International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-5, March 2019

Unsupervised Feature Learning on Big Data Based on Deep Learning with Weighted Softmax Regression

Sangram Keshari Swain, Srinivas Prasad, M. Vamsi Krishna

Abstract: ***Deep learning is the recent technology which is effectively applied to feature learning in image classification, feature learning and language processing. However, recent deep learning models process with vector space which produces failure when learning features from non-linear distribution of heterogeneous data. Large amount of labeled data can be required for supervised learning of recurrent neural networks (RNN). These large volume of training samples are time consuming and high cost. One direction of addressing this problem is to extract features from unlabeled data. This paper proposes a deep computation model for feature learning on big data to learn underlying data distribution using Deep Recurrent Neural Network based weighted Softmax regression (DRNNWSR) with no need of labeled instances. The proposed approach is moderately simple, however achieves accuracy comparable to that of more advanced techniques. The proposed strategy is significantly easier to train, contrasted with existing neural system strategies, making less prerequisites on manually labeled training data. It is additionally appeared to be impervious to over fitting. We give comes about on some outstanding datasets, specifically STL-10, Caltech-256, and Caltech 101 and CIFAR-10. The results show that the proposed system obtains really high order accuracy and is superior to the present techniques for the broad dataset. Because of learning features adaptively, the proposed system diminishes the need of tedious and makes data classification more efficient. Our numerical results demonstrated good convergence when compared to the different datasets for different classifiers.***

Keywords: ***Unsupervised Learning, Recurrent Neural Network, Big Data, Softmax Regression, Deep Learning.***

I. INTRODUCTION

The large volume of data gathered with various distributed devices refers to Big Data. These big data are classified based on the specific types of datasets which contain formless data [5]. The big data include three characteristics namely, volume, velocity, and variety. Here, the volume indicates the storage size of the big data, the speed of analyzing and processing is termed as velocity, and various data gathering sources are considered as varieties [6].

Manuscript published on 30 March 2019.
*Correspondence Author(s)
Sangram Keshari Swain, Research Scholar, Centurion University of Technology and Management, Odisha, India.
Srinivas Prasad, Professor, Department of Computer Science & Engineering, K L University, Andhra Pradesh, India.
M. Vamsi Krishna, Associate Professor, Department of Computer Science & Engineering, School of Engineering & Technology, Parlakhemundi Campus, Centurion University of Technology and Management, Odisha, India

The big data are created and gathered together from various fields and on IoT systems, by using the object abstraction layer the gathered data sets are transmitted from the object layer to service management layer. The upper layers will get some services like decision support and prediction from the service management layer. It provides such services by analyzing the data received from the object layer. For clients, the upper layers will offer interface, it provides an intelligent service for big data [7]. The bygone learning method was not able to manage huge amount of heterogeneous information available on distributed domains [8]. It needs high speed and storage devices to handle these maximum number of varied information [9]. Usually, the Big Data Analysis (BDA) denotes a method of information gathering, move in the centralized cloud data centers, pre-processing, assessment, and visualization [10-12]. The classical deep learning methods like deep belief networks, deep convolutional neural networks and stacked auto encoders learn the features of some file types like image, text, and audio. It was understood as a single type feature learning [13]. For feature learning, various multi-modal feature learning methods are introduced recently and some popular multi-modal methods are multi-modal deep neural networks and deep Boltzmann machines. Initially, these multi-modal learning methods learn the features of every assorted information example and learned characteristics are integrated as a solitary vector to provide the combined illustration of heterogeneous data model. With the help of unlabelled data, the multi-modal methods will learn the representation of single modal [14].

Contribution: Sparse representation and RNN is fused in the proposed approach DRNNWSR for the learning of robust feature from vast data in internet. High-level structures were learned by the feature extraction because of the scarcity proficiency. Large nonlinear structure of data was handled by using Weighted Softmax Regression (WSR) as a classifier and the result of this Softmax regression was properly adjusted by back propagation algorithm. This fine-tuning make the full process more robust and improve the classification performance also it helps the learning rate faster and avoids the gradient diffusion problem as well as local extrema problem. We have conducted a number of experiments on unlabelled large dataset to illustrate the efficiency of this projected method and the result produce by this experiment shows that DRNNWSR method obtain high accuracy in cluster feature learning.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

Organization: The article is structured as follows. Section 2 shows the associated work of unsupervised feature learning procedures. In Section 3, the arrangement and proposal of DRNNWSR was presented. In Section 4 experimental procedure, outcomes of some experiments and their exploration are presented in Section 4. Section 5 contains the conclusion part.

2. Related work

Here, some of the works based on feature learning on big data which are related to our work is illustrated.

Min Chen *et al* [15] presented a method for learning the features of medical image data. The researches of this article introduced a framework based on deep learning for feature learning of medical image with unlabelled data. This framework efficiently learn features of medical images with the minimal amount of labeled data. The method proposed in this article is used for some task like classification, similarity check, and lung nodule recognition. From the experimental results of this methods shown that their method is much better than other methods. The performance of auto-encoder approach was good, but it was affected by the quality of data. This issue is caused by ignoring the role of expert. On their future work, they planned to overcome this issue by integrating the domain knowledge and data-driven feature. The task like image retrieval, image classification and object detection and segmentation are handled on big data with various methods. Jamil Ahmad *et al* [16] presented an article for classifying the features of multimedia big data. Here the authors, extract the features and transform it into hash codes by using the sensitivity property. This hash codes uses neighbor search procedures for the retrieval of multimedia data from big data. Two steps were used in this approach. Initially on first step, the introduced feature selection method chooses the features which they have maximum diversity. They found more than 1800 features from the 4000 features of images on their experiment. On the second step of the proposed system, with the selected feature the fast Fourier transform is estimated and higher frequencies will be binarizes with mean frequency. The method of this article aims to describe the selected multimedia feature as a signal and the feature vector is estimated with the FFT. The performance of this method earned higher performance when compared to other classic methods. The researchers of this article telling that the performance of this method is weak only when using the smaller hash codes and additionally it is not suitable for sparse features, otherwise this method is better and executes good for deep features. They planned to make their method to support sparse features and improve the performance on their future work. Adrian Barbu *et al* [17] proposed a method for learning from large dataset. The method used in this research was put on for improving different loss function, arrangement, ranking and application in regression. The researchers telling that their method is very simple and easy for implement. The over fitting problem solved by enforcing the second order prior on piecewise linear response function. The experimental evaluation of this paper was carried out with some existing methods and they yield better performance on regression, classification and ranking and their method is highly scalable and efficient.

Qingchen Zhang *et al* [18] proposed big data learning of cloud computing deep computational model with crowdsourcing. In this article, the over fitting of deep computation model was avoided by distribution function. This function fix the dropout values for every hidden layers. The suitable information from the big data sets of crowd source are selected with outsourcing selection algorithm, which selects the data samples based on the highest entropy. The main advantages of the model in this article was, preventing the over fitting of hidden layers, crowdsourcing technique and the parameter training. These can build an efficient distribution function and also it will monitor the available training samples. The results of this model showed that the offered model got higher performance than other existing crowdsourcing algorithms. The framework introduced in this paper successfully avoided the over fitting and efficiently collecting the samples of parameter training for IoT big data feature learning. On their future work, some best method are planned to apply on their model for improving the performance. Dacheng Tao *et al* [19] presented an outline for improvement of image superiority in big data images. The method recommended for this article, takes five factors of image into account for the enhancement process. The authors of this article extracts more than 15 features from the image and they estimated the visual quality of image with the regression module. These are analyzed and learned from big data sets. This big data sets are much larger in size than the similar image data sets. The researchers of this article conducted experimental setup of their model with nine datasets and performance are measured in three categories. The performance of their method is much better than FR and RR algorithms. But the IQA of the proposed method is slightly lower than FR and RR algorithms. On their feature work, they improve the IQA frame for increasing the performance of their method.

3. Proposed methodology

In this method cluster feature learning was used to obtain feature description in more advanced level and the complete plan of projected Deep Recurrent Neural Network based weighted Softmax regression (DRNNWSR) was presented in this section. Unsupervised feature knowledge was used to detect the hidden features spontaneously from the outsized dataset without depending the supervisory signal and the identified features were used to obtain illustration which simplify the successive supervised learning (e.g. Classification of object).Unsupervised learning has lot of advantages than the supervised learning among them one of the most important advantage was learning the consistent patterns from unlabelled datasets which are complimentary and easy to acquire. Two steps were performed for this, the first step was to prepare a set of limited riddles from the unlabelled data and the second step was to separate the selected image into patches and encrypt them with feature vector which uses the learnt filter bank [20].

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

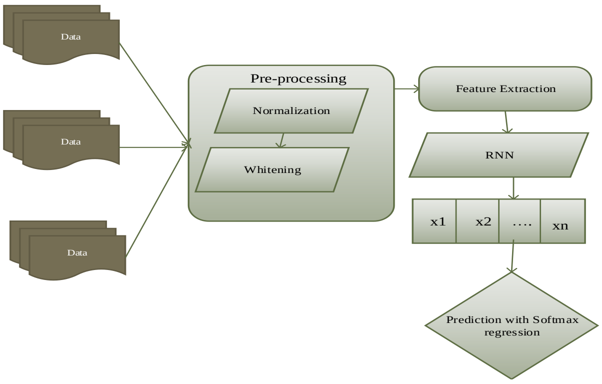


Fig. 1: System framework

3. 1 System overview

There are number of layers in deep learning (DL) and each layer has the capacity to generate non-linear response of data that obtained from the input layer. DL performance was copied from the signal processing mechanism done by human brain and neurons and the results obtained from this DL attracts lot of researchers towards it than the other old machine learning techniques. The proposed work consists of two phases. In the first phase, Recurrent Neural Network, an unsupervised feed forward neural network, is used to directly learn features from heterogeneous dataset. In the second phase, weighted Softmax regression is employed to classify the dataset based on the learned features. In order to test our design the input unlabelled dataset was fed to the first layer and then the output was obtained from the last layer. Amongst the first and fifth(last) layer, several layers distillate the features from the previous section and fed that extracted features to next section and finally the received output was given to weighted Softmax regression.

Problem formulation:

Object image set was represented as $I=\{I_1,I_2,...,I_N\}$, the objective was to remove discriminative characters of the images in terms of their subordinate categories. Let, $M=(F,C,S)$ denote a model that consist of part features $F=(F_1,...F_k)$, part centre locations $C=(C_1,...C_N)$ and part sizes $S=(S_1,...S_N)$, where K is the number of parts. For each image I^m we further denote $C^m=(C^m_1,...C^m_N)$ and $S^m=(S^m_1,...S^m_N)$. The location and size of each part are represented by the normalized coordinates with respect to the image size, i.e., $C_i,S_i\in[0,1]\times[0,1]$. Based on this, discovering object parts can be written as a constrained minimization programming problem:

$$(F^*,C^*,S^*)=\arg\min J(F,C,S), \quad (1)$$

$$\text{Subject to } 0\le S_i\le 1, i=1,2,...K, \quad (2)$$

$$0\le C_i\pm(1/2)S_i\le 1, i=1,2,...K, \quad (3)$$

Where, $J(F,C,S)$ is the objective function model and condition 2, 3 are introduced to impose entry-wise inequalities.

3.2 Pre-processing

Here the accommodating influence was obtained by pre-processing procedures like normalization and fading.

Normalization: Normalization in image processing was used to modify the intensity rate of pixel. In some applications like pictures with reduced contrast due to glare normalization transforms a 3-dimensional RGB image $I:\{X\subseteq R^n\}\rightarrow\{Min,...,Max\}$ with intensity values in the range (Min, Max), into a new image $I_N:\{X\subseteq R^n\}\rightarrow\{newMin,....,newMax\}$ with intensity values in the range (newMin, newMax).

The following formula was used to carry out the linear normalization in RGB digital image [21],

$$I_N=(I-Min)\frac{newMax-newMin}{Max-Min}+newMin \quad (4)$$

Whitening: In this whitening method the images that are already processed was given to Zero-Phase Component Analysis (ZCA). ZCA was a decorrelation process to remove the initial and next order data and impact the learning stage to capture advanced order data and the co-variance matrix 'M' was specified to the identity matrix I. Then the matrix M was given as

$$MM^T=(n-1)I \quad (5)$$

Where ‘n’ represent the number of ‘w’ input vectors. The model covariance which is the maximum likelihood evaluation (MLE) was used for the large value of n in Equation.5 was likely to be nearer to actual co-variance matrix. ZCA whitening tends to find out the set of regular, definite filters which convert 'w' to 'M' respectively and the results of processed image was given by,

$$\boldsymbol{M}=\boldsymbol{U}_z\boldsymbol{w} \quad (6)$$

Where, w was considered to be null concentrated and the term U_z is computed by the following condition,

$$U_z\alpha\left(ww^T\right)^{-1/2} \quad (7)$$

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The outer product ww^T is symmetric and orthogonally diagonalizable. Hence, applying eigenvalue decomposition to ww^T, we obtain,

$$ww^T = V \wedge V^T, \quad (8)$$

Where, V are the Eigen vectors and $\wedge$ are the Eigen values of ww^T. Raising both sides of Eq.5 to the power -1/2, we get,

$$(ww^T)^{-1/2} = V \wedge^{-1/2} V^T \quad (9)$$

Combining Eq. 9 and Eq. 7, we obtain an expression for U_z as,

$$U_z \alpha V \wedge^{-1/2} V^T \quad (10)$$

The matrix, ww^T was still in hostile condition so a regularization term $\in$ was supplemented to the eigenvalues and the resultant was not randomly measured by the inverse smaller eigenvalues. Hence, the expression for the filters of a standard ZCA whitening algorithm is given by,

$$U_z \alpha V(\wedge + \in)^{-1/2} V^T \quad (11)$$

Applying U_z on w, whitened patches were introduced with perfect covariant matrix I.

3.3 Feature extraction

An unsupervised feature learning is proposed to learn the features independently over the volume of data. The hidden-to-hidden connections models the short term time dependency without consider time delay-taps. Back propagation through time based procedure is used for train the data iteratively. RNNs are considered as a very deep networks include shared parameters at every layer when expanded in time. This outcomes in the problem of vanishing gradients and has motivated the exploration of second-order methods for unsupervised pre-training and deep architectures. A representation of feature is modelled with recurrent layers and these features are feed to classification. This method consider as the simplest possible type of RNN and it is easy to train and implement. RNN feeds each window as an edge by edge procedure to the recurring level and feature extraction was done with the outputs, unseen conditions of regular units in every edge from the consecutive windows. This RNN obtains best features using this RNN than the convolutional layers. A common structure is characterized for a layer to extract local features from a sequence dataset. Let us consider an unlabelled dataset 'u' whereas each sample in the dataset was an unlabelled data and to explain this one data 'x' was chosen as an example from the entire dataset. Number of character in the series was represented as 1 and the amount of features of each frame in dataset was represented as 'k' therefore the k size becomes k x 1.One-dimensional feature vector of length k was available to each object but this data was suggested as multi-dimensional case by providing the data to one-dimension. Window sequence was generated from the sequence x and each window consist of successive frames. Then, each window is of size $k \times r_1$, and we take a shift of r_2 object between the starting points of two consecutive windows. Next we have to proceed on feature abstraction function f on several window to obtain a group of n features of several window and the n size feature vector that defines the window 'w' was indicated as f (w). Additional series x' was created by providing f for several window and then the window of size $n \times p_1$ was created from the order x' for pooling as same as the way the windows are created from the series x, by shifting the p_2 frames among the proceeding points of two successive windows. Max-pooling was performed through each frames to transfer the window size $n \times p_1$ to vector size $n \times 1$ and the pooled series contains local feature of series x mined with the function f was given to the classifier.

A sequence $(x_1, x_2, ...x_t)$ was taken by the simple recurrent neural network and produces a sequence $(h_1, h_2, ...h_t)$ of hidden states and a sequence $(y_1, y_2, ...y_t)$ of outputs in the subsequent method [22]:

$$h_t = \sigma(W_{xh} x_t + W_{hh} h_{t-1} + b_h) \quad (12)$$

$$y_t = W_{hy} h_t + b_y \quad (13)$$

Where, σ is the logistic sigmoid function, W contains the weight matrices and b are biases.

3.4 Classification

The learned features can then be used repeatedly for classification of any subclass. This method was fully focused on feature learning stage and the single classification algorithm was used by every experiment along with this algorithm weighted softmax regression classifier was used for feature learning. First the weight of the network was provided by pre-training the network layer by layer then the finest network model number of hidden layers and hidden nodes were determined by improving the scarce constraints. Moreover SR was used to categorize the facial expression features and GD method was used to train the SR optimal model parameter. Finally the overall weight of RNN was clearly tuned by Back- Propagation (BP) algorithm to improve the performance and robust nature of entire network of deep learning.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

The Softmax classifier used in this approach was a development of logical classifier which was used to implement the multi-class classification. Logic classifier was more suitable to the problems created by binary classification and also reduce the problems produced by nonlinear classification. This classification provide probability as output and the final category was determined by threshold. In this the obtained probability was compared with threshold and transform it into binary classification problem. The formula shown below represent the expression for logical function [23],

$$y_\theta(x) = g(\theta^T x)\frac{1}{1+e^{-\theta^T x}} = p(y=1\,|\,x;\theta) \qquad (14)$$

Where, y(x) represent the probability 1, and t represent the perfect limit.

Loss function was minimized by optimizing the limits which was done by repeated adjustments and the equation for loss function was shown below,

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\}\log\frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\right] \qquad (15)$$

The extension term for multiple Softmax regression classifier was shown in equation.1

$$y_\theta(x^{(i)}) = \begin{bmatrix} p(y^{(i)} = 1\,|\,x^{(i)};\theta \\ p(y^{(i)} = 2\,|\,x^{(i)};\theta \\ . \\ . \\ p(y^{(i)} = k\,|\,x^{(i)};\theta \end{bmatrix} = \frac{1}{\sum_{j=1}^{k} e^{\theta_j^T x(i)}}\begin{bmatrix} e^{\theta_1^T x(i)} \\ e^{\theta_2^T x(i)} \\ . \\ . \\ e^{\theta_k^T x(i)} \end{bmatrix} \qquad (16)$$

Here the output rate was represented as 'k' dimensional vector class, and $\theta_1, \theta_2, \theta_3, \ldots \theta_k$ represent the standard limits.

For several class, probability of output j represent the probability of data object. In this the classification was achieved by the highest probability rate of category and the features learned by deep recurrent neural network was sorted by SR., for training set: $\{x_1, y_1, \ldots x_m, y_m\}$, it has $y_i \in \{1,2,\ldots k\}$. K=10 categories was proposed in our classification difficulties such as (airplane, bird, car, cat, deer, dog, horse, monkey, ship, truck). The probability of each category was identified by using the function $y_\theta(x)$

The matrix constraints for this model was defined as follows,

$$\theta = \begin{bmatrix} \theta_1^T \\ \theta_2^T \\ . \\ . \\ \theta_k^T \end{bmatrix} \qquad (17)$$

Eq. 18 shows the cost function for SR. Decay element was added to the cost function to correct the peak values of constraints and the cost function equation was modified as,

$$J(\theta) = -\frac{1}{m}\left[\sum_{i=1}^{m}\sum_{j=1}^{k} 1\{y^{(i)} = j\}\log\frac{e^{\theta_j^T x^{(i)}}}{\sum_{l=1}^{k} e^{\theta_l^T x^{(i)}}}\right] + \frac{\lambda}{2}\sum_{i=1}^{k}\sum_{j=0}^{n}\theta_{ij}^2 \qquad (18)$$

A standard gradient-based optimization method was expressed in the form

$$\theta_j = \theta_j - \alpha\nabla_{\theta_j} j(\theta)(j=1,2,\ldots k) \qquad (19)$$

Using the above iterative scheme, the weighted Softmax regression classification model is optimize. Let us assume the data with m samples and the reduced performance index was declaims as shown below,

$$J(w,b) = \left[\frac{1}{m}\sum_{i=1}^{m} j(w,b,x_i y_i)\right] + \frac{\lambda}{2}\sum_{l=1}^{n_l-1}\sum_{i=1}^{s_l}\sum_{j=1}^{s_{l+1}}(w_{ji}^{(l)})^2 \qquad (20)$$

In this the first term in the formula was mean square error and the next one represent weighted decay. The following expression was obtained by calculating the activation rate of each layer in network without considering the output layer,

Table 1: Accuracy

	DRNNWSR	CANN	CNN	MCCNN	AE
STL-10	97	95	90	91	92
CIFAR-10	94	93	88	91	89
CALTECH-101	96	95.045	90	90	86
CALTECH-256	95	94	89	91	88

Table 2: F-measure

	DRNNWSR	CANN	CNN	MCCNN	AE
STL-10	94	93	89	88	90
CIFAR-10	96	95	85	90	88
CALTECH-101	92	91	86	87	86
CALTECH-256	93	92	88	89	87

*Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)*

$$a^{l+1} = f\left(w^{(l)}a^{(l)} + b^{(l)}\right) \tag{21}$$

After performing the repeated calculation in residual error, the pertinent formula was specified in the form,

$$\delta_i^{nl} = \left(\sum_{j=1}^{s_{l+1}} w_l^{ji}\delta_i^{(l+1)}\right) f'(z_i^{nl}) \tag{22}$$

The updated formulas for i = 1, 2...,m, will be in the form,

$$\Delta w^l = \Delta w^l + a_j^l \delta_i^{l+1} \tag{23}$$

$$w^l = w^l - a\left[\left(\frac{1}{m}\right)\Delta w^l\right] + \lambda w^l \tag{24}$$

$$\Delta b^l = \Delta b^l + \delta_i^{l+1} \tag{25}$$

$$b^l = b^l - a\left[\left(\frac{1}{m}\right)\Delta b^l\right] \tag{26}$$

II. EXPERIMENTAL SETUP AND EVALUATION

The results obtained during the classification on various large datasets are reported for matching the proposed method with other existing feature learning techniques. The results are verified with the different quantity of grouped datasets. Specially, the STL is suitable for the unsupervised learning, because it includes massive set of 100,000 samples which are unlabelled. We gathered data for replacing the training data from the unlabelled subset of STL-10. We enlarged the size of image from 32×32 pixels to 64×64 pixels for creating the quantity of illustrated objects which similar to the additional datasets while examining on the CIFAR-10. Another dataset Caltech-101 includes 102 classes totally, which are 101 different categories of objects and an extra background class. Likewise, another dataset Caltech-256 is also gathered. In the MATLAB instrument, the suggested work is imitated. The huge data for both input and output are managed by the Mat file by collecting the complete data within that file. During the process of simulation, minor quantity of data is stored into the system memory and it is said to as streaming.

4.1 Results

The numerical values obtained from experimental evaluation are summarized in below Table 1 to 5.

Table 3: Recall

	DRNNWSR	CANN	CNN	MCCNN	AE
STL-10	92	91	90	87	92
CIFAR-10	95	94	88	91	89
CALTECH-101	92	90	85	88	84
CALTECH-256	91	90	87	89	86

Table 4: Precision

	DRNNWSR	CANN	CNN	MCCNN	AE
STL-10	92	91	90	89	92
CIFAR-10	93	92	90	91	87
CALTECH-101	91	90	87	88	86
CALTECH-256	91	90	87	89	90

Table 5: AUC

	DRNNWSR	CANN	CNN	MCCNN	AE
STL-10	0.97	0.93	0.89	0.88	0.9
CIFAR-10	0.96	0.95	0.85	0.9	0.88
CALTECH-101	0.95	0.91	0.86	0.87	0.86
CALTECH-256	0.93	0.92	0.88	0.89	0.87

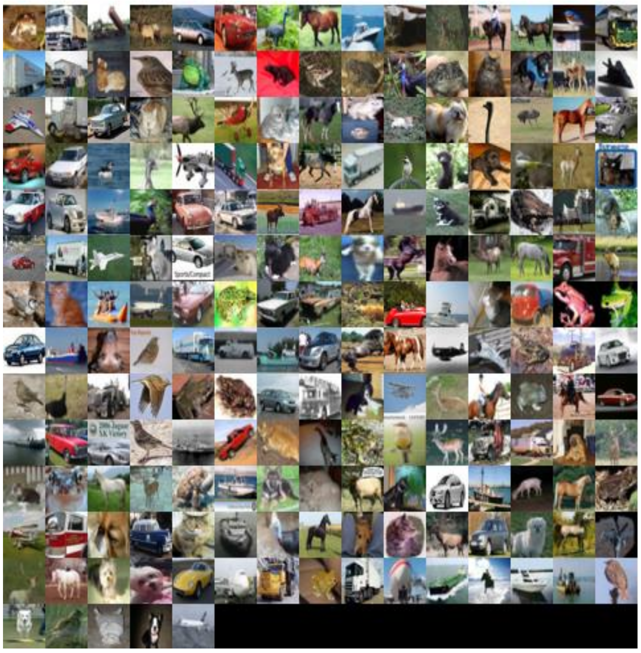


Fig.2 input image

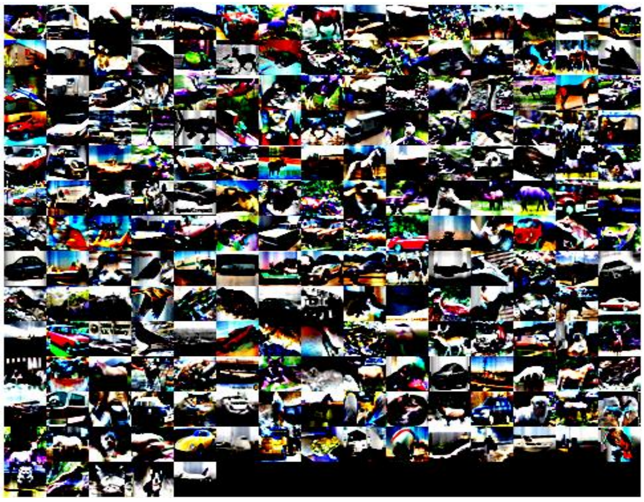


Fig. 3: normalized image

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)

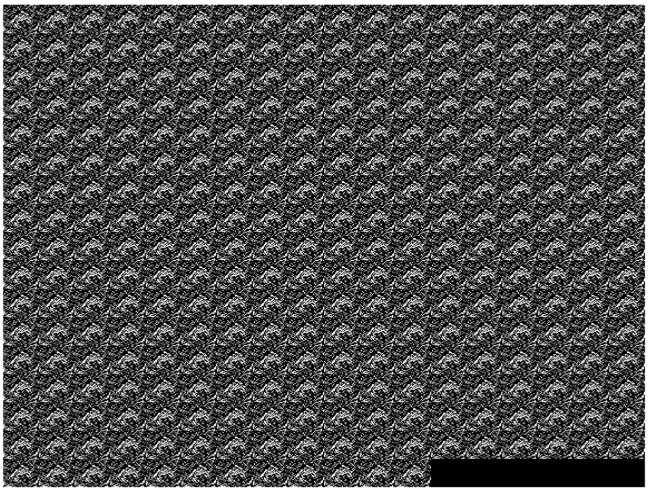


Fig. 4: Whitening image

The input image sample is represented in figure 2 which is from STL-10 database and figure 3, 4 represents the normalized image and whitening image respectively. The basic testing procedure is described in this section. A single layer of features train for each unsupervised learning algorithm and the features are extracted either from raw data or whitened data. The experimentation consider 100, 200, 400, 800 and 1200 features properly. Training and testing properly done to get the accurate classification.

4.2 Performance evaluation

Performance metrics

False Positive Rate, False Negative Rate, True Positive Rate and True Negative Rate are some of the processes recycled to determine the efficiency of various machine learning approaches.

Sensitivity is the probability that a sample data test is positive,

$$Sensitivity = \frac{TP}{TP + FN} \tag{27}$$

Specificity is the probability that a sample data test is negative,

$$Specificity = \frac{TN}{TN + FP} \tag{28}$$

The probability of the sample data test is correctly performed is called precision.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{29}$$

In Fig.5, the outcomes are displayed regarding the comparison of the categorization execution of the convolutional neural network (CNN), Convolutional Auto encoder Neural Network (CANN), MCCNN and auto encoder (AE) with the unlabelled dataset. The categorization execution of the suggested process is not lower than both the CNN and MCCNN technique. The incorporation of the unsupervised feature learning and supervised fine tuning can considerably develop the execution is confirmed by the assessment.

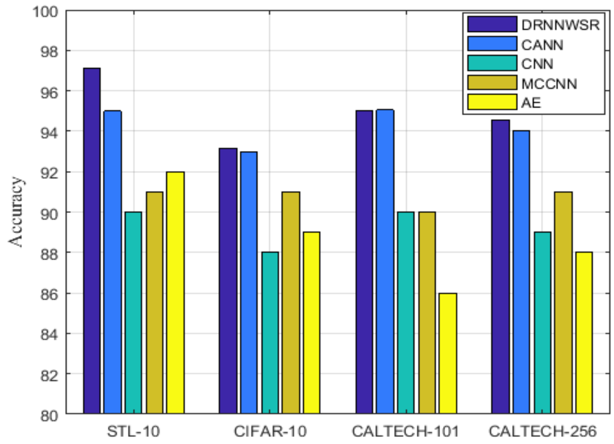


Fig. 5: Analysis of accuracy

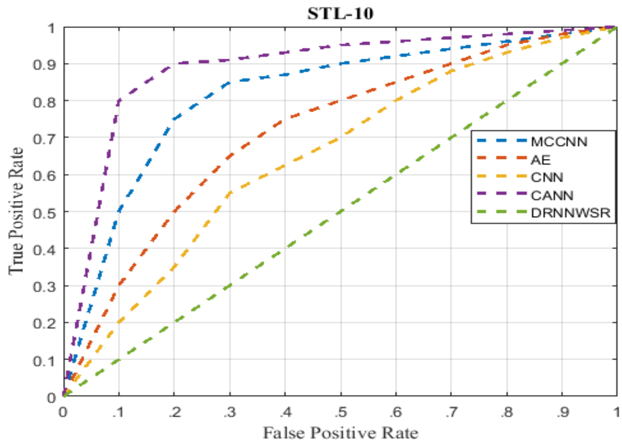


Fig. 6: STL-10 dataset

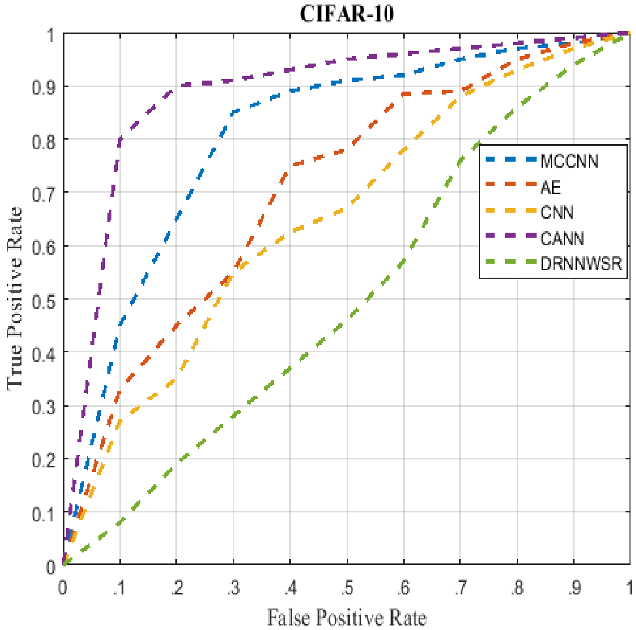


Fig. 7: CIFAR-10 dataset

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

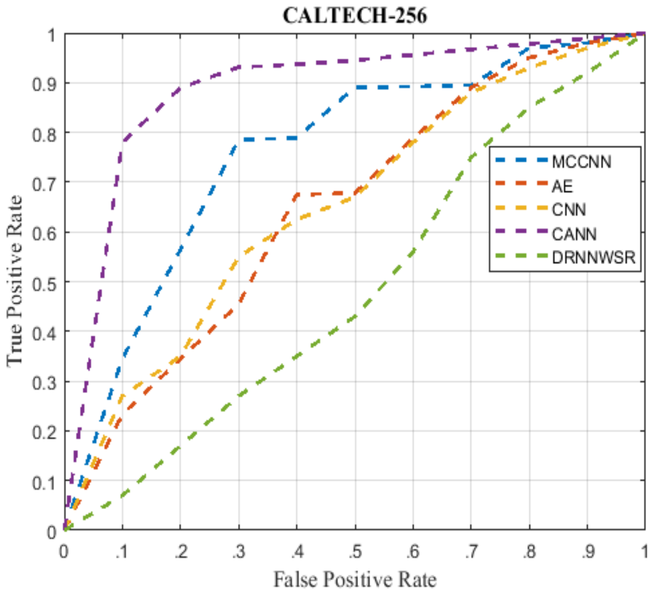


Fig. 7: CALTECH-256 dataset

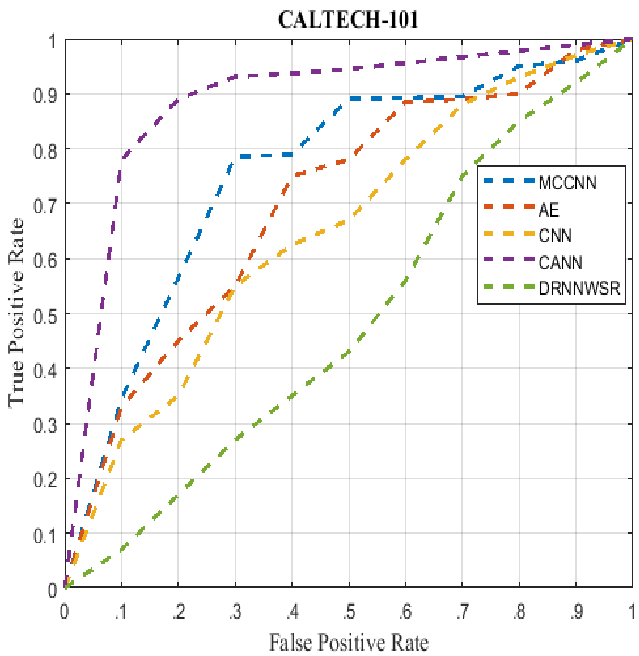


Fig. 8: CALTECH-101 dataset

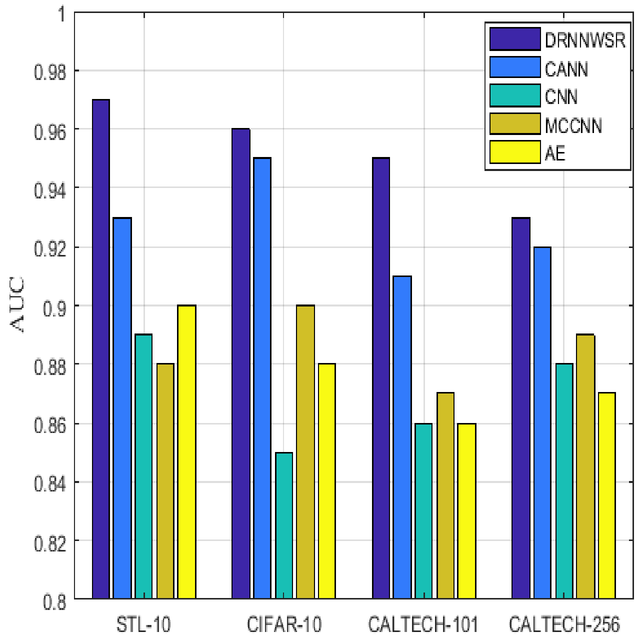
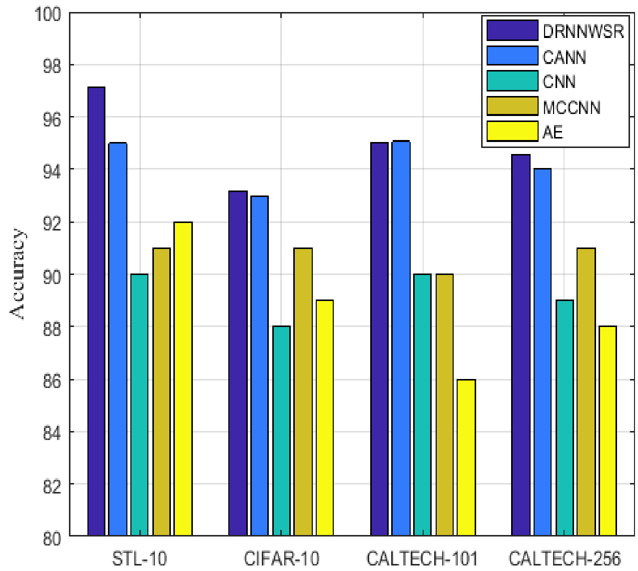
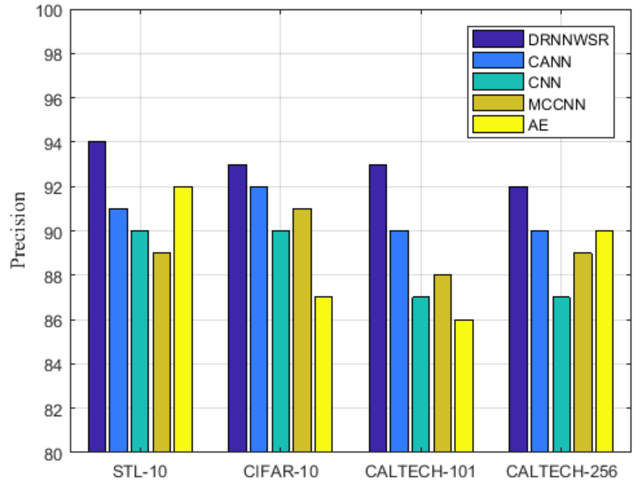


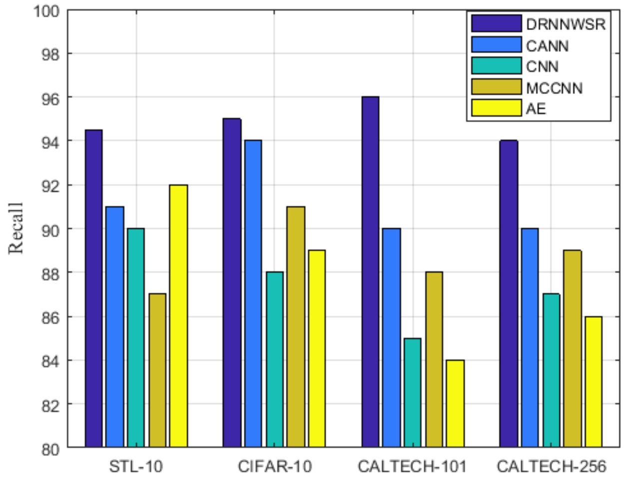
Fig. 9: AUC analysis for various database





Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-8 Issue-5, March 2019



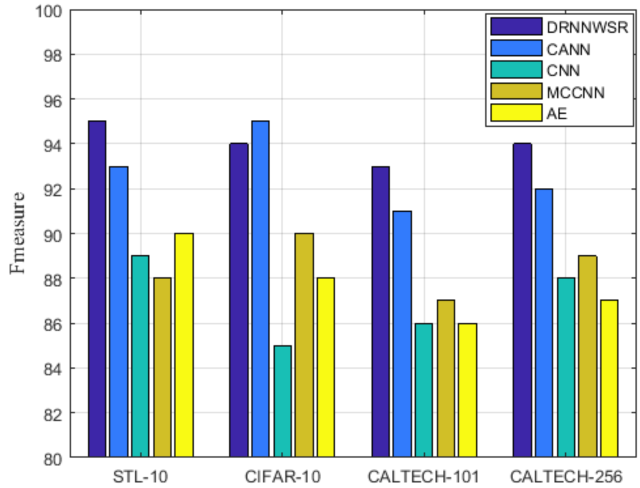


Fig. 10: Classification performance

Investigations are carried out for the development of deep learning and detection of object on various dataset which dedicate heterogeneous dataset. The extensive experiments are conducted using various database by unsupervised feature learning representation to enhance the performance of proposed system which is shown in figure 5 to 10. We have indicated all the more critically that these components can, truth be told, be as essential as the unsupervised learning calculation itself while affirming the fundamental finding that more highlights and thick extraction are valuable. The real reason is that this model can remove the highlights which is incorporated into the heterogeneous information with the layer rising.

III. CONCLUSION

In this paper we considered the application of unsupervised feature learning for large scale data. We exhibited that classification accuracy can be significantly enhanced by feature learning of various datasets. The Deep Learning has the benefit of possibly delivering a solution to address the data assessment and learning difficulties discovered in the huge volumes of input data as resisted to feature manufacturing algorithms and additional conventional machine learning. Specifically, it supports as a result of segregating complex data representations from the extensive volumes of unsupervised data. For the big data analytics, this creates it as an essential tool that contains information research from the huge gatherings of raw data which is commonly unsupervised and unclassified. Deep learning method proposed in this model was used for the feature learning of big data. The MATLAB tool 2016a version executes the suggested work. The executions are estimated and contrasted based on precision, misclassification charge. The weighted Softmax regression model in terms of the deep Recurrent Neural Network attains greater categorization precision than the multi-modal deep learning model and the deep calculation model.

Future work: In future, Fault detection and classification through unsupervised learning will be considered. This work was enlarged to implement the deep learning framework to complete together the tasks of grouping and error detection and by using this framework error recognition system was designed.

REFERENCES

1. Ranjan, Rajiv, Dhavalkumar Thakker, Armin Haller, and Rajkumar Buyya. "A note on exploration of IoT generated big data using semantics." (2017): 495-498.
2. Chin, Wen-Long, Wan Li, and Hsiao-Hwa Chen. "Energy Big Data Security Threats in IoT-Based Smart Grid Communications." IEEE Communications Magazine 55, no. 10 (2017): 70-75.
3. Papalkar, Rahul R., Pravin R. Nerkar, and C. A. Dhote. "Issues of concern in storage system of IoT based big data." In Information, Communication, Instrumentation and Control (ICICIC), 2017 International Conference on, pp. 1-6. IEEE, 2017.
4. Wu, Zhipeng, Zhaozong Meng, and John Gray. "IoT-Based Techniques for Online M2M-Interactive Itemized Data Registration and Offline Information Traceability in a Digital Manufacturing System." IEEE Transactions on Industrial Informatics 13, no. 5 (2017): 2397-2405.
5. Cai, Hongming, Boyi Xu, Lihong Jiang, and Athanasios V. Vasilakos. "IoT-based big data storage systems in cloud computing: perspectives and challenges." IEEE Internet of Things Journal 4, no. 1 (2017): 75-87.
6. Ahmad, Awais, Murad Khan, Anand Paul, Sadia Din, M. Mazhar Rathore, Gwanggil Jeon, and Gyu Sang Choi. "Toward modeling and optimization of features selection in Big Data based social Internet of Things." Future Generation Computer Systems (2017).
7. Zhang, Qingchen, Laurence T. Yang, Zhikui Chen, and Peng Li. "High-order possibilistic c-means algorithms based on tensor decompositions for big data in IoT." Information Fusion39 (2018): 72-80.
8. An Adaptive Dropout Deep Computation Model for Industrial IoT Big Data Learning with Crowdsourcing to Cloud Computing
9. Manogaran, Gunasekaran, R. Varatharajan, Daphne Lopez, Priyan Malarvizhi Kumar, Revathi Sundarasekar, and Chandu Thota. "A new architecture of Internet of Things and big data ecosystem for secured smart healthcare monitoring and alerting system." Future Generation Computer Systems (2017).
10. Plageras, Andreas P., Kostas E. Psannis, Christos Stergiou, Haoxiang Wang, and B. B. Gupta. "Efficient IoT-based sensor BIG Data collection–processing and analysis in smart buildings." Future Generation Computer Systems (2017).
11. Wang, Yichuan, LeeAnn Kung, and Terry Anthony Byrd. "Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations." Technological Forecasting and Social Change 126 (2018): 3-13.
12. ur Rehman, Muhammad Habib, Ejaz Ahmed, Ibrar Yaqoob, Ibrahim Abaker Targio Hashem, Muhammad Imran, and Shafiq Ahmad. "Big Data Analytics in Industrial IoT Using a Concentric Computing Model." IEEE Communications Magazine 56, no. 2 (2018): 37-43.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

13. Zhang, Qingchen, Laurence T. Yang, Zhikui Chen, Peng Li, and M. Jamal Deen. "Privacy-preserving double-projection deep computation model with crowdsourcing on cloud for big data feature learning." IEEE Internet of Things Journal (2017).
14. Gao, Lianli, Jingkuan Song, Xingyi Liu, Junming Shao, Jiajun Liu, and Jie Shao. "Learning in high-dimensional multimedia data: the state of the art." Multimedia Systems 23, no. 3 (2017): 303-313.
15. Chen, Min, Xiaobo Shi, Yin Zhang, Di Wu, and Mohsen Guizani. "Deep features learning for medical image analysis with convolutional autoencoder neural network." IEEE Transactions on Big Data (2017).
16. Ahmad, Jamil, Khan Muhammad, Jaime Lloret, and Sung Wook Baik. "Efficient Conversion of Deep Features to Compact Binary Codes using Fourier Decomposition for Multimedia Big Data." IEEE Transactions on Industrial Informatics (2018).
17. Barbu, Adrian, Yiyuan She, Liangjing Ding, and Gary Gramajo. "Feature selection with annealing for computer vision and big data learning." IEEE transactions on pattern analysis and machine intelligence 39, no. 2 (2017): 272-286.
18. Zhang, Qingchen, Laurence Tianruo Yang, Zhikui Chen, Peng Li, and Fanyu Bu. "An Adaptive Droupout Deep Computation Model for Industrial IoT Big Data Learning with Crowdsourcing to Cloud Computing." IEEE Transactions on Industrial Informatics (2018).
19. Gu, Ke, Dacheng Tao, Jun-Fei Qiao, and Weisi Lin. "Learning a no-reference quality assessment model of enhanced images with big data." IEEE transactions on neural networks and learning systems 29, no. 4 (2018): 1301-1313.
20. Ren, Jimmy SJ, Wei Wang, Jiawei Wang, and Stephen Liao. "An unsupervised feature learning approach to improve automatic incident detection." In Intelligent Transportation Systems (ITSC), 2012 15th International IEEE Conference on, pp. 172-177. IEEE, 2012.
21. Prabhushankar, Mohit, Dogancan Temel, and Ghassan AlRegib. "Generating adaptive and robust filter sets using an unsupervised learning framework." In Image Processing (ICIP), 2017 IEEE International Conference on, pp. 3041-3045. IEEE, 2017.
22. Keren, Gil, and Björn Schuller. "Convolutional RNN: an enhanced model for extracting features from sequential data." In Neural Networks (IJCNN), 2016 International Joint Conference on, pp. 3412-3419. IEEE, 2016.
23. Chen, Luefeng, Mengtian Zhou, Wanjuan Su, Min Wu, Jinhua She, and Kaoru Hirota. "Softmax regression based deep sparse autoencoder network for facial emotion recognition in human-robot interaction." Information Sciences 428 (2018): 49-61.

Authors Profile

Mr. Sangram Keshari Swain is a Research Scholar in the Department of Computer Science and Engineering, Centurion University of Technology & Management, Bhubaneswar, Odisha, India. He is currently working as an Assistant Professor in the Department of Computer Science and Engineering, Centurion University of Technology & Management, Bhubaneswar, Odisha, India. His research interests are Data Mining, Big Data, Data Analytics, Machine Learning and IoT.

Dr. Srinivas Prasad received his Ph.D. degree in Computer Science and Engineering from Utkal University, Bhubaneswar, Odisha, India. He is currently working as a Professor in the Department of Computer Science and Engineering, KL University, Andhra Pradesh, India. His research interests are Data Mining, Big Data Analytics and Machine Learning.

Dr. M. Vamsi Krishna received his Ph.D. degree in Computer Science and Engineering from Centurion University of Technology & Management, Bhubaneswar, Odisha, India. He is currently working as a Professor in the Department of Computer Science and Engineering, Centurion University of Technology & Management, Bhubaneswar, Odisha, India. His research interests are Database, Data Mining and Big Data Analytics.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)