Journal of Biomedical Informatics 102 (2020) 103376

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Compositional framework for multitask learning in the identification of cleavage sites of HIV-1 protease

Deepak Singh*, Dilip Singh Sisodia, Pradeep Singh

Department of Computer Science and Engineering, National Institute of Technology, Raipur, C.G, India

ARTICLE INFO

Keywords:
HIV-1 protease
Multifactorial evolution
Multitask learning
Multiple Kernel learning
Protein encoding

ABSTRACT

Inadequate patient samples and costly annotated data generations result into the smaller dataset in the biomedical domain. Due to which the predictions with a trained model that usually reveal a single small dataset association are fail to derive robust insights. To cope with the data sparsity, a promising strategy of combining data from the different related tasks is exercised in various application. Motivated by, successful work in the various bioinformatics application, we propose a multitask learning model based on multi-kernel that exploits the dependencies among various related tasks. This work aims to combine the knowledge from experimental studies of the different dataset to build stronger predictive models for HIV-1 protease cleavage sites prediction. In this study, a set of peptide data from one source is referred as 'task' and to integrate interactions from multiple tasks; our method exploits the common features and parameters sharing across the data source. The proposed framework uses feature integration, feature selection, multi-kernel and multifactorial evolutionary algorithm to model multitask learning. The framework considered seven different feature descriptors and four different kernel variants of support vector machines to form the optimal multi-kernel learning model. To validate the effectiveness of the model, the performance parameters such as average accuracy, and area under curve have been evaluated on the suggested model. We also carried out Friedman and post hoc statistical test to substantiate the significant improvement achieved by the proposed framework. The result obtained following the extensive experiment confirms the belief that multitask learning in cleavage site identification can improve the performance.

1. Introduction

Infectious diseases being serious health and social challenge that is being faced by the world today. Due to its epidemic effect and non-existence of complete curable medicine, the problem remains a critical threat to sustainable development [1]. The human immunodeficiency virus (HIV) is one such infectious disease which is commonly supposed to be the causative agent of Acquired Immunodeficiency Syndrome (AIDS). To better navigate this infectious rate of HIV related diseases, the recent report of UNAIDS [2] and WHO [3] on the global statistics of HIV estimate around 36.9 million people are living with HIV in the year 2017, and approximately 1.8 million people were found newly infected in the same year. HIV is the actual infectious organism that if left untreated destroys a person's immune system. HIV utilizes a part of a human immune system called a CD4 cell to replicate the virus, and this eventually kills the CD4 cell of which are in limited amounts. Due to the HIV lifecycle, there is no cure however it can be treated. HIV infected cells can become latent, which means that they don't make new HIV viruses right after infection, but they could start at any point [4].

Highly active antiretroviral therapy (HAART) is a combination of drugs is taken to keep HIV at bay [5]. Each of the drugs inhibits HIV at a different point in its life cycle, usually by interfering with enzymes the virus needs for replication. The different drug classes include Protease inhibitors, Reverse transcriptase inhibitors, Integrase inhibitors, and Fusion inhibitors. These inhibitors prevent the proper functioning of protease by binding the active enzymes. HAART has been a major turning point in the management of HIV related disease [5]. Regardless of the successes in disease treatment and reduction in HIV related mortality, several challenges continue steadily to hamper the protease inhibitor therapies. The quick emergence of drug resistance is the most important concern since it renders current remedies ineffective and for that reason compels the researcher to keep efforts in the design of inhibitors that may effectively combat drug resistance [6]. Thus, for an effective design of inhibitors, the understanding and prediction of HIV-1 proteases activity in protein is of paramount importance [7].

Machine learning is concerned with the automatic acquisition of

* Corresponding author at: Department of Computer Science and Engineering, National Institute of Technology, G.E. Road Raipur, Chhattisgarh 492001, India.
E-mail addresses: dsingh.phd2016.cs@nitrr.ac.in (D. Singh), dssisodia.cs@nitrr.ac.in (D.S. Sisodia), psingh.cs@nitrr.ac.in (P. Singh).

https://doi.org/10.1016/j.jbi.2020.103376
Received 30 July 2019; Received in revised form 19 December 2019; Accepted 8 January 2020
Available online 11 January 2020
1532-0464/

models from the extracted features, as well as utilization of such models for intelligent inference and prediction, is often very useful in the interpretation of protease cleavage site identification [8]. Over the past years, several machine learning approaches have been proposed in the literature to deal with the classification of cleaved sites. These techniques broadly encompass two phases: peptide feature descriptors phase and learner phase [9]. Numerous feature extraction techniques were proposed by many researchers to accomplish the objective [10]. These feature extraction techniques exploit the various properties associated with the amino acids, some of the feature extraction methods use the amino acid sequence property [11] while some use the physicochemical property [12]. There are other techniques also exist which utilizes the sequence homology about proteins, evolutionary properties and genetic programming based feature extraction [13]. However, the collective formation of features by integrating them from the diverse properties of the amino acid seems to be underutilized and needs consideration for the improvement of feature quality.

In the learner phase, support vector machine and their kernel usage have shown incredible potential in the identification of cleavage sites [14]. Though, the performance of the model found superior, it still fall short when there is scarcity in data. In general, the scarcity of high-quality labeled data is a major issue in peptide analysis. The annotations of cleaved and non-cleaved sites can be identified by many biochemical experiments. However, the experimental methods are time-consuming and labor intensive, and thus, available data are limited and sparse. Using the sparse and limited quantity of data for supervised learning can only result in low generalization capability. To deal with the difficulty, many correlated tasks can be learned concurrently instead of independently which improves the generalization of the discovered models [15,16]. On peptide dataset, we observe that peptide databases with annotations from multiple sources are related to each other and carries few similarities. This observations inspire us to explore whether it is promising to disseminate the annotated knowledge across diverse sources such that predictions can be improved.

Knowledge transfer across the multiple related task can be achieved by Multitask learning (MTL). It utilizes the domain-specific information among the training cases of related tasks [17]. MTL is accomplished by learning multiple tasks in parallel and exploiting common shared knowledge within the tasks. Multitask learning in cleavage site prediction could improve the generalization by leveraging the common knowledge among the different datasets. However, this notion of multitask learning in the protease cleavage site analysis has certainly been unexplored. Therefore, in this paper, we propose an approach of exploiting multitask learning technique following the principles of common related feature sharing to enhance the performance in protease cleavage site prediction. Common feature sharing will be used to improve predictive performances, the swiftness of learning, and the intelligibility of discovered models.

The proposed model poses the problem of multitask learning as that of learning a shared kernel, constructed by combining a given set of base kernels broadly termed as multiple kernel learning (MKL). In the recent past, multiple kernel learning [18,19] techniques have been extensively exploited in many application, where multiple kernels are used instead of engaging one specific kernel function and its corresponding parameters. However, there is not yet a mechanism that is capable of guiding optimal kernel selection. Since choosing an ideal kernel with suitable feature is very important to the success of any multi-kernel method. Therefore the evolutionary based multi factorial optimization (MFO) [20] is used here that exploits the common shared features within the multiple tasks to model relationships between the features and kernels without the overhead of kernel selection and their parameters.

Another important component of the proposed framework is the integration of three genres of feature descriptors which includes seven different protein feature extraction techniques. Integration of features form the various feature extraction techniques could definitely aid the diverse property of amino acid which in response will be helpful for the learner [21]. But this integration introduces the curse of dimensionality problem. Handling a large number of features, it is important to identify the informative features amongst all, and this can be effectively achieved by the dimension reduction technique [22,23]. Large numbers of dimension reduction techniques have been evaluated in the biomedical field with promising results [24]. Here, the framework proposes a novel wrapper based feature selection method which is inspired from recently proposed principal component analysis (PCA) based Normalized Projection Error (NPE) [25] method. Since the data in multitask learning have diverse distribution, therefore, PCA based method is preferred over the other techniques.

The aim of this work is to contribute a novel framework that resolves the identified issues of data scarcity, kernel selection, feature selection and multiple kernels learning together in an ensemble approach that improves the performance of HIV-1 protease cleavage site identification. To achieve the objective, this paper uses seven feature descriptor schemes, feature selection technique, multi-kernel learner and the theory of factorial evolution combined to form an ensemble framework. Thus, the primary contribution of the paper can be summarized in following six points:

- As far as we know, it is the first multitask learning framework for the cleavage sites prediction that combines feature integration, feature selection, multi-kernel learning, and multifactorial evolutionary algorithm into a single composite framework.
- Each kernel has a different approach of data handling, and the appropriate kernel selection in such classification problem is a crucial decision, the use of proposed multi-kernel learning aids the capability of handling data without the overhead of kernel choice.
- To explore the possibility of diversified properties, we utilize the features obtained from the sequence, physicochemical and structural feature extraction techniques (seven feature descriptor techniques) and integrate them to improve the prediction performances.
- Not all the features are capable enough to fit into the classification of protease cleavage site models, and the removal of such features from the pool is vital which is effectively be achieved by feature selection techniques. A novel wrapper based feature selection based on normalized projection error is proposed to eliminate the redundant features.
- Optimal feature set preparation for each kernel is achieved by multifactorial evolutionary algorithm. This algorithm firstly tries to find out the optimal set of a feature from the pool of integrated features (fusion of feature from seven feature extraction technique) and then pair the selected features to the ideal kernel (any one from the four kernels).
- Extensive Benchmark dataset experiments and comparison with fifteen states of the art (7 multitask learning and 7 multi-kernel techniques) for the validation of the proposed model.

The rest of the paper is organized as follows; Section 2 reviews the previous work on the HIV-1 protease prediction, multitask learning and Multifactorial optimization. Section 3 describes the proposed methodology by dividing the section into five subsections: Section 3.1 illustrates the seven feature descriptor techniques, Section 3.2 introduce the multi-kernel learning. Section 3.3 describes feature selectionby normalized projection error, and Section 3.4 illustrates proposed multi-kernel model using factorial evolutionary approach. The experimental setup and result discussion are reported in Section 4. The concluding remark on the performed experiment by the proposed model and its comparison with the other state-of-the-art techniques is depicted in Section 5.

2. Preliminary studies and related work

2.1. HIV-1 cleavage sites classification

HIV-1 protease cleavage site classification is of paramount importance for the design and development of effective protease inhibitors [26]. Researchers have extensively used machine learning models for evolving the cleavage sites prediction techniques. There are various approaches however, they can be broadly categorized into feature descriptor phase and classification phase [10].

Feature descriptor broadly categorizes into structure-based, physicochemical-based and sequence-based; these descriptor maps the peptide amino acids into various residue property [21]. Furthermore, structure-based methods are exercised commonly by secondary structure elements (SSE) and solvent accessibility (SA). Amino acid database [27] based Physicochemical feature descriptors [28,29] uses the intrinsic physicochemical properties for encoding the sequences. There are many variants of physicochemical encoding schemes like average physicochemical encoding [12] and weighted physicochemical encoding [30] which exploits the fixed length of the sequence efficiently. Among the sequence-based features, Dipeptide composition (DipC), amino acid composition (AAC) [31], and Pseudo amino acid composition (PseAAC) [32] are the techniques that can better represent the amino acids in peptide. OETMAP encoding [33], genetic programming [34], are the techniques which exploit both sequence and physicochemical based features. Recently genetic programming based approach is proposed for the cleavage site prediction problem [35] where a new amino acid sequence encoding is accomplished by spatial and structural features.

In the classification phase Artificial neural network (ANN) [36,37], Decision tree (DT) [38], Rule-based predictive model [39], Support vector machine (SVM) [40,8], and variable context markov chains [41] are examined on the peptide classification problem. Among all these classifiers, the performance of linear SVM with orthogonal encoding [42] has been declared the best [14]. Furthermore, the role of ensemble models are also investigated where multiple classifiers [29] trained on different encoding schemes and found improved results than the single classifiers. In [13] genetic algorithm is used to train the multi-classifier system with optimal feature encoding method. The concept of ensemble classifier is further extended in [9,43] where the pairing of feature and base classifier is done by a genetic algorithm. The results obtained by the optimal ensemble model is promising with the advantage of autonomous kernel and feature encoding selection.

2.2. Multi-Task learning

MTL algorithms are majorly categories into the three forms: instance-based, feature-based and parameter-based. Instance-based MTL recognizes valuable data instances in a task for the other related tasks and then shares knowledge through the discovered instances. Feature-based MTL aims to understand common features among different tasks in an effort to share knowledge. Parameter-based MTL uses model parameters to cooperate in learning the various other tasks [44].

To begin with the feature-based approach, multi-layer feedforward neural network [45], which is one of the feature transformation strategies, is among the earliest model for multi-task learning. In [46], the radial basis function network, which has only one hidden layer, is prolonged to MTL by greedily identifying the framework of the hidden layer. Silver et al. [47] propose a context-based multi-task neural network which includes only one output layer shared by different tasks but includes a task-specific context as yet another input. Not the same as multi-layer feedforward neural systems which are connectionist versions, the multi-task feature learning (MTFL) technique [48,49] is formulated beneath the regularization framework. Multi-task sparse coding [50] technique is proposed to learn a linear transformation on features. In parameter-based MTL, there are five main approaches: low-rank approach [51], task clustering approach [52], task relation learning approach [53], dirty approach [54], and multi-level approach [55].

2.3. Multifactorial optimization

Multifactorial Evolutionary Algorithm (MFEA) [56] is originated as a computational analogy of the bio-cultural types of multifactorial inheritance [57]. The salient features of the MFEA is a multitasking environment which effectively combines the biological transmission of the cultural blocks from parents to their offspring [58]. Recently, a number of research efforts based on MFEA have proved highly effective in sharing the knowledge to achieve the multi-task learning [59,60]. Among them, multi-task learning for neural networks that evolves modular network topologies was one of the earliest models to realize the evolutionary multitasking [61]. Multifactorial genetic programming (MFGP) [62] based ensemble learning, which introduces multifactorial evolution to Genetic programming (GP) for ensemble learning. Multiple GP classifiers evolve concurrently with one single population to reduce the computational cost required in ensemble learning. Each of the tasks is defined by particular network topologies that may also be considered as modules which represent the elementary unit in knowledge transfer. An extension of MFEA in neural network learning is proposed on extreme learning machine in [63] where a modular training technique is used to evolve the modular topologies of an extreme learning machine. The evolutionary extreme learning machine and multi-task modular training are combined efficiently, and each task is defined with a different number of hidden neurons.

3. Methodology

We consider multitask learning for HIV-1 protease cleavage site prediction by learning across multiple data sources that have data distribution differences between the datasets. Formally, cleavage site prediction problem can be defined as follows: let $\mathscr{T} = \{(x_i, y_i)\}_{i=1}^{N}$ be a peptide training instances of length N where $x_i \in \mathscr{P}$ (amino acid sequence of length eight) and $y_i \in \{unclevage, cleavage\}$ represents the binary class non-cleave and cleaved site. A classification model is a mapping from instances of peptide x to predicted classes y: $f(X)$: $\Re^d \mapsto \{unclevage, cleavage\}$.

Before we describe our proposed multitask approach, we briefly introduce the protein feature descriptors techniques, multi-kernel learning, normalized projection error, and functioning of a multifactorial evolutionary algorithm that lay the foundations for our approach.

3.1. Protein feature descriptors

Several protein feature extraction methods are studied in various protein analysis problem. Feature descriptor techniques maps amino acids sequences into an effective mathematical expression. The mapping of the alphabets is attained to reflect the coherent properties that can truly correlate with the attribute to be predicted [32]. Here, we considered three different genres of feature descriptors which includes composition information, physicochemical property, and structural/functional property. They further bear various methods individually, here we scrutinize seven feature descriptors techniques. Table 1 shows the characteristics of seven feature extraction techniques.

3.2. Multi-Kernel learning

Kernel methods in learning give a structured approach to employ a linear algorithm in a transformed feature space. This transformation is typically nonlinear or mapping data to a higher dimensional space. In this work, we consider four popular kernels which are Linear, Polynomial, Gaussian, and Sigmoidal.

Table 1
Characteristics of protein feature descriptors.

Feature Extraction Method	Genre	Description	Feature Length
Pseudo Amino Acid Composition (PseAA) [32]	Sequence-based	"Maps the peptide into (20 + λ) real values where λ denotes the maximum distance between the two amino acids. (λ = 15)".	20 + λ
Amino Acid Composition (AAC) [31]	Sequence-based	"The amino acid composition is defined as the percentage or fraction of amino acid in the protein sequence".	20
Dipeptide composition (DipC) [32]	Sequence-based	"This technique counts the frequency of a couple of amino acids in a peptide. It is also referred to as 2-gram".	400
Quasi-Residue Couple [64]	Physicochemical based	"It takes physicochemical (hydrophobicity, polarizability, steric property, isoelectric point, volume, and polarity) property of the protein. In the current setting, we consider order $m \leq 3$".	1200
Substitution matrix representation (SMR) [65]	Sequence-based	"Amino acid index database has predefined substitution matrix which replaces the amino acid with the matrix value".	400
Secondary structure elements (SSE) [21]	Structure-Based	"Takes a peptide (amino acid sequence) and matches it with either of the three secondary structural shapes which are helix, strand, and Coil".	24
Solvent accessibility (SA) [21]	Structure-Based	"Three feature descriptors including buried or exposed class, relative solvent accessibility and absolute surface accessibility".	24

Table 2
Properties of HIV-1 Protease Benchmark Dataset.

Name	Characteristics	Description
Data_746	Linearly Separable	Data has been collected from various literature from 1988 to 2003 including 362 octamers from Cai & Chou 1998
Data_1625	Linearly Separable	Data has been collected for 16 years (1990 to 2005) including 362 octamers from Cai & Chou 1998
Data_Schilling	Non-Linearly Separable	Independent dataset (corrected version of the Schilling and Overall 2008 dataset)
Data_Impens	Linearly Separable	Data has been collected from various literature from 2004 to 2012 (comprises data from the four publications)

$$\mathscr{K}_{lin}(X_i, X_j) = \langle X_i, X_j \rangle \quad \mathscr{K}_{poly}(X_i, X_j) = (\langle X_i, X_j \rangle + 1)^q$$

$$\mathscr{K}_{rbf}(X_i, X_j) = \mathbf{exp}(-\|X_i - X_j\|_2^2/s^2) \quad \mathscr{K}_{sig}(X_i, X_j) = \mathbf{tanh}(X_i X_j)$$

Selecting the kernel function $k(\hat{A}\cdot,\hat{A}\cdot)$ and its parameters can be an important concern in training. Therefore, multiple kernel learning (MKL) [18] strategies have been proposed, where multiple diverse kernels actively participate rather than selecting one particular kernel function and their corresponding parameter.

$$\mathscr{K}_c(X_i, X_j) = f_c(\{\mathscr{K}_m(X_i, X_j)\}_{m=1}^{p}) \tag{1}$$

Where the kernel combination function $f_c(.)$ can be linear or non-linear function. We engaged linear combination of the weighted kernels listed in Eq. (2). where the kernel weights are selected by the factorial evolution algorithm. More details about the combination of kernels could be referred from [18].

$$\mathscr{K}_c(X_i, X_j) = \sum_{m=1}^{p} \{\eta_m \mathscr{K}_m(X_i, X_j)\} \tag{2}$$

Combining kernel will aid the benefits of structure diversity, since different kernels match different ideas of similarity, and rather than seeking for the best, a learning technique will make a combination of them. Utilizing a specific kernel could be a way to introduce bias, and forming multi-kernel model by utilizing many diverse kernels could be possible solution to avoid the bias. Second benefit is accomplishing data diversity; different kernels could be handling the various data representations perhaps from different sources or methods. Since they are the data with different representations, they have different methods of similarity corresponding to the kernels. Hence, combining kernels is normally one feasible method to mix multiple information resources.

3.3. NPE based feature reduction

Normalized Projection Error ($\mathscr{N}_{pe}$) can be formally defined by considering the subset of features $f = \{f_1, f_2, \cdots, f_{n_s}\}$ and its transformed features $tf = \{tf_1, tf_2, \cdots, tf_{n_s}\}$ from a method$M$. Let eigen values $ev = \{\lambda_1, \lambda_2, \cdots, \lambda_{n_s}\}$ is associated with transformed feature tf such that tf_i has $\lambda_i$eigen value and $\lambda_1 > \lambda_2 > \cdots > \lambda_{n_s}$. These eigen values signifies the amount of error and while projecting the data from $\mathscr{L}$ to $\mathscr{L} - 1$ dimension the normalized projection error of the feature set $f_1, f_2, \cdots, f_{\mathscr{L}}$ is defined as

$$\{\mathscr{N}_{pe}\}_{1,\cdots,\mathscr{L}}^{\mathscr{L},\mathscr{L}-1} = \frac{\lambda_{\mathscr{L}}}{\sum_{i=1}^{\mathscr{L}} \lambda_{\mathscr{L}}} \tag{3}$$

If the predefined threshold value is greater than the $\{\mathscr{N}_{pe}\}_{1,\cdots,\mathscr{L}}^{\mathscr{L},\mathscr{L}-1}$ then the reduction of the feature set from $\mathscr{L}$to $\mathscr{L} - 1$ could be considered. More detailed about the NPE can be referred from [25].

3.4. Factorial evolution based multi-task learning for HIV-1 protease cleavage site prediction

The protease cleavage sites prediction is a standard binary classification problem. On analysing the benchmark data (Table 2), there are few number of samples and could be insufficient for the model training. Due to limited samples in the datasets, the conventional learning algorithms tend to underperform. Therefore, utilizing auxiliary dataset could be favourable in such cases.

In this paper, we engaged multifactorial optimization (MFO) to facilitate multitask learning in predicting the cleavage sites. We incorporated seven feature descriptor techniques and integrated the features to cover the various inherent properties of the protein. Due to integration the size of the training dataset increases, subsequently strengthen the training and prediction processes. Conversely, the increased data size may slow the training phase. Therefore, we additionally apply the feature reduction process with the help of normalized projection error. Framework also considered the performance gain produced by multi diverse kernel model. But, different kernels have different characteristics corresponding to the given input data therefore, we utilize the factorial evolution algorithm whose primary objective is to produce the optimal data for each different kernels. Thus, together with dimensionality reduced features and multi-kernel model effectively governed by multi factorial evolution algorithm in the framework to realize the multitasking in cleavage site prediction. Specifically, our aim is to optimize the two complementary objectives

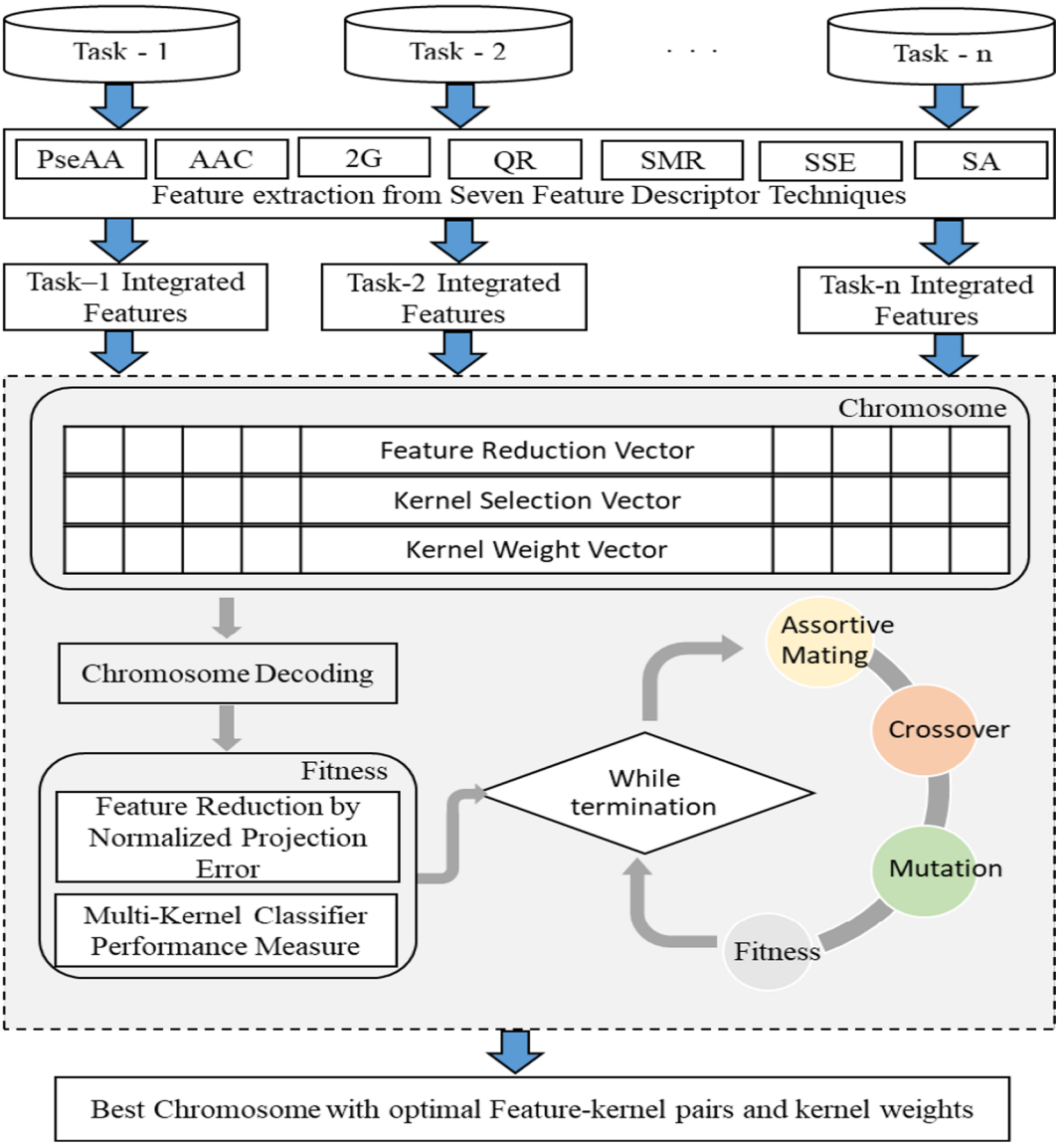


Fig. 1. Process Flow Diagram.

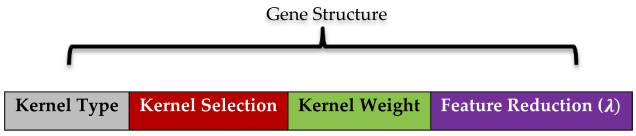


Fig. 2. Structure of the Gene.

- Minimizing the number of features that can be effectively employed in the cleavage site prediction.
- Maximizing the Area under the curve values, such that the performance of the model improves.

The process flow diagram of the proposed model is represented in Fig. 1. In the figure, a task represents the different data source. These data sources carry multiple octapeptide instances. To realize multitask learning framework, the proposed framework uses a multi-kernel learning model. However, the kernel selection, optimal features for the kernel training and the weight of the kernels are the factors that are being administrated by multifactorial optimization (MFO) algorithm. MFO performs concurrently evolutionary search for the optimal solution on multiple different tasks. The novelty in comparison to the other multitask learning method is the optimal selection of individual kernel and its parameters in the multi-kernel model which has the potential to handles multiple data source of varied distribution. In this implementation, instead of initializing the whole individual as single task learning model, here each chromosome is partitioned and represented by the subtask. Additionally, a characteristic of MFO is the vertical cultural transmission that maintains multitask learning in the chromosome throughout the evolutionary process. With the multi-kernel requirement of having diverse and accurate kernels, the method is able to exploit this concurrent learning by optimal feature selection, instead of using a large pool of features. Before delving into the procedure of the MFEA, it is important to define the individual representation (see Fig. 2 and Fig. 3).

The chromosome structure of the proposed model is divided into n parts named as gene where $n$is the number of independent training datasets. Each partition is bearing identical structure and its formation is shown in Fig. 2. The considered length of the gene is four with hybrid encoding (mixture of real and binary values). The first element in the gene structure indicates kernel type (linear at position 1, polynomial at position 2, Gaussian at position 3 and sigmoid at position 4). The second position indicates the presence and absence of the corresponding kernel. The third element depicts the kernel weight which is

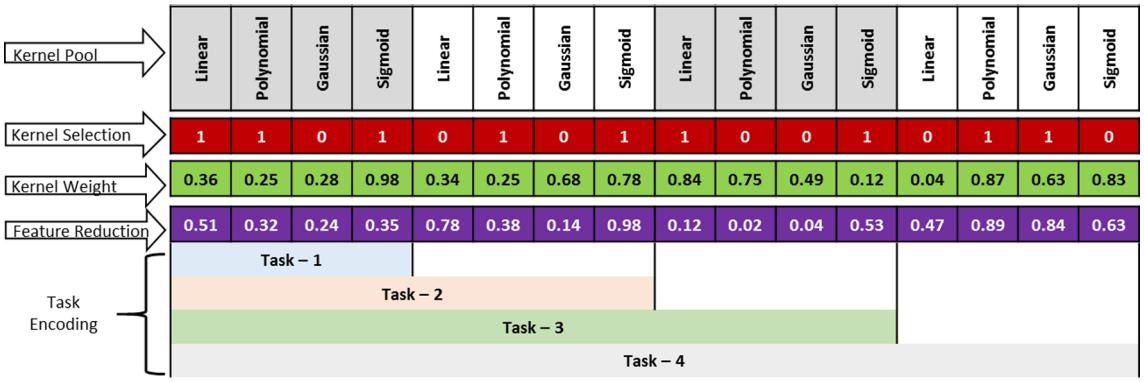


Fig. 3. Chromosome Representation of the proposed model.

used when multiple kernel combined with weights (ηin Eq.2). Lastly feature reduction factor (λ)is a threshold value for the selection of features from the Normalized Projection Error (NPE) model. Thus the dimension of the chromosome is limited by three vectors (Kernel selection - binary vector, Kernel weight – real vector of values ranging {0, 1}, Feature reduction – real vector) of length $n \times 4$ (number of task × number of unique kernel).

Based on the type of formulation required to represent the solution, the individual formed by the algorithm is represented in Fig. 3. In the figure, the three colored vector (red for kernel selection, green for kernel weight and violet for feature reduction) shows a random chromosome of length 16 (4 number of task ×4 number of kernels). The kernel selection vector carries binary values which indicates that total nine kernels are selected for the formation of multi-kernel learners. However, this kernel is further grouped by the tasks, where the three kernels (linear, polynomial and sigmoid) will be handling the first task (first data source with corresponding weights and feature selection threshold values). Task − 2 will be handled by five kernels (1-linear, 2-polynomial and 2-sigmoid with adjacent weights and threshold values). To learn Task – 3, seven kernels are selected by the chromosome (2-linear, 2-polynomial and 3-sigmoid). And for the Task-4, all the nine selected kernels will be exercised in the learning (2-linear, 3-polynomial, 1-gaussian and 3-sigmoid with their weights and feature selection threshold values).

The fitness function used in the proposed model for the assessment of each individual is the area under curve (AUC) value achieved by the corresponding multi-kernel techniques. MFO mechanisms split the individual into different skill groups or task; each has its own objectives. However, in our case, we assign each task with a single common objective given in Eq. (4).

$$\textbf{\textit{maximize}} \quad \textbf{\textit{fit}} = \textbf{\textit{AUC}} \tag{4}$$

In the first step, the initialization of the chromosomes were done in a random manner. Each chromosome is partitioned by the number of tasks, so the evaluation of each task in the individual is achieved by the fitness function Eq. (4). In this case the dimensions of decision space of different tasks in the MFO problem is set four (example shown in Fig. 3), where the MFO problem contains four tasks (independent data sources), whose dimensions are four (kernels) each respectively. In the evaluation of Task-1, only the first part (the first 4 values of the three vectors) of the individuals signifies the multi-kernel model, while the remaining 12 values are not used for the evaluation of the first task. Similarly, Task-2 evaluation is accomplished by considering the first eight values of the three vectors, and the remaining eight values remained unused. As a result, the four dimension of Task-2 cannot be used for exchanging the information with Task-1. Furthermore, Task-1 dimension values can be used for sharing the knowledge from Task-1 to Task-2. In this way, a good solution of Task-1 could be considered good for Task-2 because the task are related and the solution of the model belongs to the similar search space.

Algorithm 1 (*Multi-Factorial Evolution for Multi-Task Learning*).

Procedure MFO_MTL(***np***, ***nd***, ***nt***, ***rmp***, ***Task***)
Input: *np*, *nd*, *nt*, *rmp*, *Task*
np: *number of population nd : maximum dimension amongst the tasknt: number of generations*
rmp: *mating probability Task*: *Task list*
1. $itr \leftarrow 0$ // *iteration variable*
2. *for* $i = 1$ *to np do*
3. *generate chromosome* C_i // *random bit string of length*nd
4. $pop_{itr}(i) \leftarrow C_i$ // *population initialization*
5. *endfor*
6. *for each* P_i *in* pop_{itr}
7. $Skill_i \leftarrow mod(i + k) + 1$
8. *for each* t_j *in Task*
9. $Factcost(i, j) \leftarrow EvaluateCost(P_i, t_j)$
10. *endfor*
11. *end*
12. $Scalerfitness_{itr} \leftarrow ComputeFitness(pop_{itr})$
13. *while termination criteria or* $itr < nt$
14. *if* $rand > rmp$ *then* // *mating criteria*
15. $pop_{child} \leftarrow Crossover(pop_{itr})$
16. *else*
17. $pop_{child} \leftarrow Mutation(pop_{itr})$
18. *endif*
19. *for each* C_j *in* pop_{child}
20. $Skill_j \leftarrow ComputeSkill(pop_{child}, Task)$
21. $Factcost(j) \leftarrow EvaluateCost(pop_{child}, Task)$
22. *endfor*
23. $pop_{itr} \leftarrow pop_{child} \cup pop_{itr}$
24. $Scalerfitness_{itr} \leftarrow ComputeFitness(pop_{itr})$
25. $pop_{itr} \leftarrow Select\_np\_Elitist(pop_{itr})$
26. $itr \leftarrow itr + 1$
27. *end*
28. *return fittest individual*

To assess the individual task, a skill factor in initialized and evaluated. Each individual is evaluated only for the associated task (the assigned skill factor). For the fitness of the individual, K-fold cross-validation [66] method is used which evaluate the predictive performances of the obtained model. 10-fold cross validation is chosen here on the ground that it may often attain the smallest variance and mean squared error in the estimation of prediction errors. The integrated features of the dataset are partitioned into 10 subsets of training data (90%) is used in learner as a training purpose, and testing data (10%) is used for the model evaluation. The average AUC of the 10 subsets of such partitions is regarded as the fitness of the chromosome to evolve into optimal multitask learning model. In the proposed model, we considered the fusion of seven different protein feature descriptors

techniques, four varied kernels of support vector machines and three number of tasks which correspond to the benchmark peptide dataset (3 for training model generation and 1 for model evaluation) were simulated during the experiment. The individual kernel-feature pairs are then exploited by MFEA to find out the best multi-kernel model.

To conclude with the justification of the proposed system, the reason for the multi-kernel model formation is due to the observed accuracy of the multi-kernel learner found superior when compared with the individual kernel model in many applications [67]. Moreover, the performance of the RBF and polynomial kernel is noticeably competitive with the linear kernel in the protease site cleavage identification [14]. Therefore, in the optimal formation of the multi-kernel model, we consider the four kernels with varied features. However, to find the optimal model from this large search space of feature and kernels, exhaustive search would have been computationally costlier. So as a solution multifactorial evolutionary algorithm is exploited in this work to get the optimal combination. More interestingly and importantly, the theory of transfer and sharing of knowledge across the task potentially lead to the discovery of effective multitask learning model. Thus through MFO, it is possible that the genetic material created within a particular group turns out to be useful for another task as well. Lastly, the notion of solving the problem through multitask learning overcomes the data scarcity problem.

Algorithm 2 (*Fitness Computation***).**

Procedure ComputeFitness **(*octamers*, *pop*)**
1.*for* $i = 1$ *tonp*
2.*for* $j = 1$ *to ntask*
3. $Classifier_{mkl} \leftarrow Decode(pop(i))$ // *Decoding for kernel selection*
4. $Feature_{select} \leftarrow Decode(pop(i))$ // *Decoding for feature selection*
5. $Data_{bags} \leftarrow Prepare\_each\_kernel(Feature_{select}, Classifier_{mkl})$
6. $Data_{train}, Data_{test} \leftarrow K\_fold\_validation(Data_{bags})$
7. $Classifier_{mkl} \leftarrow Training(Data_{train})$
8. $Perf\_measure \leftarrow Testing(Classifier_{mkl}, Data_{test})$
9.*end*
10.*end*
11. $ComputeAUCofClassifier_{mkl}model$
12.***return*** *AUC*

Although there are many sophisticated methods of combining classifiers with a fusion of various feature descriptors technique exist, the use on HIV-1 protease classification is very limited. To achieve the desired model we have designed two algorithms represented in the form of pseudo code. The algorithm 1 is the primary MFO based multi-kernel model formulation, and algorithm 2 is the pseudo code for the calculation of fitness for each individual. Algorithm 1 outlines the main steps of the proposed multitask learning framework, which is an implementation of the MFEA algorithm described in [58] obtained by customized individual representation for the formation of weighted multiple-kernel classifiers. The termination criteria set for the MFO algorithm is the maximum number of iterations and function tolerance with a value of 0.001.

4. Experimental setup and results

The experimental session aims to compare the generalization accuracy of peptide classification when there are multiple sources employed on the learning algorithm. Here we evaluate the considered hypotheses of multiple related data sources in making the HIV-1 protease cleavage site classification better. The performance criterion for the evaluation of the proposed model are evaluated on the six measures including Accuracy, F-Measure, Specificity, Sensitivity, Area under the receiver operating characteristics curve (AUC), and Mathews correlation (MCC) which is defined as follows:

$$Accuracy = \frac{\mathscr{T}_p + \mathscr{T}_n}{\mathscr{T}_p + \mathscr{T}_n + \mathscr{F}_p + \mathscr{F}_n} \quad Sensitivity = \frac{\mathscr{T}_p}{\mathscr{T}_p + \mathscr{F}_n}$$

$$Specificity = \frac{\mathscr{T}_n}{\mathscr{T}_n + \mathscr{F}_p} \quad F - Measure = \frac{2.\mathscr{T}_p}{2.\mathscr{T}_p + \mathscr{F}_p + \mathscr{F}_n}$$

$$MCC = \frac{\mathscr{T}_p \times \mathscr{T}_n - \mathscr{F}_p \times \mathscr{F}_n}{\sqrt{(\mathscr{T}_p + \mathscr{F}_p) + (\mathscr{T}_p + \mathscr{F}_n) + (\mathscr{T}_n + \mathscr{F}_p) + (\mathscr{T}_n + \mathscr{F}_n)}}$$

Here $\mathscr{T}_p$and $\mathscr{T}_n$ denotes the number of correctly predicted cleaved and non-cleaved sites, and $\mathscr{F}_p$and $\mathscr{F}_n$ denotes the incorrectly predicted cleaved and non-cleaved sites.

To validate, we demonstrate the proposed model on the standard benchmark data set obtained from UCI machine learning repository [68]. The standard benchmark carries four different datasets which includes Octamers with 746 instances (401 cleaved and 345 non-cleaved) [40], Octamers with 1625 instances (374 cleaved and 1251 non-cleaved) [39], Octamers with 3272 instances (434 cleaved and 2838 non-cleaved) [69] and Octamers with 947 instances (149 cleaved and 798 non-cleaved) [70]. Various characteristics of the benchmark datasets are listed in Table 2.

Our focus in this paper is to introduce the multitask learning in the peptide cleavability classification that can deal with the data scarcity and dataset drift problems. Thus, to make the experimental results more convincing, we form the extended dataset by varied combinations of benchmark data. The various combinations of the four benchmark datasets are listed in Table 3. A total of seven compound data is prepared that were exercised for the performance evaluation of the proposed multitask learning model. The table is organized with four columns where, first column list the various combination of benchmark data, second column list the unique number of instance present after the conflict removal steps. Third column displays the number of non-cleavage entries and fourth column list the number of cleavage entries. The benchmark data has many overlaps and conflicts, therefore during the formation of extended dataset, we followed the pre-processing steps in [14] for the filtration of conflict entries.

The parameter values considered during the experiment are reported in Table 4. These parameters of the considered algorithms are initialized with the listed values for the evaluation of individual as well as proposed classifiers. We utilize ten-fold cross-validation to estimate the various performance measure of each experiment. For multitask learning validation, the training set is formed by two or more than two datasets, and one of the datasets is reserved as the testing set. Further, ten runs of a random shuffle of the test dataset are examined and reported in the tables.

4.1. Peptide dataset analysis

In this section, explorative analyses and overview visualizations of the peptide dataset is presented. They mostly highlight correlations between features. Here we used heatmap to graphically visualize the matrix data by representing individual values with different colors. It is the most common method to show structures of the data. Heat maps are versatile in that both x-axis and y-axis could be organized to explore a

Table 3
Characteristics of the Extended Dataset.

Benchmark Combinations	Unique Octamers	Non-Cleaved sites	Cleaved sites
746_1625	1707	1287	420
Impens_746	1693	1142	551
Schilling_746	4018	3182	836
Impens_1625	2572	2048	524
Schilling_1625	4877	4068	809
Impens_746_1625	2654	2085	569
Schilling_746_1625	4959	4105	854

Table 4
Considered Parameter Values for the Experimentation.

Method	Parameters
Multi FactorialEvolutionary Algorithm	Generation = 500, Population = 50, Dimensions = 12 probability of individual learning = 1, Selection = 'roulette wheel' Crossover = 'simulated binary crossover', Mating probability = 0.3; standard deviation for gaussian = 0.02, Mutation = 'Gaussian'
Multi Kernel Learning	Kernel = 3 (linear, polynomial, gaussian) Gaussian sigma = 2, Polynomial degree = 2 Normalized kernel = 'True', Normalized data = 'True' Regularization Parameter = 1, Optimization method = 'SMO' Tau for SMO = 1e-3, Threshold = 1e-3 Combination Method = {convex, linear} Regularization norm = 1, (RBMKL Rule = 'mean')
Multi Task Learning	Function tolerance = 1e-5, Initial point = 0 Terminate flag = 'ON', Maximum iteration = 1500 Lambda_1 = { 1 : −0.01 : 0.01 }, Rho_1 = 0.01, Lambda_2 = { 2 : −0.05 : 0.05 }

specific hypothesis and allow an easy summary of data with a third quantitative dimension captured in different colors. The transformation and visualization of multi-dimensional peptide data in a single heatmap can provide a concise but comprehensive presentation of amino acids distribution under different peptide positions.

In the heatmap (Fig. 4 and Fig. 5), the x-axis represents the amino acids, and the y-axis denotes the peptide subsites (P1 – P8). The color represents the observed frequency of the amino acids. Green color represents the average abundance for that unique amino acid across the corresponding sites. A yellow color indicates the highest frequency of the amino acids and a light green and light yellow color indicates the average appearance of the amino acids. The magnitude of the color change correlates with the magnitude of change in the frequency of the amino acids at the specific peptide sites.

From the Figs. 4 and 5, we can see the distinction between the cleavage and non-cleavage sites. Fig. 4(a) denotes the cleavage sites of Data_746 data, and 4(b) presents the non-cleavage sites of the Data_746. Similarly, 4(c) realizes the cleavage sites of Data_1625 and 4(d) represents the non-cleavage sites of Data_1625. Due to the high overlapping peptides amongst these two datasets, the non-cleavage sites are highly symmetric. Despite the commonalities, the cleavage sites amino acids frequency of both datasets differs. Fig. 5(a) and (b) exhibit the cleavage sites and non-cleavage sites amino acid frequencies and Fig. 5(c) and (d) unveil the relation of the amino acids frequency specific to the peptide sites. Heat maps of both the dataset show the variation in the amino acid distribution in cleavage and non-cleavage sites; the difference can be seen across the dataset. Therefore, the sequence of amino acids in benchmark peptide datasets exhibit the dissimilarity across the datasets.

4.2. Results and discussion

We performed extensive quantitative experiments on benchmark and extended dataset using a multitude of different multitask and multi-kernel learning techniques in order to answer the following questions:

- How does the proposed multitask model perform in terms of classification performance when multiple independent peptide datasets are collectively used as training data?
- How does the peptide cleavability classification handle by the state-of-the-art multitask learning methods in comparison to the proposed model?
- Since the core of the proposed model is multi-kernel; thus the performance accountability of the proposed model relative to the state-of-the-art multi-kernel methods is arguable.
- What is the relationship between the different protein feature descriptor, multiple kernels and predictive performance efficiency?

To answer the first question, we comprehensively evaluated the proposed multitask learning model on the standard and extended dataset. The predictive performance of proposed ensemble model is summarized in Table 5. The table is organized in four columns where the first column reports the names of the dataset, the second column depicts a number of specific kernels required in the optimal model, and the last two column lists the performances measure accuracy and AUC. The kernels column is further specified by the contribution of the individual kernel. The values in the kernel column are expressed in the percentages (U / T where U is the number of kernels employed in the ensemble and T is the number of total kernels). The values listed in the table are the outcome of the best chromosomes.

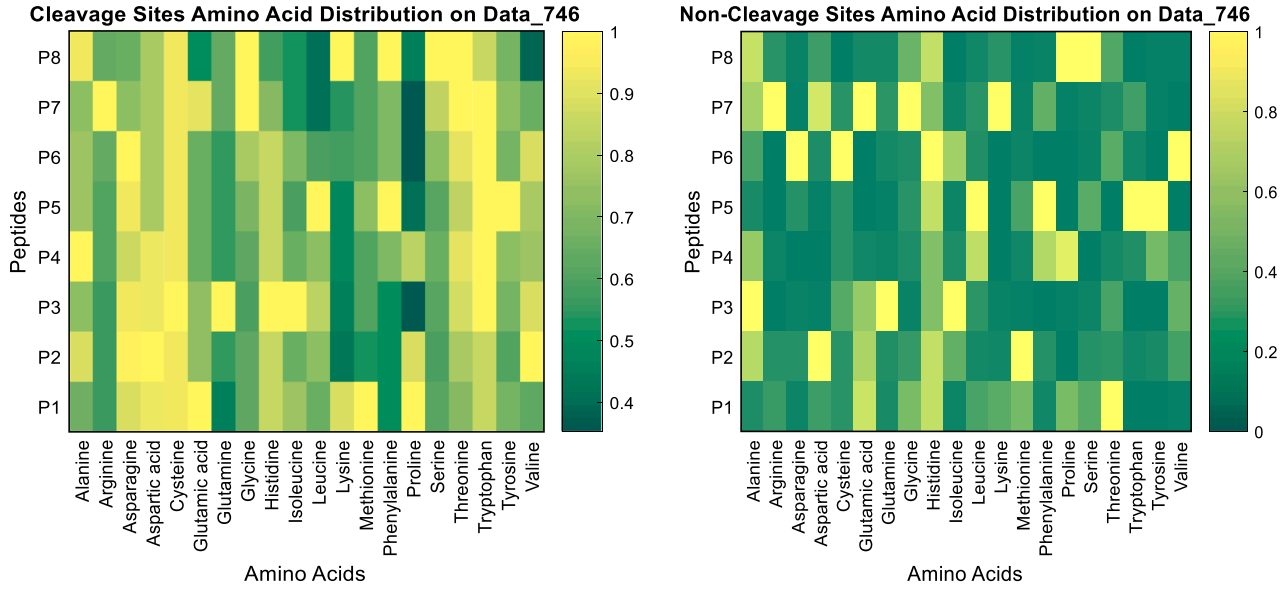
To further investigate, we performed an experiment by varying the number of learning tasks in the proposed model. These tasks are denoted by the individual learning dataset. In Table 6, we report the Accuracy and AUC obtained by combining one or more than one learning task. The first column of the table represents the name of the test dataset, a second column which is further portioned into the number of training data sources. In the experiment, we have considered 1, 2 and 3 data sources for training the multitasking model which is different from the test dataset.

Based on the observation of Tables 5 and 6, We can highlight the following points:

- The result reported the highest accuracy of 96.65 and highest AUC value to be 0.996 in *Data_1625*.
- Polynomial Kernel Contribution is the highest (30.81%) amongst the four in the design of multi-kernel model.
- Despite the involvement of linear and nonlinear data in training and testing process, the performance of the proposed model remains unaffected and continues prediction with the high predictive rate.
- As there is increase in data sources, there is also a performance boost in the proposed model. The average accuracy and AUC obtained 93.69 and 0.9638 when the data source is one. With two data sources, the average accuracy is 94.03 and AUC is 0.9658.With data source three the average accuracy is 95.20 and AUC is 0.9725.

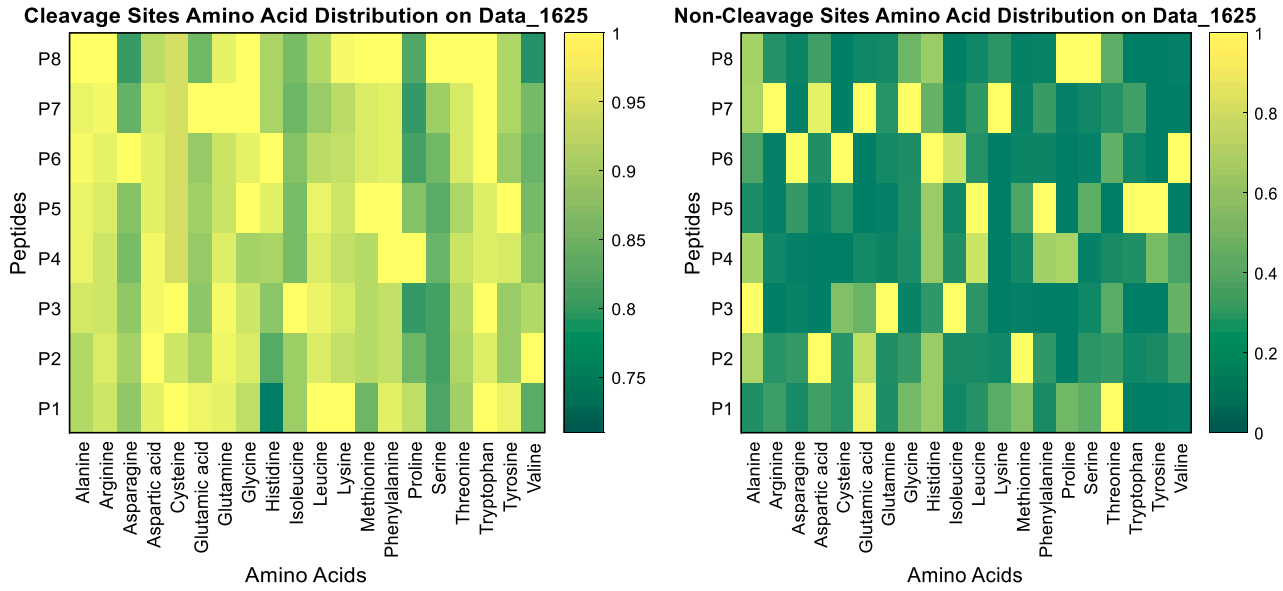
Receiver operating characteristic curve (ROC) play important role for performance evaluation of binary classifiers. Precisely, it is a graphs used for visual evaluation of classifier performances. The two-dimensional ROC space is normally traversed along the axes sensitivity and specificity. ROC presents the better estimation of the accuracy. AUC is the area spanned by the ROC curve, and it correlates the quality of classification.

Fig. 6 presents the ROC curve and AUC values of the proposed model on all the benchmark datasets. From the figure it is observed that AUC of multitask learning model on Data_1625 is highest in comparison



(a) Heatmap chart of the cleavage sites of Data_746

(b) Heatmap chart of the non-cleaved sites of Data_746



(c) Heatmap chart of the cleavage sites of Data_1625

(d) Heatmap chart of the non-cleaved sites of Data_1625

Fig. 4. Heatmap chart of Data_746 and Data_1625 dataset for cleavage (a, c) and non-cleavage sites (b, d).

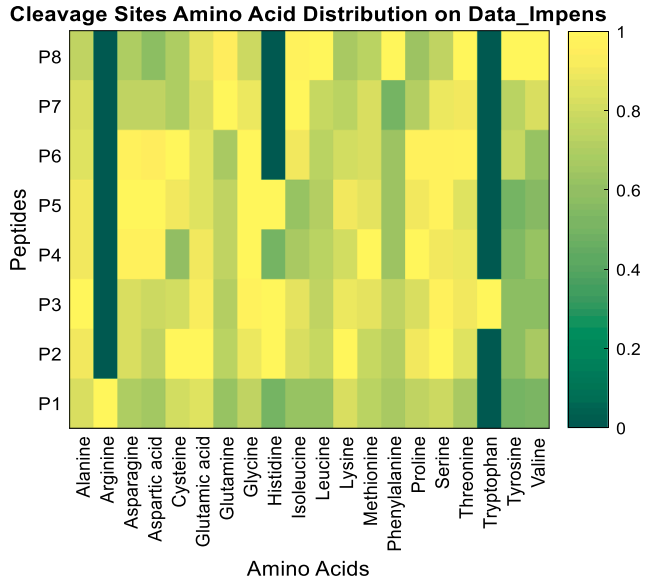
to the four benchmark data. Overall, the observed AUC values suggest that proposed multitask learning model has substantial potential in predicting the cleavability of HIV-1 protease sites with higher performance measure.

4.3. Multi-Task learning performance comparison

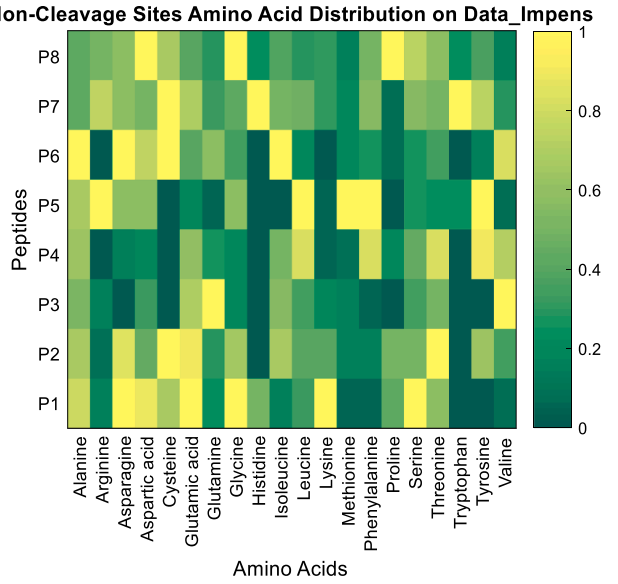
To answer the second question, we considered eight current state-of-the-art multitask learning techniques and performed comprehensive experiment on standard peptide benchmark dataset. To the best of our knowledge, it is the first attempt to apply these eight multitask learning techniques on the peptide dataset to identify the cleavage sites. These include Robust Multitask (RMTL) [54],Dirty Multi-Task Learning (Dirty) [71], Convex multi-task feature learning (MTLF) [49], Joint feature learning (JFL) [72], Trace-Norm Regularized Learning (Low Rank) [73], Sparse Structure-Regularized Learning (Least Lasso) [74], and Multi-task Lasso (Multi Lasso) [75]. MALSAR [76] toolbox has been used in the implementation.

Table 7 outline the performances comparison of eight multitask learning techniques. The first column depicts the test data and the remaining three benchmark data is used as a training data. From the table we observed the following major trends:

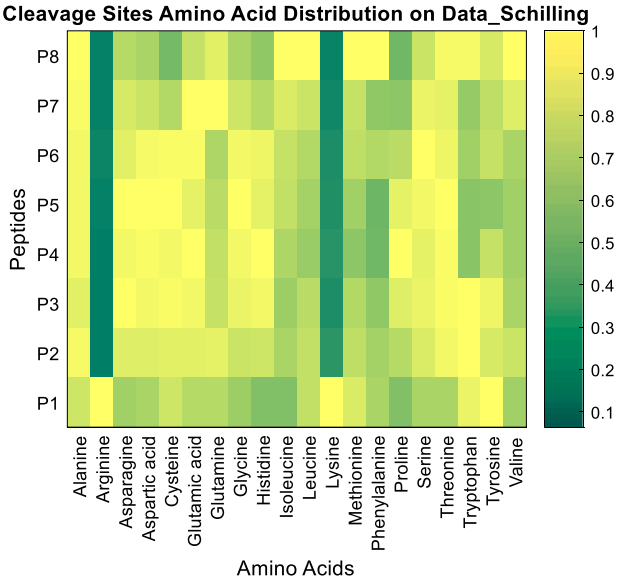
- Proposed model achieves the highest performance values on *Data_1625* dataset.
- A descent amount of improvement is observed over the state-of-the-art multitask learning techniques.
- The consistency of the proposed model can be observed from the performance measures across each benchmark datasets.



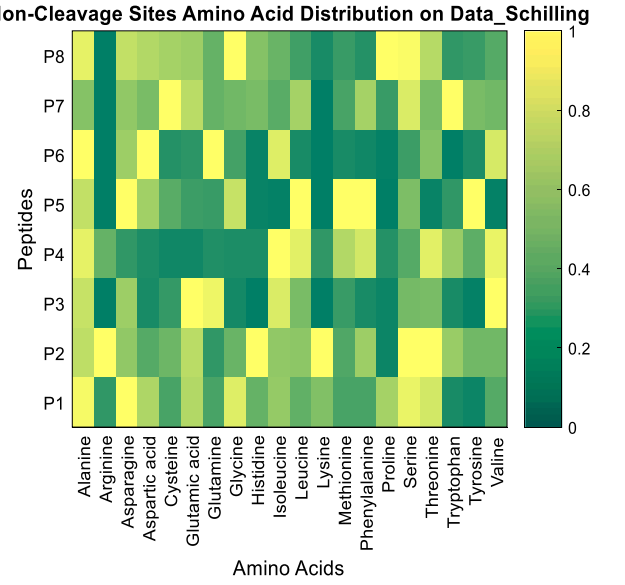
(a) Heatmap chart of the cleavage sites of Data_Impens



(b) Heatmap chart of the non-cleaved sites of Data_Impens



(c) Heatmap chart of the cleavage sites of Data_Schilling



(d) Heatmap chart of the non-cleaved sites of Data_Schilling

Fig. 5. Heatmap chart of Data_Imepns and Data_Schilling dataset for cleavage (a, c) and non-cleavage sites (b, d).

Table 5
Classification Performance of the proposed model (10-Fold).

Dataset	Kernels				Total Number of Kernels	Accuracy	AUC
	Linear	Polynomial	RBF	Sigmoid			
Data_746	30.00	30.00	10.00	30.00	7	96.54	0.9924
Data_1625	37.50	25.00	25.00	12.50	6	96.65	0.9963
Data_Schilling	27.27	18.18	27.27	27.27	8	95.77	0.9741
Data_Impens	28.57	28.57	28.57	14.29	5	94.64	0.9633
746_1625	27.27	36.36	9.09	27.27	8	96.29	0.9824
Impens_746	30.00	30.00	10.00	30.00	7	94.39	0.9648
Schilling_746	12.50	25.00	50.00	12.50	6	93.59	0.9524
Impens_1625	37.50	37.50	12.50	12.50	6	94.61	0.9673
Schilling_1625	25.00	33.33	33.33	8.33	9	94.71	0.9739
Impens_746_1625	25.00	50.00	12.50	12.50	6	95.48	0.9629
Schilling_746_1625	33.33	25.00	25.00	16.67	9	94.57	0.9674

Table 6
Performance of the proposed model with Varied Task Size (Data sources).

Test Dataset	Number of Task (Data Sources)					
	1		2		3	
	Accuracy	AUC	Accuracy	AUC	Accuracy	AUC
Data_746	96.13	0.9854	96.39	0.9848	96.54	0.9951
Data_1625	93.78	0.9823	94.56	0.9868	96.65	0.9963
Data_Impens	94.35	0.9715	95.05	0.9720	95.77	0.9832
Data_Schilling	92.54	0.9498	92.60	0.9593	94.64	0.9791
746_1625	93.26	0.9618	94.05	0.9752	96.29	0.9824
Impens_746	92.45	0.9674	94.11	0.9638	94.39	0.9648
Schilling_746	92.03	0.9593	93.12	0.9573	93.59	0.9524
Impens_1625	94.35	0.9554	93.63	0.9459	94.61	0.9673
Schilling_1625	93.35	0.9634	93.09	0.9605	94.71	0.9739
Impens_746_1625	93.57	0.9435	93.58	0.9441	95.48	0.9629
Schilling_746_1625	94.82	0.9617	94.17	0.9690	94.57	0.9674

The effectiveness of the feature descriptor in the peptide identification is shown in Table 8. The performance measure AUC for seven protein feature descriptor is compared with the proposed model. Since our model integrated all these features therefore, a single value of AUC is mentioned in the last column of the table. From the table we highlight the following points:

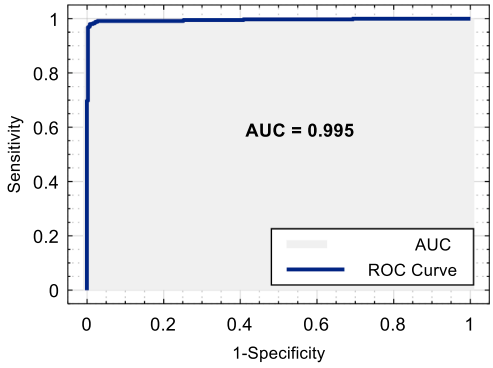
- Proposed model predict the cleavable sites with higher performance ratio when compared with all feature descriptor.
- The improvement of AUC over the sequence based feature descriptor (0.9884–0.786) is 20%, for physicochemical based feature (0.9884–0.7267) is 26%, for secondary structure based (0.9884–0.7312) is 25%. All the listed values of the corresponding features are averaged over the benchmark dataset.

These results correspond well with the justification of our first and second hypothesis question. Therefore, this suggests that the proposed method in predicting HIV-1 protease cleavage sites classification is robust with respect to shift in the dataset distribution, and this also indicates that multi-kernel model along with optimal feature pairing can improve the overall prediction rate.

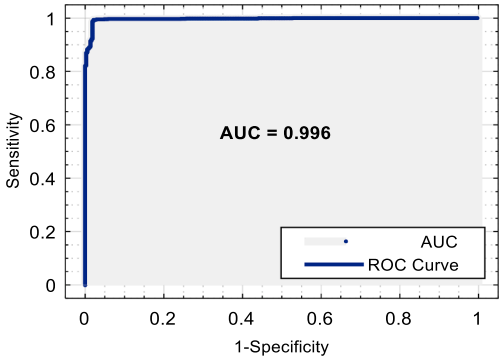
4.4. Multi-Kernel learning performance comparison

The third question is answered by the following experiment where we consider seven multi-kernel learning algorithm from the MKL toolbox [18]. This includes Alignment-Based (ABMKL), Centred-Alignment based (CABMKL), Group lasso (GLMKL), Generalized multiple kernel (GMKL), Localized multiple kernel (LMKL), Non-linear multiple kernel (NMKSVM) and Rule-Based multiple kernels (RBMKL).

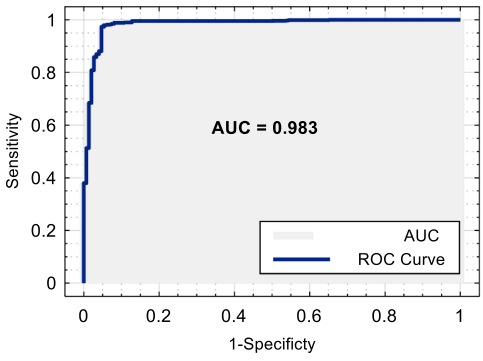
Table 9 compares the performance measure of seven multi-kernel method along with the proposed model for attaining the multitask



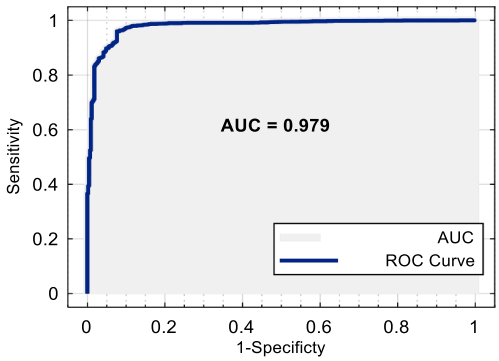
(a) Characteristics curve (ROC) on Data_746



(b) Characteristics curve (ROC) on Data_1625



(c)Characteristics curve (ROC) on Data_Impens



(d) Characteristics curve (ROC) on Data_Schilling

Fig. 6. Receiver operating characteristics Curve (ROC) and AUC of the proposed model.

Table 7
Performance Comparison with the state-of-the-art Multitask Learning Techniques.

Test Data	Measures	RMTL	Dirty MTL	JFL	Least Lasso	Multi Lasso	Low Rank	MTLF	Proposed
Data_746	Specificity	59.93	74.87	65.07	62.64	66.81	50.87	59.93	**93.32**
	Sensitivity	75.22	80.3	75.71	75.78	76.54	73.27	75.22	**97.09**
	F-Measure	50.76	72.2	59.17	55.23	61.64	33.48	50.76	**95.04**
	Accuracy	56.84	73.06	62.47	59.79	64.34	47.05	56.84	**96.54**
	MCC	31.65	54.91	39.37	62.64	42.24	41.25	31.65	**90.33**
Data_1625	Specificity	60.36	74.84	65.02	64.89	66.61	50.59	83.43	**94.85**
	Sensitivity	84.43	87.1	82.76	81.84	85.63	74.28	82.25	**91.74**
	F-Measure	61.87	78.61	67.95	63.23	70.03	44.82	82.81	**93.16**
	Accuracy	81.23	87.14	82.77	85.02	83.82	77.11	87.57	**96.65**
	MCC	37.77	60.72	44.34	54.89	48.66	37.55	65.67	**86.54**
Data_Impens	Specificity	57.74	66.7	68.65	70.63	66.55	52.01	82.05	**91.97**
	Sensitivity	84.57	85.13	85.65	78.21	86.98	92.4	85.33	**93.47**
	F-Measure	59.7	71.17	73.26	73.45	71.23	49.76	83.50	**92.34**
	Accuracy	86.27	88.49	89.02	87.75	88.71	84.9	92.76	**95.77**
	MCC	32.72	48.44	51.57	70.63	49.48	18.48	67.25	**85.42**
Data_Schilling	Specificity	50.46	67.21	73.67	68.67	61.67	51.48	82.74	**94.89**
	Sensitivity	93.42	81.30	80.03	75.41	83.45	89.97	84.29	**91.84**
	F-Measure	47.39	71.37	76.29	71.25	65.56	49.44	83.49	**93.23**
	Accuracy	86.86	89.67	90.22	88.48	89.03	87.11	92.57	**94.64**
	MCC	48.95	46.42	53.33	68.67	39.51	45.38	67.02	**86.68**

Table 8
AUC of Feature Descriptors in Multitask Learning Techniques.

Test Data	Fea. Descriptor	RMTL	Dirty MTL	JFL	Least Lasso	Multi Lasso	Low Rank	MTLF	Proposed
Data_746	2G	0.8123	0.8714	0.8277	0.8277	0.8277	0.8382	0.7711	**0.9951**
	PseAA	0.4611	0.4759	0.5563	0.5550	0.5563	0.4638	0.4611	
	QRC	0.5295	0.6019	0.5188	0.5188	0.5188	0.6810	0.5429	
	SMR	0.7692	0.7692	0.8345	0.8338	0.8345	0.7692	0.7692	
	AAC	0.8134	0.8034	0.7618	0.8524	0.6813	0.7168	0.8517	
	SSE	0.8026	0.7953	0.7707	0.8431	0.6773	0.7271	0.8451	
Data_1625	2G	0.8443	0.8710	0.8276	0.8276	0.8276	0.8563	0.7428	**0.9963**
	PseAA	0.7692	0.7815	0.7742	0.7748	0.7742	0.7705	0.7692	
	QRC	0.7920	0.8154	0.7871	0.7871	0.7871	0.8375	0.7988	
	SMR	0.4611	0.4611	0.6421	0.6421	0.6421	0.4611	0.4611	
	AAC	0.8165	0.6554	0.6778	0.6379	0.4733	0.6376	0.4573	
	SSE	0.8097	0.8138	0.7890	0.8456	0.4420	0.7561	0.8309	
Data_Impens	2G	0.8627	0.8849	0.8902	0.8902	0.8902	0.8870	0.8490	**0.9832**
	PseAA	0.8427	0.8427	0.8448	0.8448	0.8448	0.8427	0.8427	
	QRC	0.8627	0.8722	0.8680	0.8680	0.8680	0.8881	0.8669	
	SMR	0.8427	0.8427	0.8807	0.8807	0.8807	0.8427	0.8427	
	AAC	0.8205	0.7452	0.7353	0.6617	0.5183	0.6511	0.5000	
	SSE	0.8533	0.8263	0.7633	0.8967	0.8108	0.7999	0.4337	
Data_Schilling	2G	0.8686	0.8967	0.9022	0.9022	0.9019	0.8903	0.8710	**0.9791**
	PseAA	0.8674	0.8600	0.6840	0.6855	0.6815	0.8674	0.8674	
	QRC	0.8808	0.8744	0.8735	0.8735	0.8735	0.8891	0.8756	
	SMR	0.8674	0.8686	0.8875	0.8884	0.8881	0.8674	0.8674	
	AAC	0.8350	0.7759	0.7474	0.7131	0.5003	0.6897	0.4645	
	SSE	0.8090	0.6306	0.6685	0.6109	0.5088	0.6134	0.5000	

learning in cleavage site prediction problem. First column of the table represents the name of the test data, we have considered the other three remaining benchmark dataset for the training. From the table we can infer the following key points:

- With the reported results, the proposed techniques outperform the baseline multi-kernel model for multitasking in protease cleavage site prediction. The mean accuracy of the proposed model is 95.90% in comparison to 89.28 of RBMKL (best average accuracy amongst the multi-kernel model).
- Highest performance is observed in *Data_1625* with 96.65% accuracy in comparison to 95.62 of NMKSVM (best amongst the multi-kernel learning for *Data_1625*).
- The proposed model consistency can be inferred from the reported performance measure. It is observed that the proposed algorithm has superior performance than the other multi-kernel model in every performance measure.

To answer the fourth question, the results of the various feature descriptors in conjunction with the multi-kernel techniques are summarized in Table 10. Test data are mentioned in the first column, and rows represent the multi-kernel techniques. As the proposed model is a composite model which carries multiple feature descriptors and various kernels, thus it is impractical for testing the individual feature descriptor performance. However, for the sake of comparison, we have included the performance of the proposed model in the last column. On observation, we can infer the following points:

- With the feature integration from many descriptors, the performance improvement achieved by the proposed model is superior to the other state-of-art multi-kernel and encoding techniques.
- AUC values obtained by the proposed composite model on benchmark data has the mean AUC of 0.9884. The improvement over sequence-based feature descriptor (0.9884–0.723) is 26%, for physicochemical based feature (0.9884–0.833) is 15%, for secondary

Table 9
Performance Comparison with the state-of-the-art Multi-kernel Learning Techniques.

Test Data	Measures	ABMKL	CABMKL	GLMKL	GMKL	LMKL	NMKSVM	RBMKL	Proposed
Data_746	Specificity	93.13	51.45	71.29	62.64	50.01	59.16	91.97	**93.32**
	Sensitivity	93.16	73.31	78.68	75.78	23.06	74.32	93.47	**97.09**
	F-Measure	93.15	33.74	67.64	55.23	31.56	49.63	92.34	**95.04**
	Accuracy	93.16	47.18	69.17	59.79	46.11	56.03	92.49	**96.54**
	MCC	82.41	81.69	86.68	86.54	84.12	90.33	82.55	**90.33**
Data_1625	Specificity	94.53	94.44	94.39	93.85	50.03	93.32	94.72	**94.85**
	Sensitivity	88.12	87.54	91.84	90.44	38.46	90.09	88.1	**91.74**
	F-Measure	90.61	90.14	93.23	92.61	43.48	93.04	90.64	**93.16**
	Accuracy	92.74	92.31	95.02	94.95	76.92	95.62	92.74	**96.65**
	MCC	80.26	79.53	77.07	84.31	67.73	72.71	84.51	**86.54**
Data_Impens	Specificity	73.76	72.34	71.3	70.63	50.03	51.34	77.01	**91.97**
	Sensitivity	72.33	68.47	78.64	78.21	42.13	92.31	73.37	**93.47**
	F-Measure	73.03	70.14	74.08	73.45	45.73	48.45	74.93	**92.34**
	Accuracy	85.22	82.37	87.96	87.75	84.27	84.69	85.64	**95.77**
	MCC	82.16	78.74	76.87	82.32	65.37	72.19	83.51	**85.42**
Data_Schilling	Specificity	73.51	75.21	68.71	68.67	50.03	50.64	78.21	**94.89**
	Sensitivity	69.54	68.54	75.1	75.41	43.37	76.77	71.68	**91.84**
	F-Measure	71.18	70.87	71.19	71.25	46.45	47.82	74.16	**93.23**
	Accuracy	85.36	84.08	88.39	88.48	86.74	86.83	86.25	**94.64**
	MCC	80.16	83.45	85.68	85.32	86.03	81.54	82.66	**86.68**

structure based (0.9884–0.722) is 26%. All the AUC values of the corresponding features are averaged over the benchmark dataset

The convergence rate analysis of factorial evolution in generating multitask learning model is performed and outlined in Figs. 7 and 8. This convergence chart is the outcome of change in the fitness with respect to iterations. From the analysis it can be said that factorial evolutionary converges faster and hence holds superior convergence capability for multitask learning models. Individually, the convergence chart of Data_746 is shown in Fig. 7(a), Data_1625 is demonstrated in Fig. 7(b), Data_Impens is shown in Fig. 8(a) and Data_Schilling is depicted in Fig. 8(b). It is observed from the extensive experimental studies that performance of multitasking evolutionary algorithm is accurate as well as consistent for peptide classification.

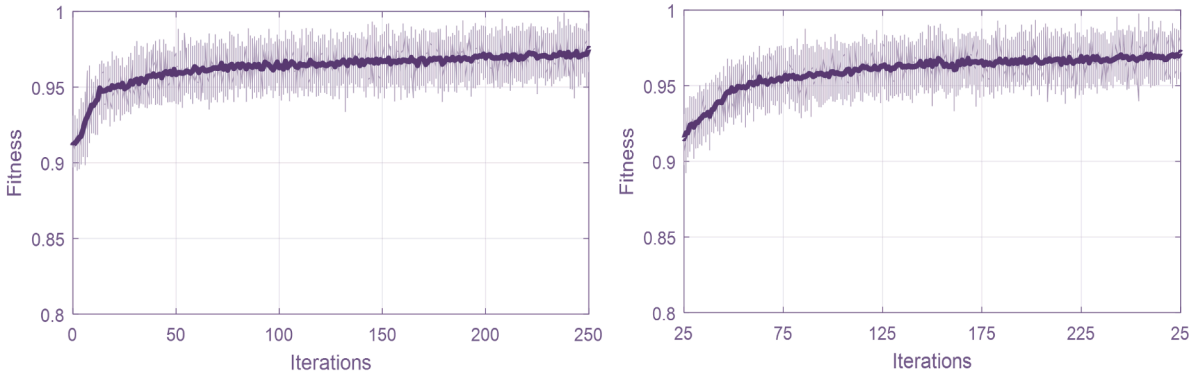
Further, the Figs. 7 and 8 affirm the quicker convergence rate in Data_746 and Data_1625 in comparison to the Impens and Schilling dataset. It is noticed from the charts that the MFO explored the search space in the beginning, but after 15 iterations most of the individuals converge towards the optimal solution. Hence the MFO offers a consistent performance.

Table 11 outline the AUC performance comparison for final evaluation of multitask learning model. The performance values listed in the Table 11 is referred from [35]. On comparison with the state of the art HIV protease cleavage site classification methods, it is found that there is an increase in the predictability ratio from the current best is achieved by the evolutionary based MKL model.

The proposed multitask learning model is composed of multiple diverse kernels with best suited features. The runtime efficiency of the proposed model can be divided into generation and application time. In the model generation time, the time for the evolution of best kernel with parameters and optimized feature for the formation of weighted multi-kernel learner. Model generation takes substantial amount of time whereas application time refers the time to label the single peptide instance. Thus, in Table 12 we report the run time comparison of the proposed model with the other state of the art methods. The reported timings are the average of 10 runs of model generation and application

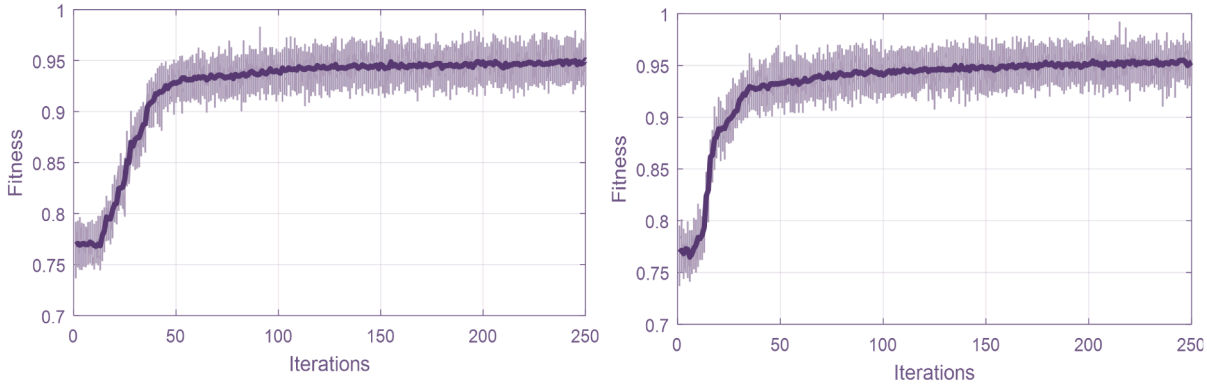
Table 10
AUC Effect of Feature Descriptor in Multi-kernel Learning Techniques.

Test Data	Fea. Descriptor	ABMKL	CABMKL	GLMKL	GMKL	LMKL	NMKSVM	RBMKL	Proposed
Data_746	2G	0.8976	0.5146	0.7126	0.6289	0.5000	0.5841	0.9398	**0.9951**
	PseAA	0.9344	0.6373	0.9209	0.9035	0.3846	0.9002	0.8712	
	RC	0.9137	0.4639	0.7562	0.6523	0.4348	0.5866	0.8964	
	SMR	0.9416	0.7760	0.8665	0.8283	0.7692	0.8080	0.9188	
	AAC	0.8205	0.7452	0.7353	0.6617	0.5183	0.6511	0.8205	
	SSE	0.7868	0.6109	0.5681	0.4573	0.4573	0.7846	0.7488	
Data_1625	2G	0.9185	0.9196	0.9318	0.9287	0.5000	0.8384	0.9416	**09,963**
	PseAA	0.9267	0.9227	0.9305	0.9280	0.2306	0.8544	0.9464	
	RC	0.9195	0.9203	0.9301	0.9274	0.3156	0.8269	0.9429	
	SMR	0.9208	0.9209	0.9303	0.9276	0.4611	0.8285	0.9436	
	AAC	0.9000	0.8932	0.8810	0.9163	0.8329	0.8557	0.9000	
	SSE	0.8807	0.6916	0.7262	0.4645	0.4645	0.7643	0.7466	
Data_Impens	2G	0.8789	0.6438	0.6770	0.5000	0.5000	0.8923	0.6927	**0.9833**
	PseAA	0.8857	0.9079	0.9109	0.4337	0.4337	0.7283	0.9131	
	RC	0.8807	0.6916	0.7262	0.4645	0.4645	0.7643	0.7466	
	SMR	0.9456	0.9019	0.9098	0.8674	0.8674	0.8472	0.9138	
	AAC	0.8457	0.8513	0.8565	0.8565	0.8565	0.8698	0.9240	
	SSE	0.8551	0.9333	0.8716	0.4213	0.4213	0.7549	0.9154	
Data_Schilling	2G	0.8090	0.6306	0.6685	0.6109	0.5088	0.6134	0.5000	**0.9791**
	PseAA	0.8389	0.8046	0.7016	0.8303	0.4601	0.7798	0.4213	
	RC	0.8165	0.6554	0.6778	0.6379	0.4733	0.6376	0.4573	
	SMR	0.9082	0.8691	0.8448	0.8691	0.8437	0.8638	0.8427	
	AAC	0.7522	0.8030	0.7571	0.7571	0.7571	0.7654	0.7327	
	SSE	0.8665	0.8283	0.7692	0.8080	0.9188	0.8327	0.8665	



(a) Fitness curve on Data_746 (b) Fitness curve on Data_1625

Fig. 7. Fitness Change with respect to iterations on Data_746 and Data_1625.



(a) Fitness Curve on Data_Impens (b) Fitness Curve on Data_Schilling

Fig. 8. Fitness Change with respect to iteration on Data_impens and Data_Schilling.

time. The model generation time and application time are listed in column two and three respectively.

4.5. Statistical test

Statistical test is a fundamental step to examine the performances of various algorithms. Thus, non-parametric statistical test were performed for the comparison between different multitask and multi-kernel learning models. We used multiple tests for the comparison to have the sufficient reliability. Friedman [79], Bonferroni-Dunn [80], and post hoc [81] are the wider adaptable test to find the differences among the performance of the algorithms. All these tests have used the reported mean accuracy for examining the proposed and state of the art multitask learning techniques.

Friedman's non-parametric test is used to detect the distribution differences among the set of methods (more than two). It ranks the algorithms by averaging the rank of over all the datasets. The best algorithms is indicated by lowest average rank. Based on the individual rank, Friedman test detect the significant differences. Table 13 reports the statistical test results. Post computation of Friedman statistic ($=30.743$) with the $p-value$ 0.00602 at a confidence level of 0.05 rejects the null-hypothesis.

Freidman test confirms the difference in the distribution of the classifier results. Therefore, further pairwise test is performed to seek the best among the fifteen algorithms. Bonferroni-Dunn and Post hoc and test is applied to find the significant differences between the algorithms. Both the tests perform a series of pairwise comparison to evaluate the two algorithms. On the basis of results, both the test rejects

Table 12
Runtime analysis and comparison.

Methods	Generation Time	Application Time
Nanni et al. [34]	124.23 s	0.023 s
Gok et al. [33]	178.24 s	0.013 s
Rognavaldsson et al. [14]	111.61 s	0.007 s
Singh et al. [43]	1.08 E+04 s	1.877 s
Proposed Model	1.32 E+04 s	2.009 s

Table 11
Performance comparison with the State of the art method.

Dataset	Nanni et al. [34]	Gok et al. [33]	Shen et al. [77]	Rognavaldsson et al. [14]	Fathi and Sadeghi [35]	Singh et al. [43]	Shayanfar et al. [78]	Proposed Model
746	0.892	0.894	0.929	0.885	0.910	0.98	0.98	**0.995**
1625	0.886	0.870	0.891	0.856	0.895	0.99	0.988	**0.996**
Schilling	0.926	0.919	0.753	0.925	0.927	0.94	0.97	**0.983**
Impens	0.910	0.898	0.687	0.908	0.912	0.98	0.934	**0.979**
Avg.	0.904	0.895	0.815	0.894	0.911	0.972	0.968	**0.982**

Table 13
Non-parametric Statistical Test Performance.

Learning Techniques	Ranking	*Z*	*p*	*Holm*	P_{Bonf}	P_{Holm}
RMTL	10.625	2.964635	0.00303	0.004545	0.042426	0.033335
Dirty MTL	5.5	1.343968	0.178959	0.025	2.505422	0.535097
JFL	6.25	1.581139	0.113846	0.0125	1.593848	0.535097
Least Lasso	8.125	2.174066	0.0297	0.005556	0.415803	0.267302
Multi Lasso	6.5	1.660196	0.096875	0.01	1.356251	0.535097
Low Rank	12.25	3.478505	0.000504	0.004167	0.007059	0.006051
MTLF	5.625	1.383496	0.166513	0.016667	2.331178	0.535097
ABMKL	8.125	2.174066	0.0297	0.00625	0.415803	0.267302
CABMKL	12.5	3.557562	0.000374	0.003846	0.00524	0.004866
GLMKL	5.5	1.343968	0.178959	0.05	2.505422	0.535097
GMKL	6.625	1.699724	0.089183	0.008333	1.248559	0.535097
LMKL	14	4.031904	0.000055	0.003571	0.000775	0.000775
NMKSVM	9.25	2.529822	0.011412	0.005	0.159769	0.11412
RBMKL	7.875	2.095009	0.03617	0.007143	0.506383	0.267302
Proposed	**1.25**	–	–	–	–	–

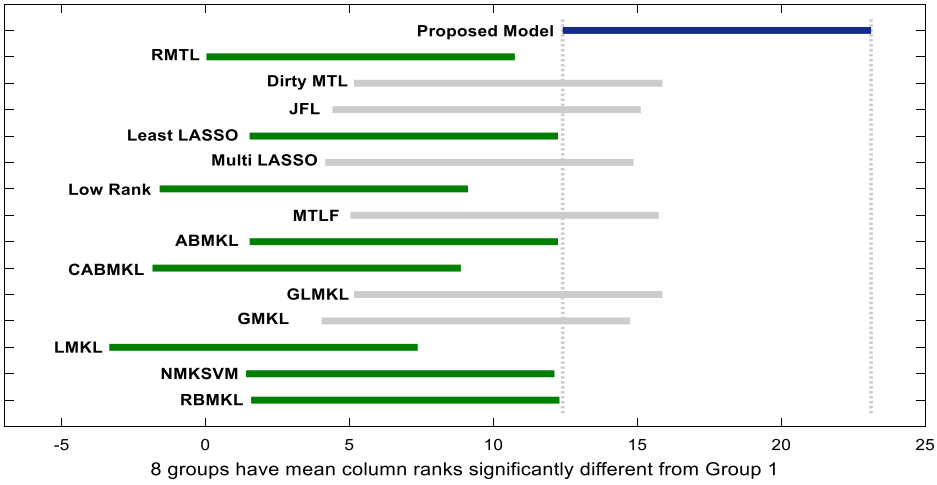


Fig. 9. Pairwise comparison test with interactive graph.

the null hypothesis. Bonferroni-Dunn's test rejects hypotheses that have an unadjusted p-value ≤0.003571 while Holm's test rejects those hypotheses that have an unadjusted p-value ≤0.005. The level of significance used in the test is 5%. Fig. 9 depicts the pairwise comparison with the help of interactive graph where blue horizontal line represents the proposed model and eight green horizontal line represents the significant different models. With these observation, we can state that the proposed model with the mean rank values of 1.25 is best in the peptide cleavability site prediction.

5. Conclusion

In this paper, a novel multitask learning model for HIV-1 protease cleavage sites prediction is proposed that can extract the knowledge from multiple auxiliary data sources. The use of multitask learning model improves the generalization performance especially where data scarcity problem persist. The proposed model is the composition of multi-kernel learner, factorial evolution and feature selection to obtain multitask learning framework. In contrast to the conventional multitask learning models, the optimal feature selection with efficient multi-kernel utilization are the additional advantages. The best kernel-feature pairs are generated by multi factorial evolutionary algorithm with the objective of seeking higher area under the curve (AUC). For the experimental observation of the propose model, four benchmark datasets were considered. Additionally, to overcome the problem of high bias we also combined these data to produce different sets of training instances. We compared our method with fifteen state-of-the-art techniques that include multitask learning and multi-kernel methods. The promising results obtained after extensive experimentation confirms the belief that multitask learning in the cleavage site identification could be used to alleviate the data scarcity problem. Furthermore, the improvement achieved by the proposed method is validated by the non-parametric test. The outcome of the statistical test proves the proposed model as most successful approach of predicting the cleavage sites. The optimal configuration of diverse kernel-feature pair by factorial evolutionary algorithm are the prime reasons for the improved performance. An area of future research investigation is to extend the proposed model as a generalized framework to solve the other domain problems.

CRediT authorship contribution statement

Deepak Singh: Conceptualization, Methodology, Writing - original draft. **Dilip Singh Sisodia:** Writing - review & editing, Supervision. **Pradeep Singh:** Writing - review & editing, Supervision.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] U.R. Acharya, P. Dua, Machine learning in healthcare informatics, Springer Berlin Heidelberg, 2014 http://doi.org/0.1007/978-3-642-40017-9.

[2] UNAIDS, UNAIDS Fact Sheet November, (2016). http://www.unaids.org/sites/default/files/media_asset/UNAIDS_FactSheet_en.pdf.
[3] World Health Organization, (2016). http://www.who.int/gho/hiv/en/.
[4] R.C. Gallo, L. Montagnier, The discovery of HIV as the cause of AIDS, N. Engl. J. Med. 24 (2003) 2283–2285.
[5] Y.-S. Chen, A comprehensive identification-evidence based alternative for HIV/AIDS treatment with HAART in the healthcare industries, Comput. Methods Programs Biomed. 131 (2016) 111–126, https://doi.org/10.1016/J.CMPB.2016.04.001.
[6] K.R. Bisaso, G.T. Anguzu, S.A. Karungi, A. Kiragga, B. Castelnuovo, A survey of machine learning applications in HIV clinical research and care, Comput. Biol. Med. 91 (2017) 366–371, https://doi.org/10.1016/J.COMPBIOMED.2017.11.001.
[7] H. Li, R.W. Omange, F.A. Plummer, M. Luo, A novel HIV vaccine targeting the protease cleavage sites, AIDS Res. Ther. 14 (2017) 10–14, https://doi.org/10.1186/s12981-017-0174-7.
[8] A. Lumini, L. Nanni, Machine learning for HIV-1 protease cleavage site prediction, Pattern Recognit. Lett. 27 (2006) 1537–1544, https://doi.org/10.1016/j.patrec.2006.01.014.
[9] D. Singh, P. Singh, D.S. Sisodia, Evolutionary based optimal ensemble classifiers for HIV-1 protease cleavage sites prediction, Expert Syst. Appl. 109 (2018) 86–99, https://doi.org/10.1016/j.eswa.2018.05.003.
[10] L. Nanni, Comparison among feature extraction methods for HIV-1 protease cleavage site prediction, Pattern Recognit. 39 (2006) 711–713, https://doi.org/10.1016/j.patcog.2005.11.002.
[11] D.T. Barkan, D.R. Hostetter, S. Mahrus, U. Pieper, J.A. Wells, C.S. Craik, A. Sali, Prediction of protease substrates using sequence and structure features, Bioinformatics. 26 (2010) 1714–1722, https://doi.org/10.1093/bioinformatics/btq267.
[12] W. Huang, C. Tung, H. Huang, S. Hwang, S. Ho, ProLoc: prediction of protein subnuclear localization using SVM with automatic selection from physicochemical composition features, BioSystems 90 (2007) 57–581.
[13] L. Nanni, A. Lumini, A genetic approach for building different alphabets for peptide and protein classification, BMC Bioinformatics 9 (2008) 45, https://doi.org/10.1186/1471-2105-9-45.
[14] T. Rognvaldsson, L. You, D. Garwicz, State of the art prediction of HIV-1 protease cleavage sites, Bioinformatics 31 (2015) 1204–1210, https://doi.org/10.1093/bioinformatics/btu810.
[15] M. Iqbal, B. Xue, H. Al-Sahaf, M. Zhang, Cross-domain reuse of extracted knowledge in genetic programming for image classification, IEEE Trans. Evol. Comput. 21 (2017) 4, https://doi.org/10.1109/TEVC.2017.2657556.
[16] R. Caruana, Multitask learning, Mach. Learn. 28 (75) (1997) 41–75, https://doi.org/10.1023/A:1007379606734.
[17] R. Chandra, Y.S. Ong, C.K. Goh, Co-evolutionary multi-task learning for dynamic time series prediction, Appl. Soft Comput. J. 70 (2018) 576–589, https://doi.org/10.1016/j.asoc.2018.05.041.
[18] M. Gönen, E. Alpaydın, Multiple Kernel Learning Algorithms, J. Mach. Learn. Res. 12 (2011) 2211–2268.
[19] X. Zhang, W. Chao, Z. Li, C. Liu, R. Li, Multi-modal kernel ridge regression for social image classification, Appl. Soft Comput. J. 67 (2018) 117–125, https://doi.org/10.1016/j.asoc.2018.02.030.
[20] L. Zhou, L. Feng, J. Zhong, Y.S. Ong, Z. Zhu, E. Sha, Evolutionary multitasking in combinatorial search spaces: A case study in capacitated vehicle routing problem, 2016 IEEE Symp, Ser. Comput. Intell. SSCI 2016 (2017), https://doi.org/10.1109/SSCI.2016.7850039.
[21] O. Singh, E.C. Su, Prediction of HIV-1 protease cleavage site using a combination of sequence, and physicochemical features, BMC Bioinformatics 17 (2016), https://doi.org/10.1186/s12859-016-1337-6.
[22] B. Xue, M. Zhang, S. Member, W.N. Browne, A survey on evolutionary computation approaches to feature selection, IEEE Trans. Evolut. Comput. 2007 (2015) 1–20, https://doi.org/10.1109/TEVC.2015.2504420.
[23] E. Hancer, B. Xue, M. Zhang, Differential evolution for filter feature selection based on information theory and feature ranking, Knowledge-Based Syst. 140 (2018) 103–119, https://doi.org/10.1016/j.knosys.2017.10.028.
[24] C. Lazar, J. Taminau, S. Meganck, D. Steenhoff, A. Coletta, C. Molter, V. de Schaetzen, R. Duque, H. Bersini, A. Nowe, A survey on filter techniques for feature selection in gene expression microarray analysis, IEEE/ACM Trans. Comput. Biol. Bioinformatics. 9 (2012) 1106–1119, https://doi.org/10.1109/TCBB.2012.33.
[25] C.A. Murthy, Bridging feature selection and extraction: compound feature generation, IEEE Trans. Knowl. Data Eng. 29 (2017) 757–770, https://doi.org/10.1109/TKDE.2016.2619712.
[26] N. Pagano, P. Teriete, M.E. Mattmann, L. Yang, B.A. Snyder, Z. Cai, M.L. Heil, N.D.P. Cosford, An integrated chemical biology approach reveals the mechanism of action of HIV replication inhibitors, Bioorganic Med. Chem. (2017), https://doi.org/10.1016/j.bmc.2017.03.061.
[27] Shuichi Kawashima, M. Kanehisa, AAindex: amino acid index database, Nucl. Acids Res. 28 (2000) 374–374.
[28] A. Kidera, Y. Konishi, M. Oka, T. Ooi, H.A. Scheraga, Statistical analysis of the physical properties of the 20 naturally occurring amino acids, J. Protein Chem. 4 (1985) 23–55.
[29] L. Nanni, A. Lumini, MppS: An ensemble of support vector machine based on multiple physicochemical properties of amino acids, Neurocomputing 69 (2006) 1688–1690, https://doi.org/10.1016/j.neucom.2006.04.001.
[30] P. Mundra, M. Kumar, K.K. Kumar, Valadi K. Jayaraman, B.D. Kulkarni, Using pseudo amino acid composition to predict protein subnuclear localization: Approached with PSSM, Pattern Recognit. Lett. 28 (2007) 1610–1615.
[31] K.C. Chou, Some remarks on protein attribute prediction and pseudo amino acid composition, J. Theor. Biol. 273 (2011) 236–247, https://doi.org/10.1016/j.jtbi.2010.12.024.
[32] K.-C. Chou, H.-B. Shen, Recent progress in protein subcellular location prediction, Anal. Biochem. 370 (2007) 1–16, https://doi.org/10.1016/J.AB.2007.07.006.
[33] M. Gök, A.T. Özcerit, A new feature encoding scheme for HIV-1 protease cleavage site prediction, Neural Comput. Appl. 22 (2013) 1757–1761, https://doi.org/10.1007/s00521-012-0967-5.
[34] L. Nanni, A. Lumini, Using ensemble of classifiers for predicting HIV protease cleavage sites in proteins, Amino Acids 36 (2009) 409–416, https://doi.org/10.1007/s00726-008-0076-z.
[35] A. Fathi, R. Sadeghi, A genetic programming method for feature mapping to improve prediction of HIV-1 protease cleavage site, Appl. Soft Comput. J. 72 (2018) 56–64, https://doi.org/10.1016/j.asoc.2018.06.045.
[36] T. Rögnvaldsson, L. You, Why neural networks should not be used for HIV-1 protease cleavage site prediction, Bioinformatics 20 (2004) 1702–1709, https://doi.org/10.1093/bioinformatics/bth144.
[37] H. Liu, X. Shi, D. Guo, Z. Zhao, Feature selection combined with neural network structure optimization for HIV-1 protease cleavage site prediction, Biomed. Res. Int. (2015) 11.
[38] S. Jaeger, S.S.-S. Chen, Information fusion for biological prediction, J. Data Sci. 8 (2010) 269–288.
[39] Aleksejs Kontijevskis, Jarl ES Wikberg, J. Komorowski, Computational proteomics analysis of HIV-1 protease interactome, Proteins Struct Funct. Bioinforma. 68 (2007) 305–312.
[40] L. You, D. Garwicz, T. Rögnvaldsson, Comprehensive bioinformatic analysis of the specificity of human immunodeficiency virus type 1 protease, J. Virol. 79 (2005) 12477–12486.
[41] H. Oğul, Variable context Markov chains for HIV protease cleavage site prediction, BioSystems 96 (2009) 246–250.
[42] N. Qian, T.J. Sejnowski, Predicting the secondary structure of globular proteins using neural network models, J. Mol. Biol. 202 (1988) 865–884, https://doi.org/10.1016/0022-2836(88)90564-5.
[43] D. Singh, P. Singh, D.S. Sisodia, Evolutionary based ensemble framework for realizing transfer learning in HIV-1 Protease cleavage sites prediction, Appl. Intell. (2018), https://doi.org/10.1016/j.eswa.2018.05.003.
[44] Y. Zhang, Q. Yang, Survey Multi-Task Learn. (2017) 1–20.
[45] R. Caruana, Multitask learning, Learn. to Learn. (1998) 95–133, https://doi.org/10.1109/TCBB.2010.22.
[46] X. Liao, L. Carin, Radial Basis Function Network for Multi-task Learning, Nips. (2005).
[47] D.L. Silver, R. Poirier, D. Currie, Inductive transfer with context-sensitive neural networks, Mach. Learn. 73 (2008) 313–336, https://doi.org/10.1007/s10994-008-5088-0.
[48] A. Argyriou, T. Evgeniou, M. Pontil, A.A.T. Evgeniou, M. Pontil, Multi-task feature learning, Adv. Neural Inf. Process. Syst. 2007, pp. 41–48, , https://doi.org/10.1007/s10994-007-5040-8.
[49] A. Argyriou, T. Evgeniou, M. Pontil, Convex multi-task feature learning, Mach. Learn. 73 (2008) 243–272, https://doi.org/10.1007/s10994-007-5040-8.
[50] A. Maurer, M. Pontil, B. Romera-Paredes, Sparse coding for multitask and transfer learning, 28 (2012). http://arxiv.org/abs/1209.0738.
[51] R.K. Ando, Tong Zhang, R. Yahoo, A framework for learning predictive structures from multiple tasks and unlabeled data, J. Mach. Learn. Res. 6 (2005) 1817–1853.
[52] S. Thrun, S. Thrun, J. O'Sullivan, J. O'Sullivan, Discovering structure in multiple learning tasks: The TC Algorithm, 5–5, Proc. Thirteen. Int. Conf. Mach. Learn. Vol. 28 (1996).
[53] T. Evgeniou, M. Pontil, Regularized multi–task learning, Proc. 2004 ACM SIGKDD Int Conf. Knowl. Discov. Data Min. - KDD '04, 2004, p. 109, , https://doi.org/10.1145/1014052.1014067.
[54] J. Chen, J. Zhou, J. Ye, Integrating low-rank and group-sparse structures for robust multi-task learning, Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discov. Data Min. - KDD '11, 2011, p. 42, , https://doi.org/10.1145/2020408.2020423.
[55] P. Jawanpuria, J.S. Nath, A convex feature learning formulation for latent task structure discovery, Vol. 1, 2012, pp. 137–144.
[56] A. Gupta, Y.S. Ong, L. Feng, Multifactorial evolution: toward evolutionary multitasking, IEEE Trans. Evol. Comput. 20 (2016) 343–357, https://doi.org/10.1109/TEVC.2015.2458037.
[57] J. Rice, C.R. Cloninger, T. Reich, Multifactorial inheritance with cultural transmission and assortative mating. I. Description and basic properties of the unitary models, Am. J. Hum. Genet. 30 (1978) 618–643.
[58] C.R. Cloninger, J. Rice, T. Reich, Multifactorial inheritance with cultural transmission and assortative mating. II. a general model of combined polygenic and cultural inheritance, Am. J. Hum. Genet. 31 (1979) 176–198.
[59] R. Sagarna, Y.S. Ong, Concurrently searching branches in software tests generation through multitask evolution, 2016 IEEE Symp Ser. Comput. Intell. SSCI 2016, 2017, https://doi.org/10.1109/SSCI.2016.7850040.
[60] R.T. Liaw, C.K. Ting, Evolutionary many-tasking based on biocoenosis through symbiosis: A framework and benchmark problems, 2017 IEEE Congr. Evol. Comput. CEC 2017 - Proc. 2017, pp. 2266–2273, , https://doi.org/10.1109/CEC.2017.7969579.
[61] R. Chandra, A. Gupta, Y.S. Ong, C.K. Goh, Evolutionary multi-task learning for modular training of feedforward neural networks, Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics). 9948 LNCS (2016) 37–46, https://doi.org/10.1007/978-3-319-46672-9_5.
[62] Y.W. Wen, C.K. Ting, Learning ensemble of decision trees through multifactorial genetic programming, 2016 IEEE Congr Evol. Comput. CEC 2016, 2016, pp. 5293–5300, , https://doi.org/10.1109/CEC.2016.7748363.

[63] Z. Tang, M. Gong, M. Zhang, Evolutionary multi-task learning for modular extremal learning machine, 2017 IEEE Congr. Evol. Comput. CEC 2017 - Proc. 2017, pp. 474–479, , https://doi.org/10.1109/CEC.2017.7969349.
[64] L. Nanni, A. Lumini, D. Gupta, A. Garg, Identifying bacterial virulent proteins by fusing a set of classifiers based on variants of Chou's Pseudo amino acid composition and on evolutionary information, IEEE/ACM Trans. Comput. Biol. Bioinforma. 9 (2012) 467–475, https://doi.org/10.1109/TCBB.2011.117.
[65] X. Yu, X. Zheng, T. Liu, Y. Dou, J. Wang, Predicting subcellular location of apoptosis proteins with pseudo amino acid composition: Approach from amino acid substitution matrix and auto covariance transformation, Amino Acids 42 (2012) 1619–1625, https://doi.org/10.1007/s00726-011-0848-8.
[66] R. Kohavi, A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, Int. Jt. Conf. Artif. Intell. 14 (1995) 1137–1143, https://doi.org/10.1067/mod.2000.109031.
[67] L. Duan, I.W. Tsang, D. Xu, Domain transfer multiple kernel learning, IEEE Trans. Pattern Anal. Mach. Intell. 34 (2012) 465–479.
[68] C.L. Blake, C.J. Merz, UCI Repository of machine learning databases, Univ. Calif. (1998).
[69] Oliver Schilling, C.M. Overall, Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites, Nat. Biotechnol. 26 (2008) 685–694.
[70] T. Rögnvaldsson, T.A. Etchells, L. You, D. Garwicz, I. Jarman, P.J.G. Lisboa, How to find simple and accurate rules for viral protease cleavage specificities, BMC Bioinformatics 10 (2009) 149, https://doi.org/10.1186/1471-2105-10-149.
[71] A. Jalali, P. Ravikumar, S. Sanghavi, C. Ruan, A dirty model for multi-task learning, nips. (2010) 1–9. https://papers.nips.cc/paper/4125-a-dirty-model-for-multi-task-learning.pdf.
[72] A.A.T. Evgeniou, M. Pontil, Multi-task feature learning, Adv. Neural Inf. Process. Syst. 19 (2007) 41.
[73] S. Ji, J. Ye, An accelerated gradient method for trace norm minimization, Proc. 26th Annu Int. Conf. Mach. Learn. - ICML '09, 2009, pp. 1–8, , https://doi.org/10.1145/1553374.1553434.
[74] R.S. Society, Regression Shrinkage and Selection via the Lasso Author (s): Robert Tibshirani Source: Journal of the Royal Statistical Society . Series B (Methodological), Vol. 58, No. 1 Published by: Wiley for the Royal Statistical Society Stable, 58 (2018) 267–288.
[75] X. Chen, X. Shi, X. Xu, Z. Wang, R. Mills, C. Lee, J. Xu, A Two-Graph Guided Multi-task Lasso Approach for eQTL Mapping, Ece.Ubc.Ca. XX (2012) 208–217. http://www.ece.ubc.ca/~xiaohuic/publications/AISTATS-final-2012.pdf.
[76] J. Zhou, J. Chen, J. Ye, User's Manual MALSAR: Multi-tAsk Learning via StructurAl Regularization, Arizona State Univ. (2012). http://www.malsar.org.
[77] H.-B. Shen, K.-C. Chou, HIVcleave: a web-server for predicting human immunodeficiency virus protease cleavage sites in proteins, Anal. Biochem. 375 (2008) 388–390, https://doi.org/10.1016/j.ab.2008.01.012.
[78] N. Shayanfar, V. Derhami, M. Rezaeian, Deep recurrent neural networks in HIV-1 protease cleavage classification, Int. J. Data Min. Bioinform. 19 (2017) 298–311, https://doi.org/10.1504/IJDMB.2017.091364.
[79] T. Hastie, R. Tibshirani, J. Friedman, Elements Statist. Learn. (2009), https://doi.org/10.1007/b94608.
[80] Y. Hochberg, A sharper bonferroni procedure for multiple tests of significance, Biometrika. 75 (1988) 800–802, https://doi.org/10.1093/biomet/75.4.800.
[81] A. Benavoli, G. Corani, F. Mangili, Should we really use post-hoc tests based on mean-ranks? J. Mach. Learn. Res. 17 (2016) 1–10.