International Journal of Computer Applications (0975 - 8887)
Volume 184 - No.18, June 2022

Review of Deep Learning: Architectures, Applications and Challenges

Ankit Sirmorya
ML Engineering Manager, Amazon

Milind Chaudhari
Sr. Cloud Consultant, Slalom

Suhail Balasinor
Information Security Architect, Mini-Circuits

ABSTRACT

Deep Learning is a continuously evolving subset of machine learning techniques. New technology has provided solutions to a wide range of complex problems that were once unsolvable due to limitations in human intelligence. Since its conception, several DL architectures have been developed, including recursive neural networks, recurrent neural networks, artificial neural networks, and convolution neural networks. Many of their contributions have been in the area of computer vision, natural language processing, sequence generation, etc. Despite their increasing popularity, many individuals cannot see the bigger picture or comprehend these techniques. In this paper, the various deep learning models are described, as well as how they work. In addition, the article explains a few prominent DL models and their relevance in contemporary technology. As with every rapidly changing technology, DL has some limitations. These limitations are mitigated to some extent in this paper. Further, it emphasizes their continued development, the challenges they face, and the possibilities for future research in their fields

Keywords

Deep Learning, Machine Learning

1. INTRODUCTION

An important part of Deep Learning(DL) is that it mimics the information processing patterns found in the human brain. To map inputs to labels, DL uses a huge amount of data. DL is designed using multiple layers of algorithms, each providing a unique interpretation of the data it has been presented with[1,2].
Conventional ML techniques require several sequential steps to accomplish the classification task, specifically pre-processing, feature extraction, a wise feature selection, and learning. In addition, feature selection affects the performance of machine learning techniques. The selection of features may be biased, leading to incorrect class distinctions. Contrary to conventional ML techniques, DL can automate the learning of feature sets for multiple tasks[1,3].
The growth and evolution of Big Data have made DL an incredibly popular ML algorithm in recent years[4, 5]. This is still in continuous development regarding novel performance for many ML tasks[6, 7, 8] and has simplified the improvement of many learning fields [9, 10], such as image super-resolution[11], object detection[12, 13], and image recognition [14, 15]. Yoshua Bengio, Yann LeCun, and Geoffrey Hinton were among the three pioneers[16] in the field of deep learning who have been awarded the 2019 Turing Award. The ability of DL to enhance human life includes improving the accuracy of diagnosing natural disasters [17], discovering new drugs [18], and detecting cancer [19, 20]. With the 2020 outbreak of the novel coronavirus (COVID-19)[21-23], DLs will have an increasingly critical role to play in early diagnosis.
Several reasons have led to this widespread adoption of DL, among them: i) DL is robust to changes in input data because we do not need precisely designed features and we learn optimized features automatically ii) It is possible to generalize DL techniques in different applications by using a method known as transfer learning, which utilizes knowledge gained from solving one problem to solve another related problem iii) The DL techniques are highly scalable. For example, Microsoft's ResNet [24] consists of 1202 layers and is frequently applied at a supercomputing scale.

2. DEEP LEARNING MODELS

2.1 Artificial Neural Network(ANN)

Artificial Neural Networks ANNs are neural networks whose architecture is modeled after the brain. They typically consist of hundreds of simple processing units (neurons) which are wired together in a complex communication network.[122]. Each neuron has its own input, from which it receives communications from other neurons, and also it has a function f (activation function) through which it transforms its own global input into output[123]. After receiving the output it then transfers it to the next layer for further processing. The connections between the nodes can modify themselves over time to optimize the output.
After structuring the network in different layers it is then trained on a dataset. To start with, Process the initial weights are chosen randomly. Then, the training, or learning, begins. There are two approaches to training- supervised and unsupervised.[124]Supervised training involves a mechanism of providing the network with the desired output either by manually "grading" the network's performance or by providing the desired outputs with the inputs. Unsupervised training is where the network has to make sense of the inputs without outside help.[122]
These learning methods have been applied to many problems such as the modeling of performance and exhaust emissions of DI engines using emulsified diesel fuel [79], estimation of evaporation in hilly areas [80], and stock market prediction [81]. A study shows that these artificial neural networks are able to perform cognitive tasks at a high level of efficiency, much like the human brain.[82]

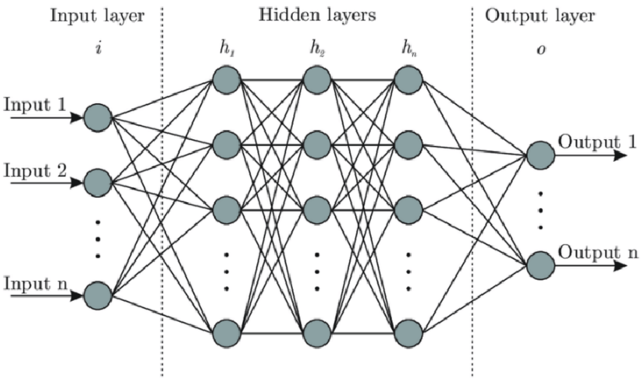

Fig. 1. ANN model

Figure 1[165] illustrates a neural network that takes n inputs and produces n outputs. In addition, there are n hidden layers and an input and output layer that are interconnected.

2.2 Recursive neural networks (RvNN)

RvNN is a deep learning model created by applying the same set of weights recursively over a structured input. By utilizing compositional vectors[27], RvNN can make predictions in a hierarchical structure and classify the outputs. Objects with randomly shaped structures, such as graphs or trees, are processed by the RvNN architecture. A variable-size recursive data structure is used to generate a fixed-width distributed representation. The network is trained by back-propagation through the structure (BPTS) learning [28]. BPTS uses a similar technique to the general-back propagation algorithm and can support tree-like structures. As a result of auto-association, the network is trained to regenerate the input-layer pattern at the output layer. In the context of NLP, RvNN is highly effective.
In addition to making predictions in a hierarchical structure, RvNN can also classify the outputs using compositional vectors[27]. In the RvNN architecture, objects with randomly shaped structures, such as graphs or trees, can be processed (repeated). RvNN plays an important role while deciding the next appearing word in a sentence where the sequence of order matters and the next word depends on the preceding sentence. Using this approach, a variable-size recursive-data structure is converted into a fixed-width distributed representation. An integrated back-propagation through structure (BTS) learning system is used to train the network[28]. BPTS tracks the same technique as the general-back propagation algorithm and allows for tree-like structures to be created. The auto-association method trains the network to re-create the input-layer pattern at the output layer. NLP applications of RvNN are highly effective. As Socher et al. stated [29], their RvNN architecture enables the processing of inputs from different modes simultaneously. They demonstrate two ways of classifying natural language sentences: splitting sentences into words and images of nature, as well as breaking each image up into various segments of interest. RvNN constructs a syntactic tree based on a pair of likely scores. Additionally, RvNN calculates a score that indicates how likely each pair of units is to merge. Next, the two pairs with the highest scores are merged into a composition vector. RvNN's tree structure's root is the compositional vector for the entire area. There are several applications in which RvNN has been applied[30-32].

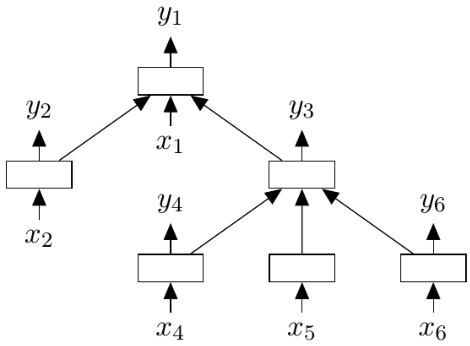

Fig. 2. RvNN model

Figure 2[166] illustrates a simple recursive neural network architecture. A RvNN is a hierarchical model which can take n inputs and these are then further given to immediate nodes for processing and then finally fed to the output layer to generate an output.

2.3 Recurrent neural networks(RNN)

RNN is mostly used in the context of speech processing and natural language processing [33, 34]. RNNs use sequential data in their network. The embedded structure in the data sequence provides valuable information, so this feature is essential to a range of applications. Understanding the context of a sentence, for example, is crucial to understanding the meaning of a particular word within it. Paccanu et al.[35] proposed three types of deep RNNs: Hidden-to-Hide, Hidden-to-Output, and Input-to-Hidden. Combining these three techniques results in a deep RNN that reduces learning difficulty and brings the benefits of a deeper network.

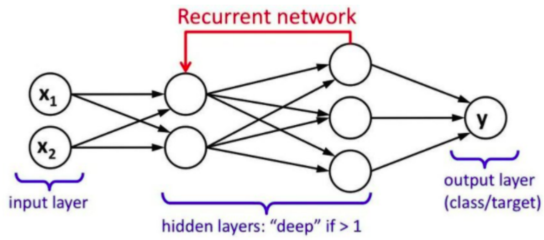

Fig. 3. RNN model

However, RNN's sensitivity to the exploding gradient and vanishing problems represent one of the main issues with this approach[36]. More specifically, during the training process, the reduplications of several large or small derivatives may cause the gradients to exponentially explode or decay. With the entrance of new inputs, the network stops thinking about the initial ones; therefore, this sensitivity decays over time. To solve this problem we use LSTM(Long Short Term Memory) cells that keep track of information and its long-term dependencies[37].

In figure 3 [121] the second hidden layer sends feedback to the previous layer creating a loop, allowing information to persist, i.e the network not only learns from the input but also from the previous output.

2.4 Convolutional neural networks

CNN is the most famous and commonly used algorithm in DL, and the main advantage of CNN over its predecessors is that it automatically identifies the relevant features without any human intervention [38]. CNN works on a hierarchical basis that helps to identify and extract features easily. Three key benefits of CNN have been identified by Goodfellow et al. [39]: equivalent representations, sparse interactions, and parameter sharing. CNN's are based on neurons found in animal and human brains. To maximize the use of 2D input structures like image signals, the CNN makes use of shared weights and local connections, unlike conventional fully connected (FC) networks. The variations in images in image recognition, for example, include illumination, lighting, viewpoint, scale, deformation, occlusion, and intra-class variations[40]. Hubel and Wiesel in 1959[41] investigated how a biological brain processes visual information, finding that the visual cortex processes different stimuli using a hierarchy of different types of cells. At first, simple cells respond to oriented edges and light orientation, and complex cells respond to light orientation and movement, while hypercomplex cells respond to movement with an endpoint. The first neural layers in facial recognition are responsible for detecting low-level features such as lines, edges, noses, and ears. Lastly, complex layers detect high-level features such as facial structure. The CNN architecture consists of a number of layers.
Convolutional Layer: In CNN architecture, the most significant component is the convolutional layer. It consists of a collection of convolutional filters (so-called kernels). The input image, expressed as N-dimensional metrics, is convolved with these filters to generate the output feature map.[135]
Pooling Layer: The main task of the pooling layer is the subsampling of the feature maps. These maps are generated by following the convolutional operations. In other words, this approach shrinks large-size feature maps to create smaller feature maps.[135]

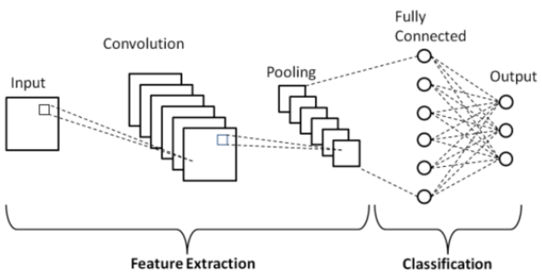

Fig. 4. CNN model

The figure 4 [120] illustrates a basic convolution neural network. The convolutional layer aims to learn feature representations of the inputs. Pooling layer that seeks to downscale the image. Generally, it is followed by a fully connected layer which is made up of a set of neurons densely connected to the succeeding layer (in this case, the output layer). The set of neurons represent features of an image extracted by the convolution and pooling layer and the output layer based on the features classifies the input into different classes.

2.5 General Adversarial Networks (GAN)

Generative Adversarial Networks (GAN) is a deep learning framework introduced by Ian Goodfellow in 2014[85]. GAN incorporates two different neural networks: a generator and a discriminator. The generator is intended to capture the distribution of some target data (e.g. distributions of pixel intensity in images), i.e. to uncover how the real data is structured. Discriminator assists the training of generators by studying the data generated by a generator in relation to the underlying data (i.e. identifying the differences between generated and original data), thereby facilitating the learning of the distribution that represents the underlying data.[126]. Generators takes Noise: A z-dimensional vector of completely random values as an input. Various outputs are produced based on different noise vectors, all belonging to the same class (i.e. same distribution of data). A discriminator is a binary classifier that distinguishes between real and fake images generated by the generator. It has been demonstrated that GANs have been able to solve many problems, such as creating an image from a description[86], getting a high-resolution image from a low-resolution image, generating characters [87], and creating a painting from a real-world image.
Different variants of GANS serve different purposes. The fully connected GANs that were proposed in the original paper were applied to data sets such as MNIST and CIFAR [85]. Deep Convolution GANs or DCGANs use Deep convolutional networks for both the generator and discriminator and are a slightly modified version of GANs[88]. A conditional GAN is similar to a GAN, but alongside the noise, conditions, or another parameter can also be used to modify the output, for example, a class i.e a specific distribution of data to which the input belongs[89]. In the years to come, many variants of GANs will be introduced and researched and are expected to have a prominent impact on deep learning. GAN has gained massive attraction in computer vision [127,128,129], feature representation [130], and more recently in Natural Language Processing (NLP) tasks: Document Modeling [131], Dialogue Generation [132], Sentiment Analysis [133], and Domain Adaptation [134].

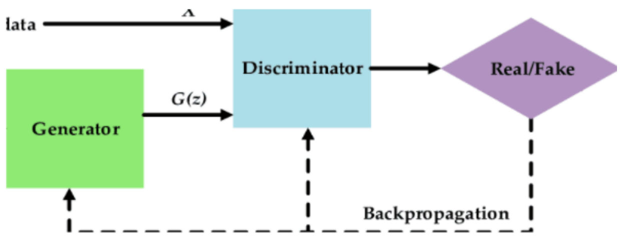

Fig. 5. GAN model

Figure 5 [118]shows a basic GAN which has a generator and discriminator. The generator generates G(z) as an output, which is then fed to the discriminator along with the real data. It's the job of the discriminator to distinguish between Real images and those generated by the generator. The discriminator then provides feedback to the generator to improve it's output.

2.6 AutoEncoders(AE)

An Autoencoder is an unsupervised neural network model used for representation learning, e.g., feature selection or dimension reduction. The network compresses the input into a meaningful representation and then decompresses it to produce an output closely related to the original input[101]. Recently, a variety of (unsupervised) representation learning algorithms has been proposed based on the idea of autoencoding where the goal is to learn a mapping from high-dimensional observations to a lower-dimensional representation space such that the original observations can be reconstructed (approximately) from the lower-dimensional representation[101]. They were first introduced in the article"Learning Internal Representations by Error Propagation"[90] and since then have been changed in different ways. Autoencoders can achieve a much better two-dimensional representation of array data, when an adequate amount of data is available [153]. An AE is performed by three components - an encoder that converts an input into a code (a meaningful representation), a code, and a decoder that attempts to bring the code back to its original form. The auto-encoder is widely used in many unsupervised learning tasks, e.g., dimensionality reduction, feature extraction, efficient coding, generative modeling, denoising, anomaly or outlier detection, etc. [108]
In the encoder, the code section that is the minimalist representation of the input is called the bottleneck. Compared to other layers in the network, it contains the fewest neurons. It is the number of neurons that constitute the bottleneck that determines the quality of the output.[108] In general, more neurons lead to higher quality results. The bottleneck is an effective way to introduce regulation into the network and also reduce the dimensionality of the data[103].

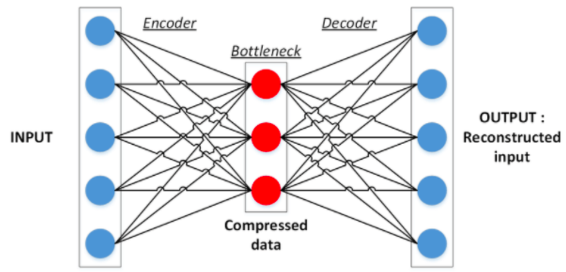

Fig. 6. Autoencoder model

Figure 6[106] shows an autoencoder consisting of two sub neural networks, encoder and decoder. The encoder takes input x and processes it to form a bottleneck vector z (The compressed Data). The Bottleneck vector is then fed to the decoder which constructs an input x'

2.6.1 Variational Autoencoders. A Variational Autoencoder is a type of likelihood-based generative model. It consists of an encoder that takes in data x as input and transforms this into a latent representation z, and a decoder that takes a latent representation z and returns a reconstruction x'.[135]. Rather than the bottleneck vector z,VAE generate a mean vector , and a standard deviation vector, which are fed to a sample latent vector. A sample latent vector is selected from the distribution and fed into the Decoder[103]. The loss function of VAE differs from that of the basic autoencoders. IT consists of two parts: a generatives loss and a latent loss.[110].A Generative loss compares the output with the model input. This can be the losses we use in the autoencoders, such as L2 loss. Latent loss compares the latent vector with a zero mean, unit variance Gaussian distribution. The loss we use here will be the KL divergence loss. A Variational Auto-encoder (VAE) has been widely used in a variety of applications, such as image generation [111], [112], dialog generation [113], [114], [115], and disentangled representation learning [116], [117].

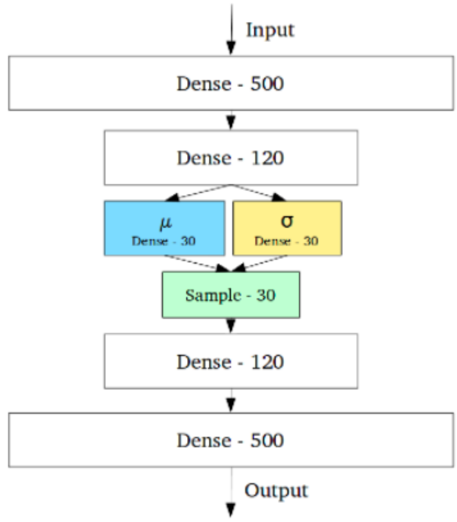

Fig. 7. Variational Autoencoder with a mean, std,dev and sample layer

Figure 7[168] shows a simple Variational Autoencoder which passes an input through the encoder. Two vectors are then output by the encoder, which are fed into a sample latent vector and then passed to a decoder.

2.6.2 Sparse Autoencoders. The Sparse Autoencoder is a type of autoencoder that generates a minimalistic representation of input through sparsity. Those neurons whose output is close to 1 are active, whereas those whose output is close to 0 are inactive[108]. The encoder has two perform two basic tasks, the autoencoder architecture should be able to reconstruct inputs well (i.e. reduce reconstruction errors), and other, it should generalize the low representation into something meaningful. This is achieved by introducing sparsity. [103]. In order to utilize the inner structure of data, an additional regularization, the sparsity constraint on the hidden units, is developed. By using L1 regularization or KL divergence, the sparsity constraint can be imposed. To improve the sparsity and the result, [107] proposed a K-sparse encoder, which is an autoencoder with linear activation function, wherein hidden layers only the k highest activities are kept. In [109], the use of sparse encoders for Feature Extraction and Assessment of Locomotive Adhesion Status is discussed. It proposes a sparse autoencoder deep neural network with dropout to diagnose the wheel-rail adhesion state of a locomotive. This deep neural network can significantly reduce the adverse effect of overfitting, making the learned features more conducive to classification and identification.

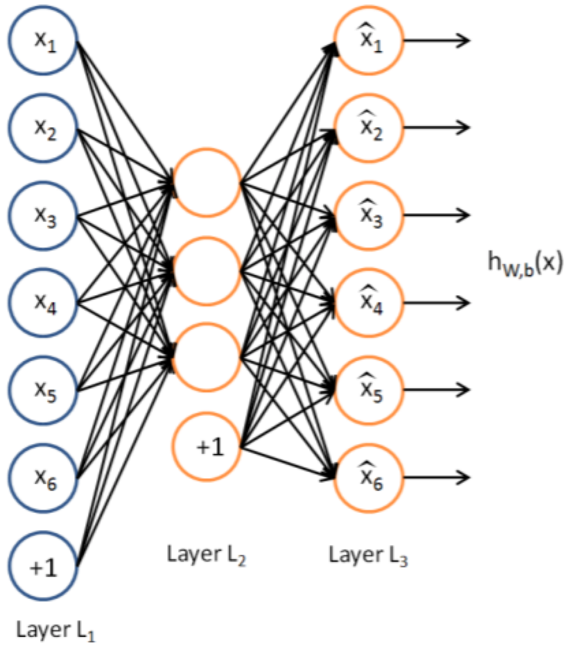

Fig. 8. Sparse Autoencoder model

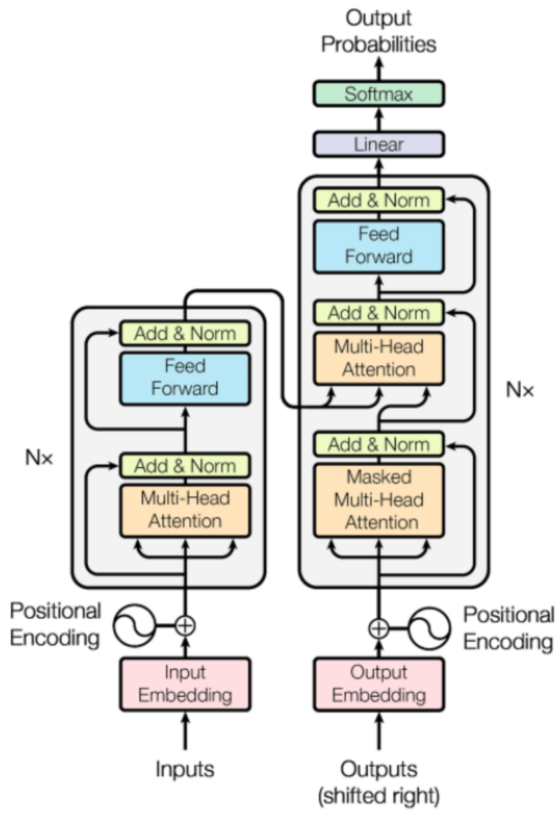

Fig. 9. Transformer model

Figure 8[167] shows a simple sparse autoencoder. The central layer(the bottleneck) has 4 out of which only one neuron has been activated and thus is only the one that will contribute towards the output.

2.7 Transformer

Transformer is a deep learning model that works on the principle of self-attention. In this case, the model is capable of enhancing some parts of the input data and diminishing others. A transformer was first presented in [154]. The article also described Sequence-to-Sequence (Seq2Seq), a neural net that converts one sequence of elements into another.[155] A Transformer is a neural network that can transfer one sequence to another with the help of an Encoder and Decoder. The Encoder is on the left and the Decoder is on the right. Both Encoder and Decoder are composed of modules that can be stacked on top of each other multiple times. It contains A Multi head attention and a feed forward layer, which is represented by Nx.[155] Most competitive neural sequence transduction models have an encoder-decoder structure [159, 159, 161]. Here, the encoder maps an input sequence of symbol representations (x1, ..., xn) to a sequence of continuous representations z = (z1, ..., zn). Given z, the decoder then generates an output sequence (y1, ..., ym) of symbols one element at a time. At each step the model is auto-regressive [160], consuming the previously generated symbols as additional input when generating the next.

Figure 9 shows a transformer[156]. The encoder is composed of a stack of N = 6 identical layers. Each layer has two sub-layers. The first is a multi-head self-attention mechanism, and the second is a simple, position wise fully connected feed-forward network.The multi-head mechanism implements h heads that receive a (different) linearly projected version of the queries, keys and values each, to produce h outputs in parallel that are then used to generate a final result.[164] We employ a residual connection [162] around each of the two sub-layers, followed by layer normalization [163].The decoder is also composed of a stack of N = 6 identical layers. In addition to the two sub-layers in each encoder layer, the decoder inserts a third sub-layer, which performs multi-head attention over the output of the encoder stack. Similarly to the encoder, we use residual connections around each layer, followed by layer normalization. As part of the first sublayer, the previous output of the decoder stack is augmented with positional information and is subjected to multi-head self-attention. Unlike the encoder, which is designed to attend to all words in the input sequence regardless of position in the sequence, the decoder is modified to pay attention to only the preceding words. Therefore, the output of a word at position i can only be predicted based on the known outputs of the words that precede it in the sequence. In the multi-head attention mechanism (which implements multiple, single attention functions in parallel), this is achieved by introducing a mask over the values produced by the scaled multiplication of matrices Q and K.[164]The Transformer [154] has rapidly become the dominant architecture for,any machine learning applications, surpassing alternative neural models such as convolutional and recurrent neural networks in performance.

3. LIMITATIONS OF DEEP LEARNING AND THEIR MITIGATIONS

3.1 Lack of training data

To achieve a well-behaved performance model, DL needs a large amount of data, so as the data increases, the model becomes even more well-behaved. However, sometimes there is not enough data to use DL directly [42].
Two methods can be used to properly address this issue. First, we employ the concept of transfer learning (TL) after collecting data from similar tasks. To address the lack of training data problem, the TL method is currently used [43, 44], which is highly efficient in addressing the undersized dataset problem. Training TL models with large volumes of data are the mechanism behind TL. Next, the model is fine-tuned for training on a small dataset of requests.
The second method involves data augmentation [45]. When used to augment image data, this task is extremely helpful, since the image translation, mirroring, and rotation do not change the image labels. On the other hand, it is important to take care when using this technique with some types of data, such as bioinformatics. As an example, when mirroring an enzyme sequence, the output data may not reflect the true sequence.

3.2 Imbalanced Data

When DL is trained with imbalanced data, undesirable results may result. The majority of biological data is imbalanced, as the number of negative samples is significantly greater than that of positive samples[47,48]. As an example, normal X-ray images are very large compared to COVID-19-positive images. To resolve this issue, we use the following techniques. As a first step, the correct criteria for evaluating the loss must be employed, along with the prediction outcome. Despite the imbalanced data, the model should perform well on small and large classes alike. Accordingly, the model should employ the area under the curve (AUC) as the resultant loss as well as the criteria[46]. The second step is to employ the weighted cross-entropy loss, as it will ensure the model will perform well with small classes if the cross-entropy loss is still preferred. The K-Fold Cross Validation method is also used in some models. During model training, both down-sampling and up-sampling can be done simultaneously.

3.3 Interpretability of data

Occasionally, DL techniques are analyzed to act as a black box. In many fields, such as bioinformatics [49], it is important to interpret DL, which is used to identify valuable motifs and patterns. For disease diagnosis, it is important not only to know the disease diagnosis or prediction results of a trained DL model but also how to enhance the certainty of the predictions, as the model makes decisions based on these verifications [50]. This can be accomplished by giving each portion of the example a score of importance. This solution employs back-propagation techniques or perturbation-based approaches [51]. The perturbation-based approaches modify a portion of the input and observe the impact on the model output [52–53]. Despite its high computational complexity, this concept is easy to understand. However, back-propagation-based techniques propagate signals from the output back to the input layer to check their importance. In [54], these techniques have been proven useful. Model interpretability can take on various meanings depending on the scenario.

3.4 Uncertainty scaling

Overfitting
It is common for the final prediction label to not be the only label required when employing DL techniques to achieve the prediction; the score of confidence for each question is also desired. The confidence score is a measure of how confident the model is in its predictions [55]. No matter the application scenario, the score of confidence is a crucial attribute because it prevents believing in unreliable and misleading predictions. The confidence score in biology reduces the resources and time spent proving the incorrect predictions. The uncertainty scaling is commonly used in healthcare or similar applications; it is used to evaluate automated clinical decisions and the reliability of machine learning-based disease diagnosis [56, 57]. The score of probability (achieved from the softmax output of the direct-DL) often does not reflect the true scale of probability since overconfident prediction can come from different DL models [58]. The probabilities of the softmax output require post-scaling to be reliable. Several techniques have been introduced for determining the probability score in the correct scale, including Bayesian Binning into Quantiles (BBQ) [59], isotonic regression [60], histogram binning [61], and the legendary Platt scaling [62]. DL techniques have recently been enhanced with temperature scaling that is more efficient than the other techniques.

3.5 Overfitting

As a result of the large number of parameters involved, which are correlated in a complex way, DL models have a high potential for overfitting at the training stage. This reduces the model's ability to achieve good performance on the tested data [63, 64]. Problems like this are not limited to one specific field, but also involve multiple tasks. It is important to fully consider and accurately handle this problem when developing DL techniques. DL overcomes critical overfitting problems due to the implied bias of the training process, as recent studies suggest [64-66]. Nonetheless, techniques to handle overfitting must be developed. The available DL algorithms that ease the overfitting problem can be categorized into three categories based on their characteristics. First, the model architecture and model parameters must both be taken into account; in this case, weight decay [67], batch normalization [68], and dropout [69] are the most common approaches. As a general regularizer, weight decay [67] is widely used in almost all DL algorithms. The second class investigates inputs to models, such as data corruption and data augmentation [45]. Due to the lack of training data, the learned distribution does not reflect the real distribution, resulting in the overfitting problem. The training data is augmented by data augmentation. Marginalized data corruption, on the other hand, improves the solution solely by augmenting the data. The final class analyzes the output of the model. In a recent proposal [70], the overconfident outputs were penalized for regularizing the model. RNNs and CNNs have been regularized using this technique.

3.6 Underspecification

A group of scientists at Google identified a new challenge for 2020, called underspecification [71]. In real-world applications such as computer vision, medical imaging, natural language processing, and medical genomics, ML models, including deep learning models, often show poor performance. The weak performance is due to underspecification. Small modifications can force the model to achieve a completely different solution as well as lead to different predictions in deployment domains. Various techniques can be used to address underspecification issues. As one example, you can design ”stress tests” to see how well a model performs on real-world data and to identify any potential issues. However, if the process is not understood correctly the model can work inaccurately. According to the team, the challenge was to develop stress tests that are well-matched to the application requirements and provide a high level of coverage of possible failure modes. In underspecified applications, ML predictions may be less accurate and may require some reconsideration. As ML will serve a range of applications, including imaging and driverless cars, it is important to pay attention to this issue.

3.7 Anamoly Detection

While analyzing real-world data sets, one of the most common tasks is identifying instances that differ from all others. These instances are known as anomalies, and the goal of anomaly detection (also known as outlier detection) is to detect them[104]. Traditional, deep learning algorithms were developed to analyze and solve simpler problems. However, when applied to complex problems, the algorithms present different challenges. In the image (e.g. medical images) and sequence datasets, the performance of traditional algorithms in detecting outliers is suboptimal since they are unable to capture complex structures in the data. In the case of large volumes of data to be analyzed, let us say to gigabytes, then it becomes nearly impossible for traditional methods to scale and find outliers on such a large scale [94]. Recent progress in deep learning has shown great success in addressing these complexities in a wide range of applications, however popular deep learning techniques are inapplicable to anomaly detection due to some unique characteristics of anomalies, such as the rarity, heterogeneity, and boundless nature of anomalies[105].

4. APPLICATIONS OF DEEP LEARNING

4.1 Image Processing and Convolution Neural Network

Nowadays, deep learning is on the verge of changing our everyday life as we know it. With the emergence of its many cutting edge applications powered by big tech companies like Amazon with Alexa and its checkout-less supermarket Amazon Go. Google with Google home and its DeepMind team efforts to automate the healthcare system with AI applications. Other notorious modern applications of deep learning are autonomous vehicles with autopilot which are commercialized
Recently, CNN has taken some medical imaging classification and localization tasks to different levels from traditional diagnosis to automated diagnosis with tremendous performance. Bharati et al. [72] used a chest X-ray dataset for detecting lung diseases based on a CNN. In 2020, CNNs are playing a vital role in the early diagnosis of the novel coronavirus (COVID-2019). CNN has become the primary tool for automatic COVID-19 diagnosis in many hospitals around the world using chest X-ray images[73-75]. Shin et al. [76] employed stacked auto-encoders on 78 contrast-improved MRI scans of the stomach area containing the kidneys or liver. Temporal and spatial domains were used to learn the hierarchical features. Based on the organs, these approaches achieved detection accuracies of 62–79 percent. Sirazitdinov et al. [77] presented an aggregate of two convolutional neural networks, namely RetinaNet and Mask R-CNN for pneumonia detection and localization. .Success of image processing and its expansion to numerous fields of applications like medical,engineering and remote sensing has also paved its way to application in agriculture. Agriculture sector has initiated to work in combination with the innovative information technology using techniques like image processing, computer vision and machine vision which could mimic the decisions of subject matter specialist. These techniques produce an output for the input image acquired through anyof the imaging techniques. The results are informative and supportive for farmers, agro based industries and marketers.[153]

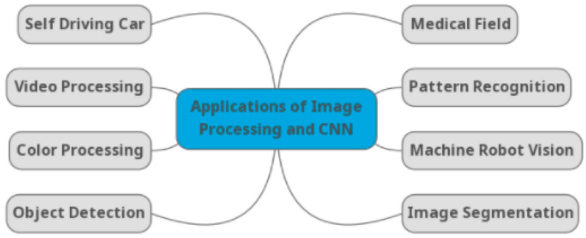

Fig. 10. Applications in image processing

4.2 Natural Language Processing

Natural Language Processing (NLP) is a tract of Artificial Intelligence and Linguistics, devoted to making computers understand the statements or words written in human languages. A few of the researched tasks of NLP are Automatic Summarization, Coreference Resolution, Discourse Analysis, Machine Translation, Morphological Segmentation, Named Entity Recognition, Optical Character Recognition, Part Of Speech Tagging, etc. [136]Due to its wide range of applications, Natural Language Processing is an emerging field of technology. It can be used to read radiology reports, extract text from images, and many more. According to [83], publication types can be categorized into 6 categories based on NLP and radiology information. This paper shows amazing results when it comes to clinical data extraction using NLP. BERT[84] is a model of machine learning for language translation developed by Google. By adding the surrounding text to establish context, it helps computers understand the meaning of the ambiguous text. It has been demonstrated that Financial Technology (FinTech) has been a rapidly-growing topic in the past five years, according to the statistics offered by Google Trends. [100] It suggests that NLP can be applied in the financial domain, for instance, to extract data from financial reports and financial announcements in order to determine a customer’s reliability. In addition to text extraction and text summarization, NLP can also provide a question and answer feature. To answer a question, first, it must be analyzed. Questions can be classified by their type: factual, opinion, or summary. In the case of a question that arises during ongoing interaction, the solution should be obtained from the online resource. Normally, the answer is presented in some kind of form [101]. In early 1980s computational grammar theory became a very active area of research linked with logics for meaning and knowledge’s ability to deal with the user’s beliefs and intentions and with functions like emphasis and themes.[136]By the end of the decade the powerful general purpose sentence processors like SRI's Core Language Engine (Alshawi,1992) [137] and Discourse Representation Theory (Kamp and Reyle,1993) [136] offered a means of tackling more extended discourse within the grammatico-logical framework.g . In recent years, various methods have been proposed to automatically evaluate machine translation quality by comparing hypothesis translations with reference translations.[136] Examples of such methods are word error rate, position-independent word error rate (Tillmann et al., 1997) [139], generation string accuracy (Bangalore et al., 2000) [140], multi-reference word error rate (Nießen et al., 2000) [141], BLEU score (Papineni et al., 2002) [142], NIST score (Doddington, 2002) [143]

Fig. 11. Applications in NLP

4.3 Medicine

With the advent of Deep Learning, clinical applications have been widespread. [91] Deep learning has proven useful for extracting information from clinical reports to solve complex drug classification and protein folding problems. [92] has demonstrated that deep learning can extract information from images and other medical reports to diagnose diseases. For example, image processing was first applied to clinical data using deep learning to predict Alzheimer disease and its variations [149,150]. CNNs have also been used in medical domains to detect cartilage clusters and predict osteoarthritis risk in low-field knee MRI scans [151]. In light of the fact that a great deal of clinically relevant information is embedded in unstructured data, natural language processing plays a vital role in extracting valuable information that can be used for decision-making, administration reporting, and research[95]. The Linguistic String Project - Medical Language Processor (LSP-MLP) of diseases. New York University is the first.The longest-running large-scale project on NLP in medicine. The LSP-MLP aims at enabling clinicians to obtain and summarize information regarding signs and symptoms, drug dosage, and response, to identify potential side effects of medications, and to highlight or flag data items [96]. [97] suggest an automatic segmentation method based on Convolutional Neural Networks (CNN), exploring small 3 * 3 kernels, which might be used to analyze the large spatial and structural variability among brain tumors. In x-rays and other medical reports, the object detection algorithm in [98] allows for finding different lesion areas.The National Library of Medicine is developing The Specialist System [144][145][146][147][148]. It is expected to function as Information Extraction tool for Biomedical Knowledge Bases, particularly Medline abstracts

Fig. 12. Applications in medicine

5. CONCLUSION

The purpose of this paper is to provide an introduction to deep learning models, their working, and their applications. Those models discussed above fall into the categories of supervised learning and unsupervised learning. A variety of architectures were discussed including ANNs, CNNs, RNNs, RvNNs, GANs, and Autoencoders. The following models are considered the core architectures of deep learning today. CNN focuses on unsupervised learning and solving classification problems, whereas ANNs are simple neural networks that solve regression problems. The RNN and RvNN are slightly complex structures of neural networks, involving loops to fine-tune their outputs. Autoencoders use two neural networks: an encoder and a decoder. An autoencoder produces a minimalistic representation of the input. Variational autoencoders and GANs are generative models.

Since deep learning's advent, it has evolved and improved tremendously, but there are still many limitations that remain. A lack of training data causes many deep learning models to fail to yield better accuracy. A lot of datasets are biased and imbalanced, even with enough data. Multiple anomalies are often present in them. There has been some progress in resolving these limitations, but more work needs to be done.

Unlike conventional machine-learning and data mining techniques, deep learning can generate very high-level data representations from massive volumes of raw data. Therefore, it has provided a solution to many real-world applications. Popular frameworks in this area include TensorFlow, Caffe, and Theano. In this article, we have discussed the challenges and provided several existing solutions to these challenges.

Although deep learning can memorize a massive amount of data and information, its weak reasoning and understanding of the data make it a block-box solution for many applications. The interpretability of deep learning should be investigated in the future. Deep learning still has difficulty in modeling multiple complex data modalities at the same time. Multimodal deep learning is another popular direction in recent deep learning research. DL needs extensive datasets for training the machine and predicting the unseen data. One-shot learning and zero-shot learning have been studied in recent years to alleviate this problem. Despite all the deep learning advancements in recent years, many applications are still untouched by deep learning or are in the early stages of leveraging the deep learning techniques(e.g. Disaster information management, finance, or medical data analytics).

6. REFERENCES

(1) LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. 2015;521(7553):436–44

(2) Zhang Z, Cui P, Zhu W. Deep learning on graphs: a survey. IEEE Trans Knowl Data Eng. 2020. https://doi.org/10.1109/TKDE.2020.2981333.

(3) Shrestha A, Mahmood A. Review of deep learning algorithms and architectures. IEEE Access. 2019;7:53040–65

(4) Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. J Big Data. 2015;2(1):1.

(5) Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning, vol. 1. Cambridge: MIT press; 2016

(6) Hasan RI, Yusuf SM, Alzubaidi L. Review of the state of the art of deep learning for plant diseases: a broad analysis and discussion. Plants. 2020;9(10):1302.

(7) Shorten C, Khoshgoftaar TM, Furht B. Deep learning applications for COVID-19. J Big Data. 2021;8(1):1–54

(8) Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90.

(9) Goh GB, Hodas NO, Vishnu A. Deep learning for computational chemistry. J Comput Chem. 2017;38(16):1291–307.

(10) Li Y, Zhang T, Sun S, Gao X. Accelerating flash calculation through deep learning methods. J Comput Phys. 2019;394:153–65.

(11) Yang W, Zhang X, Tian Y, Wang W, Xue JH, Liao Q. Deep learning for single image super-resolution: a brief review. IEEE Trans Multimed. 2019;21(12):3106–21

(12) Tang J, Li S, Liu P. A review of lane detection methods based on deep learning. Pattern Recogn. 2020;111:107623.

(13) Zhao ZQ, Zheng P, Xu ST, Wu X. Object detection with deep learning: a review. IEEE Trans Neural Netw Learn Syst. 2019;30(11):3212–32

(14) Krizhevsky A, Sutskever I, Hinton GE. Imagenet classifcation with deep convolutional neural networks. Commun ACM. 2017;60(6):84–90.

(15) He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.

(16) Metz C. Turing award won by 3 pioneers in artifcial intelligence. The New York Times. 2019;27

(17) Nevo S, Anisimov V, Elidan G, El-Yaniv R, Giencke P, Gigi Y, Hassidim A, Moshe Z, Schlesinger M, Shalev G, et al. Ml for food forecasting at scale; 2019. arXiv preprint arXiv:1901.09583.

(18) Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241–50.

(19) Benhammou Y, Achchab B, Herrera F, Tabik S. Breakhis based breast cancer automatic diagnosis using deep learning: taxonomy, survey and insights. Neurocomputing. 2020;375:9–24.

(20) Wulczyn E, Steiner DF, Xu Z, Sadhwani A, Wang H, Flament-Auvigne I, Mermel CH, Chen PHC, Liu Y, Stumpe MC. Deep learning-based survival prediction for multiple cancer types using histopathology images. PLoS ONE. 2020;15(6):e0233678.

(21) Brunese L, Mercaldo F, Reginelli A, Santone A. Explainable deep learning for pulmonary disease and coronavirus COVID-19 detection from X-rays. Comput Methods Programs Biomed. 2020;196(105):608.

(22) Jamshidi M, Lalbakhsh A, Talla J, Peroutka Z, Hadjilooei F, Lalbakhsh P, Jamshidi M, La Spada L, Mirmozafari M, Dehghani M, et al. Artifcial intelligence and COVID-19: deep learning approaches for diagnosis and treatment. IEEE Access. 2020;8:109581–95.

(23) Shorfuzzaman M, Hossain MS. Metacovid: a siamese neural network framework with contrastive loss for n-shot diagnosis of COVID-19 patients. Pattern Recogn. 2020;113:107700.

(24) He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 770–8.

(25) Saeed MM, Al Aghbari Z, Alsharidah M. Big data clustering techniques based on spark: a literature review. PeerJ Comput Sci. 2020;6:321.

(26) Mnih V, Kavukcuoglu K, Silver D, Rusu AA, Veness J, Bellemare MG, Graves A, Riedmiller M, Fidjeland AK, Ostrovski G, et al. Human-level control through deep reinforcement learning. Nature. 2015;518(7540):529–33.

(27) Socher R, Perelygin A, Wu J, Chuang J, Manning CD, Ng AY, Potts C. Recursive deep models for semantic compo- sitionality over a sentiment treebank. In: Proceedings of the 2013 conference on empirical methods in natural language processing; 2013. p. 1631–42.

(28) Goller C, Kuchler A. Learning task-dependent distributed representations by backpropagation through structure. In: Proceedings of international conference on neural networks (ICNN'96), vol 1. IEEE; 1996. p. 347–52.

(29) Socher R, Lin CCY, Ng AY, Manning CD. Parsing natural scenes and natural language with recursive neural net- works. In: ICML; 2011.

(30) Louppe G, Cho K, Becot C, Cranmer K. QCD-aware recursive neural networks for jet physics. J High Energy Phys. 2019;2019(1):57.

(31) Sadr H, Pedram MM, Teshnehlab M. A robust sentiment analysis method based on sequential combination of convolutional and recursive neural networks. Neural Process Lett. 2019;50(3):2745–61.

(32) Urban G, Subrahmanya N, Baldi P. Inner and outer recursive neural networks for chemoinformatics applications. J Chem Inf Model. 2018;58(2):207–11.

(33) Batur Dinler Ö, Aydin N. An optimal feature parameter set based on gated recurrent unit recurrent neural net- works for speech segment detection. Appl Sci. 2020;10(4):1273.

(34) Jagannatha AN, Yu H. Structured prediction models for RNN based sequence labeling in clinical text. In: Proceed- ings of the conference on empirical methods in natural language processing. conference on empirical methods in natural language processing, vol. 2016, NIH Public Access; 2016. p. 856.

(35) Pascanu R, Gulcehre C, Cho K, Bengio Y. How to construct deep recurrent neural networks. In: Proceedings of the second international conference on learning representations (ICLR 2014); 2014.

(36) Glorot X, Bengio Y. Understanding the difculty of training deep feedforward neural networks. In: Proceedings of the thirteenth international conference on artifcial intelligence and statistics; 2010. p. 249–56.

(37) Gao C, Yan J, Zhou S, Varshney PK, Liu H. Long short-term memory-based deep recurrent neural networks for target tracking. Inf Sci. 2019;502:279–96.

(38) Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J, et al. Recent advances in convolutional neural networks. Pattern Recogn. 2018;77:354–77.

(39) Goodfellow I, Bengio Y, Courville A, Bengio Y. Deep learning, vol. 1. Cambridge: MIT press; 2016

(40) O. M. Parkhi, A. Vedaldi, C. V. Jawahar and A. Zisserman, "The truth about cats and dogs," 2011 International Conference on Computer Vision, 2011, pp. 1427-1434, doi: 10.1109/ICCV.2011.6126398.

(41) HUBEL DH, WIESEL TN. Receptive fields of single neurones in the cat's striate cortex. J Physiol. 1959;148(3):574-591. doi:10.1113/jphysiol.1959.sp006308

(42) Alzubaidi L, Fadhel MA, Al-Shamma O, Zhang J, Santamaría J, Duan Y, Oleiwi SR. Towards a better understanding of transfer learning for medical imaging: a case study. Appl Sci. 2020;10(13):4523.

(43) Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. In: International conference on artifcial neural networks. Springer; 2018. p. 270–9.

(44) Weiss K, Khoshgoftaar TM, Wang D. A survey of transfer learning. J Big Data. 2016;3(1):9

(45) Shorten C, Khoshgoftaar TM. A survey on image data augmentation for deep learning. J Big Data. 2019;6(1):60

(46) Wang S, Sun S, Xu J. Auc-maximized deep convolutional neural felds for sequence labeling 2015. arXiv preprint arXiv:1511.05265.

(47) Johnson JM, Khoshgoftaar TM. Survey on deep learning with class imbalance. J Big Data. 2019;6(1):27

(48) Yang P, Zhang Z, Zhou BB, Zomaya AY. Sample subset optimization for classifying imbalanced biological data. In: Pacifc-Asia conference on knowledge discovery and data mining. Springer; 2011. p. 333–44

(49) Li Y, Huang C, Ding L, Li Z, Pan Y, Gao X. Deep learning in bioinformatics: introduction, application, and perspective in the big data era. Methods. 2019;166:4–21

(50) Choi E, Bahadori MT, Sun J, Kulas J, Schuetz A, Stewart W. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. In: Advances in neural information processing systems. San Mateo: Morgan Kaufmann Publishers; 2016. p. 3504–12

(51) Ching T, Himmelstein DS, Beaulieu-Jones BK, Kalinin AA, Do BT, Way GP, Ferrero E, Agapow PM, Zietz M, Hof- man MM, et al. Opportunities and obstacles for deep learning in biology and medicine. J R Soc Interface. 2018;15(141):20170,387

(52) Pokuri BSS, Ghosal S, Kokate A, Sarkar S, Ganapathysubramanian B. Interpretable deep learning for guided microstructure-property explorations in photovoltaics. NPJ Comput Mater. 2019;5(1):1–11.

(53) Ribeiro MT, Singh S, Guestrin C. "Why should I trust you?" explaining the predictions of any classifer. In: Proceed- ings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016. p. 1135–44.

(54) Sundararajan M, Taly A, Yan Q. Axiomatic attribution for deep networks; 2017. arXiv preprint arXiv:1703.01365

(55) Platt J, et al. Probabilistic outputs for support vector machines and comparisons to regularized likelihood meth- ods. Adv Large Margin Classif. 1999;10(3):61–74

(56) Nair T, Precup D, Arnold DL, Arbel T. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. Med Image Anal. 2020;59:101557

(57) Herzog L, Murina E, Dürr O, Wegener S, Sick B. Integrating uncertainty in deep neural networks for MRI based stroke analysis. Med Image Anal. 2020;65:101790

(58) Pereyra G, Tucker G, Chorowski J, Kaiser Ł, Hinton G. Regularizing neural networks by penalizing confdent output distributions; 2017. arXiv preprint arXiv:1701.06548

(59) Naeini MP, Cooper GF, Hauskrecht M. Obtaining well calibrated probabilities using bayesian binning. In: Proceed- ings of the... AAAI conference on artifcial intelligence. AAAI conference on artifcial intelligence, vol. 2015. NIH Public Access; 2015. p. 2901

(60) Li M, Sethi IK. Confdence-based classifer design. Pattern Recogn. 2006;39(7):1230–40.

(61) Zadrozny B, Elkan C. Obtaining calibrated probability estimates from decision trees and Naive Bayesian classifers. In: ICML, vol. 1, Citeseer; 2001. p. 609–16

(62) Steinwart I. Consistency of support vector machines and other regularized kernel classifers. IEEE Trans Inf Theory. 2005;51(1):128–42

(63) Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural net- works from overftting. J Mach Learn Res. 2014;15(1):1929–58

(64) Xu Q, Zhang M, Gu Z, Pan G. Overftting remedy by sparsifying regularization on fully-connected layers of CNNs. Neurocomputing. 2019;328:69–74

(65) Zhang C, Bengio S, Hardt M, Recht B, Vinyals O. Understanding deep learning requires rethinking generalization. Commun ACM. 2018;64(3):107–15

(66) Xu X, Jiang X, Ma C, Du P, Li X, Lv S, Yu L, Ni Q, Chen Y, Su J, et al. A deep learning system to screen novel coronavirus disease 2019 pneumonia. Engineering. 2020;6(10):1122–9.

(67) Zhang G, Wang C, Xu B, Grosse R. Three mechanisms of weight decay regularization; 2018. arXiv preprint arXiv: 1810.12281

(68) Laurent C, Pereyra G, Brakel P, Zhang Y, Bengio Y. Batch normalized recurrent neural networks. In: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP), IEEE; 2016. p. 2657–61

(69) Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural net- works from overftting. J Mach Learn Res. 2014;15(1):1929–58

(70) Pereyra G, Tucker G, Chorowski J, Kaiser Ł, Hinton G. Regularizing neural networks by penalizing confdent output distributions; 2017. arXiv preprint arXiv:1701.06548

(71) D'Amour A, Heller K, Moldovan D, Adlam B, Alipanahi B, Beutel A, Chen C, Deaton J, Eisenstein J, Hofman MD, et al. Underspecifcation presents challenges for credibility in modern machine learning; 2020. arXiv preprint arXiv: 2011.03395

(72) Bharati S, Podder P, Mondal MRH. Hybrid deep learning for detecting lung diseases from X-ray images. Inform Med Unlocked. 2020;20:100391

(73) Heidari M, Mirniaharikandehei S, Khuzani AZ, Danala G, Qiu Y, Zheng B. Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms. Int J Med Inform. 2020;144:104284

(74) Al-Timemy AH, Khushaba RN, Mosa ZM, Escudero J. An efcient mixture of deep and machine learning models for COVID-19 and tuberculosis detection using X-ray images in resource limited settings 2020. arXiv preprint arXiv: 2007.08223.

(75) Abraham B, Nair MS. Computer-aided detection of COVID-19 from X-ray images using multi-CNN and Bayesnet classifer. Biocybern Biomed Eng. 2020;40(4):1436–45.

(76) Shin HC, Orton MR, Collins DJ, Doran SJ, Leach MO. Stacked autoencoders for unsupervised feature learn- ing and multiple organ detection in a pilot study using 4D patient data. IEEE Trans Pattern Anal Mach Intell. 2012;35(8):1930–43.

(77) Sirazitdinov I, Kholiavchenko M, Mustafaev T, Yixuan Y, Kuleev R, Ibragimov B. Deep neural network ensemble for pneumonia localization from a large-scale chest X-ray database. Comput Electr Eng. 2019;78:388–99

(78) Hossin M, Sulaiman M. A review on evaluation metrics for data classifcation evaluations. Int J Data Min Knowl Manag Process. 2015;5(2):1

(79) Dr.R.Parthasarathi1: ANN Based Modeling for Performance and Exhaust Emission of DI Diesel Engine using Emulsified Diesel Fuel

(80) Sushma Tamta: Estimation of Evaporation in Hilly Area by Using Ann and Canfis System Based Models

(81) Mruga Gurjar: STOCK MARKET PREDICTION USING ANN

(82) Laura E. Suarez : Learning function from structure in neuromorphic networks

(83) Arlene Casey1*, Emma Davidson2 , Michael Poon2 , Hang Dong3,4, Daniel Duma1 , Andreas Grivas5 , Claire Grover5 , Víctor Suárez-Paniagua3,4, Richard Tobin5 , William Whiteley2,6, Honghan Wu4,7 and Beatrice Alex1: A systematic review of natural language processing applied to radiology reports

(84) Jacob Devlin Ming-Wei Chang Kenton Lee Kristina Toutanova: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding

(85) Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. Bengio,Y. (2014)Generative Adversarial Nets. NIPS 2014

(86) Arjovsky M, Chintala S, Bottou L (2017) Wasserstein gan. arXiv preprint

(87) Chen Y, Lai YK, Liu YJ (2018) Cartoongan: Generative adversarial networks for photo cartoonization. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 9465–9474

(88) Alec Radford Luke Metz, Soumit chintala: UNSUPERVISED REPRESENTATION LEARNING WITH DEEP CONVOLUTIONAL GENERATIVE ADVERSARIAL NETWORKS

(89) Mehdi Mirza,Simon Osindero: Conditional Generative Adversarial Nets

(90) Rumelhart, D.E., Hinton, G.E., Williams, R.J.: Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1. chap. Learning Internal Representations by Error Propagation, pp. 318–362. MIT Press, Cambridge, MA, USA (1986). URL http://dl.acm.org/citation.cfm?id=104279.104293

(91) Jinbo Xua: Distance-based protein folding powered by deep learning

(92) Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique

(93) Alan L. Yuille Chenxi Liu, "Limitations of Deep Learning for Vision, and How We Might Fix Them", The Gradient, 2019.

(94) Raghavendra Chalapathy, Sanjay Chawla: DEEP LEARNING FOR ANOMALY DETECTION: A SURVEY,2019.

(95) Wen, Andrew,Fu, Sunyang: Desiderata for delivering NLP to accelerate healthcare AI advancement and a Mayo Clinic NLP-as-a-service implementation.

(96) Spyns, Peter. (1997). Natural Language Processing in Medicine: An Overview. Methods of information in medicine. 35. 285-301. 10.1055/s-0038-1634681.

(97) Pereira S, Pinto A, Alves V, Silva CA. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. IEEE Trans Med Imaging. 2016 May;35(5):1240-1251. doi: 10.1109/TMI.2016.2538465. Epub 2016 Mar 4. PMID: 26960222.

(98) Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the 2014 IEEE conference on computer vision and pattern recognition; 2014 Jun 23-28; Washington, DC, USA. Piscataway (NJ): Institute of Electrical and Electronics Engineers; 2014. pp. 580–7.

(99) (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 5, No. 12, 2014

(100) Chen, Chung-Chi Huang, Hen-Hsen Chen, Hsin-Hsi. (2020). NLP in FinTech Applications: Past, Present and Future.

(101) Hirschman, L., Gaizauskas, R. (2001). Natural language question answering:the view from here. Natural Language Engineering, 7, 275–300.

(102) Michael Tschannen, Olivier Bachem, Mario Lucic: "Recent Advances in Autoencoder-Based Representation Learning",2018

(103) Autoencoders, Lopez Pinaya W,Viera S et.al 2019.

(104) Chandola et al. [2007]

(105) Pang, Guansong Cao, Longbing Aggarwal, Charu. (2021). Deep Learning for Anomaly Detection: Challenges, Methods, and Opportunities. 1127-1130. 10.1145/3437963.3441659.

(106) Sublime, Jeremie Kalinicheva, Ekaterina. (2019). Automatic Post-Disaster Damage Mapping Using Deep-Learning Techniques for Change Detection: Case Study of the Tohoku Tsunami. Remote Sensing. 11. 1123. 10.3390/rs11091123.

(107) arXiv:1312.5663 [cs.LG]

(108) Dong, Ganggang Liao, Guisheng Liu, Hongwei Kuang, Gangyao. (2018). A Review of the Autoencoder and Its Variants: A Comparative Perspective from Target Recognition in Synthetic-Aperture Radar Images. IEEE Geoscience and Remote Sensing Magazine. 6. 44-68. 10.1109/MGRS.2018.2853555.

(109) Changfan Zhang, Xiang Cheng, Jianhua Liu, Jing He, Guangwei Liu, ”Deep Sparse Autoencoder for Feature Extraction and Diagnosis of Locomotive Adhesion Status”, Journal of Control Science and Engineering, vol. 2018, Article ID 8676387, 9 pages, 2018. https://doi.org/10.1155/2018/8676387

(110) Huajie Shao , Zhisheng Xiao :ControlVAE: Tuning, Analytical Properties, and Performance Analysis

(111) X. Yan, J. Yang, K. Sohn, and H. Lee, “Attribute2image: Conditional image generation from visual attributes,” in European Conference on Computer Vision. Springer, 2016, pp. 776–791

(112) M.-Y. Liu, T. Breuel, and J. Kautz, “Unsupervised image-to-image translation networks,” in Advances in neural information processing systems, 2017, pp. 700–708.

(113) A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, Dive into Deep Learning, 2020, https://d2l.ai.

(114) W. Wang, Z. Gan, H. Xu, R. Zhang, G. Wang, D. Shen, C. Chen, and L. Carin, “Topic-guided variational autoencoders for text generation,” arXiv preprint arXiv:1903.07137, 2019.

(115) Z. Hu, Z. Yang, X. Liang, R. Salakhutdinov, and E. P. Xing, “Toward controlled generation of text,” in Proceedings of the 34th International Conference on Machine Learning-Volume 70. JMLR. org, 2017, pp. 1587–1596

(116) I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, “beta-vae: Learning basic visual concepts with a constrained variational framework.” ICLR, vol. 2, no. 5, p. 6, 2017.

(117) H. Kim and A. Mnih, “Disentangling by factorising,” in International Conference on Machine Learning, 2018, pp. 2654–2663

(118) Feng, Jie Feng, Xueliang Chen, Jiantong Cao, Xianghai Zhang, Xiangrong Jiao, Licheng Yu, Tao. (2020). Generative Adversarial Networks Based on Collaborative Learning and Attention Mechanism for Hyperspectral Image Classification. Remote Sensing. 12. 1149. 10.3390/rs12071149.

(119) Feng, Jie Feng, Xueliang Chen, Jiantong Cao, Xianghai Zhang, Xiangrong Jiao, Licheng Yu, Tao. (2020). Generative Adversarial Networks Based on Collaborative Learning and Attention Mechanism for Hyperspectral Image Classification. Remote Sensing. 12. 1149. 10.3390/rs12071149.

(120) Phung, Rhee,. (2019). A High-Accuracy Model Average Ensemble of Convolutional Neural Networks for Classification of Cloud Image Patches on Small Datasets. Applied Sciences. 9. 4500. 10.3390/app9214500.

(121) Mishra, Vidushi AGARWAL, SMT PURI, NEHA. (2018). COMPREHENSIVE AND COMPARATIVE ANALYSIS OF NEURAL NETWORK. INTERNATIONAL JOURNAL OF COMPUTER APPLICATION. 2. 10.26808/rs.ca.i8v2.15.

(122) M. S. B. M. M. P. W. “Research Paper on Basic of Artificial Neural Network”. International Journal on Recent and Innovation Trends in Computing and Communication, vol. 2, no. 1, Jan. 2014, pp. 96-100, doi:10.17762/ijritcc.v2i1.2920.

(123) Grossi, Enzo Buscema, Massimo. (2008). Introduction to artificial neural networks. European journal of gastroenterology hepatology. 19. 1046-54. 10.1097/MEG.0b013e3282f198a0.

(124) Christos Stergiou and Dimitrios Siganos,“Neural Networks”

(125) Gopal, Sucharita. ”Unit 188-Artificial Neural Networks for Spatial Data Analysis.” (2000).

(126) Wang, Su. (2017). Generative Adversarial Networks (GAN): A Gentle Introduction [UPDATED].

(127) Lotter, W., Kreiman, G. Cox, D. (2015) Unsupervised Learning of Visual Structure Using Predictive Generative Networks CoRR, arXiv:1151.06380.

(128) Ledig, C., Theis, L., Huszar, F., Caballero, J., Aitken, A. P., Tejani, A., Totz, J., Wang, Z., Shi, W. (2016) Photo-realistic single Image Super-resolution Using a Generative Adversarial Network CoRR,arXiv:1609.04802.

(129) Isola, P., Zhu, J.-Y., Zhou, T., Efros, A. A. (2016). Image-to-image

(130) Translation with ConditionalAdversarial Networks. CoRR, arXiv:1611.07004.

(131) Donahue, J., Krähenbühl, P. Darrell, T. (2017) Adversarial Feature Learning CoRR, arXiv:1605.09782.

(132) Glover, J. (2016)Modeling Documents with Generative Adversarial Networks. CoRR, arXiv:1612.09122.

(133) Li, J., Monroe, W., Shi, T., Jean, S., Ritter, A. Jurafsky, D. (2017)Adversarial Learning for Neural Dialogue Generation CoRR, arXiv:1701.06547.

(134) Chen X., Athiwaratkun, B., Sun, Y., Weinberger, K. Cardie, C. (2016) Adversarial Deep Averaging Networks for Cross-lingual Sentiment Classification,. CoRR, arXiv:1606.01614.

(135) Zhang, Y., Barzilay, R. Jaakkola, T. (2017)Aspect-augmented Adversarial Networks for Domain Adaptation. CoRR, arXiv:1701.00188

(136) arXiv:1312.6114v10 [stat.ML]

(137) Khurana, Diksha, et al. ”Natural language processing: State of the art, current trends and challenges.” arXiv preprint arXiv:1708.05148 (2017).

(138) Alshawi, H. (1992). The core language engine. MIT press. Kamp, H., Reyle, U. (1993). Tense and Aspect. In From Discourse to Logic (pp. 483- 689). Springer Netherlands.

(139) Tillmann, C., Vogel, S., Ney, H., Zubiaga, A., Sawaf, H. (1997, September). Accelerated DP based search for statistical translation. In Eurospeech.

(140) Bangalore, S., Rambow, O., Whittaker, S. (2000, June). Evaluation metrics for generation. In Proceedings of the first international conference on Natural language generation-Volume 14 (pp. 1-8). Association for Computational Linguistics

(141) Nießen, S., Och, F. J., Leusch, G., Ney, H. (2000, May). An Evaluation Tool for Machine Translation: Fast Evaluation for MT Research. In LREC

(142) Papineni, K., Roukos, S., Ward, T., Zhu, W. J. (2002, July). BLEU: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting on association for computational linguistics (pp. 311-318). Association for Computational Linguistics

(143) Doddington, G. (2002, March). Automatic evaluation of machine translation quality using n-gram co-occurrence statistics. In Proceedings of the second international conference

(144) McCray, A. T., Nelson, S. J. (1995). The representation of meaning in the UMLS. Methods of information in medicine, 34(1-2), 193-201.

(145) McGray, A. T., Sponsler, J. L., Brylawski, B., Browne, A. C. (1987, November). The role of lexical knowledge in biomedical text understanding. In Proceedings of the Annual Symposium on Computer Application in Medical Care (p. 103). American Medical Informatics Association.

(146) McCray, A. T. (1991). Natural language processing for intelligent information retrieval. In Engineering in Medicine and Biology Society, 1991. Vol. 13: 1991., Proceedings of the Annual International Conference of the IEEE (pp. 1160-1161). IEEE.

(147) McCray, A. T. (1991). Extending a natural language parser with UMLS knowledge. In Proceedings of the Annual Symposium on Computer Application in Medical Care (p. 194). American Medical Informatics Association.

(148) McCray, A. T., Srinivasan, S., Browne, A. C. (1994). Lexical methods for managing variation in biomedical terminologies. In Proceedings of the Annual Symposium on Computer Application in Medical Care (p. 235). American Medical Informatics Association.

(149) Liu S, Liu S, Cai W, et al. Early diagnosis of Alzheimer's disease with deep learning. In: International Symposium on Biomedical Imaging, Beijing, China 2014, 1015–18.

(150) Brosch T, Tam R. Manifold learning of brain MRIs by deep learning. Med Image Comput Comput Assist Interv 2013;16:633–40.

(151) Prasoon A, Petersen K, Igel C, et al. Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. Med Image Comput Comput Assist Interv 2013;16:246–53

(152) Deenan, Surya Prabha SatheeshKumar, J.. (2014). Image processing methods and its Role in agricultural sector – A study. International Journal of Business Intelligents. 3. 366-373.

(153) Hinton, Geoffrey E., and Ruslan R. Salakhutdinov. "Reducing the dimensionality of data with neural networks." science 313.5786 (2006): 504-507.

(154) Ashish Vaswani, Noam Shazeer: "Attention i sall you need", arXiv:1706.03762v5

(155) Maxime: "What is a Transformer?", https://medium.com/inside-machine-learning/what-is-a-transformer-d07dd1fbec04

(156) Vaswani, Ashish Bengio, Samy Brevdo, Eugene Chollet, Francois Gomez, Aidan Gouws, Stephan Jones, Llion Kaiser, Łukasz Kalchbrenner, Nal Parmar, Niki Sepassi, Ryan Shazeer, Noam Uszkoreit, Jakob. (2018). Tensor2Tensor for Neural Machine Translation.

(157) Wolf, Thomas, et al. "Transformers: State-of-the-art natural language processing." Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations. 2020.

(158) Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. CoRR, abs/1409.0473, 2014

(159) Kyunghyun Cho, Bart van Merrienboer, Caglar Gulcehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using rnn encoder-decoder for statistical machine translation. CoRR, abs/1406.1078, 2014.

(160) Alex Graves. Generating sequences with recurrent neural networks. arXiv preprint arXiv:1308.0850, 2013.

(161) Ilya Sutskever, Oriol Vinyals, and Quoc VV Le. Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems, pages 3104–3112, 2014.

(162) Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 770–778, 2016.

(163) Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. arXiv preprint arXiv:1607.06450, 2016

(164) Stefania Cristina : "The Transformer Model", https://machinelearningmastery.com/the-transformer-model/

(165) Nurtas, Marat Baishemirov, Zharasbek Tsay, V. Tastanov, M. Zhanabekov, Zh. (2020). APPLYING NEURAL NETWORK FOR PREDICTING CARDIOVASCULAR DISEASE RISK. PHYSICO-MATHEMATICAL SERIES. 4. 28-34. 10.32014/2020.2518-1726.62.

(166) Ensar Seker: "Recursive Neural Networks(RvNN) and Recurrent Neural Network(RNN)" "https://ai.plainenglish.io/recursive-neural-networks-rvnns-and-recurrent-neural-networks-rnns-2ff6a067ad01

(167) Andrew Ng: "Sparse Autoencoder", https://web.stanford.edu/class/cs294a/sparseAutoencoder.pdf

(168) rhum Shafkat:"Intuitively Understanding Variational Autoencoders",https://towardsdatascience.com/intuitively-understanding-variational-autoencoders-1bfe67eb5daf