IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 29, 2021, accepted May 10, 2021, date of publication May 13, 2021, date of current version May 24, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3080041

iNitroY-Deep: Computational Identification of Nitrotyrosine Sites to Supplement Carcinogenesis Studies Using Deep Learning

SHERAZ NASEER[1], RAO FAIZAN ALI[2], SULIMAN MOHAMED FATI[3], (Senior Member, IEEE), AND AMGAD MUNEER[2]

[1]Department of Computer Science, University of Management and Technology, Lahore 54770, Pakistan
[2]Computer and Information Sciences, Department, Universiti Teknologi PETRONAS, Seri Iskandar 32610, Malaysia
[3]College of Computer and Information Sciences, Prince Sultan University, Riyadh 11586, Saudi Arabia

Corresponding author: Sheraz Naseer (sheraz.naseer@umt.edu.pk)

This work was supported by the Prince Sultan University, Riyadh, Saudi Arabia.

ABSTRACT In biological systems, Nitration is a crucial post-translational modification which occurs on various amino acids. Nitration of Tyrosine is regarded as nitorsative stress biomarker resulting in the formation of peroxynitrite and other reactive and harmful nitrogen species. NitroTyrosine is closely related to Carcinogenesis, tumor growth progression and other major pathological conditions including systemic autoimmune diseases, inflammation, neurodegeneration and cardiovascular disorders. Additionally, the alteration in Nitrotyrosine profile occurs well before appearance of any symptoms of aforementioned diseases making nitrotyrosine a biomarker and potential target for early prognosis of aforementioned diseases. The wet lab identification of potential nitrotyrosine sites is laborious, time-taking and costly due to challenges of in vitro, ex vivo and in vivo identification processes. To supplement wet lab identification of nitrotyrosine, we proposed, implemented and evaluated a different approach to develop tyrosine nitration site predictors using pseudo amino acid compositions (PseAAC) and deep neural networks (DNNs). Proposed approach does not require any feature extraction and uses DNNs for learning a feature representation of peptide sequences and classification thereof. Validation of proposed approach is done using well-known model evaluation measures. Among different deep neural networks, convolutional neural network-based predictor achieved best scores on independent dataset with accuracy of 87.2%, matthew's correlation coefficient score of 0.74 and AuC score of 0.91 which outperforms the previous reported scores of Nitrotyrosine predictors.

INDEX TERMS Carcinogenesis, convolutional neural network, deep features, nitration, PseAAC, PTM, recurrent neural networks, tyrosine.

I. INTRODUCTION

Cells are constantly exposed to diverse stressors under physiological conditions, which leads to dynamic changes in cellular functions. Several mechanisms are used by cells to respond to these dynamic changes including regulation of energy producing pathways, alterations of epigenetic marks, modulation of metabolic enzymes activities using metabolites and protein post translational modifications (PTMs) [1]. One such PTM is 3-nitrotyrosine (NitroY) which is formed by the substitution of a hydrogen atom with nitro group (-NO2) in any of the two carbon atoms of the phenolic ring of the amino acid Tyrosine [2]. This process is also called nitration of Tyrosine and several research publications have extensively discussed biological importance of nitration mechanisms under different conditions and pathophysiological relevance thereof in pathological settings from acute to chronic diseases [3]–[5]. NitroY is regarded as nitrosative stress biomarker, a cell state caused by the over production of nitric acid (-NO), resulting in the formation of peroxynitrite and other reactive and harmful nitrogen species [6]. Increased levels of NitroY have been reported in numerous pathological

The associate editor coordinating the review of this manuscript and approving it for publication was Vincenzo Conti.

TABLE 1. NitroY as biomarker for various diseases.

SrNo	Author	Year	NitroY as Biomarker for Disease	Findings Summary
1	Bandookwala et al. [12]	2018	Neurodegenerative diseases	The alteration in 3-Nitrotyrosine profile occurs well before any symptoms appear for neurodegenerative diseases. Reduction in 3-Nitrotyrosine levels is indicative of oxidative stress-related brain and central nervous system pathologies.
2	Korde Ch et al. [14]	2013	Cancer Related Events	NitroY is a nitrosative biomarker showing overproduction of NO, which has been suggested to modulate different cancer-related events including angiogenesis, apoptosis, cell cycle invasion, and metastasis.
3	Gochman et al. [15]	2012	Colon Cancer	iNOS expression and NitroY may be an indicator of cancer development and progression in inflammation and colon cancer in humans.
4	Allameh et al. [16]	2009	Esophagus precancer, adenocarcinoma and squamous cell carcinoma	It was demonstrated that NitroY was predominantly expressed in 72.7% of Iranian adenocarcinoma patients.
5	Samoszuk et al. [17]	2002	Breast Cancer	NitroY has possible association between inflammatory cells and reactive nitrogen species in modulating microvascular density at the edges of breast cancer.
6	Kondo et al. [18]	2002	Colorectal carcinoma	Peroxynitrite-mediated stress, expressed by Nitrotyrosine, is associated with proliferation of human metastatic colorectal carcinoma in the liver.
7	Jaiswal et al. [19]	2000	Cholangiocarcinoma	Increased NitroY levels may provide a link for prognosis of cholangiocarcinoma.
8	Kato et al. [20]	2000	Esophageal Cancer	Comparison of Nitrotyrosine expression and the pathological findings showed that there was a significant association between the expression of nitrotyrosine and depth of tumor invasion. Nitrotyrosine, a product of Nitrogen species, is expressed in esophageal squamous cell carcinoma.
9	Goto et al. [21]	1999	Gastric Cancer	High production of iNOS and nitrotyrosine in the gastric mucosa infected with H. pylori may contribute to the carcinogenesis of gastric cancer.

conditions including systemic autoimmune diseases [7]–[10], inflammation [11], neurodegeneration [12] and cardiovascular disorder [3], [13].

NitroY is shown to be positively related with carcinogenesis and tumor growth progression [14]–[21] in various studies as shown in Table 1. Korde Choudhary *et al.* illuminated the effects of nitric oxide and NitroY on various human cancers [14]. Gochman *et al.* postulated iNOS expression and NitroY as an indicator of cancer development and progression while studying colon cancer in humans [15]. Allameh *et al.* [16] demonstrated the predominant expression of NitroY in adenocarcinoma during measurement of potential causative factors of esophagus cancer progression in Iranian population. Samoszuk *et al.* [17] studied NitroY levels in human breast cancer samples of varying pathological types and concluded the association between inflammatory cells and reactive nitrogen species in modulating microvascular density at the edges of breast cancer. Kondo *et al.* observed abundant levels of NitroY in hepatocytes adjacent to human metastatic colorectal carcinoma in liver and suggested positive contribution of NitroY in proliferation of cancer cells resulting in tumor cell progression [18]. Jaiswal *et al.* showed that the excess production of nitric oxide, marked by NitroY, causes DNA damage and may provide a link for prognosis of cholangiocarcinoma, which is a highly malignant and fatal adenocarcinoma starting from biliary epithelia [19]. Kato *et al.* researched NitroY expression in esophageal squamous cell cancers and showed that the survival rate of patients with NitroY-negative cancer was significantly higher than that of patients with NitroY-positive cancer [20]. Goto *et al.* suggested the positive contribution of NitroY in carcinogenesis of gastric cancer [21].

Considering the importance of NitroY for prognosis of Carcinogenesis, a better understanding for the molecular mechanisms of NitroY requires identification of NitroY sites in peptide sequences. Although multiple large-scale in-vivo, ex-vivo, and in-vitro methods such as immunohistochemical analysis [16], chromatography [22] and mass spectroscopy [23] have been applied to detect NitroY sites, these experimental methods are time-consuming and labor-intensive. Research community has applied computational methods to solve problems in proteomics and genomics using various data science and machine learning techniques [24]–[29]. Similarly, recent research contributions have proposed use of different feature extraction techniques to identify the potential NitroY sites [30]–[33] using various machine learning algorithms. Although these contributions show promising results but most of the techniques use human-engineered features. According to Lecun *et al.* [34] human features have certain limitations as they are

laborious to learn because of absence of a feedback mechanism between learning subsystem and feature extraction subsystem. This makes it impossible to ascertain the quality of features, to develop an effective predictor, until such a candidate model is developed and evaluated. Additionally, generation of human-engineered features require domain knowledge and human intervention which is sometimes hard to come by [34]. For NitroY site identification, important **research gaps** identified by Lecun *et al.* [34] still remained open. These gaps include:

- Limited capability of conventional ML based predictors to work with raw peptide sequences
- Laborious and costly feature extraction due to requirement of human intervention and domain expertise
- Dependence of conventional ML based predictors on quality of features
- Mutual isolation of feature extraction system and classification system

Lecun *et al.* [34] proposed the use of deep neural networks, to overcome aforementioned limitations, which are studied under the discipline of deep learning.

Deep learning is the study of different deep neural network architectures (DNNs) which have enabled many breakthroughs in different scientific disciplines including computer vision [35], natural language processing [36] and information security [37], [38] to mention a few. In essence, all DNNs consist of multiple layers of basic mathematical functions, dubbed as neurons, which transform the received inputs, layer by layer, until the transformed input reaches to the output layer of the neural network which uses this transformed input to make the predictions. Each DNN layer receives input from the upper layer and translates it into some representation that subsequent layers use. Each such transformation can be considered as a new and task specific representation of input data. DNN layers transform their input non-linearly, producing hierarchically abstract, task-specific representations that are insensitive to unimportant variations, but sensitive to significant features. With appropriate DNN training, the representation generated by the last hidden layer, nearest to the output layer of DNN, is so effective in recognizing hidden patterns of input that it is used by the output layer to make predictions. Hence the DNNs provide us a means to generate efficient, task specific and effective deep features which does not require human intervention, domain knowledge and laborious feature selection process. Application of DNNs for proteomics is a fledgling area of research getting attention from proteomics research community [24], [25], [39]–[41].

The main **contribution** of this study is to propose an effective NitroY site predictor which is not plagued by the aforementioned limitations identified by [34]. By combining the Pseudo Amino Acid Composition (PseAAC) [42] with deep neural networks, we propose an improved predictor for identifying NitroY sites in proteins, which does not require feature extraction and hence removes the need for human intervention and domain expertise. Due to inherent structure of DNNs, both representation learning system and NitroY site identification system work in unison for learning a feature representation of peptide sequences and performing classifications. For this study, we used well-known DNNs, including Standard neural network (FCN), three variants of recurrent neural networks (RNNs) and convolutional neural network (CNN). We used the 5-step rule of Chou [43] for this purpose opted widely in a variety of publications [44]–[47] and shown in 1. Steps of Chou's methodology include (i) collection of benchmark dataset (ii) preprocessing of raw PseAAC sequences to make them amenable to machine learning algorithms and extraction of features with some feature engineering technique (iii) Implementation and training of prediction model (iv) evaluation of models based on results and (v) deployment of predictor using webserver. Multiple candidate DNN-based prediction models are trained, in this study, using aforementioned DNN algorithms to obtain an optimal model for identification of NitroY sites. Performance of developed models is evaluated among themselves and with reported predictors in literature using well-known parameters of model evaluation. This paper is organized as follows. Section 2 illuminates the related works and how the proposed approach in this study is different and more advantageous. Section 3 describes the research methodology of the proposed study. Section 4 provides the results and findings. Section 5 provided the discussion while the conclusion and the future work are presented in last Section.

II. RELATED WORKS

Research community has applied computational methods to solve problems in proteomics and genomics using various data science and machine learning techniques [26]–[29]. Similarly, recent research contributions have proposed use of different feature extraction techniques to identify the potential NitroY sites [30]–[33] using various machine learning algorithms. Liu *et al.* [30] proposed GPS-YNO2 with three stringencies including low, medium and high. Xu *et al.* [31] developed iNitroTyr using position-specific dipeptide propensity for PseAAC and conventional ML algorithms and provided evaluation results using well-known model evaluation metrics. Hasan *et al.* [33] devised NTyroSite using an evolutionary approach for developing sequence features and filtered top differential features to reduce classification error. NTyrosite used random-forest algorithm for classification and it was evaluated using RoC, AuC, Accuracy and MCC. Xie *et al.* [48] proposed first deep neural network based predictor DeepNitro but used only the classification capability of DNNs. They used different features extraction techniques including position specific scoring matrix and k-space spectrum encodings to create feature vectors which were then classified by DeepNitro system. Our proposed approach differs from Deepnitro because we used DNNs for both feature extraction from raw PseAAC sequence and classification thereof, hence removing the need to perform costly feature extraction process. Ghauri *et al.* proposed pNitro-Tyr [32] based on features calculated using

FIGURE 1. 5-step rule of Chou for NitroY site identification.

FIGURE 2. Adopted methodology for NitroY site identification.

statistical moments of PseAAC sequences. pNitro-Tyr used a single layer shallow Neural network for classification. The most recent contribution for NitroY site identification was PredNTS [49], proposed by Nilamyani *et al.* PredNTS used human-engineered features including, K-mer, composition of k-spaced residues, AAindex and binary encoding schemes while Random-Forest was used as classification algorithm. A fundamental limitation of all aforementioned NitroY predictor development approaches is their reliance on the quality of human-engineered features for model training. Our approach is different because we aim to develop a predictor model which can automatically learn deep feature representations from raw PseAAC sequences without expert knowledge and human intervention. We exploited the inherent capability of DNNs to learn input representation by transforming it hierarchically in response to loss score of estimated labels. Once the DNN model is sufficiently trained, the intermediate layers of DNN transform raw peptide sequences of PseAAC to meaningful deep representations and an output layer of DNN perform prediction using the deep representation learned by earlier layers. Since both, the representation learning subsystem and prediction subsystem work in unison, the optimizer uses the loss score as the feedback signal to improve both the subsystems of DNN.

III. MATERIALS AND METHODS

The approach adopted in this study is gleaned from the Chou's five-step rule. Instead of relying on human-engineered features, our methodology, as shown in Figure 2, combines the feature extraction and model training step using DNNs. For this research, Several DNN-based models were trained and evaluated using standard performance evaluators of prediction models to obtain an optimal model for predicting NitroY sites. The emphasis of this section is on the first three steps of methodology shown in Figure 2, while the last two steps of the suggested methodology are elaborated in following sections.

A. COLLECTION OF BENCHMARK DATASET

We used the advanced search and annotation capabilities of UniProt to create benchmark dataset for this analysis [50]. Quality of benchmark dataset was ensured by selecting protein sequences where NitroY was detected and investigated experimentally. Using Chou'sPseAAC [42], a peptide sequence with a NitroY positive site can be shown as follows:

$$f(P) = P_{-r}P_{-(r-1)}\dots P_{-2}P_{-1}YP_{+1}P_{+2}\dots P_{+(r-1)}P_{+r}$$

where Y reflects the positive NitroY amino acid 'Tyrosine' and P's represent the neighboring amino acid of positive site. The symbol 'r' is a sequence index, where negative indexes are the left-hand side residues and positive indexes represent the right-side neighboring residue around the NitorY site. We were able to find 231 experimentally verified proteins and extracted the positive and negative samples of length ρ from aforesaid PseAAC sequences of proteins. Based on empirical observations, the length ρ is fixed at 41 for both positive and negative samples. Positive sequences were produced by fixing the index of NitroY site at $r = 21$ and attaching twenty leftside and twenty rightside neighbor residues of the site to achieve the standard ρ length sequence. For positive samples with $\rho < 41$, symbol X was used as a dummy amino acid residue and attached on both sides of the sequence to achieve standard length ρ. Same methodology was adopted to extract negative samples from acquired protein PseAAC sequences. The above procedure resulted in 327 positive and 3039 negative samples.

To remove homologous sequences, we applied USEARCH algorithm proposed by Edgar [51] on both positive and

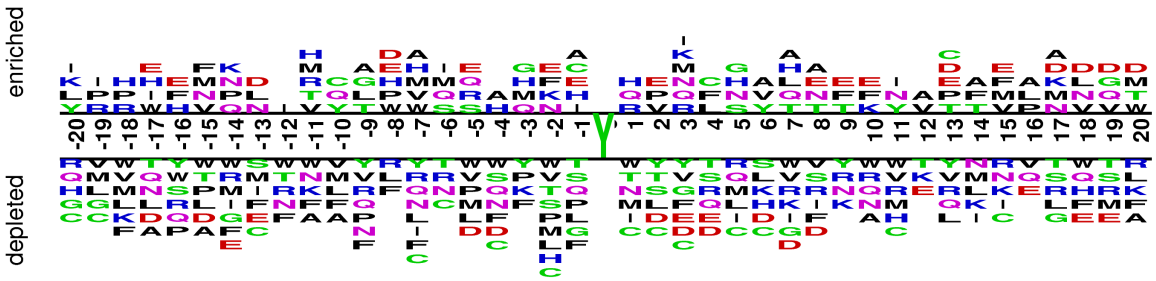


FIGURE 3. Two sample logo of NitroY sites.

TABLE 2. Encoding of amino acid used in this study.

X	A	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	U	V	W	Y
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22

negative samples. Application of USEARCH for redundancy removal on positive samples with similarity threshold of 70% resulted in severely reduced dataset of 49 samples so we chose not to remove homologous positive samples. Application of the same on negative samples resulted in 507 samples. The final benchmark dataset consisted of 327 postive samples and 507 negative samples resulting in total sample size of 834 PseAAC peptide sequences which can be represented as follows:

$$Y = Y^{+} \cup Y^{-}$$

where Y^{+} represents positive sample sequences and Y^{-} represents negative sample sequences. The class ratio between positive and negative samples was found to be 39% and 61% respectively. The dataset is available at https://mega.nz/folder/pxFhxaAJ#w8Elm5EJzPYHINEr_39aTQ. In order to help answering a question about sequence biases around Nitrotyrosine sites, a two sample logo, proposed by Vacic *et al.* [52], was generated to visualize residues that are significantly enriched or depleted in the set of NitroY fragments. The Two Sample Logo of benchmark dataset, as shown in Figure 3, contains 41 residue fragments, 20 upstream and 20 downstream, from all Tyrosines found in experimentally verified nitrated proteins. The positive sample contains 327 fragments around experimentally verified NitroY sites, while the negative sample contains all remaining non-redundant Tyrosines from the same set of proteins, 834 in total. Significant variances in the nearby tyrosines were found between the nitrated and non-nitrated sites. In the depleted position, residues Q, N, R and L were more frequently observed while in enriched region D, E and H were observed frequently. Multiple amino acid residues were found stacked at some over- or under-represented positions of the surrounding sequences suggesting small information between the positive and negative samples. The above results indicate that more abstract and task specific features are required to identify between the samples of two classes.

B. SAMPLE ENCODING

Almost all DNNs require data in the quantitative format before the neuron layers inside DNN process it. We applied a very basic quantitative encoding of PseAAC sequences, shown in Table 2, where 1st row displays the IUPAC symbols of amino acids, and corresponding entries in 2nd row show the integer used to represent the amino acid in the encoded sample. Since this encoding is the simplest possible amino acid numerical representation, it has no significant effects on the final results. The benchmark data set was split into a training set of 583 PseAAC sequences and a test set of 251 samples with a 70/30 ratio in the train set and test set. That is, for models training, 70% of the data was used, and the rest 30% was used for independent model testing. In all training and test sets, the initial 39/61 class ratio was preserved.

C. CANDIDATE DEEP MODEL TRAINING AND OPTIMIZATION

This section explains training and optimizing of DNN candidate models for predicting NitroY sites. The study conducted experiments using well-known neural network architectures such as Standard Neural Networks (FCNs), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) with simple RNN units, Gated Recurrent Units (GRU) and Long Short-Term Memory (LSTM) units respectively. For optimization of DNN candidate models, we adopted the Randomized Hyperparameter search methodology of Bergstra *et al.* [53]. Randomized Hyperparameter search offers better hyperparameters for DNNs with the limited computational budget by performing a random search over large hyperparameter space. This is achieved by randomly sampling the hyperparameters from the space and evaluating the performance of models created using these parameters. For each DNN used to identify NitroY site, the following subparagraphs provide a brief introduction and architecture.

TABLE 3. Standard neural net (FCN) architecture for NitroY identification.

Type of Neuron Layer	No. of Weights
Dense layer with 12 relu units	(41 + 1) * 12 = 504
Dense layer with 8 relu units	(12 + 1) * 8 = 104
Model Output layer with one Sigmoid unit	(8 + 1) * 1 = 09

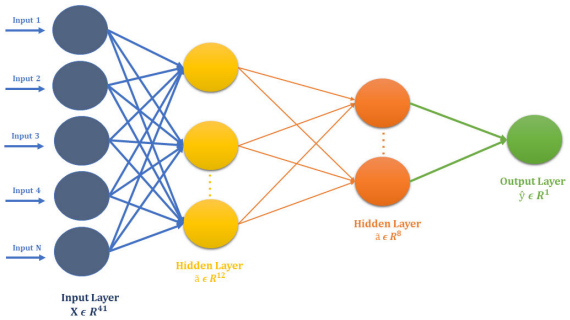


FIGURE 4. Architecture of standard neural network for NitroY site identification.

1) STANDARD NEURAL NETWORK

Standard neural networks or Fully connected neural networks (FCNs) are classic architectures of deep neural networks. FCN is said to be completely connected as each neuron in preceding layer is connected to each neuron in the next layer of neural network. The FCN is intended to approximate the learning function f^*. This function f^* can be a classifier defined by $y = f^*(\alpha, x)$ and assigns a class label y to input x. The function of the FCN is to learn the best set of parameters α such that the mapping $y = f^*(\alpha, x)$ provides best possible approximation to f^* for predicting class label y for each input x. The FCN used for the Nitrotyrosine identification is shown in figure 4. It consists of two Dense layers consisting of twelve and eight rectified linear neurons (relu) respectively. The output layer is based on a single Sigmoid neuron to perform binary classification. The FCN architecture is illustrated in Table 3. To reduce negative logarithmic loss between actual and predicted class labels, this model was optimized using stochastic gradient descent (SGD) with a learning rate of 0.05. For training the FCN, only the training set was used, which was further divided into trainset and validation set with 90/10 partition ratio. FCN and other DNNs, used in this study, were never allowed to see the test set to ascertain the generalization capability of resulting NitroY identification models. Once trained, the predictive model was independently tested on the test set, and performance was evaluated using standard performance evaluation metrics.

2) RECURRENT NEURAL NETWORKS

The inherent weakness in conventional DNNs is the lack of sharing the weights learned by individual neurons, resulting in failure to identify similar patterns occurring at different positions of sequences [54]. RNN surmount this limitation by using a looping mechanism with time steps [55]. RNNs perform computations on a series of vectors $x_1, \ldots, x_n$ using a recurrence of the from $a_t = f_\alpha(\phi_{t-1}, x_t)$ where f is an activation function, ϕ is a collection of hyperparameters used at each phase t and x_t is input at timestep t.

This research utilizes three different RNN unit types to develop candidate models for NitroY site identification, which include simple RNN units, gated recurrent units (GRU) and long-short term memory unit (LSTM). The architecture shared by these three RNNs is shown in figure 5 where the green circles show RNN units used in the network, and the red squares show different timesteps of the sequence being classified. Three different RNNs are used in this study comprising of Simple units, GRU units and LSTM units respectively. In a simple RNN neuron, the parameters controlling the connections, from the input to the hidden layer, the horizontal connection between the activations and hidden layer to the output layer, are shared. Forward pass in a simple RNN neuron can be formulated by following set of equations:

$$a^t = g(W_a[a^{t-1}, X^t] + b_a) \quad (1)$$

$$y^t = f(W_y * a^t + b_y) \quad (2)$$

where <t> denotes the current time step, g expresses an activation function, X^t represents input at timestep t, b_a describes the bias, a^t is activation output at timestep t and W_a denotes cumulative weights. This activation a^t can be used to calculate the predictions y_t at time t if desired. The architecture of the RNN model with Simple RNN cells is shown in Table 4. This model makes use of an embedding layer to project each amino acid scalar in a vector space $\mathbb{R}^{15}$ which converts the semantic relationships of amino acids, prevalent in sequences, to geometric relationships. These geometric relationships of sequence vectors are interpreted by following layers of DNN model to learn deep feature representations which in turn are appraised by output layer, consisting of a single sigmoid unit, to make predictions. In many applications, DNNs with simple RNN neurons show promising results, but these neurons are prone to vanishing gradients and show limited ability to learn long-term dependencies. To rectify this limitation of simple RNN neurons, research community has proffered many updated recurrent neuron architectures, including GRU proposed by Cho *et al.* [56] and the LSTM neurons by Hochreiter and Schmidhuber [57] to counter the problem of vanishing gradients and enable learning of long-term dependencies.

The GRU cells, proposed by Cho *et al.* [56], are superior to the Simple RNN cell in reducing vanishing gradient problems. In each stage, the GRU cell uses the storage variable $C^t = a^t$ and contains summary of all the samples passed through the cell. At each timestep t, the GRU unit considers overwriting the contents of C^t with a candidate value $\bar{C}t$. This C^t content overwriting is controlled by update gate Γ_u, which decides whether or not the contents will be overwritten. Forward pass in GRU neuron can be described as follows:

$$\bar{C}^t = tanh(W_c[\Gamma_r * C^t, X^t] + b_c)$$

$$\Gamma_r = \sigma(W_r[C^{t-1}, X^t] + b_r)$$

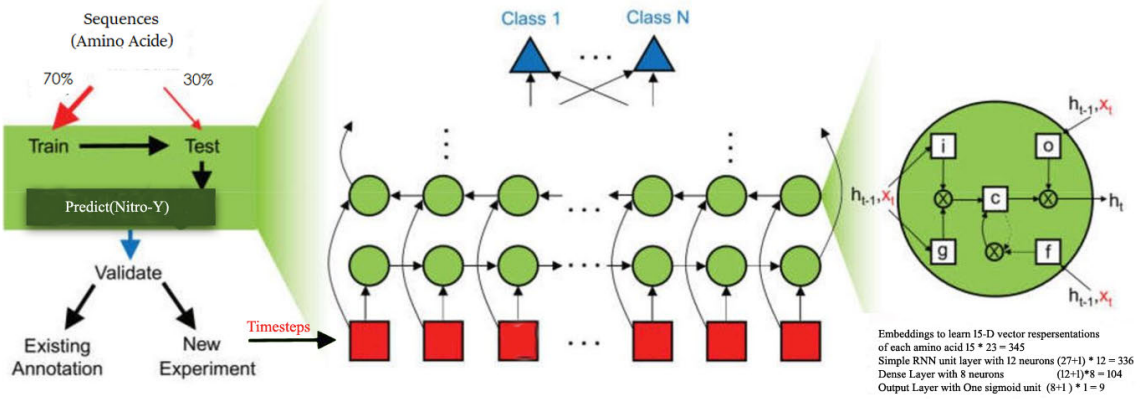


FIGURE 5. Architecture of RNNs for NitroY site identification.

TABLE 4. Simple RNN based model architecture for NitroY site identification.

Type of Neuron Layer	No. of Weights
Embeddings to learn 15-D vector representation of each amino acid	15*23 = 345
SimpleRNN layer with 12 neurons	(27+1) * 12 = 336
Dense layer with 8 neurons	(12+1) * 8 = 104
Output Layer with One sigmoid unit	(8+1) * 1 = 9

TABLE 5. GRU-RNN based model architecture for NitroY identification.

Type of Neuron Layer	No. of Weights
Embeddings to learn vector representation of Amino acids	15*23 = 345
GRU layer with 12 neurons	(80+1) * 12 = 972
Dense layer with 8 neurons	(12+1) * 8 = 104
Output Layer with One sigmoid unit	(8+1) * 1 = 9

$$\Gamma_u = \sigma(W_u[C^{t-1}, X^t] + b_u)$$
$$C^t = \Gamma_u * \bar{C}^t + (1 - \Gamma_u) * C^{t-1}$$
$$a^t = C^t$$

In the above set of equations, W_r, W_c and W_u denote respective weights, the corresponding bias terms are illustrated by b_r, b_c and b_u, X^t represents the input at timestep t, $\sigma$is the logistic regression function and a^t represents activations at time step t. For NitroY site identification, the RNN model architecture built with GRU is the same as the model based on SimpleRNN. Table 5 displays the architecture of GRU-based RNN model.

Hochreiter and Schmidhuber [57], presented LSTM with some modifications in RNN unit architecture, which is a more powerful generalization of GRU. Between GRU and LSTM neurons, the notable architectural differences are as follows:

1) For $\bar{C}^t$ computation, generic LSTM units do not use relevance gate Γ_r.
2) Instead of Update gate Γ_u, LSTM units use two different gates including Output gate Γ_o and Forget gate Γ_f. Output gate monitors the exposure of the memory cell contents C^t to compute activation outputs of LSTM unit for other hidden units in the network. Forget gate manages the amount of overwrite on C^{t-1} to achieve C^t i.e. how much memory cell content needs to be forgotten for memory cell.
3) In LSTM, the contents of the memory cell may not be equal to the activation a^t which is contrary to GRU architecture.

Except for LSTM units in recurrent layer, the RNN model built with the LSTM neurons has the same architecture as that of simple RNN and GRU models. The architecture of the NitroY site identification model developed using LSTM neurons based RNN is shown in Table 6.

3) CONVOLUTIONAL NEURAL NETWORK

CNN is a neural network structure primarily designed to analyze data with complex spatial relationships like images or videos. CNN tries to learn filters that can transform input data into the right output prediction. In addition to its capacity for handling large amounts of data, CNN can build local connections to learn feature maps, share training parameters among connections and reduce dimensions using the sub-sampling operations. These characteristics help CNN to understand the spatial features of inputs despite their locality in the input data, a property known as location invariance. The architecture of the NitroY prediction model based on CNN is shown in figure 6. The suggested CNN-based model was developed

TABLE 6. LSTM-RNN based model architecture for NitroY identification.

Type of Neuron Layer	No. of Weights
Embeddings to learn vector representation of Amino acids	15*23 = 345
LSTM layer with 12 neurons	(111+1) * 12 = 1344
Dense layer with 8 neurons	(12+1) * 8 = 104
Output Layer with One sigmoid unit	(8+1) * 1 = 9

TABLE 7. CNN based model architecture for NitroY site identification.

Type of Neuron Layer	No. of Weights
Embedding Layer to convert numeric sequence into vector sequence	23 * 15 = 345
Convolution 1D with 10 filters of size 3 with relu unit	((3 * 15) + 1)* 10 = 460
Maxpooling 1D	Not Applicable
Dropout with 25% of probability	Not Applicable
Convolution 1D with 16 filters of size 3 with relu unit	((3 * 10) + 1)* 16)= 496
Maxpooling 1D	Not Applicable
GlobalAveragePooling1D	Not Applicable
Dense Layer with 8 relu units	(16+1) * 8 = 136
Output: Output layer with one sigmoid	(8+1)* 1 = 9

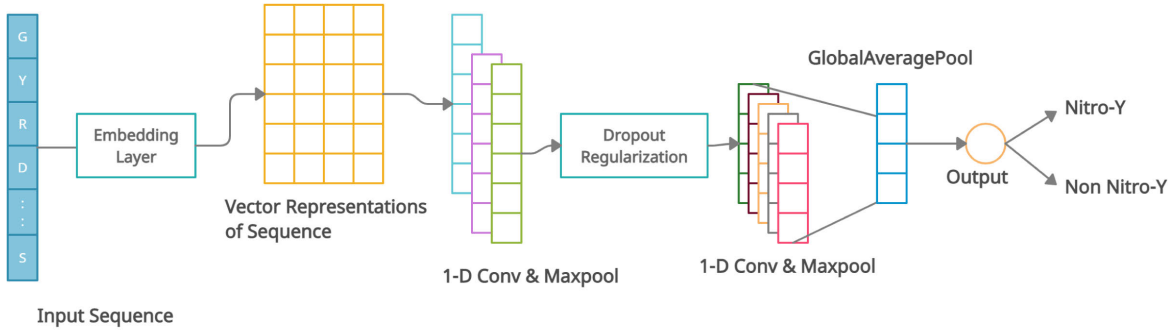


FIGURE 6. Architecture of CNN for NitroY site identification.

with an embedding layer, two convolution-maxpool blocks, a global averaging layer and an output layer of sigmoid neuron. Every sample of the peptide x with a length of $\rho = 41$ is encoded by the embedding layer in the form of tensor $X \in R^{\beta*\rho}$ where $\beta \in R^{15}$ is the representation vector of each amino acid residue in $\mathbb{R}^{15}$. One convolution layer and a maxpool layer are combined to form each convolution block. The convolution layer of both the blocks consisted of 10 and 16 1-D convolution units respectively. Every n-dimensional output sample is flattened into a 1-D array of scalars using GlobalAveragePooling operation which in turn is used by the output layer to predict the labels. A single sigmoid unit that performs binary classification is employed in the output layer to make predictions.

D. PERFORMANCE EVALUATION METHODOLOGY

This section explains the model performance evaluation metrics used in this study. Notable evaluation metrics used in this study include receiver operating characteristics curve (ROC) curve, precision-recall curve and point metrics, including mean average precision (mAP),Area under curve (AuC), accuracy, F1 measurements, and matthew's correlation coefficient (MCC). All these metrics are derived from confusion metrix which consist of four measures:

- **TP** True Positive: NitroY site predicted by model as NitroY site
- **FP** False Positive: Non-NitroY site predicted by model as NitroY site
- **FN** False negative: NitroY site predicted by model as non-NitroY site
- **TN** True Negative: Non-NitroY site predicted by model as non-NitroY site

A brief overview of these performance evaluation metrics is given in subsections below.

1) PRECISION-RECALL CURVE AND mAP

For the evaluation of prediction models, precision and recall are essential indicators. Precision measures the relevance of the positive outcomes predicted by the model while recall(sensitivity) measures the sensitivity of the model for positive samples. A high precision and recall rating imply that returned positive class predictions contain a high ratio of true positives (high Precision) while predicting the majority of positive class samples in the data set (High Recall).

Precision-Recall curve is achieved by plotting both of these metrics against each other and it evaluates the fraction of true positives among positive predictions [58]. In precision-recall space, the closer a score is to the perfect classification point (1,1), the better the predictor is and contrariwise.

Mean average precision (mAP) provides a single-digit summary of precision-recall curves, which is the area under the precision-recall curve. The higher the value of the mAP, the better the practical performance and vice versa.

2) ROC CURVE AND AUC

According to Lasko *et al.* [59], ROC is a popular measure in Bioinformatics studies to evaluate predictor models. The ROC Curve is a plot of False Positive Rate and True Positive Rate also known as recall. In a sense, the ROC curve illuminates the cost benefit analysis of the classifier under evaluation [60]. False Postive rate is defined as the ratio of false positive (FP) to total negative samples and measures the fraction of negative examples that are misclassified as positive. This is considered as the cost because any further action taken on FP result, considering it a positive prediction, is wasted. True positive rate, which is measured as the fraction of positive examples that are correctly predicted, can be considered as the benefit because the correct positive predictions, done by classifier, help solve the problem being investigated. Some important points in the ROC space are (0,0), (1,1) and (0,1). The lowest point on the left (0,0) represents models that do not predict positive samples. The contrasting strategy, represented by a point (1,1), is to unconditionally classify each positive sample. The point (0,1) expresses the perfect classification with a false positive rate of 0 and a true positive rate score of 1. For predictors, the closer the curve is to the point (0,1) in ROC space, the better the performance of the corresponding predictor and vice versa.

Sometimes it is desirable to summarize the ROC curve insights of a model to a single scalar value that shows the performance of the model. The area under an ROC curve, called the AUC, is one of these popular methods. Not only does AuC minimizes the ROC curve outcomes to a single value, it also illuminates statistical insights of the model's performance. AUC is equivalent to the probability that the classifier will rank a randomly chosen positive sample higher than a randomly chosen negative instance. Additionally AuC is also equivalent to the Wilcoxon test of ranks [60].

3) ACCURACY, F1 AND MCC SCORES

Accuracy, a popular classifier evaluation measure, highlights the fraction of results correctly classified by the model being evaluated. In situations, where an optimal combination of precision and recall is required in the form of single scalar value, F1-measure is also a popular alternative. Matthews correlation coefficient, on the other hand, is considered an effective solution overcoming the class imbalance issues prevalent in accuracy and other binary classification model evaluators. Originally developed by Matthews in 1975 for comparison of chemical structures [61], MCC was brought into limelight again by Baldi and colleagues [62] in 2000 as a standard performance metric for binary classification models with a natural extension to the multiclass case. The formulae of important measures are shown below:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$Recall(TPR) = \frac{TP}{TP + FN}$$

$$Precision(Prec) = \frac{TP}{TP + FP}$$

$$F1 - Score = \frac{2 * Prec * Recall}{Prec + Recall}$$

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

IV. RESULTS

This section explains the performance results of multiple DNN based predictors developed in this research to predict NitroY site location. All models were evaluated on test data which was not used during the predictor training phase. This was done to ensure the fairness of results and to evaluate the generalization capability of predictors being evaluated. An overview of the model evaluation parameters used in this study is given in the following subsection, which illuminates adequate discussion of the results of the evaluation. To ensure fairness, all evaluation results come from independent test samples that were not used in the training phase of DNN based models.

A. PRECISION-RECALL CURVE AND MEAN AVERAGE PRECISION

Figure 7 shows the precision-recall curve of the candidate deep models for predicting the NitroY PTM sites. As shown in Figure 7, CNN model's curve is closest to the perfect prediction point (1,1) in the precision-recall space as compared to the scores of other predictors. Although the CNN curve dips drastically at point (0.1, y) the overall performance of the curve depicts better performance. The nearest contender is FCN which lags behind the CNN from point (0.9, y) to (0.3, y) in precision-recall space but surpasses the CNN model from (0.3, y) to (0,y). All the other DNN based predictors performed poorly as compared to CNN and FCN as depicted by their respective curves in Figure 7.

The mean average precision values for all developed models are shown in the legend section of Figure 7. The CNN based model showed best score of 0.89 followed closely by the FCN with value of 0.87. We believe the reason for better map score of CNN is due to its capability to identify the sequence motifs in different positions of the feature maps learnt by convolution neurons. CNN can discover localized features which repeat themselves all over the input as their weights are shared among all locations of the input, preserving spatial locality. The latent representations learned by CNN are more sensitive to transitive relationships of input

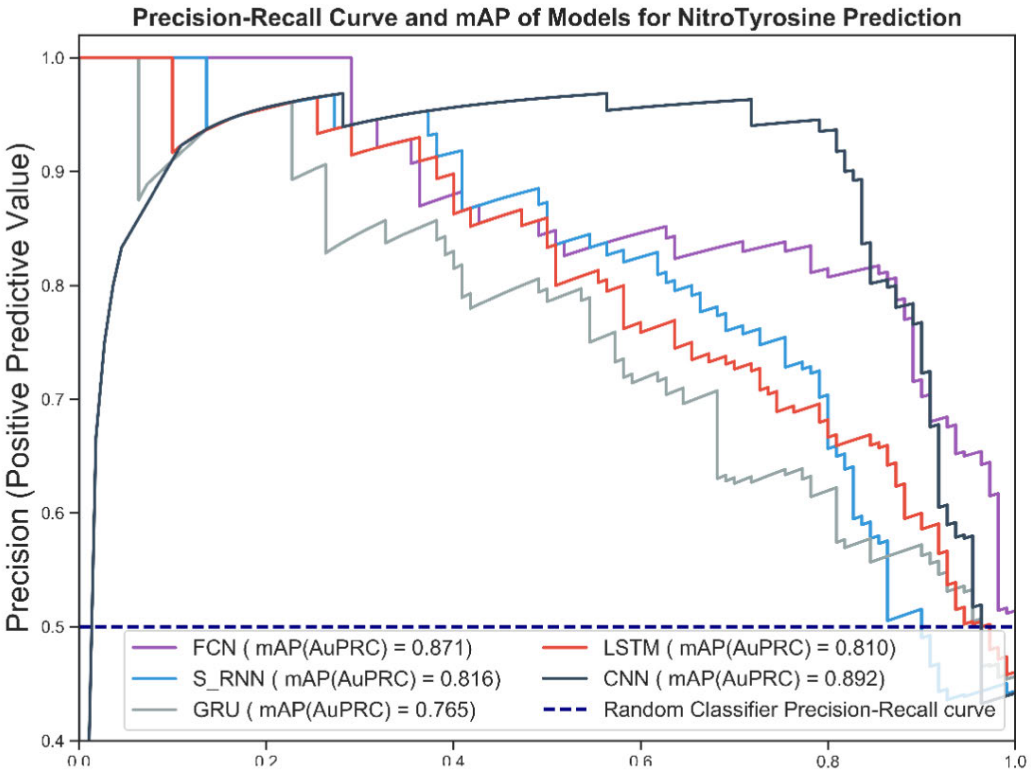


FIGURE 7. DNN based NitroY prediction models' precision-recall curves and mAP scores.

and help CNN to uncover high level relationships between input features which would otherwise be ignored by other classifiers [63]. FCN, on the other hand, does not possess this capability due to inability to share training parameters between its neurons. The least scores were shown by GRU-based model which scored 0.76 for NitroY PTM site predictions. All DNN based predictors achieved more than 75% scores.

B. RECEIVER OPERATING CHARACTERISTICS AND AREA UNDER ROC

Figure 8 shows the ROC curve of the Nitrotyrosine predictors. The results of the ROC curve corroborate the assessments demonstrated by precision-recall curve. Here too the CNN model and FCN model performed neck to neck while other models were overshadowed by the duo. Legend section of Figure 8 shows the AuC values for the models developed in this study. The CNN and FCN based prediction model showed the highest AuC value of 0.91 and 0.90 respectively while the model built with GRU obtained the least rating of 0.79. The scores achieved by remaining three models were distributed between these extreme values. The reader can notice that all the results shown in Figure 8 are above AuC score of 0.79. The comparison of the overall diagnostic accuracy of two models is frequently addressed by comparing the resulting paired AuCs using Delong's method [64] of non-parametric comparison of two or more RoC curves. We used the fast implementation of Delong's method by Sun and Xu [65] to calculate the p-values by comparing each AuC with our best performing CNN based model. Delong p-value scores of CNN and FCN are $9.78e-14$ and $8.3e-7$ respectively. We also constructed the 95% Confidence interval of AuC for CNN and FCN based predictors developed in this study. The 95% confidence Intervals for CNN and FCN are $[0.874-0.93]$ and $[0.871-0.924]$ respectively for AuC scores.

C. ACCURACY, F1-SCORE, AND MATTHEW's CORRELATION

For independent testing, Figure 9 shows the results of the accuracy scores for the NitroY prediction models developed in this study. As shown in Figure 9, the results are consistent with previously discussed evaluation metrics and CNN and FCN based deep models showed an accuracy value above 85 percent while least accuracy value of 73 percent is demonstrated by GRU based RNN model.

In situations, where an optimal combination of precision and recall is required in the form of single scalar value, F1-measure is also a popular alternative. Figure 9 indicates the predictive NitroTyrosine model's F1 value which validates the earlier evaluation demonstrated by AuC and mAP scores. The CNN model achieved optimal score of 85 percent

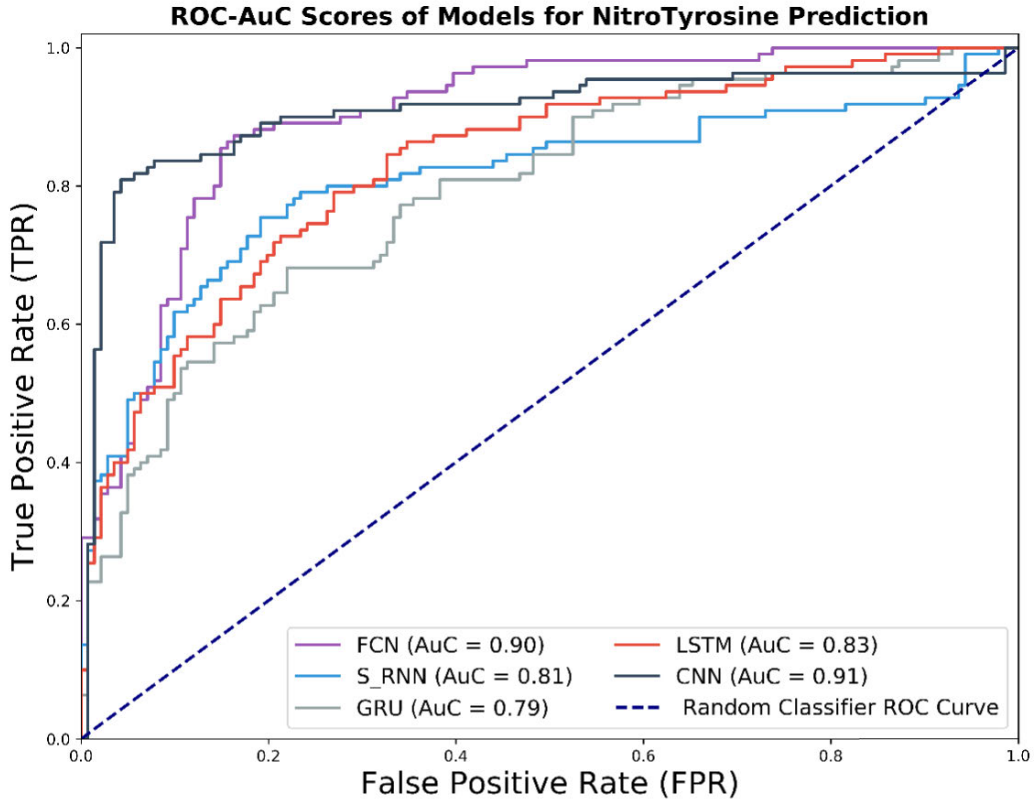


FIGURE 8. DNN based NitroY prediction models' ROC curves and AUC scores.

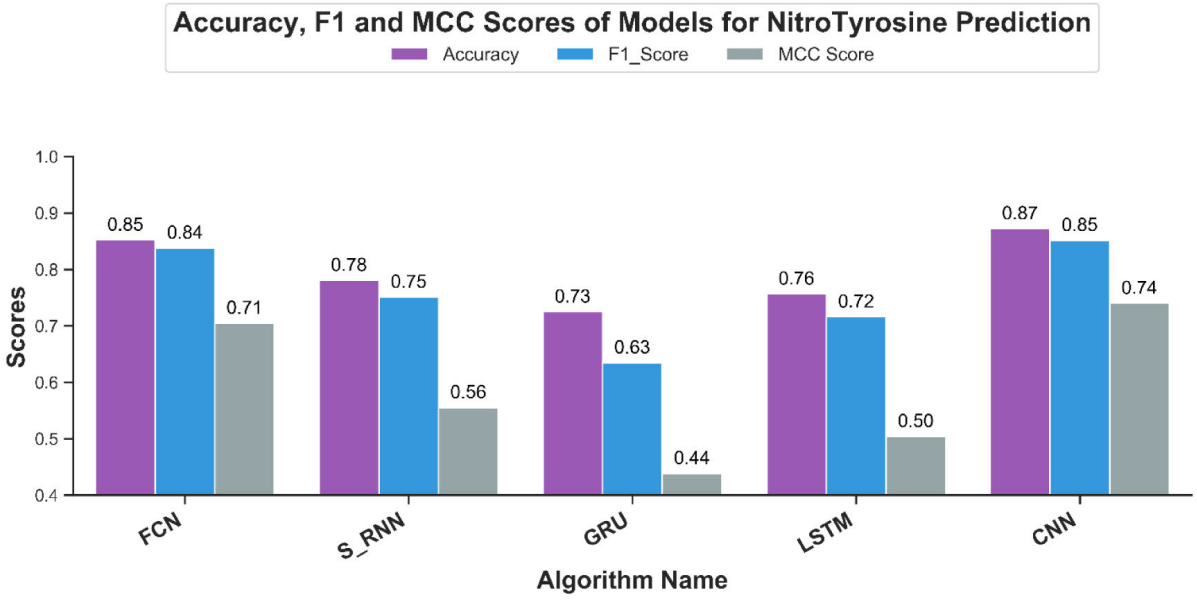


FIGURE 9. Accuracy, F1 and MCC scores of Deep NitroY site identification models.

TABLE 8. Comparison of the proposed approach with related contributions on independent test set.

Sr#	Predictor	Sensitivity	Specificity	Acc	AUC	MCC
1	Proposed CNN predictor	0.836	0.90	87.2%	0.91	0.74
2	DeepNitro [48]	Not Reported	Not Reported	Not Reported	0.67	Not Reported
3	iNitroTyr [31]	0.82	0.796	84.52%	Not Reported	0.49
4	GPS-YNO2 (Med) [30]	0.40	0.791	79.6%	Not Reported	0.21
5	NTyroSite [33]	0.60	0.801	78.2%	0.88	0.27
6	pNitro-Tyr-PseAAC [32]	0.845	0.89	88%	Not Reported	0.627
7	PredNTS [49]	0.522	0.809	76.1%	Not Reported	0.286

and the second position was attained by the FCN model with F1 score of 84 percent. The model based on GRU based RNN gave a weak rating of 63 percent.

Matthews correlation coefficient (MCC) is a more accurate statistical metric that generates a high score only if good results were obtained in the prediction in all four groups of the confusion matrix [66]. MCC-scores of all DNN models are shown in Figure 9. The reader can confirm that the CNN based model showed best performance with an MCC value of 0.74 and FCN is not far behind with an MCC value of 0.71. As depicted by previous evaluation metrics, GRU based predictor turned out to be the least suitable model for NitroY prediction with MCC score of 0.44. These result show that CNN-based predictor is the finest of all DNN based models closely followed by FCN based predictor, developed in this study for identification of NitroTyrosine sites.

V. DISCUSSION

A. COMPARISON WITH LITERATURE

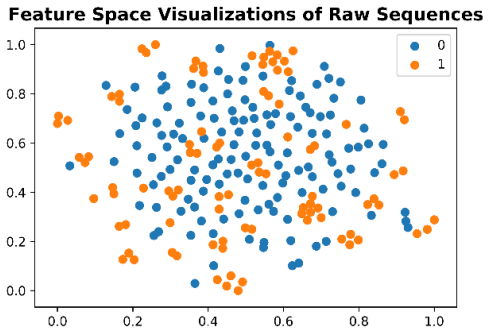
This section provides the comparison of proposed CNN based predictor on independent test set with notable contributions from literature introduced in section II. The comparison is made up of notable contribution for NitroY prediction including DeepNitro [48], iNitroTyr [31], GPS-YNO2 [30], NTyroSite [33], pNitro-Tyr-PseAAC [32] and PredNTS [49]. The comparison results are shown in Table 8. As can be seen in Table 8, the evaluation scores of proposed CNN model are shown in 1st row. The reported evaluation scores of DeepNitro, iNitroTyr and GPS-YNO2 are shown at row 3,4 and 5 respectively. From all systems mentioned in Table 8, pNitro-Tyr-PseAAC system showed comparable results to our proposed CNN predictor. Although the sensitivity and accuracy scores of pNitro-Tyr are slightly higher then the proposed CNN model, the MCC score, which is shown to be a better performance evaluator of predictors in non-balanced class problems [66], is far lower than the proposed CNN model. The higher score of MCC for CNN model shows the higher performance of proposed approach for NitroTyrosine site identification. It is relevant to mention that all NitroY predictors, shown in Table 8, rely on the quality of the features extracted or selected by the feature extraction algorithms. Models proposed in current study are different because they accept raw PSeAAC sequences as input and do not rely on the quality of features to perform better predictions. Additionally, the results shown by our optimal model are comparable to the pNitro-Tyr-PseAAC, which has shown the optimal scores as compared to other conventional feature-based predictors.

B. DEEP FEATURE SPACE VISUALIZATIONS

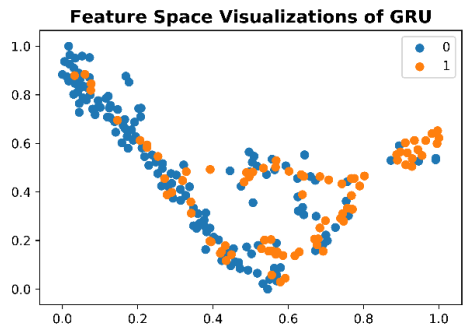
To understand the deep feature representations, learned by the non-linear transformation of iNitroY-Deep models, visualization of feature space serves as an important tool. For creating visualizations, we computed the output from penultimate layer of each trained model for test set sequences and projected this output to 2-D space using T-SNE, proposed by Maaten and Hinton [67]. T-SNE uses a non-linear statistical approach to project data from higher dimensions to lower dimensions. For t-SNE plot, perplexity is an important hyperparmeter which is related to the number of nearest neighbors that are used in other manifold learning algorithms. Larger datasets usually require a larger perplexity. Since our Dataset contains 834 samples in total with maximum 41 dimensions for raw sequences and 8 dimensions for deep representations, the recommended perplexity range in scikit-learn is 0-50. We used default perplexity value of '30', used PCA initialization and n_iter =1000 for developing the plots.

This 2-D data was plotted based on class labels to understand the distribution of sequences belonging to both classes. For plotting the visualizations, we used matplotlib and seaborn package of python. Figure 10 illustrates the feature space visualizations of PseAAC sequences and three deep feature representations. Feature space visualization for raw PseAAC sequences is shown in figure 10a. As visible in the figure, positive and negative sequences are jumbled up and no clear separation is available which means any classifier using this representation will have a hard time separating the sequences from both classes to perform predictions. Figures 10b, 10c and 10d illustrates the effect of non-linear transformations of three DNNs, used in this study, to separate both classes in respective feature space for achieving better predictions. The figures included in the manuscript belong to one least performing model i.e. GRU-RNN and two optimal models i.e. FCN and CNN models.

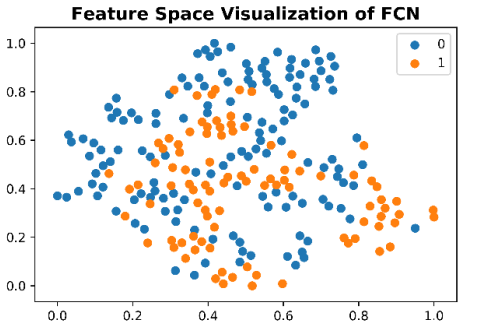
As illuminated in figure 10b, which shows the feature representation learned by GRU-RNN model, the reader can see that this model was not successful enough to separate both classes in the learned representation before passing the features to output layer and this resulted in relatively poor evaluation scores of GRU-RNN based predictor. The most successful feature representation is achieved by CNN based model, as illustrated in figure 10d. The reader can verify



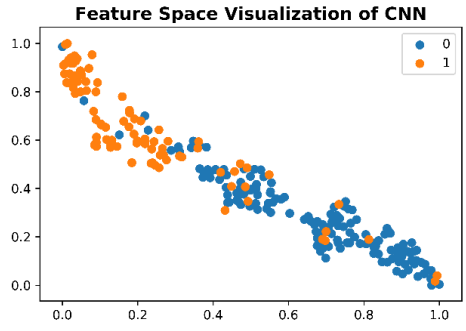
(a) Visualization of Raw NitroY sequence Space



(b) Visualization of GRU feature space



(c) Visualization of FCN feature space



(d) Visualization of CNN feature space

FIGURE 10. Feature space visualizations of deep representations for positive and negative NitroY samples.

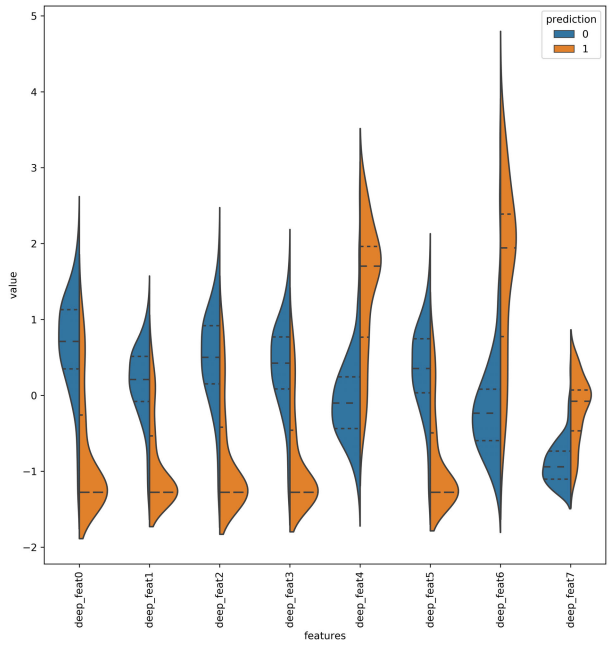


FIGURE 11. Violin plot showing the data distribution of deep features learned by CNN model.

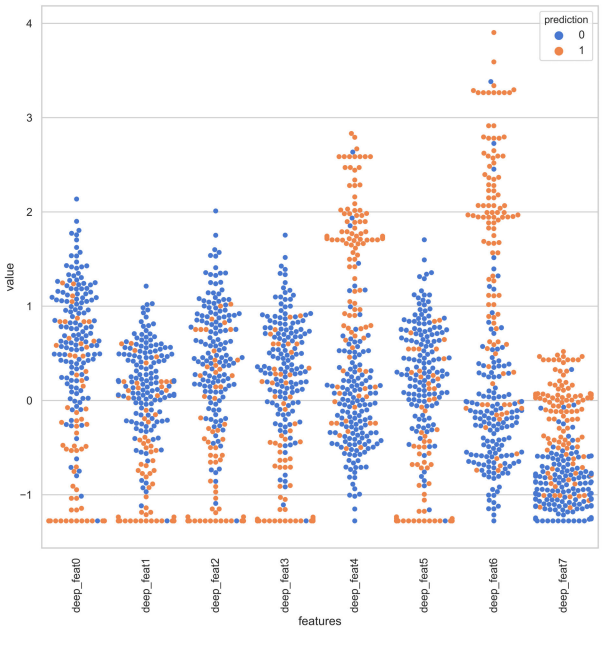
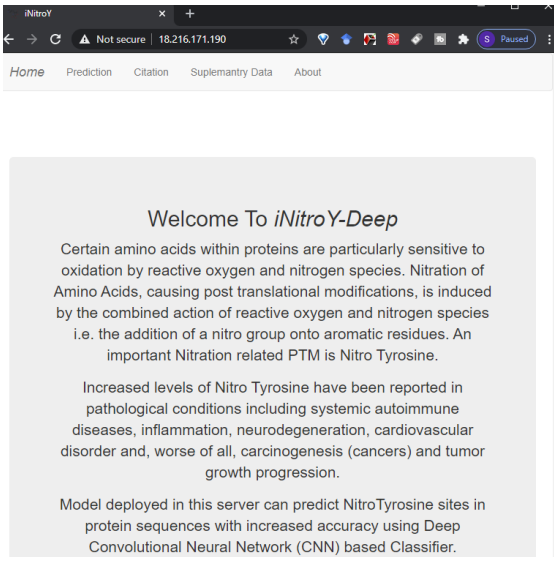
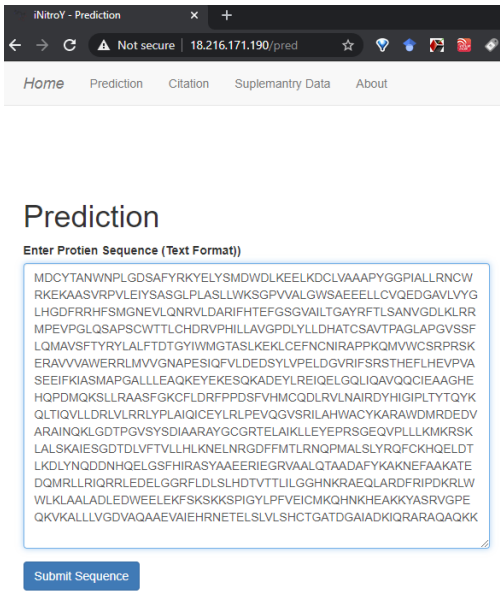


FIGURE 12. Swarm-plot showing the data distribution of positive and negative samples in CNN deep representation.

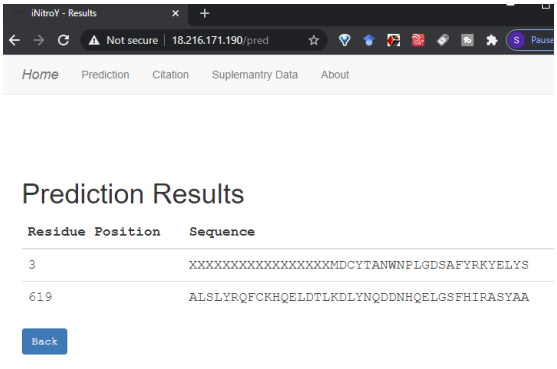
from the figure that this representation is most successful in separating the samples belonging to different classes with minimal overlap. This means any classifier consuming this representation will be able to segregate both classes with less effort and achieve better predictions. The data distribution of positive and negative samples in CNN representation



(a) iNitroY-Deep Server Homepage



(b) iNitroY-Deep submission of peptide sequence



(c) NitroTyrosine site prediction results for submitted sequence

FIGURE 13. **iNitroY-Deep Webserver functionalities for identification of NitroY sites.**

is shown as violin plot and swarm plot in Figure 11 and Figure 12. As can be seen in aforementioned figures, the CNN model was able to learn the representation in which the positive and negative samples are sufficiently separated from each other enabling better NitroY site identification by output layer. The violin plot shown in Figure 11 further corroborates this conclusion by showing minimal overlap between the positive and negative samples in data distributions of different deep features of best performing CNN based model. This fact is also illuminated by the evaluation results discussed in earlier sections.

An additional benefit of our approach is automatic learning of feature representation using stochastic gradient descent. This approach removes the need to use expensive feature engineering process. In addition, the proposed DNN based predictors developed in this contribution demonstrate only the initial step towards employing deep learning for NitroTyrosine site identification and further study will draw on the research described in this study to devise better DNN predictors for the same.

C. MODEL DEPLOYMENT AS WEB SERVICE

The final phase for Chou's 5-step rule is the deployment of the predictor in the form of a web application for making it available to the general research community. To achieve this goal, a web application was developed in this study utilizing the best performing CNN based iNitroY-Deep model. The web application can accept a peptide sample in the form of string and return the predicted Tyrosine sites which are highly likely to be nitrated. Homepage of iNitroY-Deep webserver is shown in figure 13a while figure 13b highlights the peptide sequence submission process for computing NitroY sites. Figure 12 calls attention to the result page showing the predicted tyrosine sites likely to be nitrated and the corresponding ρ length sequence of residues. The model is temporarily accessible at http://3.15.230.173/. We believe that out humble effort to improve the predictability of Tyrosine nitration sites will be of service to research community and help augment carcinogenesis studies for its prognosis.

VI. CONCLUSION

In this study, we proposed an improved, effective, and less cumbersome approach, based on deep neural networks, for identification of NitroTyrosine sites in proteins. The proposed approach makes use of Chou's Pseudo Amino Acid Composition with deep neural networks to identify Tyrosine sites susceptible to nitration. We employed well-known DNNs including Standard neural network (FCN), three variants of recurrent neural networks (RNNs) and convolutional neural network (CNN) for both the tasks of learning a feature representation of peptide sequences and performing the classifications. From all DNN based predictors in this study, the CNN based predictor reached up to the highest performance scores evaluated using well-known model evaluation metrics. The model achieved 87.2% accuracy, 0.74 Matthew correlation

coefficient (MCC) score, 0.91 AuC score and 0.90 specificity score for independent test set. The comparisons of proposed CNN based predictor with notable research contributions were performed which shows the efficacy of proposed predictor. Based on aforementioned evaluation and comparison results, it is concluded that the proposed predictor will help the research community to efficiently and accurately identify NitroTyrosine sites and enable better understanding of its role in pathological conditions including Carcinogenesis, tumor growth progression, systemic autoimmune diseases, neurodegeneration and cardiovascular disorders. The model is accessible at http://3.15.230.173/

VII. LIMITATION AND FUTURE RESEARCH

Like every experimental research work, our study also suffers from some limitations. Primary limitation of this study stems from the fact that Deep neural networks mostly work like a black box and little information is available regarding the decision-making process of various neurons working together to make predictions. Although research community is working on Explainable Artificial Intelligence (XAI) but most research contributions in (XAI) including Grad-Cam and Activation maps are targeted towards computer vision and have very limited application on sequence based predictors. Future research in this area could overcome XAI limitation discussed above. In the future we seek to develop and incorporate XAI for sequence based protein predictors to enhance their explainability.

REFERENCES

[1] E. Furuya and K. Uyeda, "Regulation of phosphofructokinase by a new mechanism. An activation factor binding to phosphorylated enzyme," *J. Biol. Chem.*, vol. 255, no. 24, pp. 11656–11659, 1980.

[2] J. M. Souza, G. Peluffo, and R. Radi, "Protein tyrosine nitration—Functional alteration or just a biomarker?" *Free Radical Biol. Med.*, vol. 45, no. 4, pp. 357–366, Aug. 2008. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0891584908002128

[3] G. Peluffo and R. Radi, "Biochemistry of protein tyrosine nitration in cardiovascular pathology," *Cardiovascular Res.*, vol. 75, no. 2, pp. 291–302, Jul. 2007.

[4] M. R. Reynolds, R. W. Berry, and L. I. Binder, "Nitration in neurodegeneration: Deciphering the 'Hows' 'nYs,'" *Biochemistry*, vol. 46, no. 25, pp. 7325–7336, 2007.

[5] C. J. van Dalen, C. C. Winterbourn, R. Senthilmohan, and A. J. Kettle, "Nitrite as a substrate and inhibitor of myeloperoxidase: Implications for nitration and hypochlorous acid production at sites of inflammation," *J. Biol. Chem.*, vol. 275, no. 16, pp. 11638–11644, 2000.

[6] I. Saimanen, D. Rahkola, V. Kuosmanen, J. Kärkkäinen, T. Selander, A. Holopainen, S. Aspinen, and M. Eskelinen, "Nitrotyrosine (NT), a nitrosative stress biomarker, plasma concentrations in gallstone disease and cancer patients," *Anticancer Res.*, vol. 39, no. 2, pp. 809–814, Feb. 2019.

[7] H. Ahsan, "3-nitrotyrosine: A biomarker of nitrogen free radical species modified proteins in systemic autoimmunogenic conditions," *Hum. Immunol.*, vol. 74, no. 10, pp. 1392–1399, Oct. 2013.

[8] H.-H. Huang, L.-Y. Chen, K.-Y. Chen, Y.-C. Lee, C.-Y. Tsai, and C.-Y. Chen, "Increased monocyte chemoattractant protein-1 and nitrotyrosine are associated with increased body weight in patients with rheumatoid arthritis after etanercept therapy," *Neuropeptides*, vol. 84, Dec. 2020, Art. no. 102100. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0143417920301189

[9] M. F. Khan and G. Wang, "Environmental agents, oxidative stress and autoimmunity," *Current Opinion Toxicol.*, vol. 7, pp. 22–27, Feb. 2018.

[10] M. A. Khan, K. Alam, M. Zafaryab, and M. M. A. Rizvi, "Peroxynitrite-modified histone as a pathophysiological biomarker in autoimmune diseases," *Biochimie*, vol. 140, pp. 1–9, Sep. 2017. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0300908417301505

[11] M. A. Pasquali, B. L. Harlow, C. N. Soares, M. W. Otto, L. S. Cohen, L. Minuzzi, D. P. Gelain, J. C. F. Moreira, and B. N. Frey, "A longitudinal study of neurotrophic, oxidative, and inflammatory markers in first-onset depression in midlife women," *Eur. Arch. Psychiatry Clin. Neurosci.*, vol. 268, no. 8, pp. 771–781, Dec. 2018.

[12] M. Bandookwala and P. Sengupta, "3-nitrotyrosine: A versatile oxidative stress biomarker for major neurodegenerative diseases," *Int. J. Neurosci.*, vol. 130, no. 10, pp. 1047–1062, Oct. 2020.

[13] M. H. Shishehbor, "Association of nitrotyrosine levels with cardiovascular disease and modulation by statin therapy," *JAMA*, vol. 289, no. 13, p. 1675, Apr. 2003. [Online]. Available: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.289.13.1675

[14] S. Korde Choudhari, M. Chaudhary, S. Bagde, A. R. Gadbail, and V. Joshi, "Nitric oxide and cancer: A review," *World J. Surgical Oncol.*, vol. 11, no. 1, p. 118, Dec. 2013. [Online]. Available: https://wjso.biomedcentral.com/articles/10.1186/1477-7819-11-118

[15] E. Gochman, J. Mahajna, P. Shenzer, A. Dahan, A. Blatt, R. Elyakim, and A. Z. Reznick, "The expression of iNOS and nitrotyrosine in colitis and colon cancer in humans," *Acta Histochemica*, vol. 114, no. 8, pp. 827–835, Dec. 2012. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0065128112000244

[16] A. Allameh, Y. Rasmi, S. Nasseri-Moghaddam, S. M. Tavangar, R. Sharifi, and M. Sadreddini, "Immunohistochemical analysis of selected molecular markers in esophagus precancerous, adenocarcinoma and squamous cell carcinoma in iranian subjects," *Cancer Epidemiol.*, vol. 33, no. 1, pp. 79–84, Jul. 2009. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1877782109000526

[17] M. Samoszuk, M.-L. Brennan, V. To, L. Leonor, L. Zheng, X. Fu, and S. L. Hazen, "Association between nitrotyrosine levels and microvascular density in human breast cancer," *Breast Cancer Res. Treatment*, vol. 74, no. 3, pp. 271–278, Aug. 2002.

[18] S. Kondo, S. Toyokuni, T. Tsuruyama, M. Ozeki, T. Tachibana, M. Echizenya, H. Hiai, H. Onodera, and M. Imamura, "Peroxynitrite-mediated stress is associated with proliferation of human metastatic colorectal carcinoma in the liver," *Cancer Lett.*, vol. 179, no. 1, pp. 87–93, May 2002. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S030438350100859X

[19] M. Jaiswal, N. F. LaRusso, L. J. Burgart, and G. J. Gores, "Inflammatory cytokines induce DNA damage and inhibit DNA repair in cholangiocarcinoma cells by a nitric oxide-dependent mechanism," *Cancer Res.*, vol. 60, no. 1, pp. 184–190, 2000.

[20] H. Kato, T. Miyazaki, M. Yoshikawa, M. Nakajima, Y. Fukai, K. Tajima, N. Masuda, S. Tsutsumi, K. Tsukada, T. Nakajima, and H. Kuwano, "Nitrotyrosine in esophageal squamous cell carcinoma and relevance to p53 expression," *Cancer Lett.*, vol. 153, nos. 1–2, pp. 121–127, May 2000. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S030438350000358X

[21] T. Goto, K. Haruma, Y. Kitadai, M. Ito, M. Yoshihara, K. Sumii, N. Hayakawa, and G. Kajiyama, "Enhanced expression of inducible nitric oxide synthase and nitrotyrosine in gastric mucosa of gastric cancer patients," *Clin. Cancer Res.*, vol. 5, no. 6, pp. 1411–1415, 1999.

[22] T. R. Larsen, N. Bache, J. B. Gramsbergen, and P. Roepstorff, "Identification of nitrotyrosine containing peptides using combined fractional diagonal chromatography (COFRADIC) and off-line Nano-LC-MALDI," *J. Amer. Soc. Mass Spectrometry*, vol. 22, no. 6, pp. 989–996, Jun. 2011. [Online]. Available: https://pubs.acs.org/doi/10.1021/jasms.8b04051

[23] X. Zhan and D. M. Desiderio, "Mass spectroscopy measurements of nitrotyrosine-containing proteins," in *Handbook of Measurement in Science and Engineering*, vol. 3. Hoboken, NJ, USA: Wiley, 2016, pp. 2431–2473.

[24] S. Naseer, W. Hussain, Y. D. Khan, and N. Rasool, "IPhosS(Deep)-PseAAC: Identify phosphoserine sites in proteins using deep learning on general pseudo amino acid compositions via modified 5-Steps rule," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Nov. 26, 2020. [Online]. Available: https://ieeexplore.ieee.org/document/9272694/

[25] S. Naseer, W. Hussain, Y. D. Khan, and N. Rasool, "NPalmitoylDeep-PseAAC: A predictor for N-palmitoylation sites in proteins using deep representations of proteins and PseAAC via modified 5-steps rule," *Current Bioinf.*, vol. 15, no. 2, pp. 294–305, Jun. 2020. [Online]. Available: http://www.eurekaselect.com/182539/article

[26] W. Hussain, Y. D. Khan, N. Rasool, S. A. Khan, and K.-C. Chou, "SPalmitoylC-PseAAC: A sequence-based model developed via Chou's 5-steps rule and general PseAAC for identifying S-palmitoylation sites in proteins," *Anal. Biochem.*, vol. 568, pp. 14–23, Mar. 2019.

[27] Y. D. Khan, N. Rasool, W. Hussain, S. A. Khan, and K.-C. Chou, "IPhosT-PseAAC: Identify phosphothreonine sites by incorporating sequence statistical moments into PseAAC," *Anal. Biochem.*, vol. 550, pp. 109–116, Jun. 2018.

[28] W. Chen, P. Feng, H. Ding, H. Lin, and K.-C. Chou, "IRNA-methyl: Identifying N6-methyladenosine sites using pseudo nucleotide composition," *Anal. Biochem.*, vol. 490, pp. 26–33, Dec. 2015.

[29] H. Lv, F.-Y. Dao, Z.-X. Guan, H. Yang, Y.-W. Li, and H. Lin, "Deep-Kcr: Accurate detection of lysine crotonylation sites using deep learning method," *Briefings Bioinf.*, vol. 21, no. 5, pp. 1842–1857, Oct. 2020.

[30] Z. Liu, J. Cao, Q. Ma, X. Gao, J. Ren, and Y. Xue, "GPS-YNO2: Computational prediction of tyrosine nitration sites in proteins," *Mol. Biosyst.*, vol. 7, no. 4, p. 1197, 2011. [Online]. Available: http://xlink.rsc.org/?DOI=c0mb00279h

[31] Y. Xu, X. Wen, L.-S. Wen, L.-Y. Wu, N.-Y. Deng, and K.-C. Chou, "INitrotyr: Prediction of nitrotyrosine sites in proteins with general pseudo amino acid composition," *PLoS ONE*, vol. 9, no. 8, Aug. 2014, Art. no. e105018. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0105018

[32] A. W. Ghauri, Y. D. Khan, N. Rasool, S. A. Khan, and K.-C. Chou, "pNitro-Tyr-PseAAC: Predict nitrotyrosine sites in proteins by incorporating five features into Chou's general PseAAC," *Current Pharmaceutical Des.*, vol. 24, no. 34, pp. 4034–4043, Jan. 2019. [Online]. Available: https://www.eurekaselect.com/167760/article

[33] M. Hasan, M. Khatun, M. Mollah, C. Yong, and G. Dianjing, "NTyroSite: Computational identification of protein nitrotyrosine sites using sequence evolutionary features," *Molecules*, vol. 23, no. 7, p. 1667, Jul. 2018. [Online]. Available: http://www.mdpi.com/1420-3049/23/7/1667

[34] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, May 2015. [Online]. Available: http://dx.doi.org/10.1038/nature14539

[35] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105. [Online]. Available: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-conv olutional-neural-networks.pdf

[36] I. Sutskever, O. Vinyals, and Q. V. Le, "Sequence to sequence learning with neural networks," , 2014, pp. 3104–3112, *arXiv:1409.3215*. [Online]. Available: https://arxiv.org/abs/1409.3215

[37] S. Naseer and Y. Saleem, "Enhanced network intrusion detection using deep convolutional neural networks," *KSII Trans. Internet Inf. Syst.*, vol. 12, no. 10, pp. 5159–5178, Oct. 2018. [Online]. Available: http://itiis.org/digital-library/manuscript/2163

[38] S. Naseer, R. F. Ali, P. D. D. Dominic, and Y. Saleem, "Learning representations of network traffic using deep neural networks for network anomaly detection: A perspective towards oil and gas IT infrastructures," *Symmetry*, vol. 12, no. 11, p. 1882, 2020.

[39] S. Naseer, W. Hussain, Y. D. Khan, and N. Rasool, "Sequence-based identification of arginine amidation sites in proteins using deep representations of proteins and PseAAC," *Current Bioinf.*, vol. 15, no. 8, pp. 937–948, Jan. 2021. [Online]. Available: https://www.eurekaselect.com/178780/article

[40] S. Naseer, W. Hussain, Y. D. Khan, and N. Rasool, "Optimization of serine phosphorylation prediction in proteins by comparing human engineered features and deep representations," *Anal. Biochem.*, vol. 615, Feb. 2021, Art. no. 114069. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S0003269720306011

[41] S. Naseer, R. F. Ali, A. Muneer, and S. M. Fati, "IAmideV-deep: Valine amidation site prediction in proteins using deep learning and pseudo amino acid compositions," *Symmetry*, vol. 13, no. 4, p. 560, Mar. 2021. [Online]. Available: https://www.mdpi.com/2073-8994/13/4/560

[42] K.-C. Chou, "Some remarks on protein attribute prediction and pseudo amino acid composition," *J. Theor. Biol.*, vol. 273, no. 1, pp. 236–247, Mar. 2011.

[43] K.-C. Chou, "Using subsite coupling to predict signal peptides," *Protein Eng.*, vol. 14, no. 2, pp. 75–79, 2001.

[44] D. Zhang, Z.-C. Xu, W. Su, Y.-H. Yang, H. Lv, H. Yang, and H. Lin, "ICarPS: A computational tool for identifying protein carbonylation sites by novel encoded features," *Bioinformatics*, vol. 37, no. 2, pp. 171–177, Apr. 2021.

[45] W.-R. Qiu, B.-Q. Sun, H. Tang, J. Huang, and H. Lin, "Identify and analysis crotonylation sites in histone by using support vector machines," *Artif. Intell. Med.*, vol. 83, pp. 75–81, Nov. 2017.

[46] Y.-W. Zhao, H.-Y. Lai, H. Tang, W. Chen, and H. Lin, "Prediction of phosphothreonine sites in human proteins by fusing different features," *Sci. Rep.*, vol. 6, no. 1, p. 34817, Dec. 2016.

[47] X. Cheng, X. Xiao, and K.-C. Chou, "PLoc-mHum: Predict subcellular localization of multi-location human proteins via general PseAAC to winnow out the crucial GO information," *Bioinformatics*, vol. 34, no. 9, pp. 1448–1456, May 2018.

[48] Y. Xie, X. Luo, Y. Li, L. Chen, W. Ma, J. Huang, J. Cui, Y. Zhao, Y. Xue, Z. Zuo, and J. Ren, "DeepNitro: Prediction of protein nitration and nitrosylation sites by deep learning," *Genomics, Proteomics Bioinf.*, vol. 16, no. 4, pp. 294–306, Aug. 2018. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S1672022918303474

[49] A. N. Nilamyani, F. N. Auliah, M. A. Moni, W. Shoombuatong, M. M. Hasan, and H. Kurata, "PredNTS: Improved and robust prediction of nitrotyrosine sites by integrating multiple sequence features," *Int. J. Mol. Sci.*, vol. 22, no. 5, p. 2704, Mar. 2021.

[50] T. UniProt Consortium, "UniProt: A worldwide hub of protein knowledge," *Nucleic Acids Res.*, vol. 47, no. D1, pp. D506–D515, Jan. 2019. [Online]. Available: https://academic.oup.com/nar/article/47/D1/D506/5160987

[51] R. C. Edgar, "Search and clustering orders of magnitude faster than BLAST," *Bioinformatics*, vol. 26, no. 19, pp. 2460–2461, Oct. 2010. https://academic.oup.com/bioinformatics/article-lookup/doi/10.1093/bioinformatics/btq461

[52] V. Vacic, L. M. Iakoucheva, and P. Radivojac, "Two sample logo: A graphical representation of the differences between two sets of sequence alignments," *Bioinformatics*, vol. 22, no. 12, pp. 1536–1537, Jun. 2006.

[53] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, no. 2, p. 305, 2012.

[54] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994. [Online]. Available: http://www.iro.umontreal.ca/lisa/pointeurs/ieeetrnn94.pdf

[55] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, p. 533, 1986.

[56] K. Cho, B. van Merrienboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: Encoder-decoder approaches," 2014, *arXiv:1409.1259*. [Online]. Available: http://arxiv.org/abs/1409.1259

[57] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997. [Online]. Available: http://www.mitpressjournals.org/doi/10.1162/neco.1997.9.8.1735

[58] T. Saito and M. Rehmsmeier, "The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0118432. [Online]. Available: https://dx.plos.org/10.1371/journal.pone.0118432

[59] T. A. Lasko, J. G. Bhagwat, K. H. Zou, and L. Ohno-Machado, "The use of receiver operating characteristic curves in biomedical informatics," *J. Biomed. Informat.*, vol. 38, no. 5, pp. 404–415, Oct. 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1532046405000171

[60] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, pp. 861–874, Jun. 2006. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S016786550500303X

[61] B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," *Biochim. et Biophys. Acta (BBA)-Protein Struct.*, vol. 405, no. 2, pp. 442–451, Oct. 1975.

[62] P. Baldi, S. Brunak, Y. Chauvin, C. A. F. Andersen, and H. Nielsen, "Assessing the accuracy of prediction algorithms for classification: An overview," *Bioinformatics*, vol. 16, no. 5, pp. 412–424, May 2000.

[63] S. Naseer and Y. Saleem, "Enhanced network intrusion detection using deep convolutional neural networks," *KSII Trans. Internet Inf. Syst.*, vol. 12, no. 10, pp. 5159–5178, 2018.

[64] E. R. DeLong, D. M. DeLong, and D. L. Clarke-Pearson, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," *Biometrics*, vol. 44, no. 3, pp. 837–845, Sep. 1988.

[65] X. Sun and W. Xu, "Fast implementation of DeLong's algorithm for comparing the areas under correlated receiver operating characteristic curves," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1389–1393, Nov. 2014.

[66] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," *BMC Genomics*, vol. 21, no. 1, p. 6, Dec. 2020.

[67] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

SHERAZ NASEER received the M.S. degree in information security and the Ph.D. degree in computer science. He received the professional certifications of IT including, CISSP, CoBit, and ITIL. He has 15 years of experience in industry and academia. He is currently working as an Assistant Professor with the Department of Computer Science, University of Management and Technology, Lahore, Pakistan. His research interests include bioinformatics, data driven information security, and anomaly detection.

RAO FAIZAN ALI received the bachelor's degree in computer science from COMSATS University Islamabad, Pakistan, and the M.Phil. degree in computer science from the University of Management and Technology, Lahore, Pakistan. He is currently pursuing the Ph.D. degree with University Technology PETRONAS, Malaysia. He has eight years of experience in teaching and research. He has been with various computer science positions in financial, consulting, academia, and government sectors. He is currently working as a Research Officer with the Department of Computer and information Sciences, University Technology Petronas, Perak, Malaysia.

SULIMAN MOHAMED FATI (Senior Member, IEEE) received the Ph.D. degree from Universiti Sains Malaysia (USM), Malaysia, in 2014. He is currently an Assistant Professor with the Information Systems Department, CCIS, Prince Sultan University, Saudi Arabia. His research interests include the Internet of Things, machine learning, cloud computing, cloud computing security, and cybersecurity. He is an Active Member in different professional bodies, as ISACA, IACSIT, IAENG, and the Institute of Research Engineers and Doctors, USA. He serves as a reviewer in many international impact-factor journals, and a TPC in different international conferences.

AMGAD MUNEER received the B.Eng. degree (Hons.) in mechatronic engineering from the Asia Pacific University of Technology and Innovation (APU), Malaysia, in 2018. He is currently pursuing the master's degree in information technology with Universiti Teknologi PETRONAS, Malaysia. He has authored several ISI and Scopus journal articles/conference papers. His research interests include machine learning, image processing, the Internet of Things, machine vision, robotics, and automation. He is a Reviewer in some international impact-factor journals, such as *Journal of Combinatorial Optimization* and several IGI global journals.