networks. *IEEE Trans. Syst., Man, Cybernetics, Part B*, 1999, **29**, 674–685.
45. Light, W. A., Some aspects of radial basis function approximation. In *Approximation Theory, Spline Functions and Applications* Springer, Dordrecht, 1992, pp. 163–190.
46. Rivas, V. M., Merelo, J., Castillo, P., Arenas, M. G. and Castellano, J., Evolving rbf neural networks for time-series forecasting with EVRBF. *Inf. Sci*., 2004, **165**, 207–220.
47. Broomhead, D. S. and Lowe, D., Multivariate functional interpolation and adaptive networks. *Complex Syst*., 1988, **2**, 321–355.
48. Vapnik, V., *Statistical Learning Theory*, Wiley New York, USA, 1998, vol. 3.

ACKNOWLEDGEMENTS. We thank the Department of Biotechnology, New Delhi for providing support under the Bioinformatics Infrastructure Facility of DBT at Maulana Azad National Institute of Technology, Bhopal.

Received 10 February 2017; revised accepted 20 August 2018

doi: 10.18520/cs/v115/i11/2063-2070