# *Deep Learning Models and Applications: A Review*

*Nadia Siddiqui, Aysha Khan, Saiful Islam, Rashid Ali*
*Department of Computer Engineering*
*Zakir Husain College of Engineering and Technology,*
*Aligarh Muslim University, Aligarh 202002, UP, India*
*nadiasid2134@gmail.com,aisha.khan9413@gmail.com,saifulislam@zhcet.ac.in,rashidaliamu@rediffmail.com*

***Abstract*—This Deep learning is a forthcoming field ofMachine Learning (ML). Deep Learning (DL) consists of several hidden layers identical to artificial neural networks. Its strategies comprises of supervised and unsupervised learning techniques to automatically learn the hierarchical representation of deep architecture. As data is becoming larger, deep learning is benefiting in this scenario to deal with such enormous amount of resources. Deep learning methodology related with non-linear transformation to abstract furthermore features in high level data representation. The recent development of architecture and techniques in deep learning shows wide application in various fields using artificial intelligence. As deep learning is becoming popular, it is of utmost importance that users should be efficiently equipped with how the model works, how it fails, how the performance is improved, etc. This paper focuses on describing the notion of deep learning along with its progression. The paper also presents some of the most popular DL models and various application areas where it can be applied. This survey can help researchers to have a proper understanding of the widely used DL models and their applications.**

**Keywords—*Artificial Intelligence, Deep Learning, Convolutional Neural Networks, Recurrent Neural Networks, Long-Short Term Memory.***

## I. Introduction

Deep learning has become one of the hottest research trends among the research community. It has arisen from the notion of human brain having multiple kinds of description with easy features at the lower levels and higher level of abstractions on top of that. It has been found that human beings organise their plans heirarchially. Human beings first learn about the basic concepts and then combine them for the representation of the abstract ones. The human brain is just like a deep neural network which consists of many layer of neurons acting as feature detectors.

Construction of a machine learning system using traditional machine leraning techniques need significant knowledge of the domain in order to construct a feature extractor to convert raw training data into appropriate internal representation so that the system could classify the different patterns in the data.

Representation learning is defined as the set of techniques in which a machine is fed with raw data and the representations needed for classification are discovered automatically. The techniques of deep learning are based on representation learning which converts the representation at one level into a representation at another higher level which is more abstract by combining simple but non-linear modules. It is the explicit set of techniques that emerge from broader class of machine learning which focuses on usage of artificial neural networks to understand formative representation of data. The effective approach to human learning and understanding is exhibited by an Artificial Intelligence (AI) first mentioned in 1940s [1].

Deep learning has been defined as the subset of machine learning, which in turn is a subset of artificial intelligence which is a way of creating machines that can think and make decisions cleverly. "The Turning Test" was first proposed in 1950s that provide satisfactory explanation on how a computer could function as human intellectual rationale [2]. Recently Artificial neural networks have exert an overwhelming guiding influence over many research domains by producing state-of-art results [3, 4] on an amble collection of big data tasks [5, 6]. As a research field, artificial intelligence has separate area in definite research sub-field.

It is at the leading edge of what machines can achieve nowadays. It is a distinctive algorithm which has gone beyond the older paradigms in the classification of voice, images, text etc. It works well in the prediction of things.

## II. Motivation For Deep Learning

First, In the past few years the data generated from the web has become very enormous. The digital information has increased in volume by nine times in a span of 5 years in 2011. It is estimated that this number will extend to 35 trillion gigabytes by 2020. The continuous generation of this vast amount of data has opened doors for the research and the industrial community to examine and utilize it properly. This data is usually collected and studied in varying domains like social networks, commerce, security [7]. So it becomes really important to analyze this data. In the past few years, deep learning has been acquired in a number of huge and complex data intensive fields. The techinques of deep learning have possibly become one of the most popular ways to mine and discover the information hidden in the data.

Although machine learning also works well with analyzing the data for varying purposes, but when the data becomes very large then the performace of machine learning degrades. Dealing with vast amounts of data, deep learning algorthims really performs well.

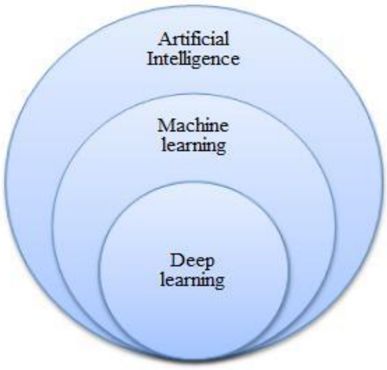


*Fig. 1. Rise of Deep Learning.*

Fig. 1 shows the evolution of deep learning, which is inspired by AI neural networks following the ML scheme to learn the data and extract features automatically.

## III. Literature Survey

In 2006, Deep Learning (DL) appeared for the first time as the research field within machine learning. The learning mechanism involves many research fields related to pattern recognition by using hierarchical learning algorithm [8]. The concept of deep learning broadly consists of two main categories: firstly multi-layer hidden neurons involve nonlinear processing and secondly, it includes either supervised or unsupervised learning techniques [9]. The current layer takes the output of previous layers as input using nonlinear processing algorithm on such multiple layers. The hierarchical structure describes the organization of data representation in order to either use them or not. While on the other hand, learning techniques like supervised or unsupervised is related to target class label, its accessibility refers to supervised system whereas its absence refers to an unsupervised system. There are several types of neural networks and layers like: Convolutional Neural Networks (CNNs), Recurrent Neural Network (RNN), Long Short Term Memory (LSTM), and may more.

In 1980s, the neocognitron was assigned to be the forerunner of ConvNets [10]. The background of Convolutional Neural Network was highlighted by task done by LeNet in 1990, which subsequently improved [11]. This Network was designed and it used images as input without any preprocessing to classify handwritten digits and recognized visual patterns. But due to the lack of sufficient training data and powerful architecture this network was failed in many complex problems. Following this, in 2012 the CNN model was developed that succeeded to reduce the error rate [12]. Over the years, CNN model becomes the most influential one that was widely used in computer vision field and was used by many by trying the variations in CNN architecture. Convolutional Neural Networks CNNs was generally inspired by visual cortex of animals [13], which used the wide structure of deep learning architecture [14]. The CNNs structure was firstly adopted for object recognition task, however now it has being observed in diverse range of computer vision as well as Natural Language Processing (NLP) tasks like: object tracking [15], posture estimation [16], handwriting recognition [17], relevant vision detection [18], action recognition [19], labeling of scene [20] and many more [21].

A family of Recurrent Neural Network (RNN) was introduced by Hopfield in 1982, which shows pattern recognition capabilities [22]. This Hopfield nets emerged as backup net to recover the corrupted versions of pattern and proved to be predecessor of Boltzmann machine and auto-encoders. In 1986, the first architecture for supervised learning was suggested in [23], where each hidden node values fed into special units in order to remember the action of previous node values. Even for standard feed forward nets, the optimization is NP-complete [27]. Over improved architecture with implementation of forward and backward propagation using packages like Theano [28] and Torch [29], rendered RNN feasible. RNN have been successfully used in image captioning area where sentences are generated from images [30, 31].

Long Short-Term Memory (LSTM) is a type of RNN. In RNN architecture the neurons have connection with all previous neurons along with input layers. RNN is more impressive for the text classification as it support the sequential flow of data or contextual view [32]. On the other hand in LSTM, the information in newer neurons is more critical than older information [33]. In various sequence prediction and for labeling task the LSTM model are successfully applied. LSTM model shows improvement over conventional RNN in language modeling. On account of context-free and context-sensitive language learning LSTM exhibit better performance [34, 35].

LSTM has been widely used in deep learning on video footage [36], because of its sequential nature. Moreover, it has been used in unsupervised video encoding [37], video captioning [38], and program execution [39].

## IV. Deep Learning Models

The learning techniques are highly improved with the combination of neural network following traditional machine learning. The major deep learning models that exhibits favorable contribution in this area are: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs) followed by Long-Short Term Memory (LSTM). The following subsections discuss these models with detailed architecture.

### *A. Conventional Neural Networks CNNs*

With the use of Convolutional Neural Networks (CNNs), deep learning has achieved the wide range of interest in multiple areas. In such structure, features are learned automatically and classification is done from original data.

Fig. 2 shows the general structure of CNNs. This consists of convolutional layer followed by classification layer.

The convolutional layer consists of three main operations: *convolution*, the *activation function* and *pooling*.Convolutional layer treat each input data with set of kernels to represent features. The output of convolutional layer also called feature map. The *convolution* can be expressed as:

$$x_j^l = \sum x_i^{l-1} * k_{ij}^{l-1} + z_j^l \quad (1)$$

Where * denoted as convolution, $x_j^l$ is j-th output map in $l$ layer, $k_{il}^{l-1}$ is convolutional kernel connecting i-th output map in $l-1$ layer and j-th output map in $l$ layer. $z_j^l$ is bias parameter, and n is the number of parameters extracted from $l-1$ layers.

The *activation function* is applied to each filtered data to converge large trainable dataset. The last operation is *pooling*, which is inserted between two convolutional layers. Their motive is to reduce the number of parameters and spatial size representation in the time of computation. During pooling, maximum and average value is computed. The last operation that is computed within convolutional layer is normalization of feature map. This is done to receive comparable neuron values.

The classification layer consists of fully connected layers followed by softmax function. The fully connected layers have set of neurons that have all connections with previous activation. The softmax function is connected to last hidden layers which compute the class score of each input class using probability distribution [40].

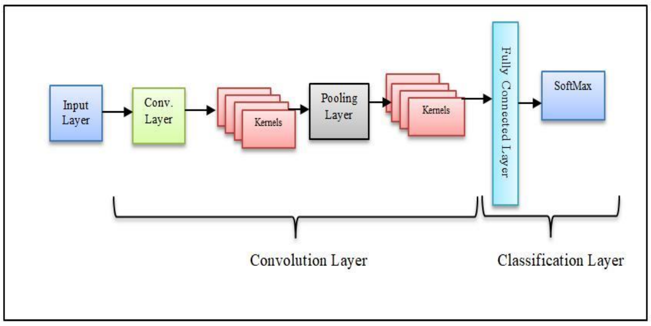


*Fig. 2. An elementary CNN model comprise of convolutional layer along with classification layer.*

The learning process in CNNs can be established by two passes: *forward pass* and *backward pass*. During *forwardpass*, input data treated with convolutional kernels thatautomatically extract features and goes to the top of the model where classification is done using softmax function. While in *backward pass*, error between predicted and ground truthvalue is computed. According to this the weights and bias can be modified and fed back into model to guide the automatically extracted features and learning mechanism established.

*B. Recurrent Neural Networks RNNs*

In conventional neural networks it is considered that all the inputs and outputs are independent of each other. But this is always not the case. Many a times it happens that the information is connected to each other and treating that information independently may not give the desired results. Examples for the prediction of next word in a sentence, it is needed to know the previous words that came before it. The main purpose behind the recurrent neural networks is to utilize the sequential information. These networks predict the output on the basis of current input and the previous information coming. RNNs generally have a memory which stores the information that has been predicted so far. The word recurrent implies that they executed the same job for each of the elements present in the sequence with the output being depended on the previous computations [24, 25].

In the Fig. 3 the RNN network applies a recurrence formula to the input vector and its previous state. To get a current new state both of these are considered. The formula for the current state is expressed in equation (2):

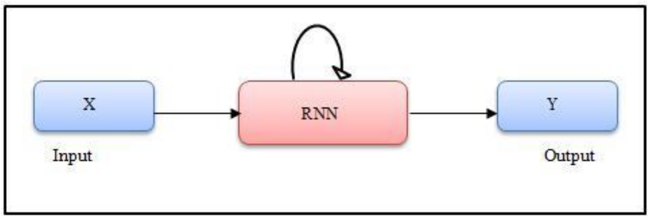


*Fig. 3. A RNN cell.*

$$h_t = f(h_{t-1}, X_t) \quad (2)$$

Where, $h_t$ is new state, $h_{t-1}$ depicts previous state, and $X_t$ is the current input. Each successive input coming here is called a time step because the input neurons are applying transformation on previous input. Considering the RNN in its simplest form, there is a *tanh* function, $W_{hh}$ is the weight at the recurrent neuron, $W_{xh}$ is the weight at the input neuron. At time t, the state can be written as:

$$h_t = tanh(W_{hh}h_{t-1} + W_{xh}X_t) \quad (3)$$

In this case, only the immediate previous state is taken into consideration by the recurrent neuron. The same equation can be unfolded into multiple states for the longer sequences. Once we calculate the current state, the output state can be expressed by following equation (4):

$$y_t = W_{hy}h_t \quad (4)$$

The output generated is compared to the actual output and then the error is produced which is then back propagated to the network in order to update the weights so as to train the network.

There is a problem with RNNs, as it can't remember the long term dependencies because of the vanishing gradient problem. We know that in a traditional neural network when the weight is updated for a particular layer it is the result of learning rate of multiple layers, which is the error from the previous layer and the input to that layer. So the error thus obtained is the result of the product of the errors from the previous layers. In the activation functions like the sigmoid functions, the value of its derivatives gets multiplied multiple times as we continue to move towards the upper layers. Because of this, the gradient completely vanishes while moving towards the starting layers making it hard to train the starting layers.

### *C. Long-Short Term Memory LSTM*

RNNs are powerful for the purpose of sequence handling to a great extent but only for the shorter contexts. When longer contexts come, the information gets lost somewhere and which in turn makes RNNs less feasible. They work well when we deal with short-term dependencies. This problem can be easily solved by the help of a variant of RNNs known as LSTMs [34].

LSTMs are known to be the extension of recurrent neural networks because of their power of carefully remembering selective information for extended periods of time. LSTMs achieve this with the help of a memory which allows them to remember or forget the information stored in the memory. This memory is known as the gated cell which takes the decision whether to store or delete the information present in it. It does this by giving importance to the information which occurs with the help of the weights learned. So LSTMs learn over a period of time which information is important and which is not. LSTM cell is composed of various components:

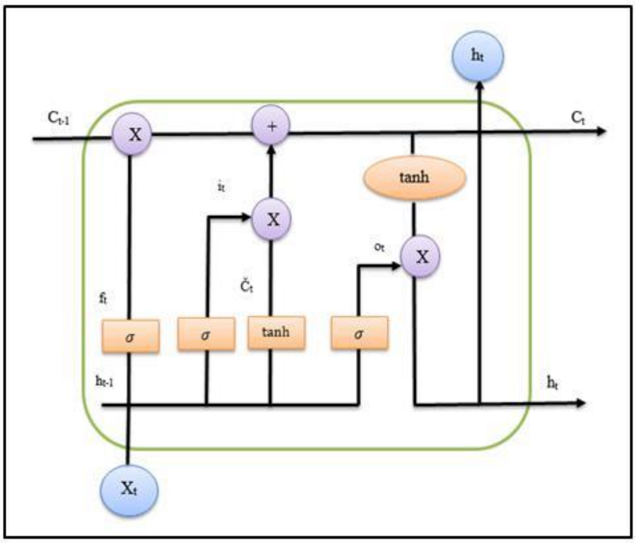


*Fig. 3. LSTM cell.*

*Forget gate:* It is responsible for eliminating theinformation which is no longer needed or is of lesser importance from the LSTM cell.

$$f_t = \sigma(W_f . h_{t-1}, x_t) + b_f \quad (5)$$

In equation (5) it can be seen that, $h_{t-1}$ is the output from previous cell, and $x_t$ is input at a particular time step. $W_f$, $b_f$ are weight matrices and biases respectively. The particular input is multiplied with these weights and biases. The sigmoid function is applied to this value which produces a vector with values from 0 to 1. This function decides which value to store and which value to remove. If it is 1, the value is kept and if it is 0, the value is removed.

*Input gate:* Whenever new information is added, then inputgate is responsible for that. It is very similar to the forget gate.

$$i_t = \sigma (W_i . [h_{t-1}, x_t] + b_i) \quad (6)$$

$$\check{C}_t = tanh(W_c . [h_{t-1}, x_t] + b_c) \quad (7)$$

This gate serves as filter for all the information from $h_{t-1}$, $x_t$.The tanh function is responsible for creating a vector which outputs values from -1 to +1 so that these values can be added to the cell state. Finally the value of the sigmoid gate is multiplied with the value of the *tanh* function and then added to the cell state. After this step, it is ensured that redundant information is removed and only relevant information is kept.

*Current state:* $C_t = f_t * C_{t-1} + i_t * \check{C}_t$ (8)

Here, $C_t$is the new update memory, and$C_{t-1}$is memory from last LSTM unit

*Output gate:* $O_t = \sigma(W_o [h_{t-1}, x_t] + b_o)$ (9)

$$h_t = o_t * tanh(C_t) \quad (10)$$

Initially a vector with values ranging from -1 to +1 is created after applying the *tanh* function to the cell state. The filter here again employs a sigmoid function. The values of this sigmoid gate are multiplied with the vector and are sent to the hidden state of the next cell or as an output.

## V. APPLICATIONS

Deep learning has exciting application around artificial intelligence and machine learning. It is applied in many areas such as automatic colorization of images, object detection using deep CNNs [15], generating realistic sounds [41], image captioning [30], text generation [42] using sequential LSTM recurrent network. The following subsections discuss several fields where deep learning has been applied: the application in respective areas of image processing, biometric, medicine and social media.

.

### *A. Image Processing*

Within the concept of pattern recognition some applications has been carried out using layering process. Face recognition problem is conducted in respective constrained environments, such as Local Binary Patterns (LBP) [43] and Local Phase Quantization (LPQ) [44]. From last three years, CNN has achieved an impressive result on Facial Recognition in Unconstrained Environments (FRUE) [45].

The unique deep learning models are used by Google, Microsoft, and Facebook for face recognition. Automatic recognition of gender and age is also established by initial features. Furthermore, a deep multi-task learning model was adopted to determine age and gender of a person from single picture accurately.

The objects and context within photograph are used to color the image using DL techniques similar to human that drive this problem. Generally, to recreate the image with addition of colors it requires large trainable CNNs with supervised layers [46].

One of the most important applications of deep learning is recognition of location based on images. In this, address can be recognized by using the contents of picture. The use of Supervised Semantics Preserving Deep Hashing (SSPDH) algorithm shows outstanding performance as compare to Visual Hash Bit (VHB) that gives accuracy of about 70% [47].

Finally, another remarkable performance of deep CNNs is highlighted in Digital Image Forensics (DIF) field, where it is widely incorporated for one of the two important techniques, i.e., source identification and temper detection. The application of CNNs in forensics field is now an emerging research area where different technique are developed including RNN, LSTM and auto-encoders which gives the accuracy of about 98.86% [48].

### *B. Biometric*

In 2009, the deep belief networks with different architecture implemented for automatic speech recognition in order to reduce the Phone Error Rate (PER). In 2012 Deep Neural Network (DNN) used by Google as core technology for development of sounds model of language, which refer as Google voice search. The Gaussian Mixture Model which has been in industry for 30 years was replaced by this DNN model. The DNN come out to be a promising technology for speech recognition at every instant of time delivered with predominant accuracy [49].

Within the framework of a Hybrid Neural Network – Hidden Markov Model (NN - HMM), a CNN method was applied that shows PER rate of about 20.07%. The smartphones developed by different vendors are tested for iris recognition. Camera resolution within smartphones undergoes iris recognition using deep convolutional neural networks that reach up to 87% of identification accuracy.

Deep learning in conjunction with biometrics characteristics are used for security purposes, especially giving access control of particular device. Nowadays, DL is continuously incorporated for development and optimization of devices for FaceSentinel, which includes biometric characteristics like iris, palm, and fingerprint recognition.

### *C. Medicine*

Deep learning in healthcare is helping a lot to medical professionals and researchers, in order to uncover hidden data in a better way. Using deep learning techniques, it provides accurate analysis of any disease to treat them resulting in better medical decision.

For drug discovery application, the first deep neural network was designed in 2015 called AtomNet, using parameters of multiple convolutional layers. Using this structure, it predicts the bioactivity of molecules which provide discovery of medicine and their development. The proposed AtomNet achieved 0.9 AUC on 57.8% of the target.

Another impressing application of DL techniques is early detection of Alzheimer's disease, which medical industry highly faced to figure out. MRI images of patients are used to perform several classification using deep learning methods [50].

Deep learning has great contribution in genomics, to understand genome and get an idea about diseases. Deep Convolutional Neural Networks (CNNs) for example capable of automatically extracting local as well as global genome data characteristics whereas Recurrent Neural Networks (RNNs) are widely adopted for sequential data recognition like speech recognition which is highly used to learn the particular DNA sequence [51].

## VI. Conclusion

In this paper, a comprehensive study on widely used deep learning models is focused. Deep learning is still a growing field to enormous number of unsolved problems, and can be applied with specific problem models. Deep learning is not just limited in the computer science field; it has gained popularity in each and every field and is excelling very well to come up with finest solutions for the problems. This quick and timely survey can help researchers rapidly learn about the key aspects of deep learning whose effect spans a broad range of domains.

It focuses on various models and how they can be used. It also highlighted the various application areas in which deep learning is applied nowadays. From the survey it was also found out that CNNs are more widely used for the image processing applications, on the other hand LSTMs and RNNs are popular among the text based applications.

## References

[1] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," The *bulletin of mathematical biophysics*, vol. 5, no. 4, 1943.

[2] A. Mosavi, A.R. Varkonyi-Koczy, "Integration of Machine Learning and Optimization for Robot Learning," in *Advances in IntelligentSystems and Computing*, vol. 519, pp. 349-355, 2017.

[3] "What I learned from competing against a convnet on ImageNet," *Karpathy.github.io*, 2018.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *CVPR*, 2016.

[5] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, and L. Fei-Fei,"ImageNet: A large-scale hierarchical image database," in *CVPR*, 2009.

[6] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z.

Huang, A. Karpathy, A. Khosla, M. Bernstein et al., “Imagenet large scale visual recognition challenge,” in *IJCV*, vol. 115, no. 3, 2015.

[7] B. Heo, K. Yun and J. Y. Choi, "Appearance and motion based deep learning architecture for moving object detection in moving camera," *2017 IEEE International Conference on Image Processing(ICIP)*, Beijing, pp. 1827-1831, 2017.

[8] A. Mosavi, A. R. Varkonyi-Koczy, “Integration of Machine Learning and Optimization for Robot Learning,” *Advances in Intelligent Systemsand Computing*, vol. 519, pp. 349-355, 2017.

[9] Y. Bengio, “Learning deep architectures for AI,” *Foundations andtrends in Machine Learning*, vol. 2, pp. 1-127, 2009.

[10] K. Fukushima, Neocognitron, “A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” *Biological cybernetics*, vol. 36, pp. 193-202, 1980.

[11] B. B. Le Cun, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard,and L. D. Jackel, “Handwritten digit recognition with a backpropagation network,” *NIPS Proceedings*, pp. 396-404, 1990.

[12] Alex Krizhevsky, Sutskever I, and Hinton G.E, “Imagenet classification with deep convolutional neural network,” *NIPS Proceedings*, 2012.

[13] D. H. Hubel and T. N. Wiesel, “Receptive fields and functional architecture of monkey striate cortex,” *The Journal of physiology*, vol. 195, no. 1, pp. 215-43, 1968.

[14] Ramachandran R, Rajeev DC, Krishnan SG, P Subathra, “Deep learning an overview,” *IJAER*, vol. 10, no. 10, pp. 25433-25448, 2015.

[15] J. Fan, W. Xu, Y. Wu, and Y. Gong, “Human tracking using convolutional neural networks, Neural Networks,” *IEEE Transactions*, vol. 21, no. 10, pp. 1610 – 23, 2010.

[16] A. Toshev and C. Szegedy, “Deep -pose: Human pose estimation via deepneural networks,” *Computer Vision & Pattern Recognition*, 2014.

[17] ]M. Jaderberg, A. Vedaldi, and A. Zisserman, “Deep features for text spotting,” *ECCV,* 2014.

[18] R. Zhao, W. Ouyang, H. Li, and X. Wang, “Saliency detection by multicontext deep learning,” *Computer Vision & Pattern Recognition*, 2015.

[19] J. Donahue, Y. Jia, O. Vinyals, J. Hoffman, N. Zhang, E. Tzeng, and T. Darrell, Decaf. “A deep convolutional activation feature for generic,” 2014.

[20] C. Farabet, C. Couprie, L. Najman, Y. LeCun, “Learning hierarchical features for scene labeling,” *PAMI,* 2013.

[21] Nithin, D Kanishka, Sivakumar, P Bagavathi, “Generic Feature Learning in Computer Vision,” *Elsevier,* vol. 58, pp. 202-209, 2015.

[22] John J. Hopfield, “Neural networks and physical systems with emergent collective computational abilities,” *Proceedings of the NationalAcademy of Sciences*, vol. 79, no. 8, pp. 2554 - 2558, 1982.

[23] Michael I. Jordan, “Serial order: A parallel distributed processing approach,” *Technical Report 8604, Institute for Cognitive Science*, University of California, San Diego, 1986.

[24] Ilya Sutskever, Oriol Vinyals, and Quoc V. Le, “Sequence to sequence learning with neural networks,” *Advances in Neural InformationProcessing Systems*, pp. 3104 - 3112, 2014.

[25] Je_rey L. Elman, “Finding structure in time,” in *Cognitive science*, vol. 4, no. 2, pp. 179 – 211, 1990.

[26] Yoshua Bengio, Patrice Simard, and Paolo Frasconi, “Learning long-term dependencies with gradient descent is difficult,” *IEEETransactions on Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[27] Avrim L. Blum and Ronald L. Rivest, “Training a 3-node neural network is NPcomplete,” *Machine Learning: From Theory toApplications, Springer*, pp. 9–28, 1993.

[28] James Bergstra, Olivier Breuleux, Fr_ed_eric Bastien, Pascal Lamblin, Razvan Pascanu, Guillaume Desjardins, Joseph Turian, David Warde-Farley, and Yoshua Bengio, “Theano: a CPU and GPU math expression compiler,” In *Proceedings of the Python for Scientifc ComputingConference (SciPy)*, vol. 4, pp. 1- 7, 2010.

[29] Ronan Collobert, Koray Kavukcuoglu, and Cl_ement Farabet, “Torch7: A matlablike environment for machine learning,” *BigLearn, NIPSWorkshop*, 2011.

[30] A. Karpathy and L. Fei-Fei, "Deep Visual-Semantic Alignments for Generating Image Descriptions," in *IEEE Transactions on PatternAnalysis and Machine Intelligence*, vol. 39, no. 4, pp. 664-676, 2017.

[31] J. Donahue *et al*., "Long-Term Recurrent Convolutional Networks for Visual Recognition and Description," in *IEEE Transactions on PatternAnalysis and Machine Intelligence*, vol. 39, no. 4, pp. 677-691, 1 2017.

[32] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent Convolutional Neural Networks for Text Classification," In *AAAI*, vol. 333, pp. 2267-2273, 2015.

[33] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neuralcomputation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[34] M. Sundermeyer, R. Schluter, and H. Ney, “Lstm neural networks ¨ for language modeling.” in *INTERSPEECH*, pp. 194–197, 2012.

[35] F. A. Gers and J. Schmidhuber, “LSTM recurrent networks learn simple context free and context sensitive languages,” *IEEE Transactions onNeural Networks*, vol. 12, no. 6, pp. 1333–1340, 2001.

[36] K. Zhang, W.-L. Chao, F. Sha, and K. Grauman, “Video summarization with long short-term memory,” in *European Conference on ComputerVision, Springer*, pp. 766–782, 2016.

[37] Nitish Srivastava, Elman Mansimov, and Ruslan Salakhutdinov. “Unsupervised learning of video representations using LSTMs,” in *Proceedings of International Conference on International Conference on Machine Learning*, vol. 37, pp. 843-852, 2015.

[38] S. Venugopalan, M. Rohrbach, J. Donahue, R. Mooney, T. Darrell, and K. Saenko, “Sequence to Sequence - Video to Text,” arXiv preprint *arXiv: 1505.00487,* 2015.

[39] W. Zaremba, I. Sutskever, “Learning to Execute,” arXiv preprint *arXiv:1410.4615*, 2014.

[40] M. Najafabadi, F. Villanustre, T. Khoshgoftaar, N. Seliya, R. Wald, and E. Muharemagic, “Deep learning applications and challenges in big data analytics,” *Springer*, vol. 2, 2015.

[41] A. Owens, P. Isola, J. McDermott, A. Torralba, Edward H. Adelson, W. T. Freeman, “Visually Indicated Sounds,” *arXiv:1512.08512[cs.CV]*, 2016.

[42] Alex Graves, “Generating Sequences with Recurrent Neural Networks,” *arXiv: 1308.0850v5[cs.NE]*, 2014.

[43] T. Ahonen, A. Hadid, and M. Pietikainen, “Face recognition with local binary patterns,” in *Computer Vision-ECCV, Springer*, pp. 469–481, 2004.

[44] T. Ahonen, E. Rahtu, V. Ojansivu and J. Heikkila, "Recognition of blurred faces using Local Phase Quantization," *19th InternationalConference on Pattern Recognition*, Tampa, FL, pp. 1-4, 2008.

[45] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Advances in neuralinformation processing systems,* pp. 1097–1105, 2012

[46] D. Varga and T. Szirányi, "Fully automatic image colorization based on Convolutional Neural Network," *2016 23rd International Conferenceon Pattern Recognition (ICPR)*, pp. 3691-3696, 2016.

[47] Huei-Fang Yang, Kevin Lin, Chu-Song Chen, “Supervised Learning of Semantics-Preserving Hash via Deep Convolutional Neural Networks,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 2, pp. 437- 451, 2018.

[48] B. Bayer, M. Stamm, “A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer,” *Proceedings of the 4th ACM Workshop on Information Hiddingand Multimedia Security – IH&NNSec’ 16*, 2016.

[49] F. Marr, “A Deep Learning Approach for Iris Sensor Model Identification,” *Pattern Recognition Letters, Elsevier*, 2017.

[50] Petersen R. C., Aisen P.S., Beckett L.A., et al, “Alzheimer’s disease Neuroimaging Initiative (ADNI) clinical characterization,” *Neurology* vol. 74, no. 3, pp. 201–209. 2010.

[51] Daniel Quang and Xiaohui Xie. Danq, “a hybrid convolutional and recurrent deep neural network for quantifying the function of DNA sequences,” *Nucleic Acids Research*, vol. 44, no. 11, pp. e107-e107, 2016.