Computers in Biology and Medicine 143 (2022) 105297

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

IL13Pred: A method for predicting immunoregulatory cytokine IL-13 inducing peptides

Shipra Jain, Anjali Dhall, Sumeet Patiyal, Gajendra P.S. Raghava *

Department of Computational Biology, Indraprastha Institute of Information Technology, Okhla Phase 3, New Delhi, 110020, India

ARTICLE INFO

Keywords:
Interleukin 13
Immunoregulatory cytokine
COVID-19
SARS-COV2
IL-4 receptors

ABSTRACT

Background: Interleukin 13 (IL-13) is an immunoregulatory cytokine, primarily released by activated T-helper 2 cells. IL-13 induces the pathogenesis of many allergic diseases, such as airway hyperresponsiveness, glycoprotein hypersecretion, and goblet cell hyperplasia. In addition, IL-13 inhibits tumor immunosurveillance, leading to carcinogenesis. Since elevated IL-13 serum levels are severe in COVID-19 patients, predicting IL-13 inducing peptides or regions in a protein is vital to designing safe protein therapeutics particularly immunotherapeutic.
Objective: The present study describes a method to develop, predict, design, and scan IL-13 inducing peptides.
Methods: The dataset experimentally validated 313 IL-13 inducing peptides, and 2908 non-inducing homo-sapiens peptides extracted from the immune epitope database (IEDB). A total of 95 key features using the linear support vector classifier with the L1 penalty (SVC-L1) technique was extracted from the originally generated 9165 features using Pfeature. These key features were ranked based on their prediction ability, and the top 10 features were used to build machine learning prediction models. Various machine learning techniques were deployed to develop models for predicting IL-13 inducing peptides. These models were trained, tested, and evaluated using five-fold cross-validation techniques; the best model was evaluated on an independent dataset.
Results: Our best model based on XGBoost achieves a maximum AUC of 0.83 and 0.80 on the training and independent dataset, respectively. Our analysis indicates that certain SARS-COV2 variants are more prone to induce IL-13 in COVID-19 patients.
Conclusion: The best performing model was incorporated in web-server and standalone package named 'IL-13Pred' for precise prediction of IL-13 inducing peptides. For large dataset analysis standalone package of IL-13Pred is available at (https://webs.iiitd.edu.in/raghava/il13pred/) webserver and over GitHub link: https://github.com/raghavagps/il13pred.

1. Introduction

Interleukin 13 (IL-13) is an immune-regulatory cytokine, primarily secreted by activated T helper-Type (Th) 2 cells, which inhibits inflammatory cytokine production [1,2]. Studies show that IL-13 is produced by diverse cell types, which include eosinophils, mast cells, basophils, smooth muscle cells, natural killer cells, and fibroblasts with varied biological functions [3,4]. The transcription of IL-13 cytokine is mainly regulated by GATA binding protein 3 transcription factors. The IL-13 has approximately 25% sequence homology with IL-4, located on human chromosome 5q31 [4]. In addition, it is shown that IL-4 and IL-13 are functionally correlated. Surprisingly, IL-13 seems to be a more promising target for designing therapeutics than IL-4 [3]. As shown in Fig. 1, IL-13 mediates several vital functions in diverse biological pathways, which include regulation of airway hyperresponsiveness, allergic inflammation, mastocytosis, goblet cell hyperplasia, tissue eosinophilia, IgE Ab production, tumor cell growth, tissue remodeling, intracellular parasitism, and fibrosis.

Some studies exhibited that alterations in IL-13 effector functions can be targeted to treat certain cancers, like B-cell chronic lymphocytic leukemia and Hodgkin's disease [5,6]. Since IL-13 inhibits tumor immunosurveillance in homo sapiens, IL-13 inhibitors can act as effective cancer immunotherapeutic candidates by activating type-1 anti--cancer defense mechanisms [7,8]. Moreover, IL-13 receptors perform a crucial role in the prognosis of cancers such as pancreatic, gastric, and colon cancer [9–12], since they interact with the tumor microenvironment by activating tumor-associated macrophages and myeloid-derived suppressor cells [13,14]. Receptors of IL-13 were found to be overexpressed in glioblastoma multiforme human samples in situ; whereas

* Corresponding author.
E-mail addresses: shipra@iiitd.ac.in (S. Jain), anjalid@iiitd.ac.in (A. Dhall), sumeetp@iiitd.ac.in (S. Patiyal), raghava@iiitd.ac.in (G.P.S. Raghava).

https://doi.org/10.1016/j.compbiomed.2022.105297
Received 11 November 2021; Received in revised form 23 January 2022; Accepted 23 January 2022
Available online 8 February 2022

Abbreviations

IL-13	Interleukin 13
AUC –	Area under the receiver operating characteristic curve
IEDB –	Immune Epitope Database
SVC-L1 –	Linear Support Vector Classifier with L1 penalty
SARS-CoV2 –	Severe Acute Respiratory Syndrome Coronavirus 2
GATA3 –	GATA Binding Protein 3
IgM –	Immunoglobulin M
IgE –	Immunoglobulin E
PAH –	Pulmonary Artery Hypertension
XGB –	eXtreme Gradient Boosting
KNN –	K-nearest neighbor
SVM –	Support vector machine
GNB –	Gaussian naive bayes
DT –	Decision tree
LR –	Linear regression
RF –	Random forest

normal specimens expressed few IL-13 binding receptor sites [15]. Moreover, increased IL-13 levels were observed in patients with pulmonary artery hypertension (PAH) as compared to non-PAH controls [16]. The IL-13 receptor α1 in myocardial homeostasis plays a pivotal role in cardiovascular diseases [17].

The IL-13 has emerged as a central regulator in airway hyperresponsiveness, fibrosis, mucus hypersecretion, and switching of B cell antibody production from IgM to IgE. Some studies highlight that the IL-13 pathway could be a promising target in the treatment of diverse asthma phenotypes [18,19], as demonstrated in the past year that anti-IL-13 drugs played a major role in controlling the 'Th2 high' asthmatic phenotype [20]. Therefore, anti-IL-13 drugs have become popular (Anrukinzumab, Lebrikizunab, and Tralokinumab) for controlling the severity of asthma [21]. However, several crucial consequences of IL-13 on non-hematopoietic cells, such as endothelial cells, smooth muscle cells, fibroblasts, epithelial cells, and sensory neurons [22–25] have been reported. Additionally, the plasma levels of IL-13 have been shown to play a significantly higher impact in COVID-19 patients [26–28]. A recent study has shown that the COVID-19 infected patients with elevated IL-13 serum levels required ventilation support as compared to others. In addition, COVID-19 patients prescribed with IL-13 inhibiting drug (Dupilumab) have shown less dreadful symptoms [26]. In another study, differential expression levels of 14 cytokines and IL-13 levels in healthy control, moderate, and severe COVID-19 patients have been reported, observing that the higher IL-13 expression level is directly proportional to the COVID-19 severity [28].

Thus, IL-13 plays a vital role in several disease progression and severity, revealing that IL-13 can act as a primary target cytokine in disease identification and monitoring, paving the way to develop a computational approach for predicting IL-13 inducing peptides in a given sequence. Therefore, it is vital to develop a prediction model devoted to classifying IL13 inducing peptides. A thorough search of the literature reveals that IL-13Pred is the first to predict the IL-13 inducing and non-inducing peptides from its amino acid sequence.

This study presents an *in-silico* method to classify the IL-13 inducing and non-inducing peptides/epitopes. We used experimentally reported IL-13 inducing and non-inducing peptides of humans from the immune epitope database (IEDB). Using this dataset, we applied various state-of-the-art machine learning classifiers and the model performance to evaluate the independent dataset to deliver a webserver and standalone version of 'IL-13Pred' for the scientific community.

2. Methods

The complete work flow of the study is depicted in Fig 2. The description of each step is provided in the following sections.

2.1. Dataset preparation and pre-processing

We extracted 343 IL-13 inducing experimentally validated human peptides from IEDB (https://www.iedb.org/) [29]. During pre-processing, we removed the peptides with lengths lower than 8 or more than 35 amino acids from the dataset. To avoid redundancy in the dataset, we removed duplicate copies of the peptides (if any). Eventually, we were left with a positive dataset of 313 IL-13 inducing peptides. Unfortunately, experimentally validated IL13 non-inducing peptides were not available in IEDB, thus creation of negative was a major challenge. We examined experiments used to identify IL-13 (or any other cytokine) inducing peptides. It was observed that researchers only report cytokine induced by a peptide, though they screen a large number of cytokines using a panel of cytokines. It means a peptide is non-inducing peptides for other cytokines except for cytokine

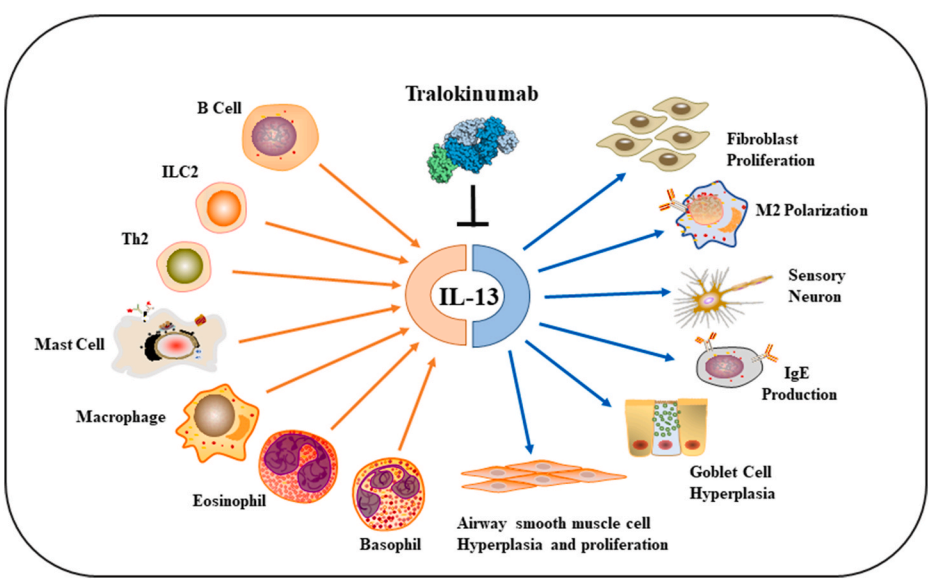

Fig. 1. Schematic representation of IL-13 activators and its functions in biological pathways.

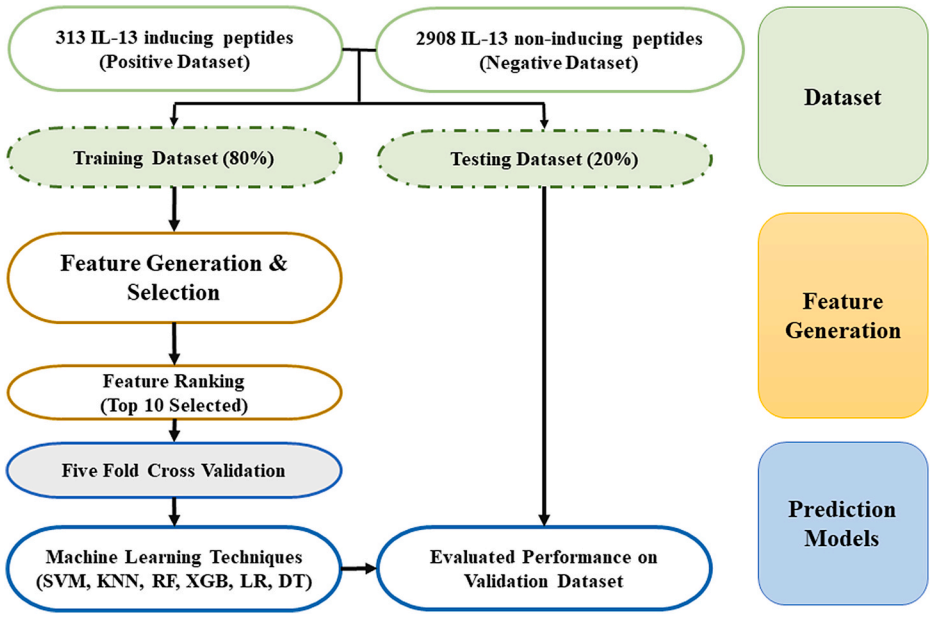

Fig. 2. Work flow depicting the algorithm of IL-13Pred method.

detected/reported in experiment. Thus, it is justified to create a negative dataset using peptides which induce other cytokines except IL-13. In this study, negative dataset comprises the peptides that induce cytokines other than IL-13 (e.g., IL1α, IL1β, TNFα, IL6, IL8, IL12, IL17, IL18). Similar approach of developing negative dataset has been adopted in previous studies for predicting IL-6 and interferon-gamma cytokines [30,31]. In addition, this approach has used for creating negative dataset for predicting specific class of proteins [32,33]. On the complied negative data set, we applied the similar pre-processing steps defined above and obtain 2908 non-inducing peptides. Finally, we proceeded with a positive dataset of 313 IL-13 inducing and a negative dataset of 2908 non-inducing unique peptides.

2.2. Feature generation

We implemented Pfeature to compute various types of descriptors using sequence information of the peptides, which generated a wide range of descriptors in a fixed vector size for a given protein or peptide sequence [34]. In addition, we used a composition-based module of Pfeature for our dataset and generated a vector of 9165 features for each sequence. We computed 15 types of composition-based features like amino acid composition, dipeptide composition, tripeptide composition, atomic composition, etc.

2.3. Ranking and features selection

To extract the key features from a larger pool of features generated from Pfeature, we used a linear support vector classifier with the L1 penalty (SVC-L1) based on the feature selection technique from the Scikit-learn package. The SVC-L1-based method implemented the support vector classifier (SVC) with linear kernel, penalized with L1 regularization [35]. This technique identified the important features from the high-dimensional feature set. SVC-L1 was a faster method when compared to other available techniques for feature selection. Using this method, we listed 95 important features from the pool of 9165 features.

Afterward, a feature-selector tool was implemented to rank the obtained key 95 features, based on their performance in classifying the IL-13 inducing and non-inducing peptides. A light gradient boosting machine based on a decision tree-based algorithm was implemented in the feature-selector tool, which calculated the rank of the feature, based on the number of times a feature was used to split the data across all trees [36]. This tool generated top-ranked features (Supplementary Table S1) to build machine learning prediction models for IL-13 inducing peptides.

2.4. Cross-validation and external validation dataset

We implemented standard protocols to build test, and evaluate our prediction models [37–42]. Then, we applied a 5-fold cross-validation and external validation technique. The complete dataset was divided into the proportion of 80:20, where 80% (i.e., 250 positive and 2326 negative) of the dataset were categorized as training, and the prevailing 20% (i.e., 63 positive and 582 negative) were designated as an independent dataset.

2.5. Machine learning models and evaluation

To develop a prediction model for classifying IL-13 inducing peptides, we implemented diverse machine learning techniques, using various classifiers such as eXtreme Gradient Boosting (XGB), K-nearest neighbor (KNN), support vector machine (SVM), Gaussian Naive Bayes (GNB), decision tree (DT), logistic regression (LR), and random forest (RF). We implemented the Scikit-learn package of Python to build these machine learning prediction models [43].

Standard norms of five-fold cross-validation were conducted, in which four-fold of the data was marked as training dataset and the remaining one-fold as testing dataset. We repeated this step five times and reported the average scores of five cycles in the result section. Then, we optimized various parameters for each classifier to obtain the best-performing models. In the result section, we reported the standard measures used to estimate the performance of the developed prediction models. We calculated both threshold-dependent and threshold-independent parameters to evaluate the performance. Sensitivity, specificity, accuracy, and Matthews correlation coefficient were calculated as threshold-dependent parameters using equations (1)- (4). Area Under the Receiver Operating Characteristic (AUC) curve was calculated as the threshold-independent measure.

$$Sensitivity = \frac{TP}{TP + FN} * 100 \quad (1)$$

$$Specificity = \frac{TN}{TN + FP} * 100 \quad (2)$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} * 100 \quad (3)$$

$$MCC = \frac{(TP*TN) - (FP*FN)}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (4)$$

Where, FP is false positive, FN is false negative, TP is true positive and TN is true negative.

2.6. Standalone and web server

We provided a user-friendly web interface named 'IL-13Pred' to predict IL-13 inducing and non-inducing peptides (https://webs.iiitd.edu.in/raghava/il13pred). The user interface of the web server was easy to navigate and has four major modules named: "Predict", "Design", "Protein Scan", and "Blast Scan". The front-end of the web interface was created using HTML5, JAVA, CSS3, and PHP scripts. This platform was compatible with most modern devices, including mobile tablets, desktops, and laptops. The standalone version of IL-13Pred could be downloaded from https://webs.iiitd.edu.in/raghava/il13pred/stand.html and GitHub link: https://github.com/raghavagps/il13pred.

3. Results

3.1. Compositional analysis

In this study, the amino acid composition for the positive and negative datasets is independently computed. Fig. 3 shows the average amino acid composition for IL-13 inducing and non-inducing peptides. The average composition of L, K, M, and N residues is higher in IL-13 inducing peptides. However, the amino acids G, P, Y, and T residues are abundant in non-inducing peptides.

3.2. Positional analysis

The amino acid positional analysis depicts the dominant residues at a particular position in IL-13 inducing and non-inducing peptides i.e., positive and negative datasets. In Fig. 4, the two-sample logo shows the important amino acid residue at a particular position and its relative occurrence in the sequence. The first eight positions represent the eight *N*-terminal residues, and the last eight positions represent the eight C-terminal residues. In IL-13 inducing peptides, we observe that Leucine (L) occurs more frequently at the 2nd, 5th, 7th,9th, 11th, and 16th positions. Asparagine (N) is mostly preferred at the 6th, 8th, 14th, and 16th positions. However, in the negative dataset, we observe that the Threonine (T) occurred at the 1st, 6th, 13th, 16th position with good abundance, and Proline (P) preferred 3rd, 5th, 14th position. In addition to these two amino acids, glycine (G) showed an interesting pattern in duplets at the 7th and 14th positions.

3.3. Machine learning models performance evaluation

To develop prediction models, we used a positive dataset of 313 IL-13 inducing and a negative dataset of 2908 non-inducing unique peptides. Moreover, we applied various machine learning techniques such as SVM, KNN, LR, DT, RF, XGB, and GNB. The performance of all the developed models for the top 95 features is reported in Table 1. The RF-based model outperforms other models with the AUC of 0.87 and 0.83 on the training and testing dataset, respectively.

Furthermore, we ranked the 95 selected features using a feature-selector library of Python. To identify the best set of features without compromising the performance, we developed the models on top 10, 20, 30, …95 features (Supplementary Tables S2 and S3). From observation, the models developed on top-10 features performed excellently without compromising the performance, compared to the models developed on 95 features. As shown in Table 2, XGB based model outperforms other classifiers with an AUC value of 0.83 and 0.80 on the training and testing dataset, respectively.

To check the robustness of the developed models, we reshuffled the dataset 10 times and calculated the performance of models on top-10 features. Table 3 presents the mean and standard deviation of performance obtained after reshuffling the data. We observed that the performance of the models was maintained even after reshuffling, signifying the robustness of the developed models. XGB based model shows a similar trend to that listed in Table 2, with an AUC of 0.82 ± 0.03.

3.4. Case study 1: IL-13 inducing peptides in proteins of SARS-CoV-2

Recent studies have reported the role of elevated IL-13 levels in COVID-19 patients [26–28]. COVID-19 proteins massively induce the release of this immunoregulatory cytokine. To identify the IL-13 inducing peptides in the SARS-CoV-2 proteins, we used the "Protein Scan" module of 'IL-13Pred' (https://webs.iiitd.edu.in/raghava/il13pred/scan.php), using the default parameters. We downloaded the SARS-CoV-2 proteins from five different countries, including India (MZ340539.1), China (MT291829.1), USA (MZ319836.1) Brazil (MZ169911.1), and Japan (LC528233.2) from NCBI (https://www.ncbi.nlm.nih.gov/datasets/coronavirus/). For USA (Geolocation: California), we identified 213 IL-13 inducing peptides out of 1259 peptides of the spike proteins of reported SARS-CoV-2 strain (as reported in

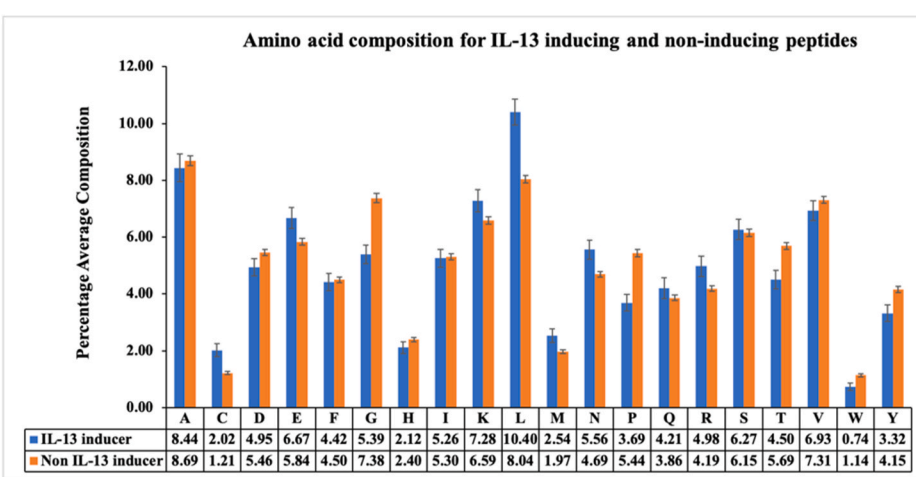

Fig. 3. Compositional analysis of amino acids in IL-13 inducing and non-inducing peptides.

Fig. 4. Two-sample logo displaying the amino acid positional preference among IL-13 positive and negative dataset.

Table 1
The performance of the machine learning models developed using top 95 features.

Classifier	Dataset	Sensitivity	Specificity	Accuracy	AUC	MCC
DT	**Training**	65.96	66.38	66.34	0.70	0.19
	Validation	51.28	66.31	64.50	0.60	0.12
GNB	**Training**	63.83	79.58	78.14	0.78	0.29
	Validation	38.46	77.43	72.71	0.61	0.12
KNN	**Training**	57.02	65.61	64.83	0.64	0.14
	Validation	50.00	69.31	66.98	0.62	0.13
LR	**Training**	73.62	73.99	73.95	0.83	0.30
	Validation	58.97	68.25	67.13	0.68	0.19
RF	**Training**	82.13	79.88	80.09	0.88	0.41
	Validation	70.51	77.07	76.28	0.83	0.34
XGB	**Training**	73.62	76.59	76.32	0.84	0.32
	Validation	69.23	73.19	72.71	0.80	0.30
SVC	**Training**	72.34	71.25	71.35	0.79	0.27
	Validation	51.28	68.08	66.05	0.62	0.13

Table 2
The performance of the machine learning models developed using top 10 features.

Classifier	Dataset	Sensitivity	Specificity	Accuracy	AUC	MCC
DT	**Training**	69.36	69.46	69.45	0.74	0.24
	Validation	60.26	71.43	70.08	0.72	0.22
GNB	**Training**	71.06	68.18	68.44	0.74	0.24
	Validation	64.10	66.31	66.05	0.73	0.21
KNN	**Training**	65.11	56.47	57.26	0.64	0.13
	Validation	60.26	55.73	56.28	0.61	0.11
LR	**Training**	64.26	61.85	62.07	0.67	0.15
	Validation	56.41	59.61	59.23	0.63	0.11
RF	**Training**	74.47	76.34	76.17	0.83	0.33
	Validation	64.10	75.13	73.80	0.77	0.28
XGB	**Training**	78.30	74.84	75.16	0.83	0.33
	Validation	71.80	73.02	72.87	0.80	0.30
SVC	**Training**	54.47	70.70	69.22	0.67	0.16
	Validation	46.15	74.25	70.85	0.64	0.15

Table 3
The performance of the machine learning models developed after reshuffling dataset 10 times using top 10 features.

Classifier	Sensitivity	Specificity	Accuracy	AUC	MCC
DT	63.65 ± 4.76	67.94 ± 2.83	67.52 ± 2.52	0.71 ± 0.03	0.20 ± 0.03
GNB	72.38 ± 4.92	63.81 ± 2.52	64.65 ± 2.30	0.74 ± 0.02	0.22 ± 0.03
KNN	64.44 ± 5.35	51.25 ± 2.00	52.54 ± 1.83	0.62 ± 0.04	0.09 ± 0.03
LR	62.22 ± 4.66	60.76 ± 1.87	60.90 ± 1.62	0.66 ± 0.02	0.14 ± 0.03
RF	74.13 ± 6.17	73.95 ± 1.41	73.97 ± 1.08	0.81 ± 0.03	0.30 ± 0.03
XGB	76.98 ± 4.98	71.13 ± 1.75	71.71 ± 1.51	0.82 ± 0.03	0.31 ± 0.03
SVC	63.81 ± 8.46	63.20 ± 3.35	63.26 ± 2.70	0.68 ± 0.04	0.16 ± 0.04

Supplementary Table S4). In addition, we identified the IL-13 inducing/non-inducing peptides in other coronavirus proteins, such as envelope protein, ORF6, ORF1ab, ORF3a, ORF7a/7b, ORF8, nucleocapsid phosphoprotein, ORF10, and membrane glycoprotein of USA strain,

Table 4
Potential IL-13 inducing peptides predicted by IL-13Pred in spike protein of USA (MZ319836.1) strain.

Epitope	Prediction score
ELDSFKEELDKYFKN	0.39
LLTDEMIAQYTSALL	0.32
KQGNFKNLREFVFKN	0.3
EIDRLNEVAKNLNES	0.3
VNIQKEIDRLNEVAK	0.27
LYRLFRKSNLKPFER	0.25
NIQKEIDRLNEVAKN	0.23
KSTNLVKNKCVNFNF	0.22
VLTESNKKFLPFQQF	0.22
IQKEIDRLNEVAKNL	0.22

which was presented in Supplementary Table S5. Table 4 represents the major peptides of spike proteins which have the capacity to induce the IL13 response. Studies have shown that antibodies that can block IL-13 receptors could be used in effectively designing vaccines [44,45]. These findings can assist the scientific community in designing a subunit vaccine against COVID-19 and other disorders that can be provoked by the induction of IL-13.

3.5. Case study 2: IL-13 inducing peptides in spike proteins of SARS-CoV-2 variant strains

In this study, we made a systematic attempt to highlight the contrast among IL-13 inducing peptides in SARS-CoV-2 variant strains. We identified and reported IL-13 inducing peptides in Alpha (B.1.1.7), Beta (B.1.351), and Delta (B.1.617.2) variants of the SARS-CoV-2 virus. We downloaded the reference Spike protein of SARS-CoV-2 from NCBI. Moreover, we manually conducted the substitution mutations reported by the Centers for Disease Control and Prevention and Indian SARS-CoV-2 Genomics Consortium using an in-house script. The Alpha variant (B.1.1.7) has a total of seven substitution mutations namely N501Y, A570D, D614G, P681H, T716I, S982A, and D1118H. The Beta variant (B.1.351) was obtained by conducting subsequent nine substitution mutations D80A, D215G, K417 N, E484K, N501Y, D614G, A701V, L18F, and R246I in reference spike protein sequence. The Delta (B.1.617.2) variant of the SARS-CoV-2 strain had nine mutations namely, T19R, T95I, G142D, R158G, L452R, T478K, D614G, P681R, and D950 N from reference protein. To identify the IL-13 inducing peptides in these curated variants of COVID-19 Spike proteins, we executed the "Protein Scan" module of 'IL-13Pred' (https://webs.iiitd.edu.in/raghava/il13pred/scan.php), using the default parameters. Initially, we computed IL-13 inducing and non-inducing peptides of all three variant strains. Subsequently, we mapped the differences of three variants across the reference protein sequence in terms of IL-13 inducing vs non-inducing properties. Table 5 represents the sequences with contrasting effects after substitution mutation in spike proteins. As shown in Table 5, a few mutations have turned on the IL-13 inducing property, and some have turned off the same. For instance, mutation at the 618th position of the spike protein in the Delta strain from amino acid Proline to Arginine has turned on the IL-13 inducing property of the mutated peptide. Similarly, a mutation from Aspartic acid to Asparagine at 950th position has changed the nature of peptides from non-inducer to IL-13 inducer.

Fig. 3 revealed similar trends of this amino acid, where the composition of Arginine and Asparagine is preferred in IL-13 inducing peptides as compared to non-inducing peptides. In a given sequence, the composition and position of the amino acid play a vital role in changing the properties and role of the proteins in variant strains, as depicted in Table 5. The complete results of three variant spike protein strains are provided in the Supplementary Tables (S6–S8).

4. Discussion

The IL-13 is involved in mucosal inflammation, which includes allergic asthma, eosinophilic oesophagitis, ulcerative colitis, and fibrosis [4]. In addition, receptors of IL-13 were found to be overexpressed in various cancer cells, serving as a promising target for cancer immunotherapy. In this study, we have shown a correlation between IL-13 levels and COVID-19 severity. A higher level of IL-13 has been reported in severe COVID-19 patients, revealing that COVID-19 patients with higher IL-13 serum levels tend to prolong ventilation support as compared to others. Moreover, IL-13 neutralization results have been shown to reduce COVID-19 severity and lung hyaluronan deposition in a mouse model. In addition, COVID-19 patients taking Dupilumab medicine suffered from a less severe disease [26].

We made a systematic approach to understand the role of IL-13 inducing peptides and provided a method to classify IL-13 inducing peptides. The experimentally validated IL-13 inducing peptides of humans from IEDB are used to develop prediction models. Datasets used were the positive dataset of 313 IL-13 inducing and a negative dataset of 2908 IL-13 non-inducing unique peptides. The amino acid composition and positional analysis were performed to present differences among IL-13 inducing and non-inducing peptides. From the compositional analysis, we observe that the amino acid leucine, lysine, methionine, and asparagine residues are preferred in IL-13 inducing peptides. However, glycine, proline, tyrosine, and threonine residues are abundantly found in non-inducing peptides. While the positional analysis shows that leucine is preferred at the 2nd, 5th, 7th,9th, 11th, and 16th position, asparagine is mostly preferred at 6th, 8th, 14th and,16th position in the positive dataset.

We also used the Pfeature tool to extract relevant features from the dataset. Using the extracted feature set, we applied well-established machine learning with tuned parameters to develop the models with the best performance. Machine learning model performance for the sets

Table 5
Changes in IL-13 inducing properties by mutation in Spike protein in SARS-CoV-2 variant strains.

SARS-Cov2 Variants	Mutation	Reference Peptide	Mutated Peptide	Score_R	Score_M	Prediction_R	Prediction_M
Alpha (B.1.1.7)	**A570D**	NKKFLPFQQFGRDIA	NKKFLPFQQFGRDID	0.05	0.11	IL-13 non-inducer	IL-13 inducer
		KFLPFQQFGRDIADT	KFLPFQQFGRDIDDT	0.03	0.06	IL-13 non-inducer	IL-13 inducer
	T716I	SNNSIAIPTNFTISV	SNNSIAIPINFTISV	0.04	0.06	IL-13 non-inducer	IL-13 inducer
	S980A	NFGAISSVLNDILSR	NFGAISSVLNDILAR	0.04	0.06	IL-13 non-inducer	IL-13 inducer
		GAISSVLNDILSRLD	GAISSVLNDILARLD	0.03	0.08	IL-13 non-inducer	IL-13 inducer
		VLNDILSRLDKVEAE	VLNDILARLDKVEAE	0.03	0.07	IL-13 non-inducer	IL-13 inducer
		LNDILSRLDKVEAEV	LNDILARLDKVEAEV	0.05	0.08	IL-13 non-inducer	IL-13 inducer
		NDILSRLDKVEAEVQ	NDILARLDKVEAEVQ	0.04	0.09	IL-13 non-inducer	IL-13 inducer
	D1118H	TQRNFYEPQIITTDN	TQRNFYEPQIITTHN	0.06	0.04	IL-13 inducer	IL-13 non-inducer
		QRNFYEPQIITTDNT	QRNFYEPQIITTHNT	0.07	0.03	IL-13 inducer	IL-13 non-inducer
		YEPQIITTDNTFVSG	YEPQIITTHNTFVSG	0.04	0.06	IL-13 non-inducer	IL-13 inducer
Beta (B.1.351)	**L18F**	LVLLPLVSSQCVNLT	LVLLPLVSSQCVNFT	0.06	0.04	IL-13 inducer	IL-13 non-inducer
		LLPLVSSQCVNLTTR	LLPLVSSQCVNFTTR	0.07	0.02	IL-13 inducer	IL-13 non-inducer
	D80A	VSGTNGTKRFDNPVL	VSGTNGTKRFANPVL	0.05	0.12	IL-13 non-inducer	IL-13 inducer
		GTNGTKRFDNPVLPF	GTNGTKRFANPVLPF	0.02	0.07	IL-13 non-inducer	IL-13 inducer
		SGTNGTKRFDNPVLP	SGTNGTKRFANPVLP	0.09	0.02	IL-13 inducer	IL-13 non-inducer
Delta (B.1.617.2)	**T19R**	LVLLPLVSSQCVNLT	LVLLPLVSSQCVNLR	0.06	0.04	IL-13 inducer	IL-13 non-inducer
	R158G	FRVYSSANNCTFEYV	FGVYSSANNCTFEYV	0.03	0.06	IL-13 non-inducer	IL-13 inducer
	P618R	TQTNSPRRARSVASQ	TQTNSRRRARSVASQ	0.02	0.08	IL-13 non-inducer	IL-13 inducer
		QTNSPRRARSVASQS	QTNSRRRARSVASQS	0.02	0.08	IL-13 non-inducer	IL-13 inducer
	D950N	DSLSSTASALGKLQD	DSLSSTASALGKLQN	0.05	0.06	IL-13 non-inducer	IL-13 inducer

#Score_R and Score_M is the score of Reference Spike protein and score of mutated Spike protein respectively.
Prediction_R and Prediction_M is the IL-13 Inducing/non-inducing prediction of the Reference Spike protein and mutated Spike protein respectively.

of features, such as 95 selected features and top 10 features are reported. The XGBoost-based model achieves a maximum AUC of 0.83 and 0.80 on the training and independent dataset, respectively using top10 features. We provided a webserver and a standalone version of 'IL-13Pred' for the scientific community to distinguish the IL-13 inducing and non-inducing peptides/epitopes. Using the protein scan module of the 'IL-13Pred' webserver, we identify the IL-13 inducing peptides in 5 different COVID-19 strains, such as India (MZ340539.1), China (MT291829.1), USA (MZ319836.1) Brazil (MZ169911.1), and Japan (LC528233.2). We also identified the IL-13 inducing regions in Spike protein of Alpha (B.1.1.7), Beta (B.1.351), and Delta (B.1.617.2) variants of the SARS-CoV-2 virus. From the prediction results, the amino acid P168R and D950 N mutations in the Delta variant induce IL-13 immunoregulatory cytokine.

5. Conclusion

The epitope/peptide is effective in inducing IL-13 and altering the immune response towards a disease state, and it has offered great importance in designing immunotherapy and vaccine. Although inducing IL-13 response in patients is a complex process that depends on various factors, an epitope/peptide is still a promising alternative in designing a vaccine or immunotherapy against any disease. Thus, IL-13 induced immunosuppression could serve as a crucial step in vaccine subunit designing. Although diverse *in silico* methods are available for T cell epitopes prediction, no computational method is available for IL-13 inducing epitopes/peptide prediction. The present study offers a direction in providing a user-friendly platform/webserver. We encourage the scientific community to use IL-13Pred to develop efficient immunotherapy and vaccine candidates by differentiating them apriori as IL-13 inducing epitopes.

Author contributions

SJ, AD, and SP collected and processed the datasets. SJ, AD, SP and GPSR implemented the algorithms. SJ, AD, and SP developed the prediction models. SJ, AD, SP and GPSR analysed the results. SP, SJ, and AD created the back-end of the web server and front-end user interface. SJ, AD, SP and GPSR penned the manuscript. GPSR conceived and coordinated the project, and gave overall supervision to the project. All authors have read and approved the final manuscript.

Data availability statement

All the datasets generated for this study are either included in this article the dataset is available at https://webs.iiitd.edu.in/raghava/il13pred/dataset.php.

Funding Source

The current work has not received any specific grant from any funding agencies.

Declaration of competing interest

The authors declare no competing financial and non-financial interests.

Acknowledgement

AD is thankful to the Department of Science and Technology (DST-INSPIRE) and SP is thankful to the Department of Biotechnology (DBT) for providing senior research fellowships. SJ, AD and SP are thankful to the Department of Computational Biology, IIIT-Delhi for infrastructure and facilities.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compbiomed.2022.105297.

References

[1] M. An, C. Ja, W. de, R. M, F. B, J. P, G. A, A. S, W. D, C. Bg, S. M, Interleukin 13, a T-cell-derived cytokine that regulates human monocyte and B-cell function, Proc. Natl. Acad. Sci. U.S.A. 90 (1993) 3735–3739, https://doi.org/10.1073/PNAS.90.8.3735.

[2] M. A, C. P, D. Jm, D. X, G. Jc, K. M, L. C, L. P, L. P, M. B, Interleukin-13 is a new human lymphokine regulating inflammatory and immune responses, Nature 362 (1993) 248–250, https://doi.org/10.1038/362248A0.

[3] W. Ta, IL-13 effector functions, Annu. Rev. Immunol. 21 (2003) 425–456, https://doi.org/10.1146/ANNUREV.IMMUNOL.21.120601.141142.

[4] M. P, R. W, Interleukin 13 and its role in gut defence and inflammation, Gut 61 (2012) 1765–1773, https://doi.org/10.1136/GUTJNL-2012-303461.

[5] S. Bf, K. U, M. Tw, Interleukin 13: a growth factor in hodgkin lymphoma, Int. Arch. Allergy Immunol. 126 (2001) 267–276, https://doi.org/10.1159/000049523.

[6] O. Y, P. Rk, Suppression of an IL-13 autocrine growth loop in a human Hodgkin/Reed-Sternberg tumor cell line by a novel IL-13 antagonist, Cell, Immunol. 211 (2001) 37–42.

[7] T. M, P. Jm, B. Ja, Role of IL-13 in regulation of anti-tumor immunity and tumor growth, Cancer Immunol. Immunother. 53 (2004) 79–85, https://doi.org/10.1007/S00262-003-0445-0.

[8] S. T, M. Y, Z. L, W. S, L. H, S. L, S. H, W. X, Z. W, Z. L, Z. J, Association between IL13 gene polymorphisms and susceptibility to cancer: a meta-analysis, Gene 515 (2013) 56–61, https://doi.org/10.1016/J.GENE.2012.11.035.

[9] T. B, S. L, M. Y, X. P, L. J, P. S, H. B, D. K, U. K, M, Endogenously expressed IL-4Rα promotes the malignant phenotype of human pancreatic cancer in vitro and in vivo, Int. J. Mol. Sci. 18 (2017), https://doi.org/10.3390/IJMS18040716.

[10] S. X, T. B, S. J, K. M, Possible roles of interleukin-4 and -13 and their receptors in gastric and colon cancer, Int. J. Mol. Sci. 22 (2021) 1–20, https://doi.org/10.3390/IJMS22020727.

[11] L. H, A. S, R. K, J. A, W. Y, L. J, M. Jl, J. G, P. E, D. Jh, Interleukin-4 and interleukin-13 increase NADPH oxidase 1-related proliferation of human colon cancer cells, Oncotarget 8 (2017) 38113–38135, https://doi.org/10.18632/ONCOTARGET.17494.

[12] C. Ya, K. J, Association of IL4, IL13, and IL4R polymorphisms with gastrointestinal cancer risk: a meta-analysis, J. Epidemiol. 27 (2017) 215–220, https://doi.org/10.1016/J.JE.2016.06.002.

[13] O.-R. S, Immune surveillance: a balance between protumor and antitumor immunity, Curr. Opin. Genet. Dev. 18 (2008) 11–18, https://doi.org/10.1016/J.GDE.2007.12.007.

[14] W. Hw, J. Ja, Alternative activation of tumor-associated macrophages by IL-4: priming for protumoral functions, Cell Cycle 9 (2010) 4824–4835, https://doi.org/10.4161/CC.9.24.14322.

[15] W. Debinski, D. Gibo, S.W. Hulet, J. Connor, G. Gillespie, Receptor for Interleukin 13 Is a Marker and Therapeutic Target for Human High-Grade Gliomas, Undefined, 1999.

[16] S.-M. Yuan, Interleukin-13 in the pathogenesis of pulmonary artery hypertension, J. Lab. Med. 43 (2019) 5–11, https://doi.org/10.1515/LABMED-2018-0323.

[17] A. U, K. D, W. A, S. A, N.-C. Y, G. N, M. N, K. T, L. N, S.N. N, B. G, M. E, K. A, K.Y. D, P. G, M. A, C. Hy, R. E, H. S, L. J, New role for interleukin-13 receptor α1 in myocardial homeostasis and heart failure, J. Am. Heart Assoc. 6 (2017), https://doi.org/10.1161/JAHA.116.005108.

[18] R. El, L. Rf, Interleukin-13 signaling and its role in asthma, World Allergy Organ. J. 4 (2011) 54–64, https://doi.org/10.1097/WOX.0B013E31821188E0.

[19] C. J, Role of interleukin-13 in asthma, Curr. Allergy Asthma Rep. 13 (2013) 415–420, https://doi.org/10.1007/S11882-013-0373-9.

[20] C. J, L. Rf, H. Na, K. Pe, P. Mv, A. JR, H. Jm, S. H, W. Lc, S. Z, M. S, E. Md, B. Sp, M. Jg, Lebrikizumab treatment in adults with asthma, N. Engl. J. Med. 365 (2011) 1088–1098, https://doi.org/10.1056/NEJMOA1106469.

[21] B. D, F. M, V. G, P. G, C. Gw, A critical evaluation of anti-IL-13 and anti-IL-4 strategies in severe asthma, Int. Arch. Allergy Immunol. 170 (2016) 122–131, https://doi.org/10.1159/000447692.

[22] B.S. Bochner, D.A. Klunk, S.A. Sterbinsky, R.L. Coffman, R.P. Schleimer, IL-13 selectively induces vascular cell adhesion molecule-1 expression in human endothelial cells, J. Immunol. 154 (1995).

[23] W. K M, IL-12/IL-13 axis in allergic asthma, J. Allergy Clin. Immunol. 107 (2001) 9–18, https://doi.org/10.1067/MAI.2001.112265.

[24] L. L, X. Y, N. A, L. Yh, F. L, M. Tr, L. D, Effects of Th2 cytokines on chemokine expression in the lung: IL-13 potently induces eotaxin expression by airway epithelial cells, J. Immunol. 162 (1999) 2477–2487. September 6, 2021, http://www.jimmunol.org/cgi/content/full/162/5/2477.

[25] O. Lk, M. Mr, F. J, W. Tm, N. H, G. Cj, C. S, T. Am, X. Az, T. Sv, L. J, G. X, Y. L, H. Sl, W. Pl, B. JR, C. Ml, B. R, S. A, B. F, H. Cs, G. Rw, M. Mj, C. Zf, H. H, D. S, L. Q, K. Bs, Sensory neurons Co-opt classical immune signaling pathways to mediate chronic itch, Cell 171 (2017) 217–228, https://doi.org/10.1016/J.CELL.2017.08.006, e13.

[26] D. An, S. Te, M. C, P. S, B. Bt, C. Rm, S. Jm, M. Jz, M. Gb, D. JR, B. Ga, S. Mg, B. Sl, A. Mm, M. C, B. Pe, P. R, Y. Mk, L. Gr, L. Jj, R. Sj, P. Md, M. Aj, D. A, M. Bj, A. Je, P. Wa, IL-13 is a driver of COVID-19 severity, MedRxiv Prepr. Serv. Heal. Sci. (2021), https://doi.org/10.1101/2020.06.18.20134353.

[27] H. C, W. Y, L. X, R. L, Z. J, H. Y, Z. L, F. G, X. J, G. X, C. Z, Y. T, X. J, W. Y, W. W, X. X, Y. W, L. H, L. M, X. Y, G. H, G. L, X. J, W. G, J. R, G. Z, J. Q, W. J, C. B, Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China, Lancet (London, England) 395 (2020) 497–506, https://doi.org/10.1016/S0140-6736(20)30183-5.

[28] Y. Y, S. C, L. J, Y. J, W. J, H. F, W. F, L. G, L. Y, X. L, P. L, Y. M, C. M, Z. H, W. W, Z. R, L. D, X. Z, W. H, Z. M, Z. Z, G. Gf, J. C, L. L, L. Y, Plasma IP-10 and MCP-3 levels are highly associated with disease severity and predict the progression of COVID-19, J. Allergy Clin. Immunol. 146 (2020) 119–127, https://doi.org/10.1016/J.JACI.2020.04.027, e4.

[29] V. R, M. S, O. Ja, D. Sk, M. S, C. JR, W. Dk, S. A, P. B, The immune epitope database (IEDB): 2018 update, Nucleic Acids Res. 47 (2019) D339–D343, https://doi.org/10.1093/NAR/GKY1006.

[30] D. A, P. S, S. N, U. Ss, R. Gps, Computer-aided prediction and design of IL-6 inducing peptides: IL-6 plays a crucial role in COVID-19, Brief, Bioinformatics 22 (2021) 936–945, https://doi.org/10.1093/BIB/BBAA259.

[31] S.K. Dhanda, P. Vir, G.P.S. Raghava, Designing of interferon-gamma inducing MHC class-II binders, Biol. Direct 8 (2013) 30, https://doi.org/10.1186/1745-6150-8-30.

[32] M. Kumar, G.P.S. Raghava, Prediction of nuclear proteins using SVM and HMM models, BMC Bioinf. 10 (2009) 22, https://doi.org/10.1186/1471-2105-10-22.

[33] W. Du, X. Zhao, Y. Sun, L. Zheng, Y. Li, Y. Zhang, SecProCT: in silico prediction of human secretory proteins based on capsule network and transformer, Int. J. Mol. Sci. 22 (2021), https://doi.org/10.3390/IJMS22169054.

[34] A. Pande, S. Patiyal, A. Lathwal, C. Arora, D. Kaur, A. Dhall, G. Mishra, H. Kaur, N. Sharma, S. Jain, S.S. Usmani, P. Agrawal, R. Kumar, V. Kumar, G.P.S. Raghava, Computing wide range of protein/peptide features from their sequence and structure, bioRxiv (2019), 599126, https://doi.org/10.1101/599126.

[35] J. Tang, S. Alelyani, H. Liu, Feature Selection for Classification: A Review, (n.d.).

[36] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, T.-Y. Liu, LightGBM: A Highly Efficient Gradient Boosting Decision Tree, (n.d.). https://github.com/Microsoft/LightGBM. (accessed September 6, 2021).

[37] G. Nagpal, S.S. Usmani, S.K. Dhanda, H. Kaur, S. Singh, M. Sharma, G.P. S. Raghava, Computer-aided designing of immunosuppressive peptides based on IL-10 inducing potential, Sci. Rep. 7 (2017), 42851, https://doi.org/10.1038/srep42851.

[38] S. Patiyal, P. Agrawal, V. Kumar, A. Dhall, R. Kumar, G. Mishra, G.P.S. Raghava, NAGbinder: an approach for identifying N-acetylglucosamine interacting residues of a protein from its primary sequence, Protein Sci. 29 (2020) 201–210, https://doi.org/10.1002/PRO.3761.

[39] N. Sharma, S. Patiyal, A. Dhall, A. Pande, C. Arora, G.P.S. Raghava, AlgPred 2.0: an improved method for predicting allergenic proteins and mapping of IgE epitopes, Briefings Bioinf. 22 (2021), https://doi.org/10.1093/BIB/BBAA294.

[40] A. Dhall, S. Patiyal, N. Sharma, N.L. Devi, G.P.S. Raghava, Computer-aided prediction of inhibitors against STAT3 for managing COVID-19 associated cytokine storm, Comput, Biol. Med. 137 (2021), 104780, https://doi.org/10.1016/J.COMPBIOMED.2021.104780.

[41] N. Sharma, S. Patiyal, A. Dhall, N.L. Devi, G.P.S. Raghava, ChAlPred: a web server for prediction of allergenicity of chemical compounds, Comput. Biol. Med. 136 (2021), https://doi.org/10.1016/J.COMPBIOMED.2021.104746.

[42] S. Jain, K. Preet, K. Malhotra, S. Patiyal, G.P.S. Raghava, A highly accurate model for screening prostate cancer using propensity index panel of ten genes, bioRxiv (2021), 436371, https://doi.org/10.1101/2021.03.22.436371, 2021.03.22.

[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, A. Müller, J. Nothman, G. Louppe, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, É. Duchesnay, Scikit-learn: machine learning in Python, J. Mach. Learn. Res. 12 (2012) 2825–2830. https://arxiv.org/abs/1201.0490v4. (Accessed 6 September 2021).

[44] T. Cs, C. C, C. Al, H. Ba, V. J, M. T, W. C, C. Ad, K. K, RPC4046, A novel anti-interleukin-13 antibody, blocks IL-13 binding to IL-13 α1 and α2 receptors: a Randomized, double-blind, Placebo-controlled, dose-escalation first-in-human study, Adv. Ther. 34 (2017) 1364–1381, https://doi.org/10.1007/S12325-017-0525-8.

[45] K. S, B. J, B. A, L. J, L. S, M. D, L. T, V. Jc, L. W, F. K, Blockade of interleukin-13-mediated cell activation by a novel inhibitory antibody to human IL-13 receptor alpha1, Mol. Immunol. 43 (2006) 1799–1807, https://doi.org/10.1016/J.MOLIMM.2005.11.001.

Shipra Jain is currently working as Ph.D. in Bioinformatics from Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India.

Anjali Dhall is currently working as Ph.D. in Bioinformatics from Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India.

Sumeet Patiyal is currently working as Ph.D. in Bioinformatics from Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India.

Gajendra P. S. Raghava is currently working as Professor and Head of Department of Computational Biology, Indraprastha Institute of Information Technology, New Delhi, India.