International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-4, April 2019 E-ISSN: 2347-2693

A Survey of Essential Methods in Deep Learning for Big Data

S.Umamageswari[1*], M. Kannan[2]

[1,2]Dept. of Computer Science and Applications, SCSVMV, Kanchipuram – 631561, Tamilnadu, India

Corresponding Author: uma.mages2008@gmail.com

DOI: https://doi.org/10.26438/ijcse/v7i4.11691180 | Available online at: www.ijcseonline.org

Accepted: 22/Apr/2019, Published: 30/Apr/2019

***Abstract* -** Big data has become an essential technology as many publicand private organizations have continuously collecteda vast amount of information regarding medical informatics, marketing, cyber security, fraud detection, and national intelligence. Deep learning is one of the remarkable machine learning techniques to find abstract patterns in Big data. Deep learning has achieved great success in variousbig data applications such as speech recognition, text understanding, and image analysis. In the field of data science, big data analytics and deep learning have become two highly focused research areas. Deep learning algorithm learns the multi-level representations and features of data in hierarchical structures through supervised and unsupervised strategies for the classification and pattern recognition tasks. In the last decade, deep learning has played a crucial role in providing the solutions for big data analytic problems. This paper provides a comprehensive survey of deep learning in Big data with the comparison of conventional deep learning methods, research challenges, and countermeasures. It also presents the deep learning methods, comparison of deep learning architectures, and deep learning approaches. Furthermore, this survey discusses the application-focused deep learning works in Big data. Finally, this work points out the challenges in big data deep learning and provide several future directions.

***Keywords*:** Big Data, Deep learning, Big data analytics, Machine learning, Deep learning architectures, and Challenges

I. INTRODUCTION

Deep learning (DL) [1] is one of the remarkable machine learning techniques, has dramatically received high popularity in a wide range of applications such as text understanding, image analysis, and speech recognition. The primary objective of the deep learning algorithms automatically extracts the multi-dimensional representations of the data. The deep learning algorithms employ the massive amount of unsupervised data to extract the complex data representation [2]. The artificial intelligence field mainly motivates the deep learning algorithms with the goal of obtaining the similar ability of the human brain for observing, analyzing, learning, and decision making, especially for complex problems. The complex challenges strive the learning system to emulate the hierarchical learning method of the human brain. Accordingly, several shallow learning architectures have been emerged, such as decision trees, support vector machines, and case-based reasoning, which fails to reach its target while attempting to extract the significant information from the complex structures and relationships in the corpus. Later, the deep learning architectures have been developed with the ability to generalize the non-local and global way of learning patterns. Deep learning plays a crucial role in the field of artificial intelligence, which extracts the complex data representations without human interference. In recent years, it is one of thepopular research topics in the field of machine learning. The deep learning architectures [3] widely employ either both the unsupervised and supervised strategies or separately unsupervised or supervised strategy to automatically learn the hierarchical representations and features for the pattern recognition and classification rather than traditional learning algorithms. The unique characteristic of the deep learning algorithm is that having the ability to exploit the unlabeled data during training. The paper is divided into 7 sections, section 2 to 5 deals with review paper outcomes related to the study work, section 6 deals with key challenges and future directions and section 7 deals with conclusion of the study.

A.Deep learning algorithms in Data mining

The deep learning methods enable the compact data representation and lead to a richer generalization. The number of possible configurations exponentially relies on the number of extracted abstract features [2]. By, the number of patterns obtained through distributed representation quickly scales with the number of learning factors compared to the local generalization based learning methods. The deep learning algorithms often lead abstract representations due to the construction of more abstract representations using a minimum number of abstract ones, which facilitates the system to be invariant during the local changes of the input data. In the pattern recognition method, learning the invariant features is the primary goal for instance, in a face recognition application, learning the invariant features of the face orientation is essential. Such invariant representations disentangle the factors of data variation, especially, in the artificial intelligence related tasks that often deal with the complicated interactions of multiple sources. In a deep architecture, learning the parameters with many hidden layers is a difficult task such as learning in neural networks. For example, at the beginning of the deep learning, the sensor data has been taken as the learning data to the first layer. By exploiting this data, the first layer is trained,and the outcome of the learned representations are fed as the learning data to the second layer. These processes are continued until obtaining the desired number of layers, which trains the whole deep network. The learned representations obtained from the last layer are used for different tasks.

To build deeper models, the existing researchers employ the unsupervised single layer learning algorithms such as Autoencoders and Restricted Boltzmann Machines (RBMs). These deep learning methods are often utilized to construct the Deep Belief Networks [4] and Stacked Autoencoders [5] using RBMs. The Autoencoders contain three layers such as input, hidden, and output layers. It attempts to learn the input in the hidden layer of the network, which also facilitates the input data reconstruction in the output layer of such intermediate representations. RBMs are the most popular version of the Boltzmann machine, containing one visible layer and one hidden layer. Two factors restrict that there is a connection between units from different layers alone and there is no interaction between the units within the same layer. The Boltzmann machine is widely trained by the Contrastive Divergence algorithm [6].

B. Deep learning algorithms in Big data analytics

Deep learning algorithms play a crucial role in learning the features from large amount of unsupervised data, which is more attractive while extracting the significant patterns and representations from Big data [2]. In deep learning model, the useful characteristics of the learned abstract representations have been used in different aspects. The characteristics involve i) simple linear models efficiently work with the knowledge of more abstract and complex data representations, ii) the growth of automatic data representation extraction enables the application to various data types such as textual, image, audio, and so on, and iii) the higher levels of abstraction and representation of the raw data leads the system to obtain the semantic and relational knowledge. Deep learning algorithms are more suitable to address the issues related to the large and heterogeneous characteristics of the Big data analytics. The existing shallow learning hierarchies fail to analyze and understand the high complexity of the data patterns. Hence, the deep learning method has been emerged due to its inherent utilization of the massive amount of available data. The extracted data representations of deep learning have been widely used for semantic indexing, decision making, information retrieval, and other different applications in big data analytics.

*Semantic indexing***:** In big data environment, effectively storing and retrieving the information is an emerging problem. Notably, for very large-scale data that includes text, image, audio, and video, that are massively collected from disparate sources such as social networks, shopping and marketing systems, security systems, fraud detection, defense systems, and cyber traffic monitoring. The massive amount of data associated system necessitates the semantic indexing to efficiently and quickly perform the knowledge discovery for example in the search engines. Deep learning is the most crucial approach in generating high-level data representations that to be used for semantic indexing rather than exploiting raw input data for indexing in the big data environment.

Deep generative model [7] learns the binary codes for documents in which the first layer and last layer of the deep learning network refers the word count vector and the learned binary code respectively. Binary codes require the relatively minimum storage space and also facilitate the quick searching thus, binary code based document retrieval ensures the accurate as well as faster results than the semantic-based analysis. Semantic hashing technique [8] is the category of describing the document in one word of memory that contains semantically similar documents. It facilitates the effective information retrieval on large-scale document collections concerning the retrieval time, which are independent of the size of the document. Semi-supervised learning based deep learning approach [9] learns the parameters of the deep learning model based on both the supervised and unsupervised data, which is more beneficial for the large-scale environment. It leads an efficient computation while generating representations for indexing as well as using indexing by reducing the computation time and storage. Vector space model based word representation model [10] employs artificial neural networks to learn the distributed representation of high-quality word vectors from the huge datasets that contain hundreds of millions of words and millions of distinct words.

*Discriminative tasks and semantic tagging:*In big data analytics, deep learning algorithms extract the complicated nonlinear features and consequently, exploit the simple linear models to perform the discriminative tasks. The main advantages of the deep learning methods for big data analytics involve i) the deep learning based feature extraction adds the non-linearity while analyzing the data and ii) applying relatively linear analytical models on the extracted features improves the computational efficiency. The discriminative analysis is the crucial process in big data analytics to facilitate the tagging in terms of semantic tagging of the data. Data tagging is the process of semantically indexing the data corpus, which differs from the term ‘semantic indexing’. Semantic indexing directly utilizes the deep learning of abstract representations for data indexing. Whereas, the semantic tagging considers the abstract representations as the features to perform the discriminative task of data tagging.

Imagenet classification model [11] employs the deep convolutional neural networks to improve the image search in the image object recognition. It demonstrates the significance of the deep learning method in the image object recognition. Also, the deep learning methods have been exploited to learn very high-level features for detecting the images. Accordingly, Google and Stanford [12] learn very high-level features such as the face detection through unsupervised learning method in a large-scale environment. In the perspective of image tagging, recursive neural network based approach [13] predicts the tree structure of the images in multiple modalities, which obtains the precise results in segmenting and annotating the complex images. It has the ability to the hierarchical tree structure of the image scenes, which parses the natural scenes and natural language sentences. Recurrent neural networks construct the significant search space through deep learning, which facilitates the data based Web design and Web search [14]. Action scene recognition and video data tagging applications rely on an independent variant analysis of deep learning while learning the invariant spatiotemporal features from video data. By exploiting the stacking as well as convolution deep learning techniques, the video tagging approach learns the hierarchical representations.

II.APPROACHES ON VARIOUS DEEP LEARNING MODELS

In big data mining, machine learning methods have been utilized as either supervised or unsupervised learning algorithms. The deep learning algorithm is one of the remarkable machine learning algorithms for providing the solutions for big data analytics. Neural networks are also a type of the learning algorithm that creates the underlying support for the most of the deep learning methods. In the past few years, various Deep learning models have been developed primarily, the four deep architectures, including Stacked Auto-Encoder (SAE), Deep Belief Network (DBN),Convolutional Neural Network (CNN), and Recurrent Neural Network (RNN) [15].

A. Stacked auto-encoder

In recent years, a large-scale deep unsupervised learning model [16] has been utilized to train a deep architecture by exploiting the sparse deep autoencoder, local receptive fields, local contrast normalization, and pooling. Autoencoder is another type of Artificial Neural Network (ANN), termed as autoassociator. It has been widely used for the dimensionality reduction in a large-scale data, which is an unsupervised deep learning model [17, 18]. In recent years, autoencoders have been utilized to learn the generative data models [19]. Deep Stacking Network (DSN) [20] incorporates several neural networks associated with a single hidden layer, which is a modified deep architecture. This network formation is based on the raw vector data of the stacked modules and the output of the previous modules in which module refers to the specialized neural network. Later, most of the researchers focus on utilizing the Tensor Deep Stacking Network (T-DSN) [21] as the deep architecture based on the DSN. To support the domain adaptation, several unsupervised and semi-supervised methods have been presented. Denoising auto-encoder [22] captures the invariant latent factors for data reconstruction. By exploiting the transformed labeled data in the source domain, it trains the classifier in which the training of denoising auto-encoder depends on both the domain information.

Multi-modal Deep Boltzmann Machine (DBM) [23] constructs a stacked Restricted Boltzmann Machines (RBM) for each data modal and consequently, integrates the text data with a real-value dense image to obtain an adequate multi-modal data representation during learning. Deep learning model [24] automatically extracts the critical features of the raw input data and learns the invariant feature hierarchies by exploring the variations in the multiple levels of representation. Stacked Extreme Learning Machine (S-ELM) [25] efficiently handles the large-scale data by transforming the vast network of ELM into a serially connected stacked ELMs. Deep computation model [26] presents a model for the complex correlations of the diverse data with the assistance of tensor model that eases the feature learning in the massive data collections. Its distributed learning paradigm is more beneficial to both the RBMs and Back Propagation (BP) algorithm.

B. Deep belief network

Deep learning algorithms exponentially meet the difficulties with the increase of the data size that is big data environment [27]. A stacked auto-encoder model is the feed-forward neural network. Auto-encoder consists of input and output layer through one or more hidden layers, which has been used to reconstruct the input and resolve the difficulties in unsupervised learning and transfer learning process [28, 29]. A deep belief network is the first successful deep learning model, which is stacked by several RBMs in contrast to the stacked auto-encoder. Deep belief networks have been widely used in various applications such as image classification [30], acoustic modeling, traffic flow prediction [31], and so on. A deep architecture [32] tackles the constraints in learning scalability, referred to as the Deep Convex Networks (DCNs). It consists of multiple layered modules in which one module includes a single hidden layer with two sets of weights in a specific neural network. The lowest module comprises two linear layers and one non-linear layer. The batch mode based learning in DCN leads the parallel training. The complementary priors are necessary to resolve the inference in densely connected belief networks. Deep neural networks and cortical networks employ the brain-inspired processor architectures as the supportive models [33]. An image retrieval method [34] is based on the DBNs and Softmax classifier while retrieving similar images from the database. By exploiting the automated feature extraction method, the standard Content-Based Image Retrieval (CBIR) algorithm retrieves the identical images based on the DBN-Softmax model that effectively provides the extraction measurement and valid representation.

C. Convolutional neural network

Deep transfer learning [35] learns and transfers the mid-level image representations through the convolutional neural network. In this deep transfer learning model, the pre-training process exploits the source and fine-tuning utilizes the labels in the target domain. Multi-task deep neural network [36] presents representation learning and extracts the domain-invariant features in deep architectures to facilitate the semantic classification and information retrieval. To effectively deal with the large-scale image classification and recognition, the recent researchers [37, 38] widely employ the convolutional neural network for feature learning. The convolutional neural network comprises three layers such as a convolutional layer, fully-connected layer, and sub-sampling layer.

Feature pooling method [39] obtains the invariance while recognizing the image transformations, which ensures the robustness against the noisy data. The performance of the different pooling methods relies on several factors involve the resolution and links between the sample cardinalities. Multi-way local pooling model [40] gathers the features together, even the features are widely dissimilar, which delivers better performance through clustering before the pooling stage. By applying receptive field learning, the image feature learning system achieves better pooling performance especially, by utilizing the concept of over-completeness. This efficient learning algorithm effectively accelerates the training process for incremental feature selection [41]. Convolutional neural network based pooling method is termed as Lp pooling [42] that obtains high accuracy on house number digit classification. Lp pooling refers to a biological model which is inspired by the complex cells. A stochastic pooling method [43] regularizes the large convolutional neural network to present a stochastic pooling procedure in each layer of the convolutional network. It randomly selects the activation in each pooling region based on the multinomial distribution, which assists in computing the deformations in a multi-layer model.

D. Recurrent neural network

Recurrent Neural Network meets the difficulty in capturing the long-term dependency due to the gradient vanishing with the back-propagation strategy while training the parameters. Several research works focus on tackling this constraint by presenting the long short-term memory to prevent the gradient vanishing. The recurrent neural networks have been used in many applications such as speech recognition, machine translation, and natural language processing [44]. Hierarchical Recurrent Neural Hashing (HRNN) approach [45] generates effective hash codes by exploiting both the spatial and semantic information, which enforces the hierarchical convolutional features to build the pyramid representation of the images. It improves the performance of the image retrieval through hash codes. An approach [46] recognizes the multivariate time series of the clinical measurements by exploiting the RNNs with Long Short-Term Memory (LSTM). It classifies the diagnosis of patients by analyzing the clinical measurements in the pediatric intensive care unit. DeepCaremodel [47] employs the hidden units of RNNs with LSTM, pooling, and word embedding, which is an end-to-end deep dynamic network. It explores the current illness states of the patients and predicts the future medical results. It can handle the irregular time events by moderating the LSTM unit with a decay effect.

III. REVIEW OF DEEP LEARNING APPROACHES: TRENDS AND PERSPECTIVES

In recent years, the deep learning architectures have been widely applied in various research perspectives such as speech recognition and acoustic modeling. Also, in the perspective of audio classification [48], and image processing area such as high-resolution remote sensing scene classification [49], handwritten classification [50], multi-category rapid serial visual presentation Brain-Computer Interfaces (BCI) [51], and single image super-resolution (SR) [52]. Moreover, the architectures of deep networks have also been applied in multi-task learning for Natural Language Processing (NLP) with the inference robustness [53].

Speech Recognition: In the past few decades, machine learning algorithms have been widely applied in areas such as acoustic modeling and Automatic Speech Recognition (ASR). Automatic speech recognition pertains to the standard classification model in terms of identifying the word sequences from the speech waveforms. Several machine learning algorithms have provided the most promising results in automatic speech recognition in which machine learning algorithms include Support Vector Machine (SVM) and Neural Network (NN). Long Short-Term Memory based deep learning model is often applied to the speech recognition process.

*Pattern recognition and Visual art processing:*The processes of pattern recognition involves regression, classification, speech tagging, and sequence labeling. Moreover, the main areas of computer vision involve the scene reconstruction, object posture estimation, object detection and recognition, image restoration, even detection, image editing, statistical learning, and video enhancement. CNN based deep learning architecture automatically selects the features of the Graphics Processing Unit (GPU) based computational resource. However, several deep neural networks fail to recognize the unrecognizable images [54]. Moreover, the deep learning methods have also been exploited in human-robot interaction [55] and Glaucoma detection systems [56] with better results.

*Natural language processing:*The earlier research works employ the neural networks to implement the language models and improves the language modeling with the help of LSTM [57]. Moreover, in a deep learning architecture, the word embedding such as word2vec model has also been used to provide the deep representation of the data. RNN can parse the sentences as well as phrases with the support of compositional vector grammar. Recursive auto-encoders detect the paraphrasing based on the sentence similarity by building the top word embedding [58].

*Drug discovery and Recommendation systems:*Deep learning research predicts the biomolecular target in a large-scale environment. AtomNet deep learning system [59] predicts the new biomolecules regarding the disease targets such as multiple sclerosis and Ebola virus, which is used for structure-based rational drug design. Recommendation systems also employ the deep learning algorithms and then, extracts the essential features for content-based music recommendation. Moreover, Multi-view deep learning model learns the user preferences from a variety of domains by exploiting the hybrid content and collaborative-based recommendations and enriches the recommendation accuracy [60].

IV. APPLICATION-FOCUSED DEEP LEARNING APPROACHES

Nowadays, Deep learning methods play a vital role in big data analytics especially internet of things analytics. In the big data analytics, collecting the raw data from the devices and applying the preprocessing is the complex task. Also, dynamically monitoring the sensor data is also expensive and challenging task on the internet of things. Deep learning methods resolve these types of constraints through various levels of representation of data. From lower level to higher level features representations, higher level features present the most significant contextual information than lower level features. This emerging deep learning methodology has been widely applied in different application fields such as natural language processing, image caption, art, machine translation, robotics, visual tracking, and object detection [27]. Deep learning methods have been used for various purposes such as medicine and biology, media and entertainment, security and defense, autonomous machine, and internet and cloud categories. The specific applications of general purpose medicine and biology field include drug discovery, diabetic grading, cancel detection and so on. The applications of media and entertainment fields include video search, video captioning, real-time translation, and so on. Security and defense systems involve video surveillance, face detection, satellite imagery, and so on. The deep learning based autonomous machine includes lane tracking, traffic sign recognition, pedestrian detection and so on. The Internet and cloud-based deep learning applications involve speech recognition, image classification, language translation, language processing, sentiment analysis, and so on. Most of the researchers introduce the different approaches to various deep learning models for a variety of applications [19]. Table 1 reviews the existing deep learning approaches based on deep learning models and application fields.

Table 1: Review of the Deep learning models applied to different domains in literature

Deep learning model	***Data***	***Approach***	***Remarks***
Stacked Auto Encode	Clinical imaging	Early diagnosis of Alzheimer’s disease with deep learning [61]	It assists in the diagnosis of Alzheimer’s disease and its prodromal stage and Mild Cognitive Impairment (MCI)
	Electronic health records	Deep patient: An unsupervised representation to predict the future of patients from the electronic health records [62]	It predicts the patient status in future by assessing the probability of the patients in developing the various diseases
		Deep Learning to predict future patient diseases from the electronic health records [63]	By exploiting the clinical records of the patients, it predicts the future diseases of the patients
	Speech Emotion Recognition	Sparse Autoencoder-based Feature Transfer Learning for Speech Emotion Recognition [64]	This single layer auto-encoder learns the emotion-specific mapping rule from the labeled data
	Real-Time Face	Coarse-to-Fine Auto-Encoder	It cascades several successive Stacked Auto-encoder

	Alignment	Networks (CFAN) for Real-Time Face Alignment [65]	Networks to detect the face landmark
	Face Recognition	Stacked Progressive Auto-Encoders (SPAE) for Face Recognition Across Poses [66]	It employs multiple shallow progressive auto-encoders to convert the non-frontal pose to frontal pose
Convolutional Neural Network	Clinical Imaging	Dermatologist-level classification of skin cancer with deep neural networks [67]	It reveals the capability of artificial intelligence in classifying skin cancer with a level of competence to dermatologists
	Genomes	Denoising genome-wide histone ChIP-seq with convolutional neural networks [68]	It handles the variety of sources of noise as well as variability by learning a mapping from sub-optimal to high-quality histone ChIP-seq data.
	Fingerprint	Deep convolutional neural network for latent fingerprint enhancement [69]	This multi-task learning consists of one common convolution part shared by two different deconvolution parts referring to the enhancement branch and orientation branch respectively
Deep Belief Network	Electronic health records	Learning vector representation of medical objects via EMR-driven non-negative restricted Boltzmann machines (eNRBM) [70]	It predicts suicide risk of mental health patients through low dimensional representations
	Speaker Recognition	Restricted Boltzmann machines for vector representation of speech in speaker recognition [71]	The RBM learns the overall speaker and session variability among the background Gaussian Mixture Models (GMM) supervectors
	Traffic accident detection	A deep learning approach for detecting traffic accidents from social media data [72]	It employs Restricted Boltzmann machines as the deep belief network for analyzing the accident related tweets
Recurrent Neural Network	Electronic health records	Doctor ai: Predicting clinical events via recurrent neural networks [73]	It exploits skip-gram embedding method to predict the diagnosis, medical conditions, medication order, and visit time
		Learning to diagnose with LSTM recurrent neural networks [46]	It classifies the diagnosis based on the clinical measurements of patients in pediatric intensive unit care
	Keyword spotting in Speech decoding	Robust discriminative keyword spotting for emotionally colored spontaneous speech using bidirectional LSTM networks [74]	It exploits Bidirectional Long Short-Term Memory (BLSTM) recurrent neural nets to decode the speech data contextually
	Human Activity Recognition	Deep, convolutional, and recurrent models for human activity recognition using wearables [75]	It captures the movement of the persons by analyzing the wearable sensors
	Time-series Applications such as Weather data and stock market prices.	Multi-context recurrent neural network for time series applications [76]	It employs three algorithms such as backpropagation, backpropagation through time, and dynamic online learning, tested on energy load forecasting and handwriting recognition

			applications

V. COUNTERMEASURES FOR BIG DATA DEEP LEARNING CHALLENGES

Even though Deep learning algorithms emphasizes more benefits and applicability for big data analytics, several characteristics associated with big data poses challenges towards the need of modifying and adapting the deep learning algorithms to resolve the issues. Accordingly, the big data deep learning requires further development, especially dealing with high-dimensional data, streaming data, and large-scale data. Big data analytics involves mining and extracting the essential representations from a huge amount of raw data for prediction, decision-making, and other inferences. The prior research works present the deep learning approaches from different perspectives to deal with the big data challenges [15].

A. Approaches for the huge amount of data

In recent years, the DistBelief framework has been developed to train the large-scale deep learning model in a parallel manner over a large number of machines. It effectively trains the large-scale models based on the combination of data and model parallelism. DistBelief framework significantly improves the training efficiency by partitioning the large deep model into the blocks [77, 78]. Several deep learning models employ the clusters of Central Processing Unit (CPU) and GPUs with the increase of training speed without compromising the deep learning accuracy. Moreover, various strategies have been developed for the data as well as model parallelism. By exploiting the forward and backward propagation, the deep learning model partitions the data and models into blocks with in-memory data [77, 79]. To alleviate the impact of the noisy labels in the big data, the deep learning models employ the semi-supervised learning method which is a more cost-effective training strategy [80, 81]. In big data analytics, handling streaming and fast-moving data is an arduous task, especially in monitoring tasks such as traffic monitoring and fraud detection. Several deep learning algorithms such as denoising autoencoders [82], deep belief networks [83], and incremental feature learning and extraction [84] deal with a large amount of continuous data. An adaptive deep belief network [83] generalizes the deep learning model to learn the online non-stationary and streaming data. It employs the generative property of deep belief network and consequently, imitates the samples from the original data, which are exploited to learn the new deep belief networks that are adaptive to the newly observed data. However, it requires the constant memory consumption. An incremental feature learning method [84] rapidly converges to the optimal number of features in massive online data streams. Moreover, it can adapt to new stream in a large-scale data while ignoring the expensive cross-validation analysis.

B. Approaches for heterogeneous data

Several multi-modal deep learning models have been presented in the learning of heterogeneous data representation. A multi-modal deep learning model [85] employs the RBMs to separately learn the features and representations of audio and video objects. Multi-source deep learning model [86] learns the non-linear representation for human pose estimation, which is one of the multi-modal deep learning model. Marginalized Stacked Denoising Autoencoder (mSDA) [87] effectively scales with a high-dimensional data, which increases the speed of the computational process than the traditional SDAs. It learns the parameters without requiring the stochastic gradient descent algorithms by marginalizing the noise during SDA training. Multi-modal deep learning [88] learns the variety of representations through audio and video data integration based application for deep learning. It reveals that the deep learning is an effective model for learning a single modality representation by unlabeled data based multiple modalities and learning shared representations across multiple modalities. Later, multimodal Deep Boltzmann Machine (DBM) integrates two different data modalities such as real-valued text data and dense image data with numerous word frequencies [23]. Even though multi-modal deep learning models achieve better performance than the stacked auto-encoders and deep belief networks in learning the heterogeneous data, they integrate the learned features in a linear way regarding each modality, which leads difficult process in capturing the complex correlations over multiple modalities. To overcome this constraint, several previous research works present a tensor deep learning a model for heterogeneous data, termed as deep computation [26, 89].

C. Approaches for real-time data

In big data feature learning, online learning methods [86, 90] are considered as the efficient method due to the non-essential of retraining the parameters on the old training objects. Even though the incremental schemes are more suitable for voluminous big data, itobtains poor performance due to the drastic change of the data distribution over the time. Another kind of incremental learning method has been used to tackle this issue with the support of structural modification. In essence, a structure based incremental auto-encoder model implements the learning model and adapts to new arriving objects by adding one or more neurons in the hidden layer [91, 92]. Online learning model learns one instance at a time, which is a sequential learning strategy that can be used to refine the model in a big data environment [93]. The big data analytics approach to often deal with the non-stationary data that is data distribution over time. Deep online learning model needs to readily parallelizable, and memory bounded to handle the piece-wise stationary data with respect the degree of correlation due to the presence of separate chunks with data in a small time interval [94, 95]. Analyzed temperature data set using machine learning algorithms [96]. Table 2 shows several existing deep learning approaches along with its merits and demerits in a big data environment.

Table 2: Comparison of existing Deep learning approaches in Big data

Title	***Method***	***Application***	***Merits***	***Demerits***
Learning cascaded deep auto-encoder networks for face alignment [28]	Global Exemplar-based Deep Auto-encoder Network (GEDAN)	Face Alignment	It achieves greater robustness against pose variations It increases the capacity of pose estimation	It has a high computation complexity of real-world applications due to its brute-force-style deep learning
Unseen noise estimation using separable deep auto encoder for speech enhancement [29]	Separable Deep Auto Encoder (SDAE).	Speech enhancement	It estimates unseen noise spectrum for speech enhancement	Achieving performance improvement is the challenging task due to the requirement of additional training data
An Improved Bilinear Deep Belief Network Algorithm for Image Classification [30]	Multiple Kernel Learning (MKL) based improved Bilinear Deep Belief Network (BDBN)	Image Classification	It optimally finds the combinations of hierarchical features pre-trained by a BDBN	It requires the large amount of training data, which increases the burden of processing
Deep architecture for traffic flow prediction: deep belief networks with multitask learning [31]	Multitask regression layer for Multi-Task Learning (MTL)	Traffic Flow Prediction	It deals with the complicated relations. It learns the features using less prior knowledge	While increasing the number of layers, it leads the error gradient vanishing
Deep convolutional neural networks for predominant instrument recognition in polyphonic music [38]	By exploiting ConvNet architecture, it aggregates multiple outputs from sliding windows over the test audio	Predominant instrument recognition in polyphonic music	It prevents 'dead' activation of initially inactive units	It increases the computational complexity
Multilingual acoustic models using distributed deep neural networks [78]	Multilingual DNN	Speech recognition	It presents the acoustic model for cross- and multi-lingual network training	It has limited scalability
Multi-source deep learning for human pose estimation [85]	Unified deep model	Human pose identification	It jointly learns the task of estimating the body locations and human detection	It lacks to handle cross-view photometric and geometric transforms
Simnest: Social media nested epidemic simulation via online semi-supervised deep learning [86]	Semi-supervised deep learning framework	Health status and intervention action identification of Social media users	It specifically identifies the ongoing events such as and disease outbreaks	It lacks to predict the event in future
Integrating Online and Offline Three-Dimensional Deep Learning for Automated Polyp Detection in Colonoscopy Videos [90]	3-dimensional fully convolutional network	Endoscopy image analysis	It improves the detection performance by effectively reducing the number of false positives	It requires a huge amount of labeled data of positivity that is polyp frame

VI. KEY CHALLENGES AND FUTURE DIRECTIONS

Even though numerous countermeasure approaches have dealt with the deep learning challenges, the deep learning researches have still encountered issues due to certain characteristics associated with Big data while adopting the deep learning algorithm for the big data environment. Deep learning research work confronts several key challenges of big data discussed as follows.

- *Big Data Preprocessing:* In a big data environment, maintaining the data quality is the challenging task due to the characteristics of the big data. Also, relying on any one data cleaning method for deep learning on big data lacks to achieve the best quality for providing the error-free data. To achieve the data quality, the data cleansing method needs to know the requirements of the organization.
- *Big Data Analytics:*Big data analytics involves the database mining, searching, and analysis. Big data mining that includes feature extraction with deep learning is a challenging area due to the data complexity and scalability. With the rise of data heterogeneity and the vast amount of data, analyzing the data by trial-and-error methods is an arduous process, which necessitates the scalable algorithms.

- *Big Data Management and Integration:*Big data management requires the rules, law, and control over data to effectively manage the data concerning the policies and rules of the data in the dynamic environment. In this scenario, the primary research problem is providing the decision-making mechanism. Also, integrating the various types of data is a complex, according to the different applications.
- *Semantic Indexing:*Owing to the massive amount of data and low storage capacity, storing the semantic indexes is essential rather than storing the data as raw data. Deep learning has been used for semantic indexing to provide better information retrieval. However, semantic indexing is still an open challenge for big data.

In future, the research on deep learning needs further analysis to resolve the specific data analysis problems in Big data. The future directions have focused on improving the machine learning, the high-level abstractions formulation, and the big data representation.

- The future direction of this work will project data from different modalities into a common latent space for effective information retrieval. Hence, applying the effective fusion methods for learning features of the diverse data types is essential to enhance the multi-modal deep learning model.
- Predict the necessary volume of input data to train the significant data representations by the deep learning algorithms. It addresses the problem of utilizing the entire Big data input corpus to train the high-level data representation patterns.
- Domain adaptation is a crucial aspect while applying the deep learning when the distribution of the training data is varied from the distribution of the test data, which targets to resolve the variability of the data types and the big data domains during big data analytics.
- With the increase of non-linearity in the Big data, understanding the characteristics of the data and predicting the mean and variance have become the crucial importance to deal with the uncertainty while applying deep learning algorithm.

VII. CONCLUSION

Nowadays, deep learning has an advantage of providing the high-level data abstraction for Big data with the aim of extracting better abstract knowledge. In contrast to the conventional feature engineering and machine learning algorithms, deep learning potentially provides the solutions for data analysis and learning problems. This paper described a brief review of deep learning in a big data environment, involving the existing deep architectures, deep learning approaches, and several countermeasures. Also, it reviews the existing deep learning approaches for the big data applications. This review reveals significant factor that learning the feature hierarchies and deep architectures are highly desirable in the big data environment. This work provides the insights to understand the various deep learning systems and the current critical challenges in the deep learning research, which paves the way to develop the deep learning algorithms under various circumstances. Finally, this work provides several possible future directions for the open research problems in big data deep learning.

REFERENCES

[1] Gheisari, Mehdi, Guojun Wang, and MdZakirulAlamBhuiyan, “A Survey on Deep Learning in Big Data”, IEEE International Conference on Computational Science and Engineering (CSE) and Embedded and Ubiquitous Computing (EUC), Vol.2, pp.173-180, 2017

[2] BengioYoshua, “Deep learning of representations: Looking forward”, In International Conference on Statistical Language and Speech Processing, Springer, pp.1-37, 2013

[3] Yu Dong, and Li Deng, “Deep learning and its applications to signal and information processing”, IEEE Signal Processing Magazine, Vol.28, No.1, pp.145-154, 2011

[4] Hinton Geoffrey E., Simon Osindero, and Yee-WhyeTeh, “A fast learning algorithm for deep belief nets”, Neural computation, Vol.18, No.7, pp.1527-1554, 2006

[5] BengioYoshua, Pascal Lamblin, Dan Popovici, and Hugo Larochelle, “Greedy layer-wise training of deep networks”, In Advances in neural information processing systems, pp.153-160, 2007

[6] Hinton Geoffrey E, “Training products of experts by minimizing contrastive divergence”, Neural computation, Vol.14, No.8, pp.1771-1800, 2002

[7] Hinton, Geoffrey, and RuslanSalakhutdinov, “Discovering binary codes for documents by learning deep generative models”, Topics in Cognitive Science, Vol.3, No.1, pp.74-91, 2011

[8] SalakhutdinovRuslan, and Geoffrey Hinton, “Semantic hashing”, International Journal of Approximate Reasoning, Vol.50, No.7, pp.969-978, 2009

[9] RanzatoMarc'Aurelio, and Martin Szummer, “Semi-supervised learning of compact document representations with deep networks”, ACM Proceedings of the 25th international conference on Machine learning, pp.792-799, 2008

[10] Mikolov Tomas, Kai Chen, Greg Corrado, and Jeffrey Dean, “Efficient estimation of word representations in vector space”, arXiv preprint arXiv:1301.3781, 2013

[11] Krizhevsky Alex, IlyaSutskever, and Geoffrey E. Hinton, “Imagenet classification with deep convolutional neural networks”, In Advances in neural information processing systems, pp.1097-1105, 2012

[12] Le, Quoc V, “Building high-level features using large scale unsupervised learning”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.8595-8598, 2013

[13] Socher Richard, Cliff C. Lin, Chris Manning, and Andrew Y. Ng, “Parsing natural scenes and natural language with recursive neural networks”, In Proceedings of the 28th international conference on machine learning (ICML-11), pp.129-136, 2011

[14] Kumar Ranjitha, Jerry O. Talton, Salman Ahmad, and Scott R. Klemmer, “Data-driven Web Design”, In ICML, 2012

[15] Chen, Xue-Wen, and Xiaotong Lin, “Big data deep learning: challenges and perspectives”, IEEE access, Vol.2, pp.514-525, 2014

[16] Raina, Rajat, AnandMadhavan, and Andrew Y. Ng, “Large-scale deep unsupervised learning using graphics processors”, ACM

Proceedings of the 26th annual international conference on machine learning, pp.873-880, 2009
[17] Schmidhuber Jürgen, “Deep learning in neural networks: An overview”, Neural networks, Vol.61, pp.85-117, 2015
[18] BengioYoshua, “Learning deep architectures for AI”, Foundations and trends® in Machine Learning, Vol.2, No.1, pp.1-127, 2009
[19] Deng Li, “Three classes of deep learning architectures and their applications: a tutorial survey”, APSIPA transactions on signal and information processing, 2012
[20] Deng, Li, Dong Yu, and John Platt, “Scalable stacking and learning for building deep architectures”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.2133-2136, 2012
[21] Hutchinson, Brian, Li Deng, and Dong Yu, “Tensor deep stacking networks”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.35, No.8, pp.1944-1957, 2013
[22] X. Glorot, A. Bordes, and Y. Bengio, “Domain adaptation for large-scale sentiment classification: A deep learning approach,” in Proceedings of International Conference on Machine Learning (ICML), pp.513–520, 2011
[23] Srivastava, Nitish, and Ruslan R. Salakhutdinov, “Multimodal learning with deep boltzmann machines”, In Advances in neural information processing systems, pp.2222-2230, 2012
[24] LeCun, Yann, “Learning invariant feature hierarchies”, In Computer vision–ECCV 2012. Workshops and demonstrations, Springer, pp.496-505, 2012
[25] Zhou, Hongming, Guang-Bin Huang, Zhiping Lin, Han Wang, and Yeng Chai Soh. “Stacked extreme learning machines”, IEEE transactions on cybernetics, Vol.45, No.9 pp.2013-2025, 2015
[26] Zhang, Qingchen, Laurence T. Yang, and Zhikui Chen, “Deep computation model for unsupervised feature learning on big data”, IEEE Transactions on Services Computing, Vol.9, No.1, pp.161-171, 2016
[27] Najafabadi, Maryam M., FlavioVillanustre, Taghi M. Khoshgoftaar, NaeemSeliya, Randall Wald, and EdinMuharemagic, “Deep learning applications and challenges in big data analytics”, Journal of Big Data, Vol.2, No.1, p.1, 2015
[28] R. Weng, J. Lu, Y. Tan, J. Zhou, “Learning cascaded deep auto-encoder networks for face alignment”, IEEE Transactions on Multimedia, Vol.18, No.10, pp.2066–2078, 2016
[29] M. Sun, X. Zhang, H.V. Hmme, T.F. Zheng, “Unseen noise estimation using separable deep auto encoder for speech enhancement”, IEEE/ACM Transactions on Audio, Speech, and Language Processing, Vol.24, No.1, pp.93–104, 2016
[30] Jie, Niu, Bu Xiongzhu, Li Zhong, and Wang Yao, “An Improved Bilinear Deep Belief Network Algorithm for Image Classification”, IEEE Tenth International Conference on Computational Intelligence and Security (CIS), pp.189-192, 2014
[31] Huang, Wenhao, Guojie Song, Haikun Hong, and KunqingXie, “Deep architecture for traffic flow prediction: deep belief networks with multitask learning”, IEEE Transactions on Intelligent Transportation Systems, Vol.15, No.5, pp.2191-2201, 2014
[32] Deng Li, and Dong Yu, “Deep convex net: A scalable architecture for speech pattern classification”, In Twelfth Annual Conference of the International Speech Communication Association, 2011
[33] Indiveri, Giacomo, and Shih-Chii Liu, “Memory and information processing in neuromorphic systems”, Proceedings of the IEEE, Vol.103, No.8, pp.1379-1397, 2015
[34] Liao Bin, JungangXu, Jintao Lv, and Shilong Zhou, “An image retrieval method for binary images based on DBN and softmax classifier”, IETE Technical Review, Vol.32, No.4, pp.294-303, 2015
[35] OquabMaxime, Leon Bottou, Ivan Laptev, and Josef Sivic, “Learning and transferring mid-level image representations using convolutional neural networks”, In Proceedings of the IEEE conference on computer vision and pattern recognition, pp.1717-1724, 2014
[36] X. Liu, J. Gao, X. He, L. Deng, K. Duh, and Y.-Y. Wang, “Representation learning using multi-task deep neural networks for semantic classification and information retrieval,” in Proc. of Conference of North American Chapter of Association for Computational Linguistics (NAACL), 2015
[37] Maggiori, Emmanuel, YuliyaTarabalka, Guillaume Charpiat, and Pierre Alliez, “Convolutional neural networks for large-scale remote-sensing image classification”, IEEE Transactions on Geoscience and Remote Sensing, Vol.55, No.2, pp.645-657, 2017
[38] Han, Yoonchang, Jaehun Kim, Kyogu Lee, Yoonchang Han, Jaehun Kim, and Kyogu Lee, “Deep convolutional neural networks for predominant instrument recognition in polyphonic music”, IEEE/ACM Transactions on Audio, Speech and Language Processing (TASLP), Vol.25, No.1, pp.208-221, 2017
[39] Boureau Y-Lan, Jean Ponce, and YannLeCun, “A theoretical analysis of feature pooling in visual recognition”, In Proceedings of the 27th international conference on machine learning (ICML-10), pp.111-118, 2010
[40] Boureau Y-Lan, Nicolas Le Roux, Francis Bach, Jean Ponce, and YannLeCun, “Ask the locals: multi-way local pooling for image recognition”, IEEE International Conference onn Computer Vision (ICCV), ,pp.2651-2658, 2011
[41] JiaYangqing, Chang Huang, and Trevor Darrell, “Beyond spatial pyramids: Receptive field learning for pooled image features”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.3370-3377, 2012
[42] Sermanet Pierre, SoumithChintala, and YannLeCun, “Convolutional neural networks applied to house numbers digit classification”, IEEE 21st International Conference on Pattern Recognition (ICPR), pp.3288-3291, 2012
[43] Zeiler, Matthew D., and Rob Fergus, “Stochastic pooling for regularization of deep convolutional neural networks”, arXiv preprint arXiv:1301.3557, 2013
[44] Chien Jen-Tzung, and Yuan-Chu Ku, “Bayesian recurrent neural network for language modeling”, IEEE transactions on neural networks and learning systems, Vol.27, No.2, pp.361-374, 2016
[45] Lu Xiaoqiang, Yaxiong Chen, and Xuelong Li, “Hierarchical recurrent neural hashing for image retrieval with hierarchical convolutional features”, IEEE Transactions on Image Processing, Vol.27, No.1, pp.106-120, 2018
[46] Lipton, Zachary C., David C. Kale, Charles Elkan, and Randall Wetzell, “Learning to diagnose with LSTM recurrent neural networks”, arXiv preprint arXiv:1511.03677, 2015
[47] Pham Trang, Truyen Tran, DinhPhung, and SvethaVenkatesh, “Deepcare: A deep dynamic memory model for predictive medicine”, In Pacific-Asia Conference on Knowledge Discovery and Data Mining, Springer, pp.30-41, 2016
[48] Lee Honglak, Peter Pham, Yan Largman, and Andrew Y. Ng, “Unsupervised feature learning for audio classification using convolutional deep belief networks”, In Advances in neural information processing systems, pp.1096-1104, 2009
[49] Hu Fan, Gui-Song Xia, Jingwen Hu, and Liangpei Zhang, “Transferring deep convolutional neural networks for the scene classification of high-resolution remote sensing imagery”, Remote Sensing, Vol.7, No.11, pp.14680-14707, 2015
[50] Kim In-Jung, and XiaohuiXie, “Handwritten Hangul recognition using deep convolutional neural networks”, International Journal on Document Analysis and Recognition (IJDAR), Vol.18, No.1, pp.1-13, 2015
[51] Manor Ran, and Amir B. Geva, “Convolutional neural network for multi-category rapid serial visual presentation BCI”, Frontiers in computational neuroscience, Vol.9, 2015
[52] Dong Chao, Chen Change Loy, Kaiming He, and Xiaoou Tang, “Image super-resolution using deep convolutional networks”,

IEEE transactions on pattern analysis and machine intelligence, Vol.38, No.2, pp.295-307, 2016

[53] Lane Nicholas D., and PetkoGeorgiev, “Can deep learning revolutionize mobile sensing?”, ACM Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications, pp.117-122, 2015

[54] Nguyen Anh, Jason Yosinski, and Jeff Clune, “Deep neural networks are easily fooled: High confidence predictions for unrecognizable images”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.427-436, 2015

[55] Kelley Richard, Liesl Wigand, Brian Hamilton, Katie Browne, Monica Nicolescu, and MirceaNicolescu, “Deep networks for predicting human intent with respect to objects”, In Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction, pp.171-172, 2012

[56] Chen Xiangyu, YanwuXu, Damon Wing Kee Wong, Tien Yin Wong, and Jiang Liu, “Glaucoma detection based on deep convolutional neural network”, In Engineering in Medicine and Biology Society (EMBC), 37th Annual International Conference of the IEEE, pp.715-718, 2015

[57] SutskeverIlya, OriolVinyals, and Quoc V. Le, “Sequence to sequence learning with neural networks”, In Advances in neural information processing systems, pp.3104-3112, 2014

[58] Socher Richard, YoshuaBengio, and Chris Manning, “Deep learning for NLP”, Tutorial at Association of Computational Logistics (ACL), and North American Chapter of the Association of Computational Linguistics (NAACL), 2013

[59] Wallach Izhar, Michael Dzamba, and Abraham Heifets, “AtomNet: a deep convolutional neural network for bioactivity prediction in structure-based drug discovery”, arXiv preprint arXiv:1510.02855, 2015

[60] Elkahky Ali Mamdouh, Yang Song, and Xiaodong He, “A multi-view deep learning approach for cross domain user modeling in recommendation systems”, In Proceedings of the 24th International Conference on World Wide Web, pp.278-288, 2015

[61] Liu Siqi, Sidong Liu, WeidongCai, Sonia Pujol, Ron Kikinis, and Dagan Feng, “Early diagnosis of Alzheimer's disease with deep learning”, IEEE 11th International Symposium on Biomedical Imaging (ISBI), pp.1015-1018, 2014

[62] Miotto Riccardo, Li Li, Brian A. Kidd, and Joel T. Dudley, “Deep patient: An unsupervised representation to predict the future of patients from the electronic health records”, Scientific reports, Vol.6, p.26094, 2016

[63] Miotto Riccardo, Li Li, and Joel T. Dudley, “Deep Learning to predict patient future diseases from the electronic health records”, In European Conference on Information Retrieval, Springer, pp.768-774, 2016

[64] Deng Jun, Zixing Zhang, Erik Marchi, and Bjorn Schuller, “Sparse autoencoder-based feature transfer learning for speech emotion recognition”, IEEE Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII), pp.511-516, 2013

[65] Zhang Jie, Shiguang Shan, MeinaKan, and Xilin Chen, “Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment”, In European Conference on Computer Vision, Springer, pp.1-16, 2014

[66] KanMeina, Shiguang Shan, Hong Chang, and Xilin Chen, “Stacked progressive auto-encoders (spae) for face recognition across poses”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1883-1890, 2014

[67] Esteva Andre, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun, “Dermatologist-level classification of skin cancer with deep neural networks”, Nature, Vol.542, No.7639, pp.115-118, 2017

[68] Koh Pang Wei, Emma Pierson, and AnshulKundaje, “Denoising genome-wide histone ChIP-seq with convolutional neural networks”, Bioinformatics, Vol.33, No.14, pp.i225-i233, 2017

[69] Li Jian, JianjiangFeng, and C-C. Jay Kuo, “Deep convolutional neural network for latent fingerprint enhancement”, Signal Processing: Image Communication, Vol.60, pp.52-63, 2018

[70] Tran Truyen, TuDinh Nguyen, DinhPhung, and SvethaVenkatesh, “Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM)”, Journal of biomedical informatics, Vol.54, pp.96-105, 2015

[71] GhahabiOmid, and Javier Hernando, “Restricted Boltzmann machines for vector representation of speech in speaker recognition”, Computer Speech & Language, Vol.47, pp.16-29, 2018

[72] Zhang Zhenhua, Qing He, Jing Gao, and Ming Ni, “A deep learning approach for detecting traffic accidents from social media data”, Transportation Research Part C: Emerging Technologies, Vol.86, pp.580-596, 2018

[73] Choi Edward, Mohammad TahaBahadori, Andy Schuetz, Walter F. Stewart, and Jimeng Sun, “Doctor ai: Predicting clinical events via recurrent neural networks”, In Machine Learning for Healthcare Conference, pp.301-318, 2016

[74] Wollmer Martin, Florian Eyben, Joseph Keshet, Alex Graves, Bjorn Schuller, and Gerhard Rigoll, “Robust discriminative keyword spotting for emotionally colored spontaneous speech using bidirectional LSTM networks”, IEEE International Conference on Acoustics, Speech and Signal Processing, pp.3949-3952, 2009

[75] Hammerla Nils Y., Shane Halloran, and Thomas Ploetz, “Deep, convolutional, and recurrent models for human activity recognition using wearables”, arXiv preprint arXiv:1604.08880, 2016

[76] Huang B. Q., Tarik Rashid, and M. T. Kechadi, “Multi-context recurrent neural network for time series applications”, International Journal of Computational Intelligence, Vol.3, No.1, pp.45-54, 2006

[77] Dean Jeffrey, Greg Corrado, RajatMonga, Kai Chen, Matthieu Devin, Mark Mao, Andrew Senior et al, “Large scale distributed deep networks”, In Advances in neural information processing systems, pp.1223-1231, 2012

[78] Heigold Georg, Vincent Vanhoucke, Alan Senior, Patrick Nguyen, Marc’AurelioRanzato, Matthieu Devin, and Jeffrey Dean, “Multilingual acoustic models using distributed deep neural networks”, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.8619-8623, 2013

[79] Coates Adam, Brody Huval, Tao Wang, David Wu, Bryan Catanzaro, and Ng Andrew, “Deep learning with COTS HPC systems”, In International Conference on Machine Learning, pp.1337-1345, 2013

[80] Weston Jason, FrédéricRatle, HosseinMobahi, and Ronan Collobert, “Deep learning via semi-supervised embedding”, In Neural Networks: Tricks of the Trade, Springer, pp.639-655, 2012

[81] Fergus Rob, Yair Weiss, and Antonio Torralba, “Semi-supervised learning in gigantic image collections”, In Advances in neural information processing systems, pp.522-530, 2009

[82] Vincent Pascal, Hugo Larochelle, YoshuaBengio, and Pierre-Antoine Manzagol, “Extracting and composing robust features with denoising autoencoders”, ACM Proceedings of the 25th international conference on Machine learning, pp.1096-1103, 2008

[83] Calandra Roberto, TapaniRaiko, Marc Peter Deisenroth, and Federico MontesinoPouzols, “Learning deep belief networks from non-stationary streams”, In International Conference on Artificial Neural Networks, Springer, pp.379-386, 2012

[84] Zhou Guanyu, KihyukSohn, and Honglak Lee, “Online incremental feature learning with denoising autoencoders”, In Artificial Intelligence and Statistics, pp.1453-1461, 2012

[85] Ouyang, Wanli, Xiao Chu, and Xiaogang Wang, “Multi-source deep learning for human pose estimation”, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,

pp.2329-2336, 2014
[86] Zhao, Liang, Jiangzhuo Chen, Feng Chen, Wei Wang, Chang-Tien Lu, and NarenRamakrishnan, "Simnest: Social media nested epidemic simulation via online semi-supervised deep learning", IEEE International Conference on Data Mining (ICDM), pp.639-648, 2015
[87] Chen Minmin, ZhixiangXu, Kilian Weinberger, and FeiSha, "Marginalized denoising autoencoders for domain adaptation", arXiv preprint arXiv:1206.4683, 2012
[88] NgiamJiquan, AdityaKhosla, Mingyu Kim, Juhan Nam, Honglak Lee, and Andrew Y. Ng, "Multimodal deep learning", In Proceedings of the 28th international conference on machine learning (ICML-11), pp.689-696, 2011
[89] Zhang Qingchen, Laurence T. Yang, and Zhikui Chen, "Privacy preserving deep computation model on cloud for big data feature learning", IEEE Transactions on Computers, Vol.65, No.5, pp.1351-1362, 2016
[90] Yu, Lequan, Hao Chen, Qi Dou, Jing Qin, and Pheng Ann Heng, "Integrating Online and Offline Three-Dimensional Deep Learning for Automated Polyp Detection in Colonoscopy Videos", IEEE journal of biomedical and health informatics, Vol.21, No.1, pp.65-75, 2017
[91] F. Bu, Z. Chen, Q. Zhang, "Incremental updating method for big data feature learning", Computer Engineering and Applications, Vol.51, No.12, pp.21–26, 2015
[92] Chen Kai, and QiangHuo, "Scalable training of deep learning machines by incremental block training with intra-block parallel optimization and blockwise model-update filtering", IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp.5880-5884, 2016
[93] Shalev-ShwartzShai, "Online learning and online convex optimization", Foundations and Trends® in Machine Learning, Vol.4, No.2, pp.107-194, 2012
[94] Bartlett Peter L, "Optimal Online Prediction in Adversarial Environments", In Discovery Science, p.371, 2010
[95] Chien Jen-Tzung, and Hsin-Lung Hsieh, "Nonstationary source separation using sequential and variational Bayesian learning", IEEE transactions on neural networks and learning systems, Vol.24, No.5, pp.681-694, 2013
[96] J.V.N. Lakshmi and Ananthi Sheshasaayee, "A Big Data Analytical approach for Analyzing Temperature Dataset using Machine Learning Techniques", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.3, pp.92-97, 2017

