Analysis of Conventional Feature Learning Algorithms and Advanced Deep Learning Models

Toshihiro Endo
Faculty of Mechanical Engineering, Tokyo Institute of Technology, Meguro City, Tokyo, Japan.
endohiro@hotmail.com

Correspondence should be addressed to Toshihiro Endo : endohiro@hotmail.com

Article Info
Journal of Robotics Spectrum (https://anapub.co.ke/journals/jrs/jrs.html)
Doi: https://doi.org/10.53759/9852/JRS202301001
Received 02 October 2022; Revised from 10 December 2022; Accepted 26 December 2022.
Available online 05 January 2023.

Abstract – Representation learning, or feature learning refers to a collection of methods employed in machine learning, which allows systems to autonomously determine representations needed for classifications or feature detection from unprocessed data. Representation learning algorithms are specifically crafted to acquire knowledge of conceptual features that define data. The field of state representation learning is centered on a specific type of representation learning that involves the acquisition of low-dimensional learned features that undergo temporal evolution and are subject to the influence of an agent's actions. Over the past few years, deep architecture has been widely employed for representation learning and has demonstrated exceptional performance in various tasks, including but not limited to object detection, speech recognition, and image classification. This article provides a comprehensive overview of the evolution of techniques for data representation learning and the research focuses on the examination of conventional feature learning algorithms and advanced deep learning models. Also presents an introduction to data representation learning history, along with a comprehensive list of available resources such as online courses, tutorials, and books. Additionally, various toolboxes are also provided for further exploration in this field. In conclusion, this article presents remarks and future prospects for data representation learning.

Keywords – Feature Learning, Feature Detection, Representation Learning, Deep Learning Models, Data Architectures, Deep Learning.

I. INTRODUCTION

Data representation learning is an important first step in many fields, including AI, biology, and finance, since it improves the efficiency of later classification, retrieval, and recommendation tasks. It is becoming more crucial and difficult to grasp the fundamental structure of data and uncover useful information from data for large-scale applications. Many different approaches to learning from data representations have been proposed over the past century and a half. K. Pearson created principal component analysis (PCA) in 1901, while Chang [1] discussed the linear discriminant analysis (LDA) that was projected in 1936. Both LDA and PCA are linear procedures. In contrast to LDA, which is a supervised algorithm, PCA does not need human oversight. There have been many other suggested expansions to PCA and LDA, such as generalized discriminant analysis (GDA), and PCA. Study on manifold learning that aims to uncover the underlying system of high-dimensional dataset, was established in the ML society in the year 2000. Manifold learning methods, such as isometric feature mapping (Isomap) and locally linear embedding (LLE), are often locality oriented, as opposed to earlier global techniques like LDA, and PCA.

To successfully utilize deep neural networks to dimensional reduction, Shrivastava et al. [2] discussed the notion of "deep learning" established in 2006. Due to their efficacy, the algorithms of deep learning are now being used in various contexts apart from AI. However, the study of artificial neural networks is a laborious process that has both fruitful and frustrating outcomes. According to Dalvi, Durrani, Sajjad, Belinkov, Bau, and Glass [3], W. Pitts and W. McCulloch launched their first artificial neurons, which defined a linear threshold unit, for neural networks in 1943; this model is now often referred to as the M-P model. In the future, Treur [4] discussed a theory of learning called Hebbian theory, which is predicated on the concept of brain plasticity. The Hebbian theory and M-P model laid the groundwork for the study of neural networks and the emergence of connectionism in the domain of artificial intelligence. According to the author, the perceptron, a 2-layer binary classification neural network, was developed by F. Rosenblatt in 1958.

However, as Dung and Mizukaw [5] pointed out, perceptrons have trouble with even the exclusive-or (XOR) problem. Before Ren, Wang, and Burkholder [6] projected the back propagation approach in the training of multi-layer perceptrons (MLP) in 1974, progress in the field of neural networks had stalled. In particular, Guo, Qiu, and He [7] demonstrated that

the algorithm of back propagation may provide valuable internal data representation within the neural network's hidden layer. Even though it was theoretically possible to train many neural networks layers using the algorithm of back propagation, there were two major problems: gradient diffusion and model overfitting. Breakthrough progress in representation-learning research began in 2006, when the deep neural networks fine-tuning and greedy layer-wise pre-training was proposed. Concerns voiced by the neural network community have been addressed. In subsequent times, several algorithms of deep learning were proposed and effectively implemented across a wide range of fields. In this study, we examine how the two main types of representation learning—contemporary deep learning and traditional feature learning —have evolved over time.

The remainder of this paper is structured as follows: Section II presents a discussion of conventional feature learning. Section III focused on advanced deep learning. In this section, two concepts are critically discussed: deep learning models and deep learning toolboxes. Lastly, Section IV presented final remarks regarding the article including directions for future research.

II. CONVENTIONAL FEATURE LEARNING

This section is dedicated to the discussion of conventional feature learning algorithms, which are categorized as "shallow" models. The primary objective of these algorithms is to acquire knowledge of data transformations that facilitate the extraction of valuable information while constructing classifiers or other predictors. **Fig 1** depicts the comprehensive arrangement of the classifications of network representation learning algorithms. Therefore, certain manual feature engineering approaches, which include image descriptors (e.g. LBP, SIFT, HOG, etc.) and document statistics (e.g. TF-IDF, etc.), will not be taken into consideration. Algorithmic formulations are typically classified as linear or nonlinear, generative or discriminative, supervised or unsupervised, and global or local. An illustration of the contrast between PCA and LDA can be made based on their respective characteristics. PCA represents a global feature, generative, unsupervised and lineal approach, while LDA represents a global, discriminative, supervised, and linear approach. This section makes use of a classification model to class the algorithms of feature learning as either local or global.

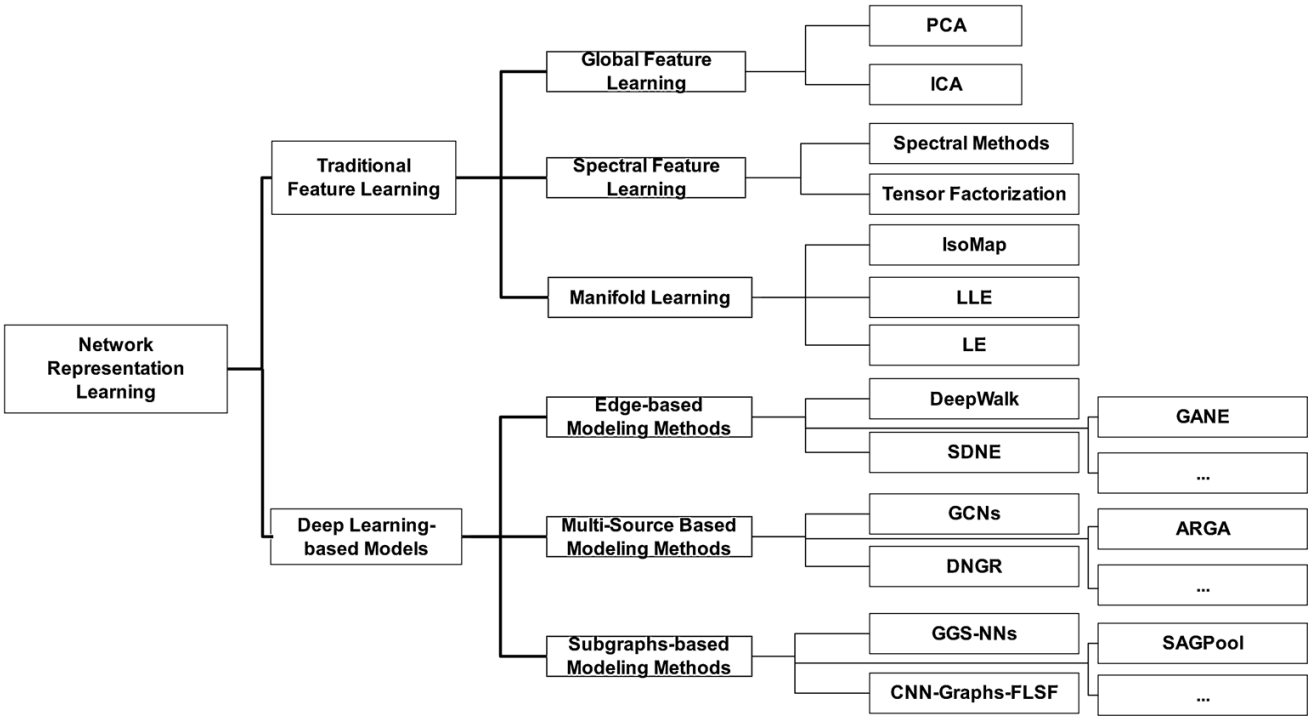


Fig 1. The Categorization of Algorithms Utilized for Network Representation Learning

During the process of learning new representations, global methods strive to maintain the globalized data in the space of learned features, whereas local approaches place more emphasis on maintaining local similarities between distinct points of data. In contrast to LDA and PCA, the locally linear embedding (LLE) algorithm is a technique for learning features based on locality. In addition, the process of uncovering the underlying manifold system available in highly dimensional data is commonly referred to as local-based manifold learning or feature learning. Gopi [8] have presented a toolbox for MATLAB used in dimensional reduction in their literature. This toolbox comprises 34 feature learning algorithms and their corresponding codes. Gou et al. [9] have recommended a comprehensive model, which is referred to as graph embedding, which aims to consolidate a diverse range of dimensional reduction techniques into a single formulation. The study conducted by Sarhadi, Burn, Yang, and Ghodsi [10] involved a comparison of three distinct types of supervised dimensionality reduction techniques for the purpose of improving handwriting recognition. Yang and

Hospedales [11] also introduced a novel framework that adopts a learning approach known as tensor representation, which handles input data as a tensor and integrates various kernels, linear, and tensor-centric dimensional reduction approaches under a criterion known as single learning.

Global feature learning

As previously stated, Principal Component Analysis (PCA) is among the initial linear feature learning algorithms. The primary objective of the PCA methodology is to identify a PCA space (W) with a reduced number of dimensions, which can be employed to convert the dataset ($X = \{x1, x2, \ldots, xN\}$) from RM (spaces that are high dimensional) to Rk (spaces of low dimension). Here, N denotes the overall number of observations and samples, which x_i alludes to i^{th} pattern, observation, and sample. All specimens possess identical dimensions, denoted as $xi \in$ RM, and stated differently, it can be observed that every sample is showed by M metrics, showing that every sample is shown as a point in spaces with M-dimension.

Principal Component Analysis (PCA) has been extensively employed for the purpose of reducing dimensionality owing to its straightforwardness. An orthogonal data is implemented to transform a category of observation of conceivably-linked metrics in the value series of linear-unlinked metrics. Classical multidimensional scaling (MDS) and PCA share some similarities. Specifically, both techniques are linear in nature and are optimized through eigenvalue decomposition. PCA and MDS exhibit a fundamental distinction in that PCA takes the data matrix as input, whereas MDS takes the distance matrix between data as input. Singular Value Decomposition (SVD) is frequently employed for optimization, in addition to Eigenvalue Decomposition. The optimization of Latent Semantic Analysis (LSA) in the context of data retrieval involves the utilization of SVD (Singular Value Decomposition) to reduce the row number while sustaining the similarity system among different columns. In this case, the rows correspond to words, while the columns correspond to documents. Kernel PCA and probabilistic PCA are two variations of PCA. Kernel PCA utilizes the kernel trick to enable nonlinear dimensionality reduction, whereas probabilistic PCA represents an iteration of PCA.

Furthermore, Yang, Heiselman, Quirk, and Djurić [12] introduced the GPLVM (Gaussian Process Latent Variable Model) as a completely non-linear, probabilistic latent variable structure based on PPCA. This model has the capability to acquire a non-lineal mapping from latent spaces to observation spaces. Song, Wang, Huang, and Tian [13] recommended a GPLVM that is discriminative as a means to incorporate supervisory data into a GPLVM model. The dimension of learned latent spaces within DGPLVM has been limited to a maximum of C-1, whereas C alludes to the class number. This is due to the fact that DGPLVM is founded on the criterion of learning of GDA or LDA. In order to tackle the aforementioned issue, Zhong, Li, Yeung, Hou, and Liu [14] have put forth the Gaussian process latent random field (GPLRF) approach. This method involves imposing a GMRF structure on the latent variables, with the basis of graphs, which are established on supervisory data. In addition to other approaches, several extensions of Principal Component Analysis (PCA) have been proposed, such as Sparse PCA, Probabilistic Relational PCA, and Robust PCA.

The Linear Discriminant Analysis (LDA) is a feature learning technique that operates under supervised conditions. Its objective is to reduce the available distant between points of data of a similar class while maximizing the distance between those belonging to different classes in the low-dimensional subspace that has been learned. The utilization of LDA in the realm of facial recognition has yielded favorable outcomes, resulting in the emergence of novel features referred to as Fisherfaces. The GDA model can be considered as the kernel version of the LDA model. Typically, the generalized eigenvalue decomposition is employed to acquire knowledge of LDA and GDA. Harris, DeCarlo, and Richter [15] have noted that the resolution of the global eigenvalue deconstruction merely serves as an estimation of the solution to the trace ratio problem in relation to the formulation of LDA. Consequently, the individuals in question converted the trace ratio issue into a sequence of trace variation issues and implemented an iterative approach to address it.

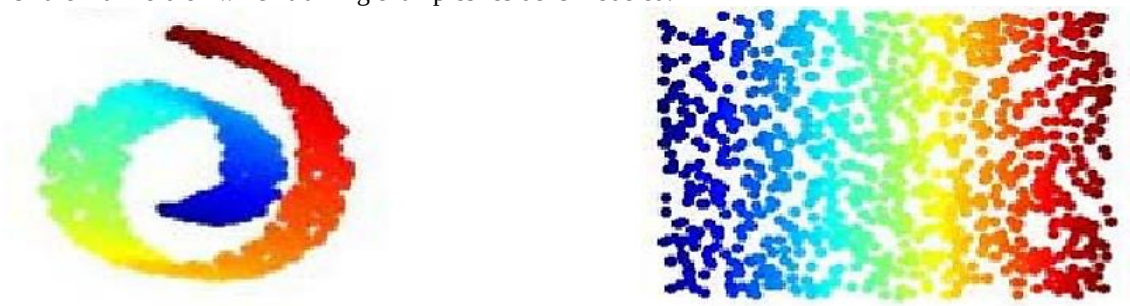
Thongkruer and Aree [16] propose a new Newton-Raphson approach for problems related to trace ratio that can be demonstrated to exhibit convergence. Dimand [17] have introduced a new technique known as relation Fisher evaluation that uses the formulation of trace ratio and effectively leverages the relational information. Lai [18] conducted an analysis of an iterative algorithm utilized in solving problems related to trace ratio. They established the essential and adequate factors and conditions needed for the availability of optimal remedies of problems related to trace ratio. In addition, there are prevailing LDA extensions, such as DGPLVM, marginal Fisher analysis (MFA), and incremental LDA.

In addition to the aforementioned feature learning algorithms, numerous other methods for feature learning exist, including independent component analysis (ICA), canonical-correlation analysis (CCA), feature extraction based on ensemble learning, multi-task feature learning, and others. In addition, a multitude of algorithms for learning tensor representations have been developed to enable direct processing of tensor data. Banerjee, Scheirer, Bowyer, and Flynn [19] presented the 2DPCA algorithm and demonstrated its superiority over PCA in the context of face recognition. Additionally, Deng, Guo, Hsu, and Mandal [20] introduced the 2DLDA algorithm, which expands upon LDA for the purpose of learning tensor representations of second order. The article referenced as [21] presents an algorithm for learning a low rank tensor representation with a large margin. Theoretical guarantees are provided for the convergence of this algorithm.

Manifold learning

Here, we discuss a class of feature-learning techniques based on locality; we name them manifold learning techniques. According to proponents of what is known as "The Manifold Hypothesis," high-dimensional data really exists on low-

dimensional manifolds inside higher-dimensional spaces. This implies that information in higher dimensions often rests on a much more nearby lower-dimensional manifold as original manifold (see **Fig 2**). Manifold Learning refers to the method through which the manifold on which training examples reside is modeled.



(a) Swiss roll (b) Original manifold

Fig 2. An Enhanced Algorithm for Manifold Learning in Data Visualization

While the majority of manifold learning approaches are classified as non-linear dimensional reduction techniques, there exist linear dimensionality reduction approaches such as MFA and locality preserving projection. It is important to note that certain algorithms for nonlinear dimensionality reduction do not fall under the category of manifold learning approaches. This is due to the fact that their objective is not to uncover the inherent structure of data with high dimensionality. Examples of such algorithms include Sammon mapping and KPCA. Two intriguing papers on manifold learning were published in the journal "Science" in the year 2000. The initial publication presents Isomap, a method that merges the Floyd-Warshall algorithm with traditional Multidimensional Scaling (MDS). Isomap is a method that calculates the distances between data points within a given neighborhood using the Floyd-Warshall algorithm. It then utilizes classic MDS to learn the lower dimensionality embeddings of information reliant of computer-based and pair-based distances.

The subsequent paper pertains to Locally Linear Embedding (LLE), a technique that incorporates the proximity details of individual points into the rebuilding weights of their respective neighbors. Subsequently, numerous diverse learning algorithms were suggested. The study conducted by Choudhury [22] integrates the concepts local tangent space alignment (LTSA) and Laplacian eigenmaps (LE). The former involves the computation of local resemblance between data based on Euclidean distance within localized tangent spaces while the latter is utilized to acquire knowledge on low dimensionality data embedding. Various manifold learning techniques were utilized by researchers to enhance the recognition of historical Arabic documents based on their shapes. These approaches yielded significant advancements compared to earlier methods. Apart from the aforementioned techniques, it is imperative to consider certain related works such as the approaches for semi-supervised learning, distance metric learning, non-negative matrix factorization, and dictionary learning. These techniques, to a certain degree, consider the prevailing data structure.

III. ADVANCED DEEP LEARNING

Machine learning approaches referred to as "deep learning" employ distinct layers to steadily extract more complex information from raw input. Within the image processing spectrum, for example, low layers could be employed to determine edges, whereas high layers could be employed to determine human-essential concepts such as faces, characters, and numbers. When viewed from a different perspective, deep learning is the process of teaching a computer to convert an input (such as an image of a dog) into an output (knowledge about dogs). That is why the concept of "deeper" or "deepest" learning makes perfect sense. The highest level of learning is achieved when information is transferred from a source to the target without any human intervention. Therefore, a deeper learning is a hybrid learning process, in which humans first learn from a source to a learned semi-object, and then computers learn from this semi-object to an ultimate learned object. With its foundation in representation learning and artificial neural networks, deep learning are categorized as a larger group of machine learning algorithms. Unsupervised, semi-supervised, and supervised learning are all viable options in this group.

Natural language processing, drug design, computer vision, speech recognition, bioinformatics, medical image analysis, materials inspection, board game programs, and climate science are just some of the areas where deep-learning architectures like deep belief networks, convolutional neural networks, deep neural networks, deep reinforcement learning, recurrent neural networks, and transformers have shown promising results. The data processing, and decentralized nodes of communication within biological frameworks served as inspiration for the development of artificial neural networks (ANNs). There are a number of ways in which ANNs vary from biological brains. With the brain system of various living organisms are analog and plastic (dynamic), ANNs tends to be symbolic and static.

The term "deep" in the aspect of deep learning alludes to the system’s various layers of computational capacity. Previous studies have indicated which a neural system featuring a solitary encrypted layer and non-polynomial activation element with unbounded dimensionality possesses the capability to function as a universal classifier. Conversely, a linear

perceptron is unable to do so. For current applications and efficient implementations, while still preserving theoretical universality under moderate circumstances, deep learning alludes to a variant, which is apprehensive with an infinite layers with bounded sizes. In deep learning, efficiency, trainability, and interpretability are prioritized above strict adherence to physiologically informed connectionist models, hence the layers may be quite diverse. Convolutional neural networks (CNNs) are the backbone of most modern deep learning models. However, other types of artificial neural networks and latent variables organized in layers can also be used in deep generative models.

Within the field of deep learning, it is observed that every level of the neural network has the capacity to acquire the ability to convert the input data into a representation that is increasingly abstract and complex. In the context of image identification, the first input could integrate the pixel's matrix. The initial layer of representation is responsible for abstracting the pixels and encoding edges. Subsequently, the second layer is tasked with composing and encoding arrangements of edges, while the third layer encodes specific facial features such as the nose and eyes. Finally, the fourth layer is responsible for recognizing the presence of a face within the image. Significantly, a deep learning algorithm has the ability to autonomously determine the optimal placement of features within each level. The necessity for manual adjustment remains despite the implementation of this approach. For instance, the utilization of distinct quantities of layers and its magnitudes can yield varying levels of abstraction.

The term "deep" in the context of "deep learning" pertains to the extent of layers involved in the process of data transformation. To be more precise, deep learning systems possess a significant depth in their CAP (credit assignment path). The CAP refers to the series of processes that occur between the initial input and final output. Causal Analysis Patterns (CAPs) are utilized to depict plausible causal associations between the input and output variables. The depth of the Capsules in a feedforward neural network corresponds to the network's depth, which is determined by the hidden layer unumber in addition to one, accounting for the parameterized output layer. Recurrent neural networks (RNNs) have the potential for an unlimited CAP depth, as signals may propagate through a layer multiple times. There is no consensus among scholars regarding the precise threshold that distinguishes shallow learning from deep learning.

However, the majority of researchers concur that deep learning entails a CAP depth that exceeds 2. It has been demonstrated that a two-layered CAP possesses the ability to serve as a universal approximator, meaning that it has the capacity to replicate any given function. Moreover, additional layers do not contribute to the network's capacity for function approximation. Deep neural networks with a capacity greater than two have been observed to possess superior feature extraction capabilities compared to shallow models. As a result, the incorporation of additional layers in the network architecture facilitates the acquisition of features in a more efficient manner.

One possible approach to constructing deep learning architectures is through the employment of greedy layer-by-layer approach. Deep learning enables abstraction disentanglement the identification of performance-enhancing features. Deep learning techniques are capable of eliminating the need for feature engineering in supervised learning tasks. This is achieved by transforming the data into condensed intermediate representations that resemble principal components. Additionally, these methods generate layered architectures that effectively eliminate the level of redundancy within representations. Unsupervised learning tasks can be subjected to the implement deep learning algorithms. The relevance of this advantage is based on the quantity of unlabeled data surpasses that of labeled data. Deep belief networks are a type of deep structure that can be trained through unsupervised learning methods.

There have been four survey articles on deep learning, which have been published in academic literature. Hernandez, Muratet, Pierotti, and Carron [23] provided an introduction to the principles, motivations, and significant deep learning approaches. Additionally, in reference [24], Espinosa, Jimenez, and Palma conducted a review of the advancements made in feature learning and deep learning from the representation learning perspective. Zhang, Sjarif, and Ibrahim [25] presented an exposition on the advancement of deep learning, as well as significant models within this field, such as convolutional neural networks and recurrent neural networks. Kim [26] conducted a retrospective analysis of the progression of artificial neural networks and deep learning over time. These survey papers provide readers with an accessible understanding of the research area and historical development of deep learning, particularly for those with an interest in the field. Several online resources are recommended for acquiring knowledge on deep learning algorithms. The initial option pertains to the Coursera course instructed by Professor Hinton. The website for the course on neural networks can be accessed at [27]. The subject matter of this course pertains to the study of artificial neural networks and their application in the field of machine learning.

The second tutorial pertains to deep learning and unsupervised feature learning, and has been developed by researchers affiliated with Stanford University. The webpage for the UFLDL Tutorial can be accessed at [28]. In addition to fundamental comprehension of deep learning algorithms and unsupervised feature learning, this tutorial incorporates numerous exercises. Therefore, it is highly appropriate for individuals who are new to the field of deep learning. The website dedicated to deep learning is the third one. The website in question can be accessed at [29].The website offers a comprehensive range of resources, including tutorials on deep learning, recommended reading materials, software tools, and datasets. The fourth item pertains to a blog that has been composed in the Chinese language [30].

The authors of [31] document knowledge in deep learning and meticulously documenting the process of coding each model. However, there exist numerous other blogs and webpages that are equally valuable and beneficial, including Wikipedia. The final item on the list is the publication entitled "Deep Learning," authored by Bengio, Professor Goodfellow, and Courville, and released by MIT Press [32]. The digital rendition of the publication is available at no cost

and can be accessed in [33]. Through the utilization of various educational resources such as courses, tutorials, blogs, and books, individuals studying or working in the deep learning field can acquire a comprehensive understanding of the theoretical intricacies associated with deep learning algorithms.

Deep learning models

In this paper, a comprehensive analysis is conducted on various deep learning models, with a particular focus on those introduced subsequent to the publication of reference [34].The resurgence of deep learning can be attributed primarily to significant advancements in three key areas: feature learning, the abundance of large-scale labeled hardware and data, particularly GPGPUs (general-purpose graphics processing units). Gu et al. [35] proposed in 2006 to employ a method of greedy layer-wise pre-training trailed by finetuning to stimulate deep neural network learning. This approach yielded superior performance compared to contemporary algorithms in the domains of MNIST handwritten digits identification and document access tasks.

Pootheri [36] introduced the concept of stacked auto-encoders and an authenticated hypothesis, which the layer-wise unsupervised training approach, which is implemented in a greedy manner, is primarily beneficial for optimization. This is achieved setting weights within a region that is proximal to a favorable local minimum, thereby generating internalized distributed representation, which is an abstraction with a higher input level, and ultimately leading to improved generalization.

The stacked denoising auto-encoders are proposed for denoising corrupted varieties of inputs and are learned locally. Using stochastic neighbor embedding (SNE) and PCA as examples, Pahuja and Prasad [37] demonstrated the efficacy of deep architectures constructed using stacked feature learning modules. Cascianelli, Cornia, Baraldi, and Cucchiara [38] proved the efficacy of the suggested strategy on handwritten text recognition tasks by applying the stretching methodology to the weights matrices between the upper consecutive layers of deep architectures constructed using stacked feature learning models. For offline handwriting identification, a tandem hidden Markov structure with deep belief networks (DBNs) is suggested as used in [39]. At the 2012 ImageNet LSVRC (ImageNet Large Scale Visual Recognition Competition), the "AlexNet" developed by Krizhevsky, Sutskever, and Hinton came out on top. Rectified linear units (ReLUs) are nonlinear activation functions that were employed in AlexNet with the dropout regularization.

AlexNet was deployed on GPUs to expedite training on 1.2 million training photos spanning 1000 categories. OverFeat, VGGNet, GoogleNet, and ResNet are only few examples of deep convolutional neural networks (CNNs) that were used in the top-performing models in the ImageNet LSVRC between 2013 and 2016. An intriguing AlexNet-based feature extraction approach was suggested in [40]. Singstad and Tavashi [41] demonstrated that activation characteristics from deep convolutional systems (such as AlexNet) learned in a completely supervised manner on a fixed, large grouping of object identification tasks may be reused to fresh, generic tasks. Therefore, we dubbed this functionality "deep convolutional activation feature" (DeCAF). Joo et al. [42] offered two difficult tasks using scanned document pictures and used DeCAF to establish a benchmark for their work. Cai, Zhong, Zheng, Huang, and Dong [43] took into account the issue of whether or not DeCAF is best for precise image classification, and they enhanced DeCAF on numerous image classification problems reliant of stretching and lowering processes. Qu, Wang, Feng, Zhang, and Yu [44] suggested a deep hashing learning algorithm based on AlexNet and VGGNet, which significantly outperformed prior hashing learning techniques to image retrieval.

The most prominent deep learning approaches include recurrent neural networks (RNNs), long short term memories (LSTMs). De-noising autoencoders (DAEs), convolutional neural networks (CNNs), and deep belief networks (DBNs). Detailed explanations of each technique are provided below, along with some noteworthy examples of its use.

Convolutional neural network (CNN)

The gap between human and machine intelligence has been narrowing rapidly thanks to the rapid development of artificial intelligence. Experts and amateurs alike work tirelessly on many facets of the area to achieve remarkable results. Computer vision is only one example of such a field. The goal of this area of study is to teach computers to see and understand the world as humans do, so that they can apply this understanding to fields as diverse as video and image recognition, image classification and analysis, media recreation, recommendation systems, natural language processing, and more. Over the course of its development and refinement, a Convolutional Neural Network has been at the center of most of the Deep Learning-based improvements to computer vision.

With the employment of learnable biases, and weights, Convolutional Neural Networks (ConvNets/CNNs) are capable of taking in input images, prioritize various objects/features within it, and then categorize them consequently. ConvNet necessitates increasingly less preparation actions compared to competing classification approaches. While the filters should be hand-engineered whenever using primitive approaches, ConvNets could be effectively trained to autonomously obtain the require filters as well as other characteristics. The architecture of ConvNets copies the operations of the human brain and its neuronal pattern connections and mimics design cues from the Visual Cortex model. The Receptive Field represents a segment of visual fields in which every neuron is more sensitive. Whenever several of these fields tend to overlap, they can safeguard the field of view.

CNN is a well-known example of DL architecture. In most cases, this method is used in fields related to image processing. **Fig 3** depicts the structure of a CNN, which integrates three different layers: fully-connected, pooling, and

convolutional layers. Each convolutional neural network (CNN) has a unique training procedure comprised of two phases: the feed-forward phase and the back-propagation phase. ZFNet, VGGNet, GoogleNet, AlexNet, and ResNet are some of the most widely used CNN designs.

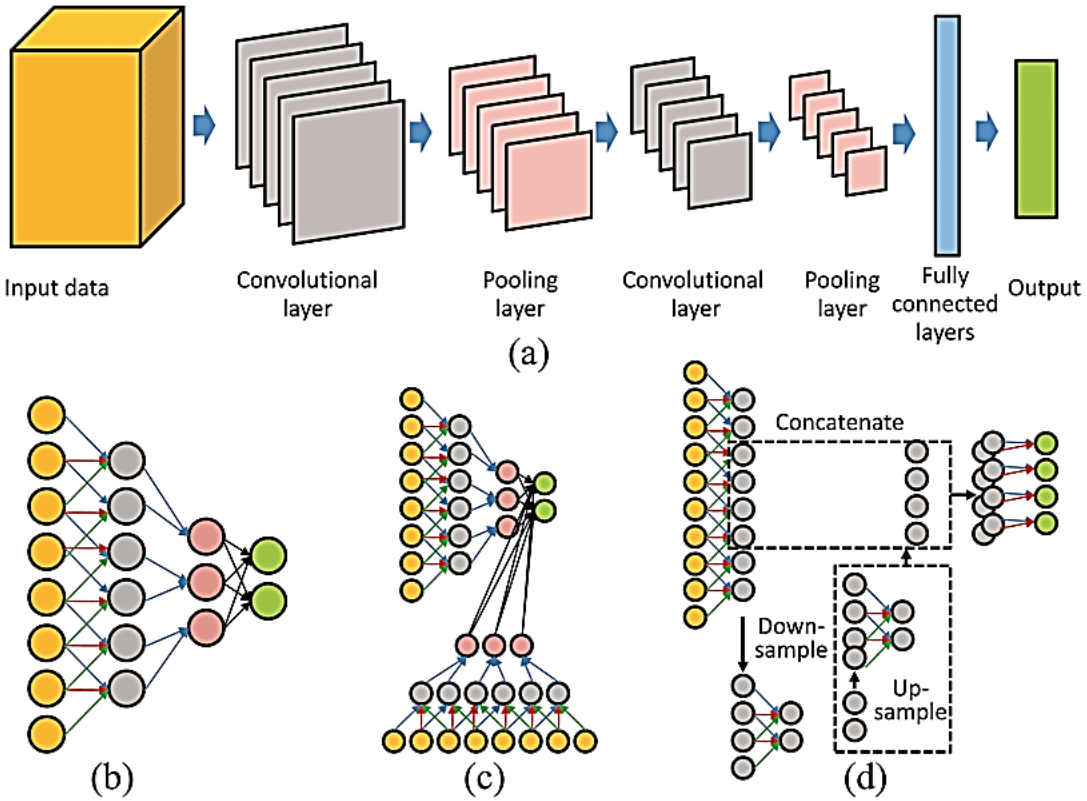


Fig 3. CNNs Architecture

CNN is most often used in image processing, although it has been employed in distinct domains (such as energy, electronics systems, computational mechanics, remote sensing, etc.) in the academic literature.

Recurrent neural networks (RNN)

A recurrent neural network (RNN) alludes to a system of ANN formulated categorically for implementation with sequence data and time series data. Regular feedforward neural networks can only properly process data points if they are completely isolated from one another. Nonetheless, neural network could be modified to account for a dependency between different points of data in case data is available in sequences where a single point of data relies on an earlier point of data. To general an upcoming output in a series, RNNs use a notion of "memory" that permits them to retain the information or states of earlier inputs.

Speech, handwriting, text, and other applications that rely on sequence and pattern recognition are ideal candidates for RNN. The recurrent calculations and cyclic connections in an RNN's structure allow it to analyze input data in a sequential fashion. By having its edges flow into an upcoming time step instead of an upcoming layer in a similar time step, an RNN is essentially an extended version of a regular neural network. The outputs are computed using state vectors that include information from all of the preceding inputs and are stored in secret units. The structure of an RNN is seen in **Fig 4**.

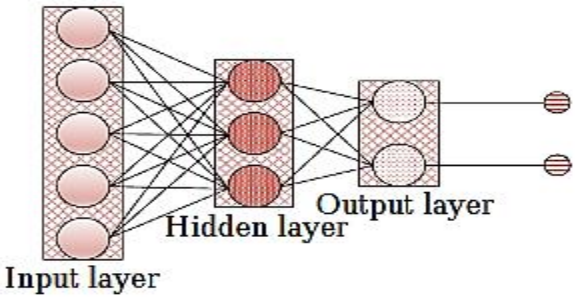


Fig 4. RNNs Architecture

RNN is a more recent deep learning technique. Because of this, there is still a lot of potential for study and exploration in the application fields. Current applications documented in the literature include in the fields of energy, expert systems, hydrological prediction, economics, and navigation.

Denoising AutoEncoder (DAE)

DAE is an asymmetrical neural network that builds off of AE to learn features from noisy datasets. The input, encoding, and decoding layers make up DAE's three primary components. Aggregating DAE allows for the extraction of meta-level characteristics. The DEA procedure produces the unsupervised Stacked Denoising AutoEncoder (SDAE) approach that could be employed for non-linear dimensional reduction. This technique uses a feed-forward neural network with a deep infrastructure integrating various hidden layers and a pre-training plan. The framework of DEA approach is shown in **Fig 5**.

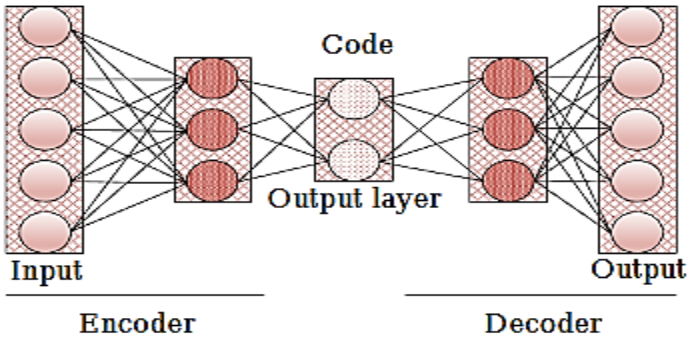


Fig 5. DEA Architecture

Researchers are gradually coming to recognize DEA as a powerful DL algorithm. Multiple fields have successfully implemented DEA with positive outcomes. Current popular uses of DEA include energy forecasting, banking, cybersecurity, fraud detection, speaker verification, and image classification.

The deep belief networks (DBNs)

Deep Belief Networks (DBNs) are an AI network architecture inspired by the concept of "deep learning." DBNs are often employed in image and voice recognition due to their ability to recognize and categorize complicated patterns in data. Although DBNs are quite intricate, they really just consist of many layers of connected neurons. Together, these modules analyze data from lower levels, allowing for more precise predictions and classifications. DBNs may be compared to a complex web of nerve fibers. Each layer is fed data from the layer below it, processes that data, and sends the result on to the next layer. This data is then used by the last layer to draw conclusions or assign labels. It is like peeling back a layer of an onion to reveal another part of the answer. DBNs are used for data learning on high dimensional manifolds.

The units inside each layer are not connected to one another, although there are connections between the layers. The directed and undirected connections of a DBN make it similar to a multi-layer neural network. DBNs have RBMs, or restricted Boltzmann machines, that are trained greedily. Each RBM layer may exchange information with the layers above and below it. This model uses numerous layers of restricted Boltzmann machines, or RBMs, as feature extractors, and is based on a feed-forward network. An RBM only has two layers—a concealed layer and an exposed one. The DBN method's structure is shown in **Fig 6**.

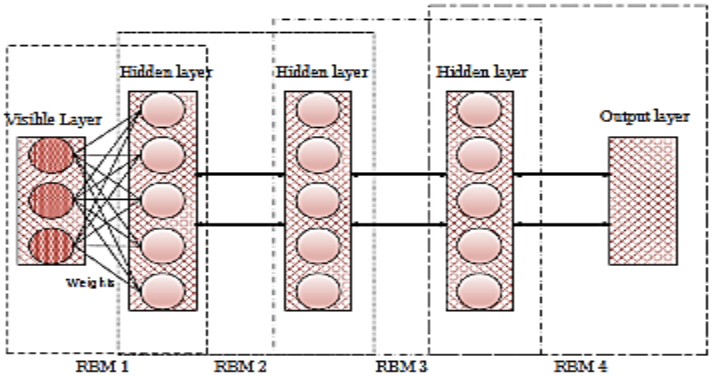


Fig 6. DBN Infrastructure

DBN has shown to be one of the most accurate and efficient deep learning algorithms. As a result, there has been a broad variety of application fields, including some very fascinating uses in various technical and scientific difficulties.

Among the public application fields are human emotion recognition, renewable energy projection, time series prediction, economic forecasting, and cancer diagnosis.

Long Short-Term Memory (LSTM)

In order to function as a general-purpose computer, LSTM is an RNN approach that takes use of feedback connections. This technique has uses in image processing as well as sequence and pattern identification. Input, output, and forget gates are the three main components of a typical LSTM network. By controlling when the input is allowed into the neuron, LSTM is able to retain the results of the previous computation. All of these decisions are made depending on the current input, which is one of the LSTM method's key strengths. The LSTM framework is seen in **Fig 7**.

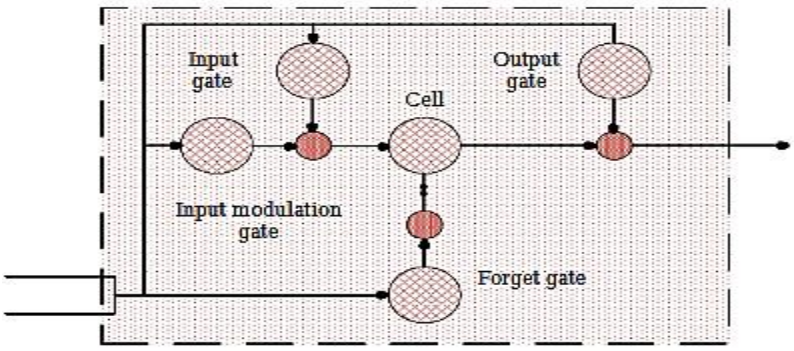


Fig 7. LSTM Architecture

LSTM has showed impressive promise in geo-logical modeling, air quality, hydrological prediction, and hazard modeling, among other environmental applications. The LSTM architecture's generalizability makes it a promising candidate in a wide variety of fields. Other areas where LSTM has found success include the modeling of solar power systems, energy demand and consumption, and the wind energy sector. As was done with machine learning techniques, further research is needed to delve into the new deep learning approaches and their potential application areas.

Deep learning toolboxes

Numerous deep learning toolboxes are frequently disseminated on the internet. Deep learning structures such as LeNet-5, DBNs, VGGNet, and AlexNet, are frequently included in the codes provided within each toolbox. The deep belief network (DBN) is a form of generative graphical structure or deep neural network that is utilized in machine learning. It is comprised of numerous layers of latent variables, also known as "hidden units," which are interconnected between the layers but not within each individual layer. A Deep Belief Network (DBN) has the ability to acquire the skill of probabilistically reconstructing its inputs when exposed to a set of examples without any supervision. The layers subsequently function as detectors of features.

Following the initial learning phase, a Deep Belief Network (DBN) may undergo additional supervised training in order to carry out classification tasks. Deep Belief Networks (DBNs) are a type of neural network architecture that can be conceptualized as a hierarchical arrangement of unsupervised networks, such as autoencoders, or restricted Boltzmann machines (RBMs). In this arrangement, the hidden layer of every network is employed as a visible layer for the respective networks. RBM is a type of energy-based generative model that is undirected. It consists of a hidden layer and a visible input layer, with interconnections between different layers but not within them. The present composition outlines a rapid unsupervised training methodology that operates on a layer-by-layer basis. Specifically, the contrastive divergence technique is employed on each sub-network sequentially, commencing with the "lowest" layer pair (wherein the lowest visible layer corresponds to the training set).

LeNet-5 is considered as one of the founding convolutional neural networks that facilitated the advancement of deep learning. LeNet-5 has been designated as the pioneering work since 1988, following extensive research and numerous successful iterations. The application of the backpropagation algorithm to practical applications was first carried out by Lecun, Bottou, Bengio, and Haffner [45]. They posited that the provision of restraints from the domain of the task could significantly enhance the network's generalization learning ability.

Hayashi [46] utilized a convolutional neural network, which was learned by back propagation approaches to accurately recognize handwritten numerical characters. This approach was then successfully employed to identify handwritten zip codes as issued by the USA Postal Service. This served as the initial model for the subsequent development of LeNet. LeCun presented a handwritten digit recognition issue in a paper during the same year. The problem was shown to be linearly separated, nonetheless, networks that are single-layered demonstrated inadequate generalization capacity. The implementation of shift-invariant feature identifiers with constrained and multi-layered network has the potential to yield high performance outcomes. The individual held the belief that the aforementioned outcomes served as evidence for the

notion that reducing the quantity of unconstrained variables within the neural network could augment the neural network's capacity for generalization.

The AlexNet architecture is a CNN, which was established by Alex Krizhevsky in collaboration with Geoffrey Hinton and Ilya Sutskever who were doctoral mentors working for Krizhevsky. In 2012, AlexNet contributed in a challenge known as the ImageNet Large Scale Visual Recognition. The neural network attained a 15% top-5 error rate that is over 10.8 percentage points lower than the second-best performer. The principal finding of the original study was that the model's depth played a crucial role in achieving its superior performance, albeit at a high computational cost. This was made possible by leveraging graphics processing units (GPUs) during the training process. It should be noted that AlexNet did not hold the distinction of being the initial rapid GPU-centric application of CNN to emerge victorious in an image recognition competition.

Jordà, Valero-Lara, and Peña's [47] study found that the implementation of a CNN on GPU resulted in a fourfold increase in speed compared to an equivalent implementation on a Central Processing Unit (CPU). Shalaby, ElShennawy, and Sarhan [48] were able to achieve superior performance compared to their predecessors by utilizing a deep CNN that was 60 times faster. From May 15, 2011 to September 10, 2012, CNN emerged victorious in a minimum of four image competitions. Furthermore, they achieved a noteworthy enhancement over the most outstanding outcome documented in the existing literature for various image databases.

The Visual Geometry Group (VGG) is affiliated with Oxford University's Science and Engineering Department. A sequence of CNN, beginning with VGG, has been introduced for utilization in face recognition and image classification. These models include VGG16 and VGG19. VGG's research on the extent of convolutional networks aimed to investigate the impact of network depth on the accuracy and precision of large-scale recognition and classification of an image. Deep-16 CNN is a form of NN infrastructure employed in deep learning. To increase the depth of network layers while minimizing the number of parameters, a compact 3 by 3 convolution kernel is employed across all layers.

The VGG model is designed to receive an input consisting of an RGB image with dimensions of 224 by 244 pixels. The training set images undergo computation of the mean RGB value, which is subsequently utilized as input for the VGG convolutional network. A filter of either 3 by 3 or 1 by 1 dimension is employed, and the convolution process is constant. The VGG architecture comprises three fully connected layers, with variations in the number of convolutional and fully connected layers determining the specific model, ranging from VGG11 to VGG19. The VGG11 architecture comprises a minimum of three fully connected layers and eight convolutional layers. The VGG19 architecture comprises a total of 16 convolutional layers at maximum. The neural network architecture includes three layers that are fully connected. Furthermore, it should be noted that the VGG network does not incorporate a pooling layer immediately after each convolutional layer. Instead, a total of five pooling layers are dispersed among various convolutional layers.

The codes may be utilized by the researchers either directly or through the development of novel models, subject to specific licensing agreements. The subsequent text provides a concise introduction to Theano, Caffe, TensorFlow, and MXNet. Theano is a software library for the Python programming language. The software exhibits a high degree of integration with NumPy, enabling users to proficiently establish, refine, and assess mathematical expressions that encompass multi-dimensional arrays. Furthermore, it has the capability to execute computationally intensive computations on Graphics Processing Units (GPUs) resulting in a performance boost of up to 140 times faster compared to Central Processing Units (CPUs). The tutorial on deep learning, which can be accessed in [49], is solely founded on Theano. Caffe is a software library designed for deep learning, written in C++ and CUDA. The software offers interfaces for command line, MATLAB, and Python.

The Caffe code exhibits efficient performance and possess the ability to seamlessly transition between GPU and CPU. TensorFlow refers to a library of software that is open source and designed for numerical computation through the use of information flow graphs. On the basis of this graph, the computational operations are symbolized by nodes, whereas the multidimensional data arrays, also known as tensors, are represented by the graph edges that facilitate their communication between the nodes. TensorFlow possesses the ability to perform automatic differentiation, which aids in the calculation of derivatives. MXNet has been collaboratively developed by multiple academic institutions and corporate entities. The software facilitates both symbolic and imperative programming paradigms and accommodates a variety of programming languages, including but not limited to C++, R, Python, Scala, Matlab, Julia, and Javascript. Overall, the velocity of MXNet codes during execution is comparable to that of codes in Caffe, and notably superior to that of TensorFlow, and Theano.

IV. CONCLUSION AND FUTURE RESEARCH

This paper provides a comprehensive review of the existing study on data representation learning, encompassing both conventional feature learning techniques and more recent advancements in deep learning. Representation learning, which is a component of decision tree representation with machine learning field, is commonly referred to as feature learning. The system employs a collection of methodologies to identify the necessary representations for detecting features or categorizing the available raw data. The existence of artificial neural networks and feature learning approaches indicates that deep learning is not an entirely novel concept. The aforementioned phenomenon can be attributed to the significant advancements in feature learning research, the increased accessibility of vast amounts of labeled data, and the development of advanced hardware. The advent of deep learning has had a significant impact not only on the field of artificial intelligence, but also on various other domains, including finance and bioinformatics, leading to notable advancements.

In regards to future research concerning deep learning, about three potential avenues of exploration include the novel algorithms and its applications, and fundamental theory. Several scholars have attempted to examine deep neural networks. Nonetheless, there exists a considerable disparity between the theoretical and practical implementation of deep learning. Despite the existence of numerous proposed deep learning algorithms, a majority of them rely on either Recurrent Neural Networks (RNNs) or deep Convolutional Neural Networks (CNNs). Hence, it is imperative to introduce innovative deep learning algorithms that can effectively address practical challenges, including transfer learning models and unsupervised models. In addition, deep learning approaches have been initially utilized in various fields. Nevertheless, in order to address complex issues, such as those encountered in natural language processing and computer vision, it is necessary to develop and implement more advanced models. It is important to note that deep learning algorithms are merely a machine learning component and should not be considered the sole means of achieving artificial intelligence. In order to address real-world issues, a variety of methodologies for intelligent data analytics are required.

Data Availability

No data was used to support this study.

Conflicts of Interests

The author(s) declare(s) that they have no conflicts of interest.

Funding

No funding agency is associated with this research.

Ethics Approval and Consent to Participate

The research has consent for Ethical Approval and Consent to participate.

Competing Interests

There are no competing interests.

References

[1]. C.-C. Chang, “Fisher’s linear discriminant analysis with space-folding operations,” IEEE Trans. Pattern Anal. Mach. Intell., vol. PP, 2023.

[2]. P. Shrivastava, Department of Electronics and Telecommunication Engineering, Graduate. The areas of interests are Machine Learning, Data Analytics, Deep Learning. Mumbai, India., K. Singh, A. Pancham, Department of Electronics and Telecommunication Engineering, Graduate. The areas of interests are Machine Learning, Data Analytics, Deep Learning, Cloud Computing. Mumbai, India., and Department of Electronics and Telecommunication Engineering, graduate. The areas of interests are Machine Learning, Data Analytics, Deep Learning, Cloud Computing. Mumbai, India., “Classification of Grain s and Quality Analysis u sing Deep Learning,” Int. J. Eng. Adv. Technol., vol. 11, no. 1, pp. 244–250, 2021.

[3]. F. Dalvi, N. Durrani, H. Sajjad, Y. Belinkov, A. Bau, and J. Glass, “What is one grain of sand in the desert? Analyzing individual neurons in deep NLP models,” Proc. Conf. AAAI Artif. Intell., vol. 33, no. 01, pp. 6309–6317, 2019.

[4]. J. Treur, “Relating an adaptive network’s structure to its emerging behaviour for Hebbian learning,” in Theory and Practice of Natural Computing, Cham: Springer International Publishing, 2018, pp. 359–373.

[5]. L. Dung and M. Mizukaw, “Designing a pattern recognition neural network with a reject output and many sets of weights and biases,” in Pattern Recognition Techniques, Technology and Applications, InTech, 2008.

[6]. K. Ren, Q. Wang, and R. J. Burkholder, “A fast back-projection approach to diffraction tomography for near-field microwave imaging,” IEEE Antennas Wirel. Propag. Lett., vol. 18, no. 10, pp. 2170–2174, 2019.

[7]. R. Guo, X. Qiu, and Y. He, “Evaluation of agricultural investment climate in CEE countries: The application of back propagation neural network,” Algorithms, vol. 13, no. 12, p. 336, 2020.

[8]. E. S. Gopi, “Dimensionality Reduction Techniques,” in Pattern Recognition and Computational Intelligence Techniques Using Matlab, Cham: Springer International Publishing, 2020, pp. 1–29.

[9]. J. Gou et al., “Discriminative and Geometry-Preserving Adaptive Graph Embedding for dimensionality reduction,” Neural Netw., vol. 157, pp. 364–376, 2023.

[10]. A. Sarhadi, D. H. Burn, G. Yang, and A. Ghodsi, “Advances in projection of climate change impacts using supervised nonlinear dimensionality reduction techniques,” Clim. Dyn., vol. 48, no. 3–4, pp. 1329–1351, 2017.

[11]. Y. Yang and T. Hospedales, “Deep multi-task representation learning: A tensor factorisation approach,” arXiv [cs.LG], 2016.

[12]. L. Yang, C. Heiselman, J. G. Quirk, and P. M. Djurić, “Class-imbalanced classifiers using ensembles of Gaussian processes and Gaussian process latent variable models,” Proc. IEEE Int. Conf. Acoust. Speech Signal Process., vol. 2021, 2021.

[13]. G. Song, S. Wang, Q. Huang, and Q. Tian, “Multimodal Similarity Gaussian Process latent variable model,” IEEE Trans. Image Process., vol. 26, no. 9, pp. 4168–4181, 2017.

[14]. G. Zhong, W.-J. Li, D.-Y. Yeung, X. Hou, and C.-L. Liu, “Gaussian process latent random field,” Proc. Conf. AAAI Artif. Intell., vol. 24, no. 1, pp. 679–684, 2010.

[15]. T. L. Harris, R. A. DeCarlo, and S. Richter, “A CONTINUATION APPROACH TO GLOBAL EIGENVALUE ASSIGNMENT11Supported by U.s. department of energy under DOE contract number DE-AC01-79ET29365,” in Computer Aided Design of Multivariable Technological Systems, Elsevier, 1983, pp. 95–101.

[16]. P. Thongkruer and P. Aree, “Power-flow initialization of fixed-speed pump as turbines from their characteristic curves using unified Newton-Raphson approach,” Electric Power Syst. Res., vol. 218, no. 109214, p. 109214, 2023.

[17]. R. W. Dimand, “Irving fisher and the fisher relation: Setting the record straight,” Can. J. Econ., vol. 32, no. 3, p. 744, 1999.

[18]. Z. J. &. F. Lai, “A convergence analysis on the iterative trace ratio algorithm and its refinements,” CSIAM Transactions on Applied Mathematics, vol. 2, no. 2, pp. 297–312, 2021.

[19]. S. Banerjee, W. Scheirer, K. Bowyer, and P. Flynn, “Analyzing the impact of shape & context on the face recognition performance of deep networks,” arXiv [cs.CV], 2022.

[20]. S. Deng, Y. Guo, D. Hsu, and D. Mandal, "Learning tensor representations for meta-learning," arXiv [cs.LG], 2022.
[21]. W. Guo and J.-M. Qiu, "A low rank tensor representation of linear transport and nonlinear Vlasov solutions and their associated flow maps," J. Comput. Phys., vol. 458, no. 111089, p. 111089, 2022.
[22]. S. D. Choudhury, "Root Laplacian Eigenmaps with their application in spectral embedding," arXiv [math.DG], 2023.
[23]. J. Hernandez, M. Muratet, M. Pierotti, and T. Carron, "Can we detect non-playable characters' personalities using machine and deep learning approaches?," Proc. Eur. Conf. Games-based Learn., vol. 16, no. 1, pp. 271–279, 2022.
[24]. R. Espinosa, F. Jimenez, and J. Palma, "Surrogate-assisted and filter-based multiobjective evolutionary feature selection for deep learning," IEEE Trans. Neural Netw. Learn. Syst., vol. PP, pp. 1–15, 2023.
[25]. C. Zhang, N. N. A. Sjarif, and R. B. Ibrahim, "Deep learning techniques for financial time series forecasting: A review of recent advancements: 2020-2022," arXiv [q-fin.ST], 2023.
[26]. L.-W. Kim, "DeepX: Deep learning accelerator for restricted Boltzmann machine artificial neural networks," IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 5, pp. 1441–1453, 2018.
[27]. S. Theodoridis, "Neural networks and deep learning," in Machine Learning, Elsevier, 2015, pp. 875–936.
[28]. "UFLDL tutorial," Stanford.edu. [Online]. Available: http://deeplearning.stanford.edu/wiki/index.php/UFLDL_Tutorial. [Accessed: 31-May-2023].
[29]. "Deep learning specialization," Coursera. [Online]. Available: https://www.coursera.org/specializations/deep-learning. [Accessed: 31-May-2023].
[30]. "AP Chinese Language and Culture past exam questions," Collegeboard.org. [Online]. Available: https://apcentral.collegeboard.org/courses/ap-chinese-language-and-culture/exam/past-exam-questions. [Accessed: 31-May-2023].
[31]. "What is deep learning? Definition, examples, and careers," Coursera, 05-Apr-2022. [Online]. Available: https://www.coursera.org/articles/what-is-deep-learning. [Accessed: 31-May-2023].
[32]. I. Goodfellow, Y. Bengio, and A. Courville, "Deep learning," MIT Press, 01-Dec-2021. [Online]. Available: https://mitpress.mit.edu/9780262035613/deep-learning/. [Accessed: 31-May-2023].
[33]. H. Schulz and S. Behnke, "Deep learning: Layer-wise learning of feature hierarchies," KI - Künstl. Intell., vol. 26, no. 4, pp. 357–363, 2012.
[34]. G. Agrafiotis, E. Makri, I. Kalamaras, A. Lalas, K. Votis, and D. Tzovaras, "Nearest Unitary and Toeplitz matrix techniques for adaptation of Deep Learning models in photonic FPGA," nldl, vol. 4, 2023.
[35]. X. Gu et al., "Hierarchical weight averaging for deep neural networks," IEEE Trans. Neural Netw. Learn. Syst., vol. PP, 2023.
[36]. S. Pootheri and G. V. K, "Localisation of mammographic masses by greedy backtracking of activations in the stacked auto-encoders," arXiv [cs.CV], 2023.
[37]. G. Pahuja and B. Prasad, "Deep learning architectures for Parkinson's disease detection by using multi-modal features," Comput. Biol. Med., vol. 146, no. 105610, p. 105610, 2022.
[38]. S. Cascianelli, M. Cornia, L. Baraldi, and R. Cucchiara, "Boosting modern and historical handwritten text recognition with deformable convolutions," Int. J. Doc. Anal. Recognit., vol. 25, no. 3, pp. 207–217, 2022.
[39]. L. Gu, L. Yang, and F. Zhou, "Approximation properties of Gaussian-binary restricted Boltzmann machines and Gaussian-binary deep belief networks," Neural Netw., vol. 153, pp. 49–63, 2022.
[40]. A. A. Barbhuiya, R. K. Karsh, and S. Dutta, "AlexNet-CNN based feature extraction and classification of multiclass ASL hand gestures," in Lecture Notes in Electrical Engineering, Singapore: Springer Singapore, 2021, pp. 77–89.
[41]. B.-J. Singstad and B. Tavashi, "Using deep convolutional neural networks to predict patients age based on ECGs from an independent test cohort," nldl, vol. 4, 2023.
[42]. K. Joo, K. Lee, S.-M. Lee, A. Choi, G. Noh, and J.-Y. Chun, "Deep learning model based on natural language processes for multi-class classification of R&D documents: Focused on climate technology classification," J. Inst. Electron. Inf. Eng., vol. 59, no. 7, pp. 21–30, 2022.
[43]. Y. Cai, G. Zhong, Y. Zheng, K. Huang, and J. Dong, "Is DeCAF good enough for accurate image classification?," in Neural Information Processing, Cham: Springer International Publishing, 2015, pp. 354–363.
[44]. W. Qu, D. Wang, S. Feng, Y. Zhang, and G. Yu, "A novel cross-modal hashing algorithm based on multimodal deep learning," Sci. China Inf. Sci., vol. 60, no. 9, 2017.
[45]. Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proc. IEEE Inst. Electr. Electron. Eng., vol. 86, no. 11, pp. 2278–2324, 1998.
[46]. K. Hayashi, "Exploring unexplored tensor network decompositions for convolutional neural networks," Brain Neural Netw., vol. 29, no. 4, pp. 193–201, 2022.
[47]. M. Jordà, P. Valero-Lara, and A. J. Peña, "cuConv: CUDA implementation of convolution for CNN inference," Cluster Comput., vol. 25, no. 2, pp. 1459–1473, 2022.
[48]. E. Shalaby, N. ElShennawy, and A. Sarhan, "Utilizing deep learning models in CSI-based human activity recognition," Neural Comput. Appl., vol. 34, no. 8, pp. 5993–6010, 2022.
[49]. "IBM Developer," Ibm.com. [Online]. Available: https://developer.ibm.com/articles/an-introduction-to-deep-learning. [Accessed: 31-May-2023].