ORIGINAL PAPER

A Survey of Deep Learning and Its Applications: A New Paradigm to Machine Learning

Shaveta Dargan[1] · Munish Kumar[1] · Maruthi Rohit Ayyagari[2] · Gulshan Kumar[3]

Abstract

Nowadays, deep learning is a current and a stimulating field of machine learning. Deep learning is the most effective, supervised, time and cost efficient machine learning approach. Deep learning is not a restricted learning approach, but it abides various procedures and topographies which can be applied to an immense speculum of complicated problems. The technique learns the illustrative and differential features in a very stratified way. Deep learning methods have made a significant breakthrough with appreciable performance in a wide variety of applications with useful security tools. It is considered to be the best choice for discovering complex architecture in high-dimensional data by employing back propagation algorithm. As deep learning has made significant advancements and tremendous performance in numerous applications, the widely used domains of deep learning are business, science and government which further includes adaptive testing, biological image classification, computer vision, cancer detection, natural language processing, object detection, face recognition, handwriting recognition, speech recognition, stock market analysis, smart city and many more. This paper focuses on the concepts of deep learning, its basic and advanced architectures, techniques, motivational aspects, characteristics and the limitations. The paper also presents the major differences between the deep learning, classical machine learning and conventional learning approaches and the major challenges ahead. The main intention of this paper is to explore and present chronologically, a comprehensive survey of the major applications of deep learning covering variety of areas, study of the techniques and architectures used and further the contribution of that respective application in the real world. Finally, the paper ends with the conclusion and future aspects.

1 Introduction

Machine learning is a subsection of Artificial Intelligence (AI) that imparts the system, the benefits to automatically learn from the concepts and knowledge without being explicitly programmed. It starts with observations such as the direct experiences to prepare for the features and patterns in data and producing better results and decisions in the future. Deep learning relies on the collection of machine learning algorithms which models high-level abstractions in the data with multiple nonlinear transformations. A deep learning technology works on the artificial neural network system (ANNs). These ANNs constantly take learning algorithms and by continuously increasing the amounts of data, the efficiency of training processes can be improved. The efficiency is dependent on the larger data volumes. The training process is called deep because the number of levels of neural network increases with the time. The working of the deep learning process is purely dependent on two phases which are called the training phase and inferring phase. The training phase includes labeling of large amounts of data and determining their matching characteristics and the inferring phase deals with making conclusions and label new unexposed data using their previous knowledge. Deep-learning is such an approach that helps the system to understand the complex perception tasks with the maximum accuracy. Deep learning is also known as deep structured learning and hierarchical learning that consists of multiple layers which includes nonlinear processing units for the purpose of conversion and feature extraction. Every subsequent
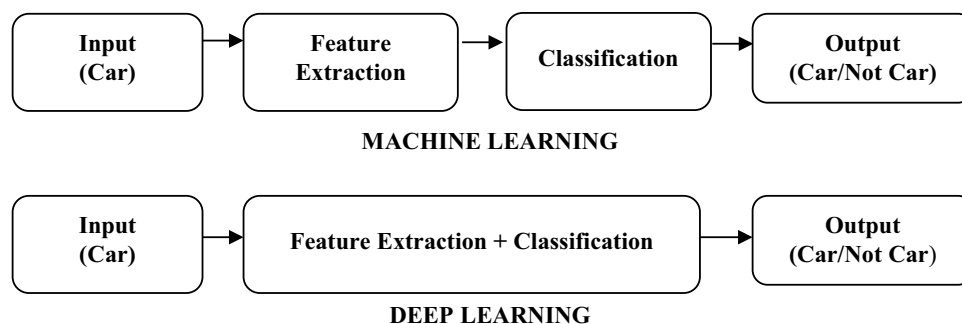
✉ Munish Kumar
munishcse@gmail.com

[1] Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda, Punjab, India

[2] College of Business, University of Dallas, Irving, USA

[3] Department of Computer Applications, Shaheed Bhagat Singh State Technical Campus, Ferozepur, Punjab, India

Fig. 1 Difference between machine learning and the deep learning

| Input (Car) | Feature Extraction | Classification | Output (Car/Not Car) |

MACHINE LEARNING

| Input (Car) | Feature Extraction + Classification | Output (Car/Not Car) |

DEEP LEARNING

layer takes the results from the previous layer as the input. The learning process takes place in either supervised or unsupervised way by using distinctive stages of abstraction and manifold levels of representations. Deep learning or the deep neural network uses the fundamental computational unit, i.e. the neuron that takes multiple signals as input. It integrates these signals linearly with the weight and transfers the combined signals over the nonlinear tasks to produce outputs.

In the "deep learning" methodology, the term "deep" enumerates the concept of numerous layers through which the data is transformed. These systems consist of very special credit assignment path (CAP) depth which means the steps of conversions from input to output and represents the impulsive connection between the input layer and the output layer. It must be noted that there is a difference between deep learning and representational learning. Representational learning includes the set of methods that helps the machine to take the raw data as input and determines the representations for the detection and classification purpose. Deep learning techniques are purely such kind of learning methods that have multiple levels of representation and at more abstract level. Figure 1 depicts the differences between the machine learning and deep learning.

Deep learning techniques use nonlinear transformations and model abstractions at a high level in large databases. It also describes that a machine transforms its internal attributes, which are required to enumerate the descriptions in each layer, by accepting the abstractions and representations from the previous layer. This novel learning approach is widely used in the fields of adaptive testing, big data, cancer detection, data flow, document analysis and recognition, health care, object detection, speech recognition, image classification, pedestrian detection, natural language processing and voice activity detection.

Deep learning paradigm uses a massive ground truth designated data to find the unique features, combinations of features and then constructs an integrated feature extraction and classification model to figure out a variety of applications. The meaningful characteristic of deep learning is the

data that uses general purpose methods, various extensive features and no intervention of human engineers. Facebook has also created Deep Text for the classification of the massive amount of data and cleaning the spam messages.

The key factors on which deep learning methodology is based are:

- Nonlinear processing in multiple layers or stages.
- Supervised or unsupervised learning.

Nonlinear processing in multiple layers refers to a hierarchical method in which the present layer accepts the results from the previous layer and passes its output as input to the next layer. Hierarchy is established among layers so as to organize the importance of the data. Here supervised and unsupervised learning are linked to the class target label. Its availability means a supervised system and absence indicates an unsupervised system. Soniya et al. [56] presented current trends, models, architecture and the limitations of deep learning. They explored some of the characteristics like learning techniques, optimization methods and tuning of these models. They also focused on the use of large datasets for the deep learning. They also discussed the challenges for the deep learning.

2 Basic Architectures of Deep Neural Network (DNN)

Different names for deep learning architectures embrace deep belief networks, recurrent neural networks and deep neural networks. DNN can be constructed by adding multiple layers which are hidden layers in between the input layers and the output layers of Artificial Neural Network with various topologies. The deep neural network can model convoluted and non-linear relationships and generates models in which the object is treated as a layered configuration of primitives. These are such feed forward networks which have no looping and the flow of data is from the input layer to the output layer. There are wide varieties of architectures and algorithms that are helpful in implementing the concept

Table 1 Years with the usage of architectures of deep learning

Year	Architecture of deep learning
1990–1995	Recurrent neural network
1995–2000	Long short term memory, convolutional neural network
2000–2005	Long short term memory, convolutional neural network
2005–2010	Deep belief network
2010–2017	Deep stacked network, gated recurrent unit

of deep learning. Table 1 depicts the year wise distribution in the architecture of deep learning.

Here, we will discuss six basic types of the deep learning architectures and these are:-

- Auto-Encoder (AE)
- Convolutional Neural Network (CNN)
- Restricted Boltzmann Machine (RBM)
- Deep Stacking Network (DSN)
- Long Short Term Memory (LSTM)/Gated Recurrent Unit (GRU) Network
- Recurrent Neural Network (RNN)

Out of these, LSTM and CNN are two of the fundamental and the most commonly used approaches.

2.1 Auto-Encoder (AE)

An Auto-encoder (AE) is a type of neural network which is based on unsupervised learning technique and uses the back propagation algorithm. The network first sets the target result values to be equal to the input values. The network tries to understand an approximation which is equivalent to the identity function. Its architecture consists of three layers which are an input, a hidden called encoding layer, and a decoding layer. The network tries to reconstruct its input, which forces the hidden layer to learn the best representations of the input. The hidden layer is used to describe a code which helps to represent the input. Auto-encoders are neural networks, but they are also closely related to PCA (Principal Component Analysis).

Some Key Facts about the Auto-encoder are:-

- Auto-encoders are neural network.
- Auto-encoders are based on the unsupervised machine learning algorithm.
- These are closely resembled with the Principal Component Analysis (PCA).
- It is more flexible than the PCA.
- It minimizes the same objective function as PCA
- The neural network's target output is its input

Although Auto-encoders are same as PCA, but the flexibility of auto-encoder is quite high. Auto-encoders allow the representation in both linear and non-linear way in the encoding whereas linear transformation is possible in PCA. Due to the network representation, Auto-encoders can be stacked and layered to produce a deep learning network.

Following are the types of Auto-encoders:

1. De-noising Auto-encoder: It is an advanced version of basic auto-encoders. To addresses the identity functions, these encoders corrupt the input and afterwards, reconstruct them. It is also called the stochastic version of the auto-encoders.
2. Sparse Auto-encoder: These auto-encoders have the learning methods that automatically extract the features from the unlabeled data. Here the word sparse indicates that hidden units are allowed to fire only for the certain type of inputs and not too frequently.
3. Variational Auto-Encoder (VAE): The variational auto-encoder consists of an encoder, decoder and a loss function. They are used for the designing of the complex models of the data that too with large datasets. It is also known as high resolution network.
4. Contractive Auto-encoder (CAE): These are robust networks as de-noising auto-encoders but the difference is that the contractive auto-encoders generate robustness in the networks through encoder function whereas de-noising auto-encoders work with the reconstruction process.

Auto-encoders are used to operate with high dimensional data and explains the representation of a set of data via dimensionality reduction. Auto-encoder (AE) uses mainly two structures, called, De-noising Auto-encoder and Sparse Auto-encoder. For De-noising Auto-encoder, it uses data from noise to experience the network weight and for Sparse Auto-encoder, they bound the activation state of hidden units. Working of an Auto-encoder considers the input and afterwards maps it to an inherent transformation with the help of nonlinear mapping.

2.2 Convolutional Neural Network (CNN)

CNN is a neural network with multiple layers and is based on the animal visual cortex. The first CNN was developed by LeCun et al. [27]. Application areas of CNN include mainly image-processing and handwritten character recognition e.g. postal code interpretation. Considering the architecture, earlier layers are used for identifying the features such as edges and the later layers are used for the recombination of features to form high level attributes of the input followed by the

Fig. 2 Architecture of Convolutional Neural Network
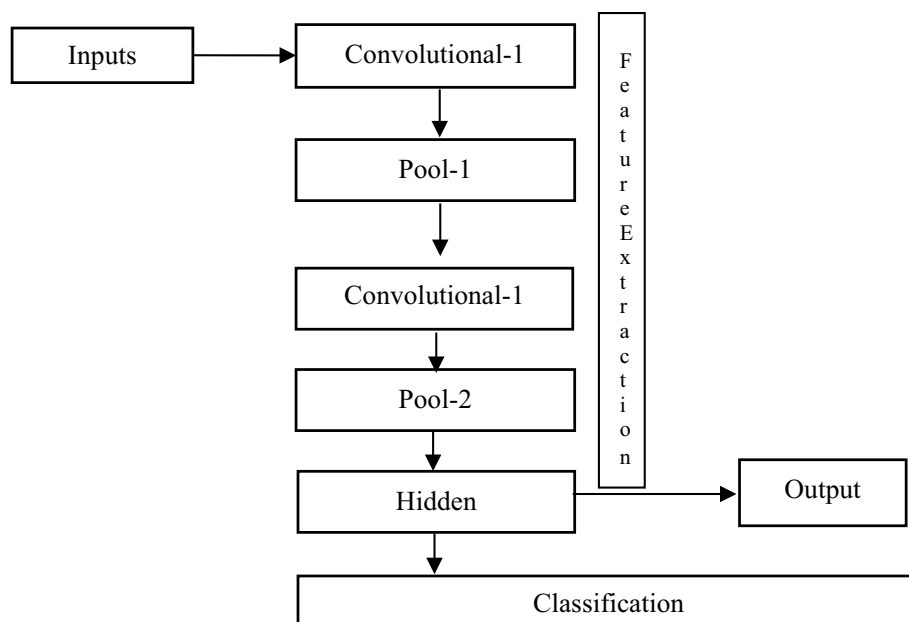
classification. Then pooling will be done, which mitigates the dimensionality of the extracted features.

The next step is to perform convolution and then again pooling, that is fed into a perfectly linked multilayer perceptron. Responsibility of the concluding layer called an output layer is to recognize the features of the image by using back-propagation algorithms. In CNN, the advantage of deep layers of processing, convolutional layer, pooling, and a fully connected classification layer reveals various applications such as speech recognition, medical applications, video recognition and various tasks of natural language processing.

CNN produces better accuracy and improves the performance of the system due to its exclusive features such as local connectivity and shared weights. It works much better than any other deep learning methods. It is the most commonly used architecture as compared to others. Figure 2 depicts the working of CNN with the flow of data from the inputs, convolutional, pooling layers, hidden layers and the outputs.

2.3 Restricted Boltzmann Machines and Deep Belief Network

Restricted Boltzmann Machine (RBM) is such an undirected graphical and modeled representation of the hidden layer, a visible layer and the symmetric connection between the layers. In RBM, there is no connection in between an input and the hidden layer. The deep belief network represents multilayer network architecture that incorporates a novel training method with many hidden layers. Here every pair of connected layers is a RBM and is also known as a stack of restricted Boltzmann machines. The input layer constitutes the basic sensory input, and the hidden layer characterizing the abstract description of this input. The job of the output layer is to only perform the network classification.

The training part is done in two stages: Unsupervised pre training and supervised fine-tuning. In unsupervised pre training, from the first hidden layer, RBM is skilled to reconstruct its input. The next RBM is qualified similar to the first one, and the first hidden layer is taken as the input and visible layer, and the RBM is worked by taking the outputs from the first hidden layer. Hence, every layer is pre skilled or pre-trained. Now when the pre training is completed, steps of supervised fine-tuning start. In this step, the nodes representing the output are marked with the values or labels so that they can help in the learning process and later on full network training is done with the gradient descent learning or back-propagation algorithm.

2.4 Deep Stacking Networks

Deep Stacking Networks (DSN) is also acknowledged as deep convex networks. DSN is different from other traditional deep learning structures. It is called deep because of the fact that it contains a large number of deep individual networks where each network has its own hidden layers.

The DSN believes that training is not a particular and isolated problem, but it holds the combination of individual training problems. The DSN is made up of a combination of modules which are part of the network and present in the architecture. There are three modules that work for the DSN. Here every module in the model is having an input zone, a single hidden zone and an output zone. Subroutines are placed one over the top of another with the input to the
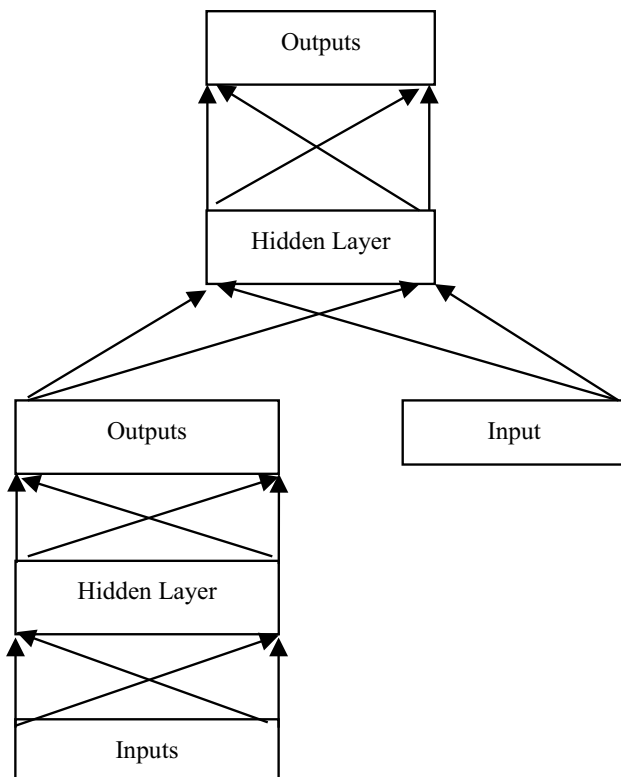
Fig. 3 Working of DSN

every module is taken as the outputs of the preceding layer and the authentic input vector. Figure 3 depicts the process of working of the layers that helps to resolve the complex classifications. In DSN, every module is trained in isolation so as to make it productive and competent with the ability to work in coordination. The process of supervised method of training is practiced as the back-propagation for each module and not for the entire network. DSNs works superior than typical DBNs making it suitable and accepted network architecture.

2.5 LSTM/GRU Network

The Long Short Term Memory (LSTM) was designed with the efforts of Hochreiter and Schimdhuber, and used for many applications. IBM selected LSTMs mainly for speech recognition. The LSTM uses a memory unit called a cell which can hold its value for a sufficient time and treats it as a function of its input. This helps the unit to memorize the last calculated value.

The memory unit or a cell is made up of three ports called gates, which control the movement of information in the unit, i.e. into the cell and out of the cell.

- The input port or the gate manages flow of new information into the memory.

- The second gate called forgets port controls and is used when an existing piece of information is forgotten and helps the cell to memorize the new data.
- The job of the output gate is again to control the information that is present in the cell and is used as the output of the cell.

The weight of the cell can be used for the controlling purpose. There is a need for the training method which is commonly called as Back propagation through time (BPTT) that enhances the weight. The method requires network output error for the optimization.

The Gated Recurrent Unit (GRU) includes two gates called as an update gate and a reset gate. Responsibility of an update gate is to tell the requirement of the contents of the previous cell for the maintenance. The reset gate describes the carrying process of previous cell contents with the new input. The GRU represents a standard RNN by initializing the reset gate to 1 and update gate to 0. Working capability of GRU model is simple as compared to the LSTM. It can be skilled in a short time and it is considered to be more efficient in terms of its execution.

2.6 Recurrent Neural Network

RNN consists of a rich set of architecture and is the basic network architecture. The important characteristic of a recurrent network is that the recurrent network has a connection that can be given as feedback into prior layers as compared to the complete feed-forward connections. It takes the previous memory of input and models the problems within time. These networks can be upgraded, skilled and expanded with standard back-propagation called as back-propagation through time (BPTT). Table 2, describes the various application areas of different architecture of deep neural networks.

3 Advanced Architectures of Deep Neural Network

Owing to many flexibilities provided by the neural network, deep neural network can be expressed by a diverse set of models. These architectures are called deep models and consist of:

- *AlexNet* The net is named for the researchers. It was the earliest deep learning architecture and was developed by Alex Krizhevsky, Geoffrey Hinton and his colleagues, who gave ground breaking research in deep learning. The architecture consists of the convolutional layers and the pooling layers which are stacked on one another and then followed by completely interlinked layers on the top. The benefits and superiority lie in the fact that the scalability

Table 2 Architectures of deep neural network and their major application areas

Architecture	Major application areas
Auto-encoder	Natural language processing, understanding compact representation of data
Convolutional neural networks	Document analysis, face recognition, image recognition, natural language processing, video analysis
Deep belief networks	Failure prediction, information retrieval, image recognition, natural language understanding
Deep stacking networks	Continuous speech recognition, Information retrieval
LSTM/GRU networks	Gesture recognition, handwriting recognition, image captioning, natural language text compression, speech recognition
Recurrent neural networks	Handwriting and speech recognition
Restricted Boltzmann machine	Collaborative filtering, classification, dimensionality reduction, feature learning, regression, and topic modeling

and the use of GPU are incomparable. AlexNet has high speed of processing and training because of the use of GPU.

- *Visual Graphic Group Net* This net was developed by the technicians at the Visual Graphics Group from the Oxford and is in pyramid shape. The model consists of the bottom layers which are wide and the top layers are deep. VGG accommodates successive convolutional layers and then the pooling layers to make the layers narrow.
- *GoogleNet* The architecture was introduced by the researchers at Google and hence the name of the Net. It involves 22 layers whereas VGG had 19 layers. Google Net is based on the novel technique which is known as the inception module. Here single layer carries multiple kinds of the feature extractors that help the network to perform better. When multiple of these inception modules are stacked one over the other, it becomes the final one. The model converges faster because of the joint and the parallel training. Training of GoogleNet is faster than VGG with small size of the pre-trained GoogleNet.
- *ResNet* Residual Network incorporates numerous successive residual modules also called as the basic building block of ResNet. The residual modules are placed on one over the other and form a successful and complete node to node network. The main benefit of ResNet is that many residual layers are capable of forming a trained network.
- *ResNeXt* It is constructed based on the concepts of ResNet with novel and enhanced architecture with improved performance.
- *RCNN (Regions with Convolutional Neural Network)* It depends upon designing a bounding box over the objects in the image and identifies the object given in the image.
- *YoLo (You only look once)* This architecture solves image detection problems. To identify the class of the object, the image is divided into bounding box and then a recognition algorithm is executed which is common for all the boxes. After identification of the classes, the boxes are merged very carefully to make a best bounding box around the objects. It is used in real time for handling day-to-day problems.

- *SqueezeNet* The SqueezeNet architecture is the most powerful architecture to select with the low bandwidth. This network architecture takes space of 4.9 MB and the inception process will take 100 MB. A fire module is used for handling the drastic change.
- *SegNet* SegNet is considered as the best model for the image segmentation problems. SegNet is a deep neural network, which is used to solve the image segmentation complexities. It is made up of an arrangement of processing layers which are called encoders, and interrelated set of decoders for pixel wise classification. The important feature of SegNet is the ability to retain very high frequency details in the segmented image. Herein the encoder network and the decoder network, pooling indices are connected. The flow of information is also straight.
- *GAN* Generative Adversarial Networks is a unique network architecture, which creates an entirely novel and different images, which are not already present in the available training dataset.

4 Characteristics of Deep Learning

Deep learning is a broad term used for the machine learning and for the artificial intelligence. Because of the following mentioned characteristics, deep learning techniques have achieved the heights of success in the variety of application areas. For example, new areas such as decision fusion, on-board mobile devices, transfer learning, class imbalance problems and human activity recognition have gained improvement in the performance and the accuracy.

So, here are the following characteristics of deep learning:

- Extensively powerful tool in many fields.
- It is purely based on neural networks with the addition of more than two layers and so called deep.
- Have strong learning ability.
- Can make use of datasets more effectively.

- Learn feature extraction methods from the data.
- Surpass human ability to solve highly computational tasks.
- Very little engineering by hand is required in deep learning.
- Optimized results.
- Deep learning networks depend upon the nature of the network structure, activation function and data representation.
- Describe highly variant features in a few parameters.
- Prediction performance can be greatly improved.
- Solve highly computational tasks.
- Capability to extract features from high dimensional sensory inputs.
- Secure and robust generalization capability and with less requirement of training data.
- Fuse the benefits of multiple features for voice activity detection.
- Stronger than machine learning model in feature representation.
- Covariance estimation can be improved for the prediction applications.
- Deep learning networks do not rely on prior data and knowledge.
- DNN has a unique representation and having innovative methods to understand the representations even with large-scale and unlabeled data.
- With high-level abstraction, these networks can extract complicated features.
- Good recognition ability approaches in the big data era.

5 Motivation to Use Deep Learning

Deep learning technology has a conception that there is nothing inherently challenging the applications to enhance the performance, e.g. handwriting recognition of the machines achieves human level of performance, same as for face recognition and the object recognition metrics. It is to be admitted that the deep learning begins from the handwriting recognition. Its architecture called CNN was created successfully in order to read handwritten postal codes. The motivation for the use of deep learning occurs from the many facts as listed below:-

- Undoubtedly, deep learning will definitely drive AI adoption into the enterprise also.
- Deep Learning is the main driver and the most essential approach to AI.
- Deep learning is a collection of methods and techniques based on artificial neural networks with multiple layers and increased functionality.

- Deep learning perceives tremendous growth because it has deep layered neural networks and the support of graphical processing units to improve the execution.
- Deep neural networks mainly include feed-forward networks with convolution and pooling layers.
- There is no sequence and inputs and outputs are independent.
- Deep neural networks achieved eminence, 4–5 years back when deep models started replacing the traditional approaches, especially in handwriting recognition, healthcare, image classification, speech recognition and natural language processing.
- Deep neural networks can be disciplined and analyzed by many researchers and academia.
- Deep learning techniques and methodologies are more accurate when skilled with large amount of data.
- NVIDIA will influence the space in 2017 because they are having the affluent deep Learning ecosystem. Intel Xeon Phi solutions are buried on influx with respect to deep learning.
- Designers will depend on meta-learning.
- Reinforcement learning will only become more creative.
- Adversarial and cooperative learning will be the king.

6 Deep Learning vs. Machine Learning

Deep learning architecture is constructed from many hidden layers and multiple neurons per layer. The multilayer architecture facilitates with the mapping of the input to higher level representation. Here we discuss the major differences that are found between two learning techniques:-

- Deep learning constructs algorithms in various layers to make an artificial neural network, which can learn and take intelligent decisions on its own, whereas machine learning needs algorithms to interpret data, learn from that data and then synthesized informed decisions.
- Deep learning takes a large amount of data while machine learning needs a small amount of data to work and arrive at a conclusion.
- Deep learning requires hardware with very high performance.
- Deep learning creates new features by its own processes and techniques, whereas in case of machine learning, features are accurately and precisely recognized by the users.
- Deep learning solves the problem on end to end basis, whereas machine learning solves it by decomposing a bigger task into smaller tasks and then combining the results.

- Deep networks are black box networks and their working is very difficult to understand because of hyper parameters and complex network design.
- Time requirement to train is much more in deep learning than in machine learning.
- Transparency is shown by machine learning methods rather than the deep learning methods.
- Accuracy rate achieved by deep learning is very satisfactory as compared to machine learning.
- Challenging and complex feature engineering phase is eliminated in the deep learning which is present in the machine learning.
- Deep networks need high-end graphical processing units which are very expensive and are skilled in sufficient time with big data.

7 Deep Learning vs. Conventional Learning

The major differences that are present between the deep learning methodologies and the conventional learning are as described below:-

- Extraction of features and their Representation

  - From the raw sensor, deep learning methods can learn features and finds the most suitable pattern for improving the recognition accuracy.
  - Conventional learning worked on the feature vectors which are manually produced and applications dependent. These features are difficult to model in complexities.

- Generalization and Diversity

  - It is possible to extract spatial, scale invariant and temporal features from the unlabeled raw data in deep learning.
  - Conventional learning used labeled sensor data. And also focus on feature selection with dimensionality reduction methods.

- Data preparations

  - In deep learning, pre-processing of the data and normalization are not mandatory.
  - Conventional learning extracts features by using sensor appearance and within the active windows.

- Temporal and Spatial changes in Activities

  - Use of hierarchical features and translational invariant features can solve the complexities present in intra-class variability's in handcrafted features.
  - Handcrafted features are not suitable and inefficient in solving the inter-class variability's and inter-class relations in the conventional learning.

- Model Training and Execution time

  - To avoid over fitting, deep learning requires large amounts of sensor dataset. It is also used for reducing high computations. Graphical Processing Unit (GPU) is used to speed up the training.
  - Less training data is required with less time for computation and memory utilization is also less in conventional training.

8 Reported Work on Various Applications of Deep Learning

The target approach of deep learning is to resolve the sophisticated aspects of the input by using multiple levels of representation. This new approach to machine learning has already been doing wonders in the applications like face, speech, images, handwriting recognition system, natural language processing, medical sciences, and many more. Its latest researches involve revealing the optimization and fine tuning of the model by using gradient descent and evolutionary algorithms. Some major challenges that the deep learning technology is facing undoubtedly are the scaling of computations, optimization of the parameters of deep neural network, designing and learning approaches. A detailed investigation in various complex deep neural network models is also a big challenge to this potential research area. The combination of fuzzy logic with deep neural network is another provoking and demanding area which needs to be explored. Numerous applications of deep learning are depicted in Fig. 4.

- Acoustic Modeling

  Mohamed et al. [37] proposed deep learning network, which contains multiple layers of features with many parameters for the phone recognition. They replaced Gaussian mixture models and used TIMITT dataset. They trained deep learning networks as a multilayered generative model. After designing features of pre-trained deep network, the next step was to perform discriminative fine tuning with the back propagation, distribution so as to adjust the features for the better prediction of probability

Fig. 4 Applications of deep learning
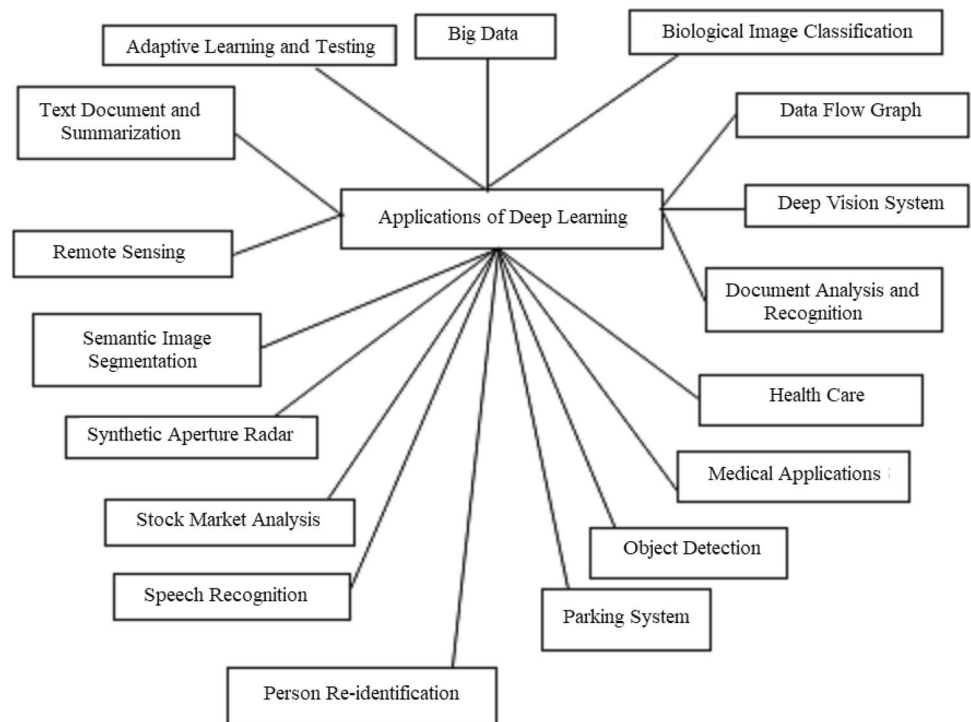
distribution. They worked on such applications of acoustic modeling where multiple layers of features were pre-trained. They explicitly exemplary the covariance structure of the input features. They were trying to reveal alternative representations of the input that helps deep neural networks to gather the relevant information in the soundwave. They also explored various ways of using recurrent neural networks for increasing the amount of past detailed information that helps in the interpretation of the future.

Ling et al. [29] presented in a very systematically way the review of the speech generation approaches. They created interest in the mind of readers to learn the existing parametric speech generation methods and also stimulated for the generation of developing new methods. They concluded in their findings that for parametric speech recognition, RBM and DBN which are called deep joint models and CRBM and DNN are better to represent the complicated and nonlinear relations. Santana et al. [54] presented a unique method for the acoustic modeling. In the presence of noise, they developed the system for the speech recognition by using deep neural network. This is the big challenge for the researchers for developing speech recognition system with the presence of noise speech signals. For their experiment, they used CNN and the recurrent architecture. CNN was used for the acoustic modeling and recurrent method with connectionist and temporary classification was used for the sequential modeling. Their method worked well as compared to the classical model such as HMM with the BioChaves datasets.

- Adaptive Testing

Chandra and Sharma [77] presented a novel approach to integrate adaptive learning rate and using Laplacian score for the updating of the weights. They considered that the neurons are important for enhancing the weights and the learning rate. These can be taken as a function of the parameter and the operations can be possible on the basis of the error gradient. Laplacian score of the neuron can be used for the incoming weights and to improve the complexities in catching the optimum value of the learning rate. This was implemented on the benchmark datasets with linear activation function and the max out. The work proved out to achieve an increase in classification accuracy. This method had a limitation that they could not use Laplacian score in an online mode. They recommended that it is better to go for the 'Exponential LP' with Rectified Linear activation function when the data was available in the streams and the batches, respectively. Xiao et al. [65] proposed a new method for adaptive testing based on the deep learning. Without manual intervention, these techniques can extract the features from the data automatically. For their work, they used DNN which proved higher accuracies rate for the failure and pass prediction. By using the features from the DNN, they developed two applications, i.e. partial testing and dynamic test ordering. These applications were used for the decision making such as pass or failure and for the dynamic test ordering. Experiments results proved improvements in accuracy and

effectiveness. Falcini et al. [17] designed a disciplined life process based on the deep learning. They designed successfully a W model that is dependent on the Software process improvement and capability determination and used DNN for the achievement of the task in addition with the traditional automotive software industry. The improvement in the W model really pushed them for achieving future aspects like fully autonomous driving.

- Automotive Industry

Luckow et al. [34] presented the applications and the tools for implementing deep neural networks in the automotive industry. They focused on the use of CNN and the computer vision uses cases. They used labeled datasets. The main contribution of this paper was the creation of the automotive dataset that helps the users to learn and automatically identify and examine the vehicle properties. They analyzed both the training time and the accuracy of the classifiers and reported that during the manufacturing process, the trained classifier were efficient and effective.

- Big Data

Today with the very fast increase in the size of data, the application communicates big scope and metamorphic possibilities for the various sectors. It widely opens up the extraordinary demands to exploit data and information for this big data prediction and analytical solutions. Chen and Lin [9] noticed and proposed that there are significant challenges in front of deep learning. These challenges are large scale, heterogeneous, disorderly labels, and non-static distribution and the requirement is to have transformation solutions. The challenges offered by big data were timely and provided many opportunities and searches for the deep learning. Gheisari et al. [18] presented that the deep learning of the Big Data analysis can produce remarkable results and helped in identifying the unknown and useful patterns with high level of abstraction which were impossible to understand.

- Biological Image Classification

Affonso et al. [3] contributed the analysis of the classification and peculiarity of wood boards by considering the images. For their experiment, they used decision tree induction algorithms, CNN, Neural Network, Nearest Neighbor and support vector machine on the used dataset. The team successfully achieved very promising results and concluded that the deep learning when applied to image processing tasks accomplished predictive performance accuracy. The accuracy also depends on the nature of the image dataset.

- Data Flow Graphs

Abadi et al. [1] put forward a TensorFlow Graph that is a part of the TensorFlow machine intelligence structure that help in understanding the machine learning architecture based on the data flow graph. They applied serial transformations for designing the standard layout for the interactive diagram. They developed the coupling on the non-critical nodes and built a clustered graph by taking the stepwise structure from the source code. They also performed edge bundling to make the expansion responsive and stable and focused on the modular composition. Ripoll et al. [46] proposed a new screening method based on deep neural network. The method was developed to assess whether a patient from the emergency should shift to cardiology. The method was studied on 1320 patients that took raw ECG signals without annotation. Learning machines, k-Nearest neighbor and classification algorithms were used. For the experiment, they selected Support vector machines with Gaussian kernel and got accuracy of 84%, sensitivity of 94% and specificity of 73%. Looks et al. [32] presented a technique known as dynamic batching which combined the diverse input graphs with dissimilar shapes but with also different nodes that were present in single input graph. The technique helped to create static graphs that follow dynamic computation graph with different size and shapes. The team also worked on the library that exhibit batch form implementation for number of models.

- Deep Vision System

Puthussery et al. [44] presented a new approach for the autonomous navigation using machine learning techniques. They used CNN to identify marker from the images and also developed a robot operating system and discovered a system based on the position of the object to orient to the marker. They also evaluated the distance and navigation towards these markers with the help of depth camera. Abbas et al. [2] proposed powerful models of deep learning that helps in implementing the real time video processing applications. They explained that in the technological era, with the powerful facilities of deep learning techniques, the real time video processing applications e.g. video entity detection, tracing and identification can be made possible with great accuracy. Their architecture and novel approach solved the problems of computational cost, number of layers, precision and accuracy. They used deep learning algorithms that give robust powerful architecture with many layers of neurons and the efficient manipulation of big video data. Sanakoyeu et al. [53] presented a framework for visual similarities with the unsupervised deep learning. They employed weak calculations of local similarities and proposed a new optimized problem for extracting batch of samples with

mutual relations. The conflicting relations were distributing over different batches and similar samples are arranged in a separate group. Convolutional neural network was used to build up relations within and between the groups and create a single representation for all the samples without labels. For the posture analysis and the classification of objects, the proposed method shows competitive performance.

- Document Analysis & Recognition

Wu et al. [78] presented a new approach for the visual and speech recognition. They proposed an R-CNN called relaxation convolutional neural network based visual and speech recognition system used to regularize CNN for the fully interlinked layers. They do not need neurons in a map to share the kernel. This architecture requires another architecture which is called alternately trained relaxation called ATR-CNN. The R-CNN increased the total number of parameters, so they used ATR-CNN for regularizing the neural network during training procedure and got the accuracy with an error rate of 3.94%. In R-CNN, neurons do not share the same convolutional kernel and ATR-CNN also used an alternative strategy for the training. For the working of ATR-CNN, they required handwritten digit dataset MNIST and ICDAR'13 competition dataset. Kannan and Subramanian [24] presented a review on the deep learning and its applicability for the optical character recognition on the Tamil script. They presented a survey on the studies done by the experts on the Tamil script. They also mentioned the steps required for better OCR development using deep learning technology with the big data analysis. LeCun et al. [27] also proposed an explanation for the handwritten ZIP code identification. They employed the back-propagation algorithm into a deep neural network with expanded training time. Nguyen et al. [39] used deep neural nets for identifying online handwritten mathematical symbols. Here they used max out dependent CNN and bidirectional long short term memory to the image patterns. These image patterns have been created from online patterns. The experiment was conducted on the CHROME database. They concluded that as compared with MQDF, the CNN network produces improved results to identify the mathematical symbols offline. BLSTM also worked better than MRF in case of online patterns. By integrating both online and offline integration methods, classification performance was improved.

Alwzwazy et al. [4] proposed a deep learning method called CNN for the handwritten Arabic pattern digit recognition. They performed their study on 45,000 samples. Deep CNN was used for the classification and proved to give more accurate results, i.e. 95.7% for the Arabic handwritten digit recognition. Xing and Qiao [79] presented an effective writer identification system by using multi stream deep CNN. It took local handwritten patches as the input

followed by training using the softmax classification loss. They designed and optimized multi stream structure with data augmentation learning and hence improved the performance of DeepWriter. For handling variable length text images, they created DeepWriter, a deep multi-stream CNN to learn a deep powerful representation for recognizing writers. Experiment was carried out in the IAM and HWDB datasets and proved the success with the accuracy rates of nearly 99%. Poznanski and Wolf [42] developed a method to solve the issues for understanding the images of a handwritten word. They used CNN to evaluate its n-gram frequency profile, which is the set of n-grams contained in the word. Frequencies for unigrams, bigrams and trigrams were estimated for the entire word. Canonical correlation analysis was used for the comparison of profiles of all words. CNN used multiple fully connected branches. After going through this process, they got more accurate results without applying much effort in atomic tasks like an image binarization and the letter segmentation.

Ashiquzzaman and Tushar [7] presented an original work on the deep learning for identifying numerals in handwritten Arabic. They used Multilayer Perceptron method of Rectified Linear Unit (ReLu) for the stimulation of neurons in the input and the hidden layer, and the softmax function was used for the classification in the outer layer. They selected CNN architecture with the activation function and the regularization layer to increase the accuracy. The proposed approach proved 97.4% accuracy rate. Ghosh and Maghari [19] proposed three most commonly used NN approaches which are deep neural network (DNN), deep belief network (DBN) and convolutional neural network (CNN). They performed the comparison on the networks by considering the factors like recognition rate, execution time and the performance. For conducting the experiment, they used random and standard dataset of handwritten digits. The results proved that out of these three NN approaches, DNN provided promising accuracy of 98.08% and the execution time of DNN is comparable with the other two algorithms. They also reported that the new approach generated an error rate of only one to two percent because of the similarity in digit shapes. Prabhanjan and Dinesh [43] presented a unique approach for the recognition of the Devanagari script. They used a deep learning approach for the feature extraction process. They used raw pixel values for the selection of features and used unsupervised restricted Boltzmann machine with deep belief network to improve the performance and the accuracy of the system. The method was suited for the character, numerals, vowels and compound characters. Experiment revealed 83.44% accuracy with unsupervised method and with the supervised method, the accuracy of 91.81% was reported.

Roy et al. [80] presented an architecture where deep belief networks were used to understand the compressed

delineation of the sequential data, and the Hidden Markov Model was used for the word recognition. For the implementation of the approach, they used RIMES and IFN/ENIT which were publicly available datasets on the Latin and Arabic languages, respectively and they also worked on the dataset in Devanagari. The experiments demonstrated that the proposed model was preferred over the MLP-HMMs tandem approaches. They combined the discriminating features of DBNs with a generative model of HMMs. By using an unsupervised pre-training algorithm and the DBN weight, they initialized the feed-forward neural network that helped in preventing over-fitting and provide better optimization of the recognition weights. Yadav et al. [67] designed a contemporary deep learning approach for character identification from the multimedia document. For the experimental study, they used diagonal feature extraction method in the convolutional layers. Then they executed genetic algorithm to feed forward network for the classification followed by the training. For their experiment, they used a dataset which consists of 360 training set data with capital letters from *A* to *Z*, small alphabets from *a* to *z*, digits from 0 to 9 and some special characters. Dataset may be taken as samples of videos and images. Studies proved that diagonal based recognition system proved with more accurate results with less training time.

- HealthCare

Loh and Then [81] proposed a new approach for heart diagnosis and management, in context with rural healthcare, and also discussed the benefits, issues and solutions for implementing deep learning algorithms. The development of rural healthcare services such as telemedicine and health applications were really required. Different solutions such as portable medical equipment and mobile technologies have been developed to find out the deficiencies present in the remote settings. Additionally, computer aided designed systems have also been used for assistive interpretation and diagnosis of medical imagery. The implementation of machine and deep learning algorithms would bring numerous benefits to both physicians and patients. The advancement of mobile technologies would expedite the proliferation of healthcare services to those residing in impoverished regions. Dai and Wang [15] proposed a framework for the healthcare application and to reduce the heavy workload of doctors and nurses by employing the advantages of the technologies of artificial intelligence. They considered that the methods of pattern recognition and the deep recognition module were sufficient enough to diagnose the health status based on deep neural networks (DNNs). They also worked on the action evaluation module, which is based on the Bayesian inference graphs and then developed a simulation environment which includes a body simulator to prepare

the body for the treatments, and the health state of the simulated patient will be changed by different interventions. So, the team worked on the body simulation module, a deep recognition module used to diagnose the bodily features and an action evaluation module used Bayesian inference graphs to maintain the record and calculate the statistical evidence. Experiment proved to be the most efficient with the increasing statistical data. For the experiment they used the dataset consisting of health state representation space of 9 body constitutional types.

- Human Activity Recognition

Nweke et al. [41] proposed human activity recognition systems for the continuous monitoring of human behaviors in the environment. For the mobile and wearable sensor-based human activity recognition pipeline, they extracted the relevant features that will influence the performance and reduced the computation time and complexity. The combination of mobile or wearable sensors and deep learning methods for feature learning really proved diversity, higher generalization, and resolved all challenging issues in human activity recognition. They presented the review on the in-depth summaries of deep learning methods for mobile and wearable sensor-based human activity recognition and presented the methods, their uniqueness, advantages and their limitations. They categorized the studies into generative, discriminative and hybrid methods and also highlighted their important advantages. The review presented classification and evaluation procedures and discussed publicly available datasets for mobile sensor human activity recognition. They reviewed the training and optimization strategies for mobile and wearable based human activity recognition. They also tested some of the publicly available benchmark datasets such as Skoda, and PAMAP2. Ignatov [22] presented an online human activity recognition and classification system based on the accelerometer. They used CNN for the implementation of the method for extracting local and statistical features. They focused the use of time series length for examining the activities. The experiment was conducted on the WISDM and UCI datasets and used 36 and 30 users respectively with labeled data. Proposed model achieved good results with less computational cost and without manual feature engineering.

- Image Recognition and Classification

Ciresan et al. [82] presented a method for image classification using deep neural networks. Their model was based on the architecture of convolutional winner take all neurons. Result was sparsely connected neurons layers. Training was given to only winner neurons. They used MNIST benchmark and successfully achieved great results which were near to

the human performance. For the traffic sign identification benchmark it outperforms humans by a factor of two. Liu et al. [31] presented different factors for the convolution neural network such as network depth, numbers of filters, and filter sizes. They implemented their approach on the CIFAR dataset. According to their experiments and observations, different factors are considered. Based on the results of their experiments on the CIFAR dataset, they found that the network depth is the first priority chosen for improving the accuracy. They achieved the same high accuracy with less complexity as compared to adding the network width. But, the weaker factor states that the excessive increase in the depth of the network may degrade the accuracy and result in the data-insufficient fitting problem. They also observed that the filter size is another important parameter for improving the accuracy. The replacement of a large filter size with stacked convolutional neural networks improves the recognition accuracy with a decrease in the time complexity. Jia [23] presented a review of the deep learning and its usefulness in the various applications. He very systematically presented the architecture, advantages and the working phenomenon of the deep learning. He also reported the uses and the accuracies of the CNN for the computer vision and the image recognition problem. Liu et al. [31] presented their study on the impact of various factors used in the convolutional neural network for image classification. They considered depth of the network, number and nature of filters, size of filters etc. and worked on the CIFAR datasets.

- Medical Applications

Vasconcelos and Vasconcwlos [60] presented how deep convolutional neural network (DCNN) based classifier can be used to deal with small and unbalanced medical data set. They used data augmentation schemes, the inclusion of the third class of lesion patterns and an awareness of diversity in committee formation. The study validates the accuracy of the technique developed. Yuan et al. [70] proposed a deep learning method in the regularized ensemble framework. It was used to handle the multi class and imbalanced problems. They used stratified sampling for the balancing of the classes and concentrate on the un-prediction caused by the base learners through regularization. For the data distribution, sampling procedure selected examples randomly and the regularization process updates the loss of function to penalize the classifiers. It also adjusts the error boundaries keeping in view the performance of the classifier. For their experiment, they used eleven dissimilar synthetic as well as real-world data sets. Their new method had successfully got the highest accuracies for the minority classes with the ensemble stability. Experimental results proved that the proposed method achieved the best accuracy and also explains the dissimilarity of the base classifiers present in the ensemble. They

explained that there is also significant reduction in the computational cost. But as the volume of training data increases, the efficiency of their method increases. Razzak et al. [45] proposed stimulating solutions and reported with very good accuracy for the health care applications such as medical imaging, image interpretation, health sector, computer-aided diagnosis, medical image processing, image fusion, image registration and image segmentation. The machine learning and artificial intelligence methods provided assistance to the doctors to diagnose and predict the disease and its risk and prevent them in time. The method helped the doctors in understanding the generic variations that lead to the occurrence of the disease. These techniques were made up of both conventional algorithms and deep learning algorithms such as Support Vector Machine (SVM), Neural Network (NN), KNN and Convolutional Neural Network (CNN), Extreme Learning Model (ELM), Generative Adversarial Networks (GANs), Recur-rent Neural Network (RNN), Long Short term Memory (LSTM), etc.

- Mobile Multimedia

Ota et al. [83] proposed a survey on the deep learning for mobile multimedia. They concluded that less complex deep learning algorithms, the software frameworks, and specialized hardware helped in the processing of deep neural network. They presented applications of deep learning in mobile multimedia with the different possibilities for real-life use of this technology. Multimedia processing and deep learning can be integrated to work with mobile devices. The earlier approach of using mobile devices just as sensor and actuator devices and the main processing and data storage services for deep learning located in servers would definitely support some applications. As mobile devices were more dominant, more applications running deep learning engines reduced the overhead of maintaining internet connectivity and also the complex server infrastructure.

- Object Detection

Ucar et al. [59] put forward a novel hybrid Local Multiple system based on Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) with the feature extraction capability and robust classification. In the proposed system, they divided first the whole image into local regions by using the multiple CNNs. Secondly, they selected discriminating features by using principal component analysis (PCA) and imported into multiple SVMs by both empirical and structural risk minimization. Finally, they tried to fuse SVM outputs. They worked on the pre-trained AlexNet and also performed object recognition and pedestrian detection experiments on the Caltech-101 and Caltech Pedestrian datasets. Their proposed system generated better results with

the low miss rate and improved object recognition and detection with an increase in accuracy. Zhou et al. [74] presented the architecture and the algorithms of deep learning in an application of object detection task. They worked on built-in datasets such as ImageNet, Pascal Voc, CoCo and deep learning methods for object detection. They created their own dataset and proved that by using CNN with the deep learning algorithm for the object detection, they got good results. The requirement of the dataset for the deep learning is large, so the applications are continually collecting large datasets. Experiment proved that the technology of deep learning is an effective tool to pass the man-made feature with the large qualitative data. Kaushal et al. [25] proposed a comprehensive survey of the object detection and tracking in videos techniques based on the deep learning. The survey included neural network, deep learning, fuzzy logic, evolutionary algorithms required for detection and tracking. In the survey, they discussed various datasets and challenges for the object detection and tracking based on the deep neural network.

- Parking System

Chen et al. [10] presented a mobile cloud computing architecture based on the deep learning that used training process and the repository in the clouds. The communication was possible with the Git protocol that helps in transmission of the data even in the unstable environment. During the driving, smart cameras in the car recorded the videos and the implementation was done on the NVIDIA Jetson TK1. Results of the experiment proved that detection rate was improved to four frames per second as compared to R-CNN. For detecting parking lot occupation, Amato et al. [5] proposed a new decentralized solution for the classification of images of a parking space when occupied directly on the smart cameras. It is built on deep convolutional neural network (CNN) suited for the smart cameras. The experiment is implemented on the two visual datasets such as PKLot and CNRPark-EXT. Actually they required a dataset that contains the images of a real parking and is collected by nine smart cameras. The images were captured on different days under in diverse weather and light conditions. They also employed a training and validation dataset for the detection of parking occupancy and performed the task in real-time directly on the smart cameras. They did not employ a central server for the experiment. The method used Raspberry Ri platform equipped with camera module. For implementing the proposed method, the server needs the binary output of the classification. They concluded that CNN received very high accuracy with the light condition variations, partial occlusions, shadows and noise.

- Person Re-identification

Cheng et al. [12] proposed the method for visual recognition, especially for person re-identification (Re-Id). They used distance metric between pairs of examples and proposed contrastive and triplet loss to enhance the discriminating power of the features with great success. They proposed a structured graph Laplacian embedding method, which can be formed and evaluated all the structured distance links into the graph Laplacian form. By integrating the proposed technique with the softmax loss required for the CNN training, the proposed method produced specific deep features by maintaining inter-personal dispersion and intra-personal compactness, which were the requirement of personal Re-Id. They used the most common and popular networks such as AlexNet, DGDNet and ResNet50. They concluded that the proposed structure graph Laplacian embedding technique was very effective for the person re-identification. Zhao et al. [72] put forward a new multiple levels strategy for feature extraction to integrate coarse and fine information coming from different layers. They also developed a multilevel triplet deep learning model called MT-net to extract multilevel features systematically. The results of the experiment, on popular datasets it was proved that the method was the most effective and robust.

- Plant Classification

Lee et al. [28] presented a new method in which deep learning methods are used for the plant classification. It helps the botanists to identify the species very accurately and quickly. From raw representations of the leaf data, they extracted useful leaf features using CNN and Deep Networks. After their extensive study, they were able to extract hybrid feature extraction models that help in improving the discriminant capabilities of the plant classification system.

- Power System Fault Diagnosis

Wang et al. [61] presented a new method for the power system fault diagnosis. The major perspective was to extract relevant data from the huge collection of un-labeled data followed by the preprocessing. The method solved three major bottlenecks which are availability of the data, to improve the local optimum and diffusion of gradients. They obtained the data of the power system from the SCADA under the administration of the power supply. Then they extracted, preprocessed and fed the data to stacked auto-encoder (SAE) for training with the hidden features to be found in the different dimensions. Then trained SAE is used for the initialization where the classifier proved the type of diagnosis. The results proved the accuracy and the feasibility of the method. Rudin et al. [47] presented a unique approach based on the CNN for the power system fault diagnosis. The goal of their research was to classify the power system voltage signal samples to

check the faulted or non-faulted state. For the dataset, they simulate a simple two bus system and balance load in three phases. At the starting of the line and between the two buses, the voltage signal is measured. At the middle of the line, the fault occurred between the buses. They used wavelet transform to extract the features. CNN was used for the learning process which learned the defective features of the power system by taking faulty and non-faulty samples.

- Radio Wireless Networks

Lopez et al. [33] proposed a method to increase the forecasting accuracies of licensed users in spectrum channels. The model was based on the long term short memory (LSTM) recurrent neural network. Results of the experiment proved that the accuracies achieved by LSTM outperformed the multi-layer perceptron network and adaptive neuro fuzzy inference systems. The method also allowed the implementation of the method in cognitive network with centralized physical topologies. Yu et al. [69] presented a framework for the spectrum prediction by taking two spectrum datasets of the real world. They employed Taguchi method for the determination of the optimized configuration and channel occupancy states of neural network, then they built LSTM for the spectrum prediction with perspectives like regression and the classification. In case of the second dataset, for the channel quality, they compared the prediction performance of MLP and LSTM. Results of the experiment proved that with the frequency bands, prediction performance varies. LSTM worked better in terms of stability and prediction with the classification aspect as compared to when taking both the regression and classification.

Wang et al. [62] put forward a systematic survey of the upbringing studies based on the deep learning based physical layer processing, redesigning of a module for the communicational system with the auto encoder. The new architecture of deep learning proved promising performance with excellent capacity and good optimization, for the communication and implementation.

- Remote Sensing

Deep learning techniques and deep nets have been successfully used in remote sensing applications in the physical models. These are complicated, nonlinear and difficult to understand and generalize. Yu et al. [69] proposed a technique of deep learning in remote sensing application for not only improving the volume and completeness of training data for any remote sensing datasets, but also uses the datasets to skill out convolutional neural network. The proposed method used three operations, which are the flipping, translation, and rotation to generate augmented data and produced a more descriptive deep model. The proposed method also introduced basic data augmentation operations to solve the data limitation problem for remote sensing image processing and contributed with potentially revolutionary changes in remote sensing scene classification. The results of the experiment significantly contributed and improved the diversity of the dataset in remote sensing. These also increased visual variability of each training in the remote sensing image by taking the intrinsic spectral and topological constraints and did not generate new information for the remote sensing image. Zhu et al. [75] used deep neural nets that put in front of the users, the various opportunities for novel areas such as supervising global changes or finding out the strategies for the decreasing the resource consumption. Deep network has always been an incredible and challenged toolbox that assists researchers in the field of remote sensing to cross the boundaries and manipulate large-scale, real-life problems with implied models. They analyzed large scale remote sensing data by considering multi-modal, multi-aspect, geo located and multi-temporal aspects.

- Semantic Image Segmentation

Chen et al. [11] made efforts to put forward a semantic image segmentation problem with the facilities of deep neural network. The study presented three main contributions to the area. First they worked on convolution with up sampled filters that help to maintain the resolution. For the prediction task, responses from the features are calculated with convolution neural network. They did not enhance the amount of computation and effectively enlarged the view of filters. Then they proposed Atrous Spatial Pyramid Pooling (ASPP) system for the segmentation at the multiple scales. It can act as a convolutional feature layer by using filters at the multiple sampling rates and capturing objects and images at multiple scales. Then the localities of the object boundaries were integrated with the procedures for the deep convolutional networks and probabilistic models. The experiment reached the invariance with the localization accuracy with both quantitative and qualitative improvements. The dataset used by the team were PASCAL- Context, PASCAL Person-Part and the CitySpaces.

- Smart City

Wang and Sng [84] proposed a method of deep learning in the video analytics of the city and was used to detect the object, tracking of the object, face recognition and the image classification. In smart cities, the task of capturing the images, videos from the sensors was required to be automatically processed and analyzed. The success of this approach was based on the fact that lots of big data were available for building a deep neural network. Advantages of Graphical

Processing Unit (GPU) would definitely reduce the training time. So, the approach of deep learning with the smart cities proved wonders as seen from the reviews by the authors.

- Social Applications

Deep learning techniques are widely used for the sentiment analysis. Araque et al. [6] proposed a deep learning technique with surface approaches and is based upon the manually extracted features. For this experiment, they designed a deep learning based sentiment classifier that is dependent on the word embedding architecture and a linear machine learning method. Results can be compared using the classifier. Then they developed two ensemble techniques and two models that were responsible for combining the baseline classifier and the surface classifiers. They employed total 7 public datasets that were extracted from the micro blogging. They proved that performance of these models was really remarkable.

- Speech Recognition

Mohamed et al. [37] proposed a novel method and used DBNs for acoustic modeling. They used standard TIMIT dataset with a phone error rate (PER) of 23.0%. They used back propagation algorithm with the network and called BP-DBN and the associative memory DBN called as AM-DBN architecture. The effect of depth of the model and the size of the hidden layer were investigated and different techniques, were adopted to reduce over fitting. Bottlenecks in the last layer of the BP-DBN also helped in avoiding the over fitting. The discriminative and hybrid generative training also contributed in preventing the over fitting in the associative memory DBN. The results given by the architecture of DBN have recorded as the best in comparison with the other. The experiment was performed on the TIMIT corpus and used 1 with 462 speaker training set. Total 50 speakers were used for model tuning and the results were shown using the 24-speaker core test set. The speech was understood using a 25-ms hamming window with 10-ms between the left edges of successive frames. For all experiments, the Viterbi decoder parameters were optimized on the development set and to compute the phone error rate (PER) for the test set.

Hamid and Jiang [21] presented an approach which used deep neural network (DNN) and Hidden Markov Model (HMM) for the speech recognition. They used Convolutional Neural Network models by updating speaker code based adaptation method that would be better for CNN structure. Noda et al. [40] proposed a novel utilization of deep neural networks in audio visual speech recognition. It is used specially in the cases when the quality of audio is damaged by the noise. Under diverse conditions, deep neural networks are able to extract latent and robust features. Their

work involved connectionist Hidden Markov Model for the noised audio visual speech recognition system. By employing auto-encoders and the CNN, they were able to achieve 65% word recognition rate under 10 decibel signal to noise ratio. Wu et al. [64] proposed a method for statistical parametric speech synthesis (SPSS) by considering adaptability and controllability with a change in speaker characteristics and speaking style. They conducted an experiment for the speaker adaptation for speech synthesis with DNN and at diverse levels. For the input, they took the low dimensional speaker specific vectors with the linguistic features which would represent the speaker identity. The model systematically analyzed the various adaptation techniques. They also found that feature transformation at the output layer worked well and the adaptation performance can be improved by combining with model based adaptation. Experimental results proved very good in terms of performance and accuracies. Experimental results clearly showed that the adaptability and listening tests of DNN generated better adaptation performance than the hidden Markov model (HMM). The method also proved that the feature transformation done at the output layer were also worked well. Serizel [55] also presented a deep learning model with Hidden Markov Model (HMM) to construct a speech recognition system with a heterogeneous group of speakers. They used DNN and vocal tract length normalization (VTLM) for the experiment. First they separately performed the experiment and then hybrid approach was used. Combination of approaches proved the improvement in the baseline phone error rate by 30 to 35% and baseline word error rate by 10%.

Xue et al. [66] presented an adaptation scheme with the deep neural network. Here discriminating cods were used which are directly fed to the pre trained DNN through connection weight. They also proposed many training methods to learn connection codes as well as the adaptation methods for every test condition. They used three methods to use the adaptation scheme based on the codes. Three ways were nonlinear feature normalization in feature space, direct model adaptation of DNN based on speaker codes and last one was joint speaker adaptive training with speaker codes. They checked the proposed method with two standard speech recognition tasks and from these two, one was TIMIT phone recognition and the other was large vocabulary speech recognition in the Switchboard task. Results proved that all three methods were quite effective with 8 to 10% of error reduction. Markovnikov et al. [36] presented a method in which the team selected DNN in automatic Russian speech recognition. They used CNN, LSTM and RCNN for their experiment. For the dataset, they used extremely large vocabulary of Russian speech and got remarkable results with 7.5% reduction in word error rate.

- Speech Music Discrimination

Papakostas and Giannakopoulos [85] presented deep visual feature extractor for performing speech and music discrimination. From the raw spectrograms, they extracted deep visual features. Different CNN based methods for audio classifications were used for the task and moreover they were compared with traditional and deep learning methods used for the handcrafted audio features. They concluded that CNN produced very promising results. For the initialization, the team used deep architecture and to work out with skill classifiers with less training data available. They used image Net dataset for the experiment and proved that parameter tuning can maintain flexibility and produced dominant results. Markovnikov et al. [36] proposed a method for speech recognition using deep neural network such as convolutional neural network, long short term memory residual network and recurrent convolutional neural network. Their experiment worked with extra-large vocabulary and successfully achieved reduction of 7.5% in error rate.

- Stock Market Analysis

Arevalo et al. [86] proposed a method for the stock market prediction, analysis and precision by using data of US Apple stock $(3 \times \{2 - 15\} + 2)$ and target output will be Stock price for 19,109 numbers of samples. They took the sampling period of 3 months and used deep neural network methods. For measuring the performance they used MSE and directional accuracy measurements. Chong et al. [13] proposed a unique method of deep learning for the stock market analysis and prediction. From the input of stock returns, they performed the experiment with the three unsupervised methods called Principal Component Analysis (PCA), Restricted Boltzmann Machine (RBM) and Auto-encoder for predicting the future market behavior. The dataset used to be Korea KOSPI 38 stock returns and target output produced would be Stock return to a number of samples 73,041 and sampling period is 4 years. They used deep neural network methods for measuring the performance. They used NMSE, RMSE, MAE, and Mutual Information (MI). From their study, they concluded that DNN perform better than a linear autoregressive model.

- Structural Health Monitoring

Salazar et al. [48] presented a machine-learning algorithm called Boosted Regression Trees, which is the core of methodology for the early detection of problems. The method includes some criteria to determine if the discrepancy between predictions and observations is normal, to calculate the realistic estimations of the model accuracy and to recognize extraordinary load combinations. Performance of non-causal and causal models is assessed to find anomalies and at the end, final method was implemented to check

and verify the results for the decision making. Salazar et al. [49] presented an effective comparison and testing prediction capabilities of the various algorithms for modeling the dam behavior with respect to displacement and leakage. Models using the concepts such as boosted regression trees (BRT), random forests (RF), neural networks (NN), multivariate adaptive regression splines (MARS) and support vector machines (SVM) are employed to predict 14 target variables with the prediction of the accuracy as compared with the statistical models. BRT stood best as compared to the RF and NN. Salazar et al. [50] presented a novel method of using promising machine learning techniques called Boosted Regression Trees (BRT) to handle four leakage flows and eight radial displacements at La Baells Dam. The goal was to explore the model interpretation, the impact of predictors was computed and the partial dependency plots were produced. The results were interpreted to draw the conclusion on dam response to the environment variables and its growth with time. Results showed that the method was working efficiently identifying dam performance and the variations with higher flexibility and reliability rate as compared to the simple regression models. Salazar et al. [51] put in front the comprehensive survey showing the usefulness of machine learning algorithm for the analysis of dam structural behavior which is based on the monitoring data. From the survey several critical issues, accuracy rates associated with the algorithms, radial and tangential displacements, leakage flow were identified. The results of this survey concluded that BRT i.e. Boosted Regression Trees is the most perfect method in terms of accuracy, association among the variables and the response of the dam with effect of time. Salazar et al. [52] presented a review on the statistical and machine learning based predictive models which are required for the dam safety analysis. They explored the state-of-the-art work with many aspects such as nature and kind of input variables, division into training sets and validation sets and hence performed the error analysis. Review concluded that firstly, other than Hydrostatic Seasonal Time Model (HST), machine learning methods are more suitable for achieving accurate results, to represent non-linear effects as well as complex interactions in between input variables and the dam response. Secondly, the papers covered only one output variable with lack of validation data. Thirdly, engineering judgments based on the experiences are critical for constructing the model, for interpreting results and decision making for the dam safety.

- Synthetic Aperture Radar

Yonel et al. [68] introduced a new deep learning environment for the contrary difficulties in an imaging and its usefulness in the passive Synthetic Aperture Radar. They considered image interpretation as a machine learning task

and utilized deep networks for the forward and inverse solutions. The team used RNN as an inverse solver which depends upon the proximal gradient descent optimization methods. The experiment performed quite well in terms of computation and reconstruction of the image. They also thought of adding some additional layers in conventional image solvers for the forward modeling and the image reconstruction. This was done to capture the non-linearity between the physical measurements. The method was best suited for the image formation problems and other real world applications in which the forward model was only partially known. The experiment proved the geometric fidelity, high contrast, reduced reconstruction errors.

• Text/Document Summarization

Zhong et al. [73] put forward a novel query oriented approach by using deep learning techniques applicable to the multi document summarization. They observed and exposed the extraction ability in the real world applications by using dynamic programming. The model contains three parts, i.e. the concept extraction, the summary generation and the reconstruction validity. Then transformation of the whole deep architecture was done by reducing the data loss in the reconstruction validation. They did not require the training stage and was the most suitable architecture for the industrial applications. For their experiment they employed three benchmark datasets DUC 2005, 2006, 2007 and proved that this method outperforms the other extraction methods. Azar and Hamey [8] presented a method of query oriented summarization on the single document by using unsupervised auto-encoder for extracting the features and did not require the query at any stage of training with small local vocabulary and reduced training and deployment computational costs. For their experiment, they used SKE and BC3 email datasets and concluded that AE provided a more discriminating feature space and improved the local term frequency.

• Voice Activity Detection

The main responsibility of a voice activity detector is to isolate the speech signal from the disturbing background noise. Zhang and Wu [87] proposed an advanced version of deep belief networks for the voice activity detection (VAD) that separate noise from the speech signals. Deep belief networks proved a sufficient model for extracting features and showing variant functions. Deep belief networks helped VAD to select and produce a different and a novel feature that sufficiently explained the benefits of acoustic features through multiple nonlinear hidden layers. In their experimental work, they used extensive AURORA dataset and seven noisy test samples of AURORA for the

performance analysis. Four signal-to-noise ratios (SNR) levels of the audio signals are selected. So, in total, 28 test samples were used for evaluation with ten different features and a linear classifier and believed that with deep learning networks, it would be possible to approach the real time detection demands of VAD. They realized that the deep model would definitely be successful in combining numerous features in a nonlinear way and the uniformity among the features.

• Word Spotting

Thomas et al. [58] proposed a word spotting system that extracts keywords in handwritten documents. They considered both the segmentation and recognition decisions. They used deep neural network with the HMM to judge the observation probabilities. The experiment was conducted on the handwritten database called the RIMES database. The experimental results proved the superiority and improved accuracies as compared to their hybrid approaches. Wicht et al. [63] presented a method that is suited for handwritten keyword spotting with the deep learning features. A new feature extraction system was generated based on the CNN. Sliding window features were skilled from the word images in an unsupervised manner. The proposed features were calculated for the template word spotting and for the learning based word spotting. Here dynamic test wrapping and HMM were used. Experiment was conducted on the three datasets with modern and historical handwriting. The proposed system clearly showed the high performance from the different baselines and showing a robust performance under all tested conditions. They presented a configuration that was stable even with the diverse data sets. Lastly, augmenting data set with synthetic distortions also produced more robust features.

Sudholt and Fink [57] proposed a novel method for the word spotting in handwritten documents. They used deep convolutional neural network for their experiment. They presented CNN model which is skilled using the proposed pyramidal histogram of characters representation and proved that the CNN architecture worked really better and proved to be a benchmark for maintaining a short training and testing time on the various datasets. They introduced pyramidal histogram of characters net which is a deep net designed for word spotting and processed input images of varying size and to predict the corresponding PHOC representation. Experiment revealed better results. The method worked better than SVMs in both Query-by-Example and Query-by String scenarios with the present datasets and proved good for Latin and Arabic script. Gurjar et al. [20] presented a new technique for the word spotting under weak supervision. The new approach used deep networks to reduce the manual effort and to retain the high performance. Under

weak supervision, they used a mixture of synthetically produced trained data and a small subset of the training partition. With less training time and annotation efforts, new method worked well with highly competitive performance. Models also suited well for integration with the consumer word spotting. Krishnan et al. [26] also employed deep CN based features for the word images and textual embedding schemes. They proposed an architecture that learnt both the text and the images of deep CNN and provided completely embedding scheme to learn the representations of the word images and the labels and to construct the state of art word image descriptor. They successfully proved the utility of the method for extracting features for word spotting and used IAM handwritten dataset with a mean average precision of 0.9509 for query-by-string based retrieval task.

- Writer Identification

Chu and Srihari [14] presented a method that used deep neural network for the writer identification. They proposed a method that was based not only on the human defined features but dependent on the automatically generated word level features given by deep neural network. Based on the word level features, they generated writing similarity that occurred in the paragraphs. This would also help in the development of writing style of a person and the differences between the writing styles of a number of persons. The method also proved how the writing styles of children changed with the age and other factors. They used CEDAR-FOX tool for getting the results and concluded that the results depend upon the size of training dataset. Dhieb et al. [16] proposed a method for online writer identification using deep neural network. They used beta elliptic model for the text independent writer identification. The method computed efficiently the writing movements of the author using profile entities. They worked on the beta impulses and elliptical arcs. Results of the feature extraction methods were used as the input to the classifier. Results obtained clearly showed the improvement and the robustness. Mohsen et al. [38] presented a new approach for identification of an author based on the deep learning. They used deep learning for the extraction of the features of the documents which represent variable sized characters. They used stack de-noising auto-encoder for the purpose of extracting features with different scenarios. And used support vector machine classifier. Zhao et al. [72] proposed a deep learning model based on multilevel triplet for person re-identification. They extracted coarse and very fine information from the layers for feature extraction. This model produced end to end training features for the execution and achieved better performance than other re-identification methods.

9 Challenges of Deep Learning

Although deep learning techniques are proving its best and has been solving various complicated applications with multiple layers and high level of abstraction. It is surely accepted that the accuracy, acuteness, receptiveness and precision of deep learning systems are almost equal or may sometimes surpass human experts. To feel the exhilaration of victory, in today's scenario, the technology has to accept many challenges. So, here is the list of challenges which deep learning has to overcome is:

- Deep learning algorithms have to continuously manage the input data.
- Algorithms need to ensure the transparency of the conclusion.
- Resource is demanding technology like high performance GPUs, storage requirements.
- Improved methods for big data analytics. Deep networks are called black box networks.
- Presence of hyper parameters and complex design.
- Need very high computation power.
- Suffer from local minima.
- Computationally intractable.
- Need a large amount of data.
- Expensive for the complex problems and computations.
- No strong theoretical foundation.
- Difficult to find the topology, training parameters for the deep learning.
- Deep learning provides new tools and infrastructures for the computation of the data and enables computers to learn objects and representations.

10 Conclusion and Future Aspects

Deep learning is indeed a fast growing application of machine learning. The rapid use of the algorithms of deep learning in different fields really shows its success and versatility. Achievements and improved accuracy rates with the deep learning clearly exhibits the relevance of this technology, clearly emphasize the growth of deep learning and the tendency for the future advancement and research. Additionally, it is very important to highlight that hierarchy of layers and the supervision in learning are the major key factors to develop a successful application with deep learning. The reason behind is that the hierarchy is essential for appropriate data classification and the supervision believes the importance of maintaining database.

Deep learning relies on the optimization of existing applications in machine learning and its innovativeness on hierarchical layer processing. Deep learning can deliver effective

results for the various applications such as digital image processing and speech recognition. During the current era and in coming future, deep learning can be executed as a useful security tool due to the facial recognition and speech recognition combined. Besides this, digital image processing is a kind of research field that can be applied in multiple areas. For proving it to be a true optimization, deep learning is a contemporary and exciting subject in artificial intelligence.

At last, we conclude here, that if we follow the wave of success, we will find that with the increased availability of data and computational resources, the use of deep learning in many applications is genuinely taking off towards the acceptance. The technology is really ephebic, young and specific and in the next few years, it is expected that the rapid advancement of deep learning in more and more applications with a great boom and prosperity e.g. natural language processing, remote sensing and healthcare will certainly achieve targets and height of triumph and satisfaction.

Future Aspects of deep learning includes:

- Working of deep networks with the sophisticated and non-static noisy scenario and with multiple noise types?
- Raising the performance of the deep networks by improving the diversity between the features?
- Compatibility of deep neural networks in the unsupervised learning online environment.
- Deep reinforcement kind of learning will be the future direction.
- With deep networks, in coming future the factors like inferences, efficiency and the accuracies will be desirable.
- Maintenance of wide repository of data.
- To develop deep generative models with superior and advanced temporal modeling abilities for the parametric speech recognition system.
- Automatically assess ECG signal through deep learning.
- Use of deep neural network in object detection and tracking in videos.
- Deep neural network with the fully autonomous driving.

It is submitted that the deep learning methodologies have got great interest in each and every area where conventional machine learning techniques were applicable. It is finally to say that deep learning is the most effective, supervised and stimulating machine learning approach. It can provide researchers a quick evaluation of the hidden and incredible issues associated with the application for producing better and accurate results.

Compliance with Ethical Standards

Conflict of interest Authors have no conflict of interest.

References

1. Abadi M, Paul B, Jianmin C, Zhifeng C, Andy D, Jeffrey D, Matthieu D (2016) Tensorflow: a system for large-scale machine learning. In: The proceedings of the 12th USENIX symposium on operating systems design and implementation (OSDI'16), vol 16, pp 265–283
2. Abbas Q, Ibrahim MEA, Jaffar MA (2018) A comprehensive review of recent advances on deep vision systems. Artif Intell Rev. https://doi.org/10.1007/s10462-018-9633-3
3. Affonso C, Rossi ALD, Vieria FHA, Carvalho ACPDLFD (2017) Deep learning for biological image classification. Expert Syst Appl 85:114–122
4. Alwzwazy HA, Albehadili HA, Alwan YS, Islam NE (2016) Handwritten digit recognition using convolutional neural networks. In: Proceedings of international journal of innovative research in computer and communication engineering, vol 4(2), pp 1101–1106
5. Amato G, Carrara F, Falchi F, Gennaro C, Meghini C, Vairo C (2017) Deep learning for decentralized parking lot occupancy detection. Expert Syst Appl 72:327–334
6. Araque O, Corcuera-Platas I, Sánchez-Rada JF, Iglesias CA (2017) Enhancing deep learning sentiment analysis with ensemble techniques in social applications. Expert Syst Appl 77:236–246
7. Ashiquzzaman A, Tushar AK (2017) Handwritten arabic numeral recognition using deep learning neural networks. In: Proceedings of IEEE international conference on imaging, vision & pattern recognition, pp 1–4. https://doi.org/10.1109/ICIVPR.2017.7890866
8. Azar MY, Hamey L (2017) Text summarization using unsupervised deep learning. Expert Syst Appl 68:93–105
9. Chen XW, Lin X (2014) Big data deep learning: challenges and perspectives. IEEE 2:514–525. https://doi.org/10.1109/ACCESS.2014.2325029
10. Chen CH, Lee CR, Lu WCH (2016) A mobile cloud framework for deep learning and its application to smart car camera. In: Proceedings of the international conference on internet of vehicles, pp 14–25. https://doi.org/10.1007/978-3-319-51969-22
11. Chen LC, Papandreou G, Kokkinos I, Murphy K, Yuille AL (2018) Deeplab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. IEEE Trans Pattern Anal Mach Intell 40(4):834–848
12. Cheng D, Gong Y, Changb X, Shia W, Hauptmannb A, Zhenga N (2018) Deep feature learning via structured graph Laplacian embedding for person re-identification. Pattern Recogn 82:94–104
13. Chong E, Han C, Park FC (2017) Deep learning network for stock market analysis and prediction: methodology, data representations and case studies. Expert Syst Appl 83:187–205
14. Chu J, Srihari S (2014) Writer identification using a deep neural network. In: Proceedings of the 2014 Indian conference on computer vision graphics and image processing, pp 1–7
15. Dai Y, Wang G (2018) A deep inference learning framework for healthcare. Pattern Recogn Lett. https://doi.org/10.1016/j.patrec.2018.02.009
16. Dhieb T, Ouarda W, Boubaker H, Alilmi AM (2016) Deep neural network for online writer identification using Beta-elliptic model. In: Proceedings of the international joint conference on neural networks, pp 1863–1870
17. Falcini F, Lami G, Costanza AM (2017) Deep learning in automotive software. IEEE Softw 34(3):56–63. https://doi.org/10.1109/MS.2017.79
18. Gheisari M, Wang G, Bhuiyan MZA (2017) A survey on deep learning in big data. In: Proceedings of the IEEE international conference on embedded and ubiquitous computing (EUC), pp 1–8

19. Ghosh MMA, Maghari AY (2017) A comparative study on handwriting digit recognition using neural networks. In: Proceedings of the promising electronic technologies (ICPET), pp 77–81
20. Gurjar N, Sudholt S, Fink GA (2018) Learning deep representations for word spotting under weak supervision. In: Proceedings of the 13th IAPR international workshop on document analysis systems (DAS), pp 7s–12s
21. Hamid OA, Jiang H (2013) Rapid and effective speaker adaptation of convolutional neural network based models for speech recognition. In: INTERSPEECH, pp 1248–1252
22. Ignatov A (2018) Real-time human activity recognition from accelerometer data using convolutional neural networks. Appl Soft Comput 62:915–922
23. Jia X (2017) image recognition method based on deep learning. In: Proceedings of the 29th IEEE, Chinese control and decision conference (CCDC), pp 4730–4735
24. Kannan RJ, Subramanian S (2015) An adaptive approach of tamil character recognition using deep learning with big data-a survey. Adv Intell Syst Comput: 557–567
25. Kaushal M, Khehra B, Sharma A (2018) Soft computing based object detection and tracking approaches: state-of-the-art survey. Appl Soft Comput 70:423–464
26. Krishnan P, Dutta K, Jawahar CV (2018) Word spotting and recognition using deep embedding. In: Proceedings of 13th IAPR international workshop on document analysis systems (DAS). https://doi.org/10.1109/das.2018.70
27. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:1–10
28. Lee SH, Chan CS, Mayo SJ, Remagnino P (2017) How deep learning extracts and learns leaf features for plant classification. Pattern Recogn 71:1–13
29. Ling ZH, Kang SY, Zen H, Senior A, Schuster M, Qian XJ, Meng HM, Deng L (2015) Deep learning for acoustic modeling in parametric speech generation: a systematic review of existing techniques and future trends. IEEE Signal Process Mag 32(3):35–52
30. Ling Y, Jin C, Guoru D, Ya T, Jian Y, Jiachen S (2018) Spectrum prediction based on Taguchi method in deep learning with long short-term memory. IEEE Access 6(1):45923–45933
31. Liu PH, Su SF, Chen MC, Hsiao CC (2015) Deep learning and its application to general image classification. In: Proceedings of the international conference on informatics and cybernetics for computational social systems, pp 1–4
32. Looks M, Herreshoff M, Hutchins D, Norvig P (2017) Deep learning with dynamic computation graphs. In: Proceedings of the international conference on learning representation, pp 1–12
33. Lopez D, Rivas E, Gualdron O (2017) Primary user characterization for cognitive radio wireless networks using a neural system based on deep learning. Artif Intell Rev: 1–27
34. Luckow A, Cook M, Ashcraft N, Weill E, Djerekarov E, Vorster B (2017) Deep learning in the automotive industry: applications and tools. In: Proceedings of the IEEE international conference on big data, pp 3759–3768
35. Makhmudov AZ, Abdukarimov SS (2016) Speech recognition using deep learning algorithms. In: Proceedings of the international conference on informatics: problems, methodology, technologies, pp 10–15
36. Markovnikov N, Kipyatkova I, Karpov A, Filchenkov A (2018) Deep neural networks in russian speech recognition. Artif Intell Nat Lang Commun Comput Inf Sci 789:54–67. https://doi.org/10.1007/978-3-319-71746-3_5
37. Mohamed A, Dahl G, Geoffrey H (2009) Deep belief networks for phone recognition. In: Proceedings of the nips workshop on deep learning for speech recognition and related applications, pp 1–9
38. Mohsen AM, El-Makky NM, Ghanem N (2017) Author identification using deep learning. In: Proceedings of the 15th IEEE international conference on machine learning and applications, pp 898–903
39. Nguyen HD, Le AD, Nakagawa M (2015) Deep neural networks for recognizing online handwritten mathematical symbols. In: Proceedings of the 3rd IAPR IEEE Asian conference on pattern recognition (ACPR), pp 121–125
40. Noda K, Yamaguchi Y, Nakadai K, Okuno HG, Ogata T (2015) Audio-visual speech recognition using deep learning. Appl Intell 42(4):722–737
41. Nweke HF, Teh YW, Al-garadi MA, Alo UR (2018) Deep learning algorithms for human activity recognition using mobile and wearable sensor networks: state of the art and research challenges. Expert Syst Appl: 1–87
42. Poznanski A, Wolf L (2016) CNN-N-gram for handwriting word recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2305–2314
43. Prabhanjan S, Dinesh R (2017) deep learning approach for devanagari script recognition. Proc Int J Image Graph 17(3):1750016. https://doi.org/10.1142/S0219467817500164
44. Puthussery AR, Haradi KP, Erol BA, Benavidez P, Rad P, Jamshidi M (2017) A deep vision landmark framework for robot navigation. In: Proceedings of the system of systems engineering conference, pp 1–6
45. Razzak MI, Naz S, Zaib A (2018) Deep learning for medical image processing: overview, challenges and the future. Classif BioApps Lect Notes Comput Vis Biomech 26:323–350
46. Ripoll VJR, Wojdel A, Romero A, Ramos P, Brugada J (2016) ECG assessment based on neural networks with pertaining. Appl Soft Comput 49:399–406
47. Rudin F, Li GJ, Wang K (2017) An algorithm for power system fault analysis based on convolutional deep learning neural networks. Int J Res Educ Sci Methods 5(9):11–18
48. Salazar F, Toledo MA, González JM, Oñate E (2012) Early detection of anomalies in dam performance: a methodology based on boosted regression trees. Struct Control Health Monit 24(11):2012–2017
49. Salazar F, Toledo MA, Morán R, Oñate E (2015) An empirical comparison of machine learning techniques for dam behaviour modelling structural safety. Struct Saf 56:9–17
50. Salazar F, Toledo MA, Oñate E, Suárez B (2016) Interpretation of dam deformation and leakage with boosted regression trees. Eng Struct 119:230–251
51. Salazar F, Oñate E, Toledo MA (2017a) A machine learning based methodology for anomaly detection in dam behaviour, CIMNE, monograph no M170, 250 pp, Barcelona
52. Salazar F, Moran R, Toledo MA, Oñate E (2017) Data-based models for the prediction of dam behaviour: a review and some methodological considerations. Arch Comput Methods Eng 24(1):1–21
53. Sanakoyeu A, Bautista MA, Ommer B (2018) Deep unsupervised learning of visual similarities. Pattern Recogn 78:331–343
54. Santana LMQD, Santos RM, Matos LN, Macedo HT (2018) Deep neural networks for acoustic modeling in the presence of noise. IEEE Latin Am Trans 16(3):918–925
55. Serizel RGD (2016) Deep-neural network approaches for speech recognition with heterogeneous groups of speakers including children. Nat Lang Eng 1(3):1–26
56. Soniya, Paul S, Singh L (2015) A review on advances in deep learning. In: Proceedings of IEEE workshop on computational intelligence: theories, applications and future directions (WCI), pp 1–6. https://doi.org/10.1109/wci.2015.7495514
57. Sudholt S, Fink GA (2017) Attribute CNNs for word spotting in handwritten documents. Int J Doc Anal Recognit (IJDAR). https://doi.org/10.1007/s10032-018-0295-0

58. Thomas S, Chatelain C, Heutte L, Paquet T, Kessentini Y (2015) A deep HMM model for multiple keywords spotting in handwritten documents. Pattern Anal Appl 18(4):1003–1015

59. Ucar A, Demir Y, Guzelis C (2017) Object recognition and detection with deep learning for autonomous driving applications. Int Trans Soc Model Simul 93(9):759–769

60. Vasconcelos CN, Vasconcwlos BN (2017) Experiment using deep learning for dermoscopy image analysis. Pattern Recognit Lett. https://doi.org/10.1016/j.patrec.2017.11.005

61. Wang Y, Liu M, Bao Z (2016) Deep learning neural network for power system fault diagnosis. In: Proceedings of the 35th Chinese control conference, 1–6

62. Wang T, Wen CK, Wang H, Gao F, Jiang F, Jin S (2017) Deep learning for wireless physical layer: opportunities and challenges. China Commun 14(11):92–111

63. Wicht B, Fischer A, Hennebert J (2016) Deep learning features for handwritten keyword spotting. In: Proceedings of the 23rd international conference on pattern recognition (ICPR). https://doi.org/10.1109/icpr.2016.7900165

64. Wu Z, Swietojanski P, Veaux C, Renals S, King S (2015) A study of speaker adaptation for DNN-based speech synthesis. In: Proceedings of the sixteenth annual conference of the international speech communication association, pp 879–883

65. Xiao B, Xiong J, Shi Y (2016) Novel applications of deep learning hidden features for adaptive testing. In: Proceedings of the 21st Asia and South Pacific design automation conference, pp 743–748

66. Xue S, Hamid OA, Jiang H, Dai L, Liu Q (2014) Fast adaptation of deep neural network based on discriminant codes for speech recognition. IEEE/ACM Trans Audio Speech Lang Process 22(12):1713–1725

67. Yadav U, Verma S, Xaxa DK, Mahobiya C (2017) A deep learning based character recognition system from multimedia document. In: Proceedings of the international conference on innovations in power and advanced computing technologies, pp 1–7

68. Yonel B, Mason E, Yazici B (2017) Deep learning for passive synthetic aperture radar. IEEE J Sel Top Signal Process 12(1):90–103

69. Yu X, Wu X, Luo C, Ren P (2017) Deep learning in remote sensing scene classification: a data augmentation enhanced convolutional neural network framework. GIScience & Remote Sens: 1–19

70. Yuan X, Xie L, Abouelenien M (2018) A regularized ensemble framework of deep learning for cancer detection from multi-class, imbalanced training data. Pattern Recogn 77:160–172

71. Zhang L, Zhang L, Du B (2016) Deep learning for remote sensing data: a technical tutorial on the state of the art. IEEE Geosci Remote Sens Mag 4(2):22–40

72. Zhao C, Chen K, Wei Z, Chen Y, Miao D, Wang W (2018) Multilevel triplet deep learning model for person reidentification. Pattern Recogn Lett. https://doi.org/10.1016/j.patrec.2018.04.029

73. Zhong SH, Li Y, Le B (2015) Query oriented unsupervised multi document summarization via deep learning. Expert Syst Appl, pp 1–10

74. Zhou X, Gong W, Fu W, Du F (2017) Application of deep learning in object detection. In: Proceedings of the IEEE/ACIS 16th international conference on computer and information science (ICIS), pp 631–634

75. Zhu XX, Tuia D, Mou L, Xia GS, Zhang L, Xu F, Fraundorfer F (2017) Deep learning in remote sensing: a comprehensive review and list of resources. IEEE Geosci Remote Sens Mag 5(4):8–36

76. Zulkarneev M, Grigoryan R, Shamraev N (2013) Acoustic modeling with deep belief networks for Russian speech recognition. In: Proceedings of the international conference on speech and computer, pp 17–24

77. Chandra B, Sharma RK (2016) Deep learning withadaptive learning rate using Laplacian score, expert systems with applications. Int J 63(C):1–7

78. Wu Z, Swietozanski P, Veaux C, Renals S (2015) A study of speaker adaptation for DNN-based speech synthesis. In: Proceedings of the interspeech conference, pp 1–5

79. Xing L, Qiao Y (2016) DeepWriter: a multi-stream deep CNN for text-independent writer identification. Comput Vis Pattern Recognit. arXiv:1606.06472

80. Roy P, Bhunia AK, Das A, Dey P (2016) HMM-based indic handwritten word recognition using zone segmentation. Pattern Recognit 60:1057–1075. https://doi.org/10.1016/j.patcog.2016.04.012

81. Loh BCS, Then PHH (2017) Deep learning for cardiac computer-aided diagnosis: benefits, issues & solutions. M Health. https://doi.org/10.21037/mhealth.2017.09.01

82. Cireşan D, Meier U, Schmidhuber J (2012) Multi-column deep neural networks for image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3642–3649

83. Ota K, Dao MS, Mezaris V, Natale FGBD (2017) Deep learning for mobile multimedia: a survey. ACM Trans Multimed Comput Commun Appl (TOMM) (TOMM) 13(3s):34

84. Wang L, Sng D (2015) Deep learning algorithms with applications to video analytics for a smart city: a survey. arXiv, preprint arXiv: 1512.03131

85. Papakostas M, Giannakopoulos T (2018) Speech-music discrimination using deep visual feature extractors. Expert Syst Appl 114:334–344

86. Arevalo A, Niño J, Hernández G, Sandoval J (2016) High-frequency trading strategy based on deep neural networks. In: Proceedings of the international conference on intelligent computing, pp 424–436

87. Zhang XL, Wu J (2013) Denoising deep neural networks based voice activity detection. In: Proceedings of the IEEE international conference on acoustics, speech and signal processing, pp 853–857