

Deep Learning Application Pros and Cons Over Algorithm

Ata Jahangir Moshayedi^{1,*}, Atanu Shuvam Roy², Amin Kollahdooz^{3,*}, Yang Shuxin⁴

¹School of Information Engineering Jiangxi University of Science and Technology, Ganzhou, China, ajm@jxust.edu.cn.

²School of Information Engineering Jiangxi University of Science and Technology, Ganzhou, China, Atanuroy911@gmail.com.

³Faculty of technology, school of engineering and sustainable development, De Montfort University, Leicester, UK. Amin.kollahdooz@dmu.ac.uk

⁴School of Information Engineering Jiangxi University of Science and Technology, Ganzhou, China, yangshuxin@jxust.edu.cn.

Abstract

Deep learning is a new area of machine learning research. Deep learning technology applies the nonlinear and advanced transformation of model abstraction into a large database. The latest development shows that deep learning in various fields and greatly contributed to artificial intelligence so far. This article reviews the contributions and new applications of deep learning. The main target of this review is to give the summarize points for scholars to have the analysis about applications and algorithms. Then review tries to investigate the main applications and uses algorithms. In addition, the advantages of using the method of deep learning and its hierarchical and nonlinear functioning are introduced and compared to traditional algorithms in common applications. The following three criteria should be taken into consideration when choosing the area of application. (1) expertise or knowledge of the author; (2) the successful application of deep learning technology has changed the field of application, such as voice recognition, chat robots, search technology and vision; and (3) deep learning can have a significant impact on the application domain and benefit from recent research with natural language and text processing, information recovery and multimodal information processing resulting from multitasking deep learning. This review provides a general overview of a new concept and the growing benefits and popularity of deep learning, which can help researchers and students interested in deep learning methods.

Keywords: Deep learning, face recognition, speech recognition, medical image recognition, character recognition,

Received on 26 November 2021, accepted on 10 February 2022, published on 18 February 2022

Copyright © 2022 Ata Jahangir Moshayedi *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](https://creativecommons.org/licenses/by/4.0/), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/10.4108/airo.v1i.19

1. Introduction

Since 2006, deep structure learning, or additional learning commonly referred to as Deep Learning (DL) or classroom-conscious learning. It has become a brand-new space for Machine Learning (ML) analysis. Many years ago, the technology developed from DL analysis had affected many different signal and data processing tasks, including the normal, new, wider scope and key aspects of ML and Artificial Intelligence (AI), and Media reports on this progress. DL allows computational models composed of multiple processing layers to learn data representations with multiple levels of abstraction. These methods have

greatly improved the technical level of speech recognition, visual target recognition, target detection, and many other fields such as drug discovery and genomics. DL uses backpropagation algorithms to discover complex structures in large data sets to instruct the machine how to change the internal parameters used to calculate the representation of each layer from the representation of the previous layer. Deep Convolutional Networks (DCNs) have made breakthroughs in processing images, video, voice and audio, while recursive networks have brought light to sequential data such as text and voice. In recent years, a series of seminars, tutorials and special issues or special conference sessions have been dedicated to in-depth learning and applications in many signal and data processing fields. The different research application fields

of DL are explored, and a chronological advancement in research has been shown in each field. This review paper represents the significance of research in DL and shows how it will soon become the world's future. Also, its advantages and disadvantages are discussed. In the next section, some accredited researches are described. The article discusses facial recognition, speech recognition, image recognition or handwriting recognition, virtual assistants, chatbots, healthcare, entertainment and music and finally, robotics. Detailed discussion is also done in facial recognition, speech recognition and healthcare sectors where DL are used more than other applications. Then finally, the popular techniques used in DL are summarized.

2. Research on the Applications of DL

2.1. DL on Facial Recognition

Liu et al. (2017) [1] proposed a novel approach towards face recognition using Convolutional Neural Networks (CNNs). They used angular softmax loss (A-Softmax loss) to enable CNNs and learn the discriminative features of a face with angular margin. They showed that A-Softmax loss could provide a suitable geometric interpretation by constraining the learned features to be discriminative on hyper-spherical manifolds, which essentially matches the prior features that the surfaces are also located on nonlinear manifolds. This connection makes A-Softmax very effective for learning facial expressions.

Luan Tran et al, (2017) [2] have proposed Disentangled Representation learning Generative Adversarial Network (DR-GAN) with three different novelties. First, in addition to image synthesis, the encoder-decoder structure of the generator allows DR-GAN to learn, generate and discriminate representations. Secondly, this representation is clearly separated from other facial changes such as poses through the pose code provided to the decoder and the pose estimation in the discriminator. Third, DR-GAN can take one or more images as input to generate a unified representation and any number of composite images. Quantitative and qualitative evaluations of controlled databases and wild databases show that DR-GAN is superior to the latest technology.

Yang et al, (2017) [3] proposed a Neural Aggregation Network (NAN) for video face recognition. The network takes a face video or a face image set (with a variable number of face images) as its input and generates a compact, fixed-size feature representation for recognition. The entire network consists of two modules. The feature embedding module is a deep CNN, which maps each face image to a feature vector. The aggregation module is composed of two attention blocks. The two attention blocks adaptively aggregate feature vectors to form a single feature in the convex hull that they span. Due to the attention mechanism, the aggregation is invariant to the image order. Their NAN uses standard classification or

validation loss for training without any additional supervision signals. They found that it automatically learns to support high-quality face images while rejecting low-quality images such as blur, occlusion, and incorrect exposure of the human face.

Ding et al, (2017) [4] present FaceNet2ExpNet, a novel idea to train an expression recognition network based on static images. They first proposed a new distribution function to simulate the high-level neurons expressing the network. On this basis, a two-stage training algorithm is designed. In the pre-training stage, they train the convolutional layer of the expression network, which is regularized by the face network; in the refinement stage, they attach the fully connected layer to the pre-trained convolutional layer and jointly train the entire network. The visualization results show that the model trained with their method captures improved high-level expression semantics. The evaluation of the four public expression databases of CK+, Oulu-CASIA, TFD and SFEW shows that their method achieves better results than other methods.

Ding et al, (2017) [5] proposed a comprehensive framework based on CNN to overcome the challenges of Video-based Face Recognition (VFR). First, to learn fuzzy robust face representations, researchers artificially blur the training data composed of clear still images to make up for the lack of real-world video training data. Using training data composed of still images and artificial blur data, CNN is encouraged to automatically learn blur-insensitive features. Secondly, to enhance the robustness of CNN features to pose changes and occlusions, they proposed a trunk branch integrated CNN model (TBE-CNN), which extracts supplements from the overall face image and the cropped patches around the facial parts information. TBE-CNN is an end-to-end model that efficiently extracts features by sharing low-level and mid-level convolutional layers between the backbone and branch networks. Third, to further improve the discrimination ability of the representation learned by TBE-CNN, they proposed an improved triplet loss function. The systematic experiment proves the effectiveness of the proposed technique. The most impressive thing is that TBE-CNN has achieved the most advanced performance on three popular video face databases (PaSC, COX face and YouTube Faces). Using the proposed technology, this research won first place in the BTAS 2016 video person recognition evaluation.

Nech et al, (2017) [6] stated some questions. If all algorithms for face recognition are very different; are large datasets the real key to more accurate facial recognition? Where does facial recognition need to improve? In a pursuit to find the answers to these sorts of questions, the researchers created a benchmark, MF2, which requires all algorithms to be trained on the same data and tested on a million scale. MF2 is a large public collection with 672K identities and 4.7 million photos, aiming to create a level playing field for large-scale face recognition. We compare our results with the results of two other large benchmarks, MegaFace Challenge and MS-Celebs-1M, in which groups can train on any private or public, and large or small

collection. Some key findings are (1) the algorithms trained on MF2 can reach the most advanced level, and the results are comparable to those trained on large-scale private sets; (2) some algorithms perform well after training on MF2; (3) aging invariance as with MegaFace, the accuracy is lower, and in future tests, it is determined that a greater age changes may be required in the adjustment of identity or algorithm.

Wang et al. (2018) [7] have stated that the traditional SoftMax loss of deep CNNs usually lacks the power of discrimination. Hence, they proposed a new loss function, the Large Margin Cosine Loss (LMCL), to realize this idea differently. They use L2 to normalize features and weight vectors to eliminate radial changes, re-express the maximum soft loss as a cosine loss, and introduce a cosine margin term to maximize the decision margin angular space further. As a result, the smallest intra-class variance and the largest inter-class variance are achieved through normalisation and maximisation of the cosine decision margin. They call this approach Cosface. In the mega face challenge, the CosFace approach shows its superiority for both the identification and verification tasks on both protocols.

Mittal et al. (2018) [8] designed a mobile application that performs real-time multiple face recognition using CNNs. Their model implemented a ResNet network. It tries to learn a mapping from face images to a compact 128-D Euclidean space where the correlation between two vectors represents the face similarity. This method optimizes the Inception model using Triplet Loss Function and curriculum learning. They used the Labeled Faces in the Wild dataset to train their images. With this training, they classified the faces extracting features afterwards. Finally, with the data augmentation, their system transforms various degrees on the photos, and finally, identification was achieved. They found that with the increasing number of members in the photos or videos, recognition precision dropped.

Cheng et al, (2018) [9] have proposed two face recognition methods based on multiple CNN classifiers. These methods have the same CNN structure but have different training sets. The first method used random sampling and replacement to train a single CNN using different training sets generated in the original data set. The second method generated different training sets by enlarging and reducing the image size of the original data set. These two methods finally used the voting method to merge a single CNN result. Experiments showed that the ensemble CNNs classifier is better than the single CNNs classifier, and face recognition accuracy is as 99.5%. Deng et al. (2019) [10] introduced Arcface that uses additive angular margin loss on Deep CNNs for facial recognition. Their experimentation showed that Arcface surpasses cosface, sphereface, centerloss, VGGFace and other opensource face recognition methods by a fair margin. They have also shown comprehensive comparisons with other types of losses in facial recognition, including each study's code.

2.2. DL on Speech Recognition

Rabiner (1997) [11] presented the current functions of speech recognition systems and how they are used in today's services and applications, and how they will evolve over time into the next generation of voice recognition services. he pointed out that there are different types of speech recognition, including spontaneous, conversational, hybrid recognition, etc. He also showed that the vocabulary size goes up to 64,000 words but comes with an error rate as high as 50%. He stated that although the main application of speech recognition would be telecommunication in the future, it could be extended to services like voice dialling, voice banking service, voice assistance, call completion agent technology, customer care computer to the phone integration voice dictation and so on.

Li et al, (2004) [12] have mentioned several challenges in speech recognition. Human and machine speech recognition has a big gap. If there is no underlying recognition technology that can provide sufficient robustness and low error performance, the mainstream adoption of speech recognition is impossible. One of these programs is a computational model based on pronunciation, the science of language, and some fundamental aspects of human linguistic communication. Based on the general principles of computational phonetics and phonology, a quantitative statistical model is established. Advanced algorithms can be developed to optimize the measurement parameters automatically. These parameters have physical significance. Another program is based on humans. All aspects of speech perception go contrary to the common acoustic speech information used in the spectral envelopes frame by frame in all the main speech recognition systems. This technique aims to develop a new processing front acoustic that does not depend on the spectral envelope and uses several fronts on the whole of the temporal and frequency plane. The researchers state that overcoming the challenge of making speech recognition systems robust in noisy acoustic environments and the challenge of creating workable recognition systems for natural, free-style speech is the ultimate challenge for the system. They also mentioned a few existing directives to these challenges and showed the architecture of Microsoft's Speech Recognition System.

Quoc et al. (2012) [13] proposed a robust speech recognition based on a binaural speech enhancement system as a pre-processing step. The system uses existing de-redundancy techniques, followed by a noise removal algorithm based on spatial masking, in which only signals from the desired direction are retained by using a threshold angle. Here, they urge consideration of adaptive calculation, where the threshold is first learned in a few noise frames and then updated frame by frame. Their robust speech recognition algorithm is based on binaural sound source separation and data loss. The simulation results show that because the target speech data obtained by the binaural sound source separation method can eliminate noise and interference, the algorithm can significantly

improve the speech recognition performance in a complex acoustic environment. The result of speech recognition in a real environment shows the effectiveness of this method. Miao et al. (2013) [14] studied the application of deep maxout network (DMN) in large vocabulary continuous speech recognition (LVCSR) tasks. They experimented on the challenging Babel corpus. They concluded that, compared with DNNs, DMNs could improve the performance of hybrid and BNF systems under limited LP conditions. Their conclusions further include that stacked denoising autoencoders (SDA)-based pre-training is effective for DMN initialization and brings benefits when DMN becomes truly deep, and DMN can be used as a sparse feature extractor to generate hierarchical high-level representations. They mentioned that in future work, they are interested in researching Restricted Boltzmann machines (RBM) for DMNs initialization, which requires a probabilistic pooling strategy to achieve a fully generative model, and they hope to extend the idea of sparse feature extraction to bottleneck features (BNF) and generate sparseness for tandem systems Bottleneck characteristics.

Based on the Gaussian mixture model and the improved N-best speech recognition algorithm, an improved lattice-based speech keyword recognition system is proposed by Xiao Xi et al, (2015) [15]. First, they used tests to evaluate different simplified structures of the Gaussian mixture model. Then, an N-best token passing algorithm is proposed based on the classic token passing algorithm and using the unique pronunciation rules of Chinese. These two modifications improve the performance of 1-best and N-best candidate speech recognition. Finally, a keyword recognition system based on the N-best lattice was developed to verify the effectiveness of these improvements.

Noda et al. (2015) [16] introduced a connectionist Hidden Markov Model (HMM) system for anti-noise AVSR. First, use a deep denoising autoencoder to obtain audio features with strong noise robustness. By preparing training data for a network with audio characteristics degraded in several consecutive pairwise steps and corresponding eigen characteristics, the network is formed to produce noise elimination characteristics from the corresponding noise degraded characteristics. Secondly, use a CNN to extract visual features from the original oral region image. By preparing the training data of CNN as a pair of original images and corresponding phoneme label output, the network is trained to predict phoneme labels from the corresponding oral region input image. Finally, a multi-stream hidden Markov model (MSHMM) is used to integrate independently trained audio and video HMMs with their respective features. By comparing the normal and mel-frequency cepstral coefficients (MFCCs) (MFCC) as HMM audio features, our single-peak isolated word recognition results show that under the 10 dB signal-to-noise ratio (SNR) of the audio signal input, the denoising MFCC can obtain a word recognition rate gain of about 65%. Attention-based recursive networks have been successfully applied to many tasks such as handwriting

synthesis, machine translation, image caption generation, and visual object classification. Researchers have introduced an extension of the attention-based recurrent network to make it suitable for speech recognition. Learning to recognize speech can be seen as learning to generate a sequence (transcription) given another sequence (speech).

Chorowski et al, (2015) [17] proposed and evaluated an end-to-end trainable speech recognition architecture based on a hybrid attention mechanism that combines content and location information to select the next location in the input sequence for decoding. An ideal feature of the model is that it can recognize utterances that are much longer than the training time. In conclusion, their work provides two new ideas for the attention mechanism: a better-standardized method to produce smoother alignments and a general principle for extracting and using features from previous alignments. They state that both methods may be applied to fields other than speech recognition. However, despite the great achievements made in the past few decades, natural and powerful human-machine voice interaction seems still out of reach, especially in challenging environments with significant noise and reverberation. To improve robustness, modern speech recognizers usually use an acoustic model based on a recurrent neural network (RNN), which can naturally take advantage of large time context and long-term speech modulation. Therefore, it is very meaningful to continue to study appropriate technologies to improve the effectiveness of RNN in processing speech signals.

Ravanelli et al. (2018) [18] modified one of the most popular RNN models, the gated recurrent unit (GRU), and proposed a simplified architecture that is very effective for ASR. The work has contributed on two fronts: First, they analyzed the role played by the reset gate and showed that there is significant redundancy with the update gate. They suggested removing the former from the GRU design, resulting in a more efficient and compact single-gate model. Then, they suggested replacing the hyperbolic tangent with a corrected linear unit activation. This change is well combined with batch normalization, which can help the model learn long-term dependencies without numerical problems. The results show that the proposed architecture light GRU reduces the epoch training time by more than 30% compared with the standard GRU and continuously improves the recognition accuracy in different tasks, input features, noise conditions, and different ASR paradigms from the standard DNN- HMM speech recognizer to the end-to-end connectionist temporal classification model.

2.3. DL in Image Recognition

Image recognition is a subclass or a derivation of computer vision and AI, representing a set of image detection and analysis methods to achieve automation of specific tasks - identify places, people, objects and many other types of items in a picture and draw conclusions through analysis [19-20].

Today, more and more people are using pictures to represent and convey information. It is convenient for us to learn much information from pictures. Image recognition is an important area of research that is widely used. For image recognition issues such as handwritten classification, we should know how to use the data to represent the image. The data here is not line pixels but image features with an advanced representation. The quality of the extraction of the characteristics is very important for the results [21-23].

Wu et al, (2015) [24] applied DL to handwritten character recognition and explored two common DL algorithms: the CNN and the deep belief network (DBN). CNN and DBN in the MNIST database and in the actual handwritten characters database. CNN and DBN in the MNIST database were 99.28% and 98.12 % accurate, respectively, and the accuracy of the classification accuracy in the database of the actual handwritten characters was 92.91% and 91.66%. Experimental results show that DL has a good ability to learn characteristics. There is no need to extract the characteristics manually; DL helps to understand the essential characteristics of the data better.

Elleuch et al, (2016) proposes a method that constitutes two classifiers – CNN and SVM. They use it to identify Arabic characters from handwritten text. CNN and dropout regularization are not only used by this paper but also in other projects such as [25-29].

Pham et al. [30] used RNN with Dropout as one of the first projects to detect handwritten text and tested with a range of handwritten databases. Sun et al, [31] used Deep LSTM to detect Chinese handwriting. Their methods have been evaluated on the CASIA-OLHWDB set of data. Compared to the most advanced method, the accuracy and precision of the test set is reduced by more than 30 % over relative error [32].

Yang et al, (2016) [33] recommend combining Neural Network Classifier with Style Transfer Mapping (STM) for unsupervised author adaptation, which only requires author-specific unmarked data and is more common and efficient than supervised adaptation. To increase the performance of neural network classifiers, they used techniques such as Dropout, reread, momentum and Deep Monitoring Strategies. Experiments on CASIA-OLHWDB, a Chinese database writing online, show that this method effectively improves classification accuracy.

Islam et al (2018) [34] presents an image-based food recognition method. They used CNNs to classify food images. The project used 16,643 images to classify a food dataset containing different food categories with 92.86% accuracy.

2.4. Virtual Assistants

Virtual Assistant, also known as AI assistant or digital assistant, is an application that understands natural language voice instructions and performs tasks for the user. These tasks are usually performed by a personal assistant or secretary and include dictating, reading texts or emails

aloud, finding phone numbers, scheduling, making phone calls, and reminding appointments end-user. Currently, popular virtual assistants include Amazon Alexa, Apple's Siri, Google Assistant, and Microsoft's Cortana digital assistant, all built into Windows Phone 8.1 and Windows 10 [35-36]. Although the definition focuses on the digital form of virtual assistants, the term virtual assistant or virtual personal assistant is also often used to describe contract workers who work from home, who usually perform administrative tasks by administrative assistants or secretaries [37-39]. Virtual assistants can also be compared with another type of consumer-oriented AI programming (called smart advisors). Smart advisors are subject-oriented, while virtual assistants are task-oriented. Virtual assistants are typically cloud-based apps to function that is nowadays integrated into almost all devices by default. Three of these apps are Siri on Apple, Cortana on Microsoft, and Google Assistant on Android. There are also devices dedicated to virtual assistance. Amazon's Alexa, Google Assistant and Microsoft's Cortana, and Apple's Siri on Apple Home are the most popular. Usually, these devices have a call signal like for Google Assistant; it's "Hey Google or Okay Google", or for Cortana, it's "Hey Cortana". The LEDs on the device indicate to the user that they are ready to receive orders. Often these are simple language requests such as "Tell me a joke," or "play some music", or "buy a shirt". These requests will be addressed and stored in the Amazon cloud. The technology that powers virtual assistants requires large amounts of data from artificial power intelligence (AI) platforms, including ML, DL and AI, natural language processing, and speech recognition platforms. When the user interacts with the virtual assistant, AI programming uses complex algorithms to learn from data input and better predict the end user's needs. Virtual assistants usually perform simple tasks for end-users, such as adding tasks to the calendar; providing information that is usually searched in a web browser; or controlling and checking the status of smart home devices, including lights, Air conditioners, ovens etc.

Users also require virtual assistants to make and receive calls, create text messages, get directions, listen to the news, and weather forecasts, find hotels or restaurants, check flight reservations, listen to music, or play games.

Page et al, (2017) [40] showed how an Artificially Intelligent Virtual Assistant Helps Students Navigate the Road to College. This paper shows a pure application of an AI-powered VA to reduce workload in real life. The paper states that they use Conversational AI to provide personalized, SMS-based outreach and guidance to thousands of college freshmen to help them complete each task that requires support, thereby effectively supporting them. The system which was tested through field trials at Georgia State University (GSU) was called Pounce. GSU promised that students receiving treatment via the Virtual Assistant showed a greater success rate in terms of pre-entry requirements. And at the same period of time, the pressure on university teachers is greatly reduced.

Kepuska et al, (2018) [41] have designed a prototype system that showcases the future of current Virtual

Personal Assistants (VPAs). Their application used a multi-modal dialogue system that processes two or more combined user input modes, such as voice, image, video, touch, manual gestures, gaze, and body movement, to design the next generation of VPAs model. The new model of VPAs will be used to increase human-machine interaction by using different technologies, such as gesture recognition, image and video recognition, voice recognition, many knowledge-based on conversations and dialogues and a common knowledge base. In addition, the new VPA system can be used in various other application areas, including education aid, medical assistance, robots and vehicles, systems for people with disabilities, home automation and secure access control.

Giancarlo Iannizzotto et al. (2018) [42] combine some of the most advanced technologies in the fields of computer vision, DL, speech generation and recognition, and AI into the virtual assistant architecture of the smart home automation system. The proposed assistant is efficient and customizable, and the implemented prototype runs on a low-cost, small-sized Raspberry PI 3 device. The system was integrated with an open-source home automation environment for testing purposes and ran for a few days while encouraging people to interact with it and proving that it was accurate, reliable, and attractive.

Dipanshu et al, (2020) [43] designed a system that can successfully use the integrated webcam to capture gestures, process them, convert them into text format, display them on the input frame, and convert them into text then into audio format when receiving a call command. The audio becomes the query of the virtual assistant, and the audio output is successfully converted to text format and displayed on the screen again. This system is built to showcase a gesture-based virtual assistant that is targeted towards people with speaking or hearing disabilities. They have used DL and TensorFlow along with OpenCV to accomplish their goal. Vishnu et al, (2021) [44] proposed a CNN-based solution that utilizes transfer learning for developing a scene recognition Android application. They have used MobileNet to implement the CNN model with the TensorFlow ML framework. To reduce the inference time, they used hardware acceleration using GPU is used in the application. They trained the system using images with occlusion, different illumination and background clutter for improving the robustness. In their system, they trained the system with the MIT Indoor Recognition dataset and then used it for testing. The main goal is to showcase how DL can be used to expedite current-generation virtual assistants' advancement further and even expand its territory into the Robotic field with AI.

Carnier et al, (2021) [45] ToraxIA is a Virtual Assistant for Radiologists Based on DL from Chest X-Ray. The system was trained with over 240000 images of chest x-rays. And the system can detect several pathological signs from them with an accuracy of about 97%. Even it was used to detect COVID-19 signs by training the system to differentiate between normal x-ray images and images from the COVID-19 dataset. Ultimately, this paper presents a virtual assistant made with DL capabilities for automatic detection

of pathological signs suggestive of respiratory disease, reducing the diagnostic error, increasing efficiency, and quality of care in the medical field, reducing efforts of radiologists.

2.5. Chatbots

Chatbots are programs that use natural languages to interact with users. This technology has been gaining popularity since the 1960s [46]. Chatbots had been in use by almost every field of business, e-commerce, organization out there. Prominent uses include as a Tool to Learn and Practice a Language, as a tool to retrieve information, as a tool to conduct business, as a tool to provide customer support etc. Where there is no voice supported virtual assistant or where users cannot use voice-based VA, Chatbots are the way to go. One recent example is Microsoft's Bing Chatbox on their search engine [47], which uses AI. And chatbots have been on the rise since 2016 at an unprecedented speed with the introduction of Microsoft's Cortana, Amazon's Alexa etc [48].

Chatbots made with traditional methods such as conditional programming are not as efficient as using AI and DL. The current-day DL method has made training models for Natural Language Processing (NLP) far easier than before, which is exactly why chatbots and similar fields like VA can take advantage of it [49].

ELIZA is one of the first Chatbots to be built as a program to pass the Turing test [50] but it was simple and totally built on a rule-based framework. Other chatbots such as A.L.I.C.E. and PARRY was also built using the same technology [36 and 49]. Jabberwocky was one of the first chatbots to be built with AI to conduct hilarious conversations [51]. Hybrid chatbots with Recurrent Neural networks (RNN) and Long/Short-Term Memory (LSTM) prove more efficient in many cases [49].

Wu et al, (2019) [52] state the different ways chatbots are built in their presentation. They show how retrieval-based chatbots are built with single turn response selection and multi-turn response selection. They infer various conclusions based on comparing the different ways DL is used in building chatbots. For example, in the case of single turn response selection, the neural tensor is a powerful matching function and that combining information from different sources can be very beneficial for matching sequence etc.

Research has been on a high horse since 2016 due to the sudden uprising for using human interaction programs, and various methods and frameworks have been in development since then to make the job even smoother. For example, some notable mentions are the DL to Respond (D2LR) model [53] with MLP matching, Coupled LSTM [54] with MAP matching etc. Conversational AI for Chit-Chat Programs or chatbots is catching on fire as it is more and more involved in mass business domains because big data technology is also on the rise [55].

2.6. DL in Healthcare

Health care is the care or improvement of health through the prevention, diagnosis, treatment, recovery, treatment of illnesses, injuries and other physical and mental injuries. Definition of anger from the book that everyone knows because it concerns our well-being. In this area, the convergence of technologies makes it easier to achieve this goal for people from all walks of life. Implications of DL in healthcare has been increasing in the past 6 years [56-58]. This has mostly happened because of the increase in Big Data analytics and research, IoT advances, and ML advancements [59 and 60]. CNNs can be trained for medical imaging thanks to the massive advancements of Computer Vision with DL. CNNs and RNNs both can be used to predict health problems [56]. DNNs can be used in genomics for research in genetics, gene splicing, generating new gene sequences etc [61]. The following table 1 shows the various applications of DL in Health care.

Table 1. A brief list of DL Applications in Healthcare [62]

Type	Application	Model	Ref.
Medical imaging	Early diagnosis of Alzheimer's disease by brain magnetic resonance imaging	Stacked Sparse AE	[63]
	Brain magnetic resonance imaging detects the diversity of Alzheimer's disease variant patterns	RBM	[64]
	Automatic segmentation of knee cartilage MRI to predict the risk of osteoarthritis	CNN	[65]
	Segmentation of multiple sclerosis lesions in multi-channel three-dimensional magnetic resonance imaging	RBM	[66]
	Ultrasonic diagnosis of breast nodules and lesions	Stacked Denoising AE	[67]
	Finding of diabetic retinopathy in retinal fundus photos	CNN	[68]
	Dermatologist skin cancer classification	CNN	[69]
	Longitudinal EHR predicts congestive heart failure and chronic obstructive pulmonary disease	CNN	[70]
	Diagnostic classification of patients in pediatric intensive care unit with clinical measurement	LSTM RNN	[71]
	DeepCare: a predictive medical dynamic memory model based on the patient's medical history	LSTM RNN	[72]
Health Record Data Processing	Deep Patient: unsupervised patient representatives who can be used to predict future clinical events	Stacked Denoising AE	[73]
	Predict future diseases from the patient's clinical status	Stacked Denoising AE	[74]
	Automatically assign diagnoses to patients based on the patient's clinical status	RBM	[75]
	Predict the suicide risk of mental health patients through the low-dimensional representation of EHR Chinese medical concepts	RBM	[76]
	Discovery and detection of characteristic physiological patterns in clinical time series	Stacked AE	[77]

Genetics	Longitudinal sequence of model serum uric acid measurement to indicate multiple population subtypes and distinguish the characteristics of uric acid from gout and acute leukemia	Stacked AE	[78]
	Doctor AI : Use the patient's medical history to predict the diagnosis and medication for follow-up visits	GRU RNN	[79]
	Deeper: An end-to-end system for predicting accidental readmissions after discharge	CNN	[80]
	Predicting disease onset from longitudinal laboratory tests	LSTM RNN	[81]
	De-Identification the identification of patient clinical records	LSTM RNN	[82]
	Predicting chromatin markers from DNA sequences	CNN	[83]
	Basset: an open-source platform for predicting DNase I hypersensitivity in multiple cell types and quantifying the effect of SNV on chromatin accessibility	CNN	[84]
	DeepBind: Predict the specificity of DNA and RNA binding proteins	CNN	[85]
	Predicting methylation status in single-cell bisulfite sequencing studies	CNN	[86]
	Estimation of the prevalence of different chromatin markers	CNN	[87]
Mobile Healthcare	Classify cancer from gene expression profile	Stacked Sparse AE	[88]
	Predict protein backbone from protein sequence	Stacked Sparse AE	[89]
	HAR detects PD patients' gait freezing	CNN/RN N	[90]
	Estimating EE using wearable sensors	CNN	[91]
	Photoplethysmography signal recognition for health monitoring	RBM	[92]
	EEG and local field potential signal analysis	CNN	[93]
	Predict sleep quality from wearable data on physical activity while awake	CNN	[94]

2.8. DL in Entertainment

Content consumption behaviour is becoming more complex and growing faster than ever. Data science, ML and AI in media and entertainment offer huge possibilities for ready companies to invest in viable long-term solutions. Companies such as Google, Amazon, and Netflix have influenced the media and entertainment sectors with their data-driven business models. Advanced data science and AI can understand customers and create value for customers by showing content that customers may like and may be relevant to them. But this change does not happen overnight; only by gradually developing into a truly data-powered organization can its full potential be realized [95]. Chou et al. (2020) [96] have designed a four-dimensional (4D) audio-visual entertainment system called Sense. The

system includes a scene recognition system and hardware modules. When users watch movies and animations at home, these modules provide users with tactile sensations. Their recognition module uses Google Cloud Vision to detect common scene elements in the video, such as fire, explosion, wind, and rain. After the system considers the target detection result through the scene recognition system, the system generates the corresponding tactile or haptic sensation. Their system seamlessly integrates DL, auditory signals, and touch to provide an enhanced viewing experience. This is one of the more recent research as to show how DL can be used in media consumption. But DL has been in use for a longer period in digital games. Digital game-based learning or DGBL is an old learning method to learn about player goals and playing patterns. This is used to analyze the in-game approaches of tutorial bots and feedback systems [97]. Popular games like ATARI, Racing, DOOM, Minecraft, and StarCraft all use DL, albeit not a perfect implementation, and can cause problems such as unwanted NPC behaviour [98], enhance user experience and make the game easier and suits it to the consumer. Prediction is the future of DL [99]. Player data is used to predict future game contents and delivered as DLC, which is the current business model for the game industry.

Gudmundsson et al. (2018) [100] showed that it is possible to predict the next most “human” action, and from that, they can update future contents of the game. They used Candy Crush Saga (CCS) and Candy Crush Saga (CCSS) player data to learn human playtesting methods and generate key metrics for future content creation. Another major advancement in the entertainment industry with the help of DL is Deepfake. It is a technology where the algorithm learns the facial characteristic of a person, and those facial characteristics are copied to a video or image of another individual. In easier words, it's refacing someone. Deepfake movies can make the movie setting more efficient, and we can watch movies without subtitles in our native language [101]. In the music and audio sector, DL is being used. Generating realistic original music is time-consuming and requires energy. Technology has made that easier. DL and midi files of music can learn music genre and generate real music [102]. Online music streaming platforms like Spotify, YouTube music, apple music rely on AI and DL for the music suggestion to their users. Still, novel ways to do the same task arises like T-RECSYS [103]. Services like Google, shazam used DL to extract music information from the audio [104].

2.9. DL in Robotics

In the past 5 years, DL technology has completely changed many aspects of computer vision and has been rapidly adopted by robotics [105-110].

Punjani et al. (2015) [111] in their research, shows that the performance of DNN in obtaining helicopter dynamics of the remote-controlled helicopter is about 60% higher than that of other methods. DL is now being actively used in

industrial robots for purposes like object identification and picking [112].

Le et al. (2019) [113] showed this field using Kinect and Denso robot. They teach the robot to identify and pick up USB packs using both offline and online methods. In the era of online purchase and e-commerce, where goods are delivered right to the doorstep, object picking or bin picking is very important; Amazon conducted the first competition of this exact challenge using DL in 2015 [114]. Outside of concentrated fields like industrial robotics, there are works to other fields like civil engineering and architecture that have been made with DL. McLaughlin et al. (2020) [115] used CNN and LIDAR to collect and analyze 3D spatial data, which a robot used to automatically quantify defects in concrete bridges. The researcher also used similar structures, i.e. mobile net v2 and deep lab v3, to automatically assess concrete delamination [116-120].

3. Conclusions

DL is raging today, as major breakthroughs in artificial neural networks have prompted companies in all industries to implement DL solutions as part of their strategy over the past few years. From chatbots in client service to image and object recognition in retail and more, DL has opened up many new and complex AI applications. In recent years, DL has been particularly appealing to many organizations as more and more publicly accessible data pre-training models have been developed. However, this does not mean that DL answers all questions related to ML. One of the chief advantages of DL is its ability to resolve complex problems that need discovering hidden patterns in the data and a deep understanding of the complex relationships between a large number of interrelated variables. DL algorithms can learn patterns on their own hidden from the data, combine them and build more efficient decision rules. DL is excellent for complex tasks that often require processing large amounts of unstructured data, such as image classification, natural language processing, speech recognition, etc. However, classical ML may be better for simple tasks involving more direct features engineering that do not require unstructured data processing.

In this paper, several fields have been described that is shining in DL – image recognition, object detection, healthcare, education, entertainment, music, robotics etc. in whole DL consist of four main network named CNNs, recurrent neural networks (RNNS), generative adversarial networks (GAN) and reinforcement learning (RL). The CNNs are mainly used in image and video applications like face recognition and object detection, RNN are applicable for text classifications and commercials like Exchange and image captions. In Computer Vision and Natural Language Processing (NLP), The Gans are mostly used for non-real images like deep fake videos, text to image conversion, and photo editing. And the last on RL is used as the basis of computational neuroscience to model decision-making

processes like self-driving cars, traffic light control, robotics. A summary of each application has been described in the following figure (1).

As the figure 1 shows, the most popular DL technique is CNN. Every field's research begins from CNN. Then other techniques are incorporated on top of it. The second most used technique is RNN or Recurrent Neural Network. CNN and RNN seem to go hand in hand in all fields. The expansion of DL has been steadily spreading over the world. Steady but not slow. As seen above how AI and DL has been touching all aspects of normal life. From education to business to industry to entertainment – very soon, all fields will be using ML to complete their fundamental tasks. And it looks like a blessing. The main advantages of DL are as follows:

Features are automatically derived and optimized to achieve desired results. Features do not need to be extracted in advance. This avoids time-consuming ML techniques.

The robustness of ML to natural variations in data.

The same neural network-based approach can be applied to many applications and different types of data.

The GPU can be used to perform large-scale parallel calculations and is expandable for large amounts of data, and when the data is large, it can provide better performance results.

The DL architecture is flexible and can adapt to new future issues.

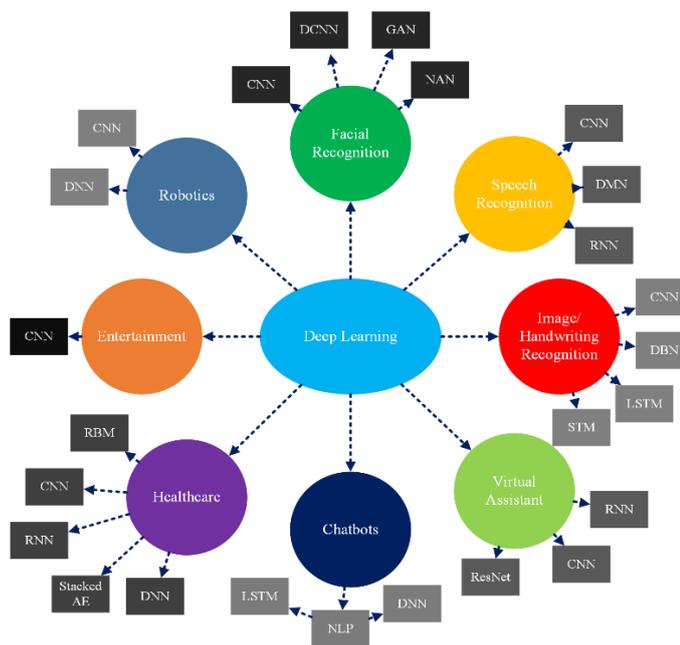


Figure 1. DL Techniques used in different fields

But nothing is without drawbacks. Some of the downsides of DL are as follows:

It takes a lot of data to be better than other technologies.

The cost of training a model is very high due to complex data models. Besides, DL requires expensive GPUs and hundreds of machines, which increases the cost for users.

There is no standard theory to guide you in choosing the appropriate DL tool, as it requires knowledge of topology, training methods and other parameters, so it is not easy to adopt it by less qualified people. It is not easy to understand outputs based on learning, and a classifier must understand. The CNN-based algorithm performs these tasks. DL is currently known as the hot topic in the field of ML and social cognition. This article discusses face recognition, speech recognition, Chinese character recognition, neural networks, and chat robots. We also discussed the future impact of these applications on society. In DL, the learning of humans and animals is unsupervised: we tend to discover the earth's structure through perception rather than by being told the name of each object. Human vision is a powerful method that continuously collects optical arrays in an intelligent, task-specific manner, using tiny, high-resolution fovea and super-large, low-resolution surrounds. We tend to expect that, in the long run, many advances in vision will be restored from systems that take square measurements from the end of training to the end and mix Conv networks and RNNs. These RNNs use reinforcement learning to let you Make up your mind anywhere. Systems that combine DL and reinforcement learning begin to measure in their infancy. Tongue understanding is another space during which DL will have a huge impact in the next few years. Ultimately, major advances in computer science can be achieved through systems that combine illustration learning with advanced reasoning. However, language and handwriting have adopted in-depth study and straightforward reasoning. DL is extensively used in the field of medical image recognition. Face recognition, recognition of character and many things. It trains a model on a given set of information to complete the specific task of new information, but ancient medical image recognition is inefficient in feature extraction and data processing, and its popularization effect is not ideal. These application domain units are changing with each passing day, and a large number of the latest skill domain units have also joined these technologies.

Acknowledgements.

This work was supported by Jiangxi University of Science and Technology, 341000, Ganzhou, P.R China, [grant number 205200100460].

References

- [1] Liu W, Wen Y, Yu Z, Li M, B Raj, Song L, Sphreface: Deep hypersphere embedding for face recognition, in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, 212–220.
- [2] Tran L, Yin X, Liu X, Disentangled representation learning gan for pose-invariant face recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017; 1415–1424.
- [3] Yang J, et al., Neural aggregation network for video face recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017; 4362–4371.

- [4] Ding H, Zhou SK, Chellappa R, Facenet2expnet: Regularizing a deep face recognition net for expression recognition. in 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition. 2017; 118–126.
- [5] Ding C, Tao D, Trunk-branch ensemble convolutional neural networks for video-based face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., 2017; 40(4):1002–1014.
- [6] A Nech, I Kemelmacher-Shlizerman, Level playing field for million scale face recognition. in Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua:3406–3415, DOI: 10.1109/CVPR.2017.363.
- [7] Wang H, et al., Cosface: Large margin cosine loss for deep face recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition, 2018, 5265–5274.
- [8] Mittal S, Agarwal S, Nigam MJ, Real time multiple face recognition: A deep learning approach. in ACM International Conference Proceeding Series, 2018, 70–76, DOI: 10.1145/3299852.3299853.
- [9] Cheng WC, Wu TY, Li DW, Ensemble convolutional neural networks for face recognition. ACM Int. Conf. Proceeding Ser. 2018; 40(4):1002–1014. DOI: 10.1145/3302425.3302459.
- [10] Deng J, Guo J, Xue N, Zafeiriou S, Arcface: Additive angular margin loss for deep face recognition. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, 4690–4699.
- [11] Rabiner LR, Applications of speech recognition in the area of telecommunications. in 1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings, 1997, 501–510.
- [12] Deng L, Huang X, Challenges in adopting speech recognition. Commun. ACM, 2004; 47(1) 69–75.
- [13] Quoc CN, Tien DT, Dang KN, Huu BN, Robust speech recognition based on binaural speech enhancement system as a preprocessing step. ACM Int. Conf. Proceeding Ser., 2012; 91–96. DOI: 10.1145/2350716.2350732.
- [14] Miao Y, Metze F, Rawat S, Deep maxout networks for low-resource speech recognition. in 2013 IEEE Workshop on Automatic Speech Recognition and Understanding, 2013, 398–403.
- [15] Xi X, Jingqian W, Improved lattice-based speech keyword spotting algorithm. J. Tsinghua Univ. Science Technol. 2015; 55(5):508–513.
- [16] Noda K, Yamaguchi Y, Nakadai K, Okuno HG, Ogata T, Audio-visual speech recognition using deep learning. Appl. Intell. 2015; 42(4):722–737.
- [17] Chorowski J, Bahdanau D, Serdyuk D, Cho K, Bengio Y, Attention-based models for speech recognition arXiv Prepr. arXiv1506.07503, 2015.
- [18] Ravanelli M, Brakel P, Omologo M, Bengio Y, Light gated recurrent units for speech recognition. IEEE Trans. Emerg. Top. Comput. Intell. 2018; 2(2):92–102.
- [19] Image Recognition : A Complete Guide,” Deepomatic. Jan. 2019, Accessed: Sep. 25, 2021. <https://deepomatic.com/en/what-is-image-recognition>.
- [20] Torkian, N, et al, Synthesis and characterization of Ag-ion-exchanged zeolite/TiO₂ nanocomposites for antibacterial applications and photocatalytic degradation of antibiotics. Environmental Research, 2022; 207:112157
- [21] Pak M, Kim S, A review of deep learning in image recognition. in 2017 4th International Conference on Computer Applications and Information Processing Technology (CAIPT), 2017; 1–3, DOI: 10.1109/CAIPT.2017.8320684.
- [22] Khan AT, Cao X, Li Z, and Li S, Evolutionary Computation Based Real-time Robot Arm Path-planning Using Beetle Antennae Search. EAI Endorsed Trans AI Robotics, 2022; 1:1–10.
- [23] Chen Z, Walters J, Xiao G, and Li S, An Enhanced GRU Model with Application to Manipulator Trajectory Tracking”, EAI Endorsed Trans AI Robotics, 2022; 1:1–11.
- [24] Wu M, and Chen L, Image recognition based on deep learning. in 2015 Chinese Automation Congress (CAC), 2015, pp. 542–546, DOI: 10.1109/CAC.2015.7382560.
- [25] Miao Y, Metze F, Improving low-resource CD-DNN-HMM using dropout and multilingual DNN training. in Interspeech, 2013; 13:2237–2241.
- [26] Shen J, Shafiq MO, Deep Learning Convolutional Neural Networks with Dropout - A Parallel Approach,” in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), 2018; 572–577, DOI: 10.1109/ICMLA.2018.00092.
- [27] Maalej R, Tagougui N, Kherallah M, Online Arabic Handwriting Recognition with Dropout Applied in Deep Recurrent Neural Networks. in 2016 12th IAPR Workshop on Document Analysis Systems (DAS), 2016; 417–421, DOI: 10.1109/DAS.2016.49.
- [28] Mansoorianfar, M, et al., Amorphous/crystalline phase control of nanotubular TiO₂ membranes via pressure-engineered anodizing. Materials & Design, 2021; 198:109314.
- [29] Mansoorianfar, M, et al., MXene-laden bacteriophage: A new antibacterial candidate to control bacterial contamination in water. Chemosphere, 2022; 290:133383.
- [30] Pham V, Bluche T, Kermorvant C, Louradour J, Dropout Improves Recurrent Neural Networks for Handwriting Recognition. in 2014 14th International Conference on Frontiers in Handwriting Recognition, 2014; 285–290, DOI: 10.1109/ICFHR.2014.55.
- [31] Sun L, Su T, Liu C, Wang R, Deep LSTM Networks for Online Chinese Handwriting Recognition. in 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR). 2016; 271–276, DOI: 10.1109/ICFHR.2016.0059.
- [32] Elleuch M, Maalej R, Kherallah M, A New Design Based-SVM of the CNN Classifier Architecture with Dropout for Offline Arabic Handwritten Recognition. Procedia Comput. Sci., 2016; 80:1712–1723. DOI: <https://doi.org/10.1016/j.procs.2016.05.512>.
- [33] Yang HM, Zhang XY, Yin F, Luo Z, Liu CL, Unsupervised Adaptation of Neural Networks for Chinese Handwriting Recognition. in 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), 2016, pp. 512–517, DOI: 10.1109/ICFHR.2016.0100.
- [34] Islam MT, Karim Siddique BMN, Rahman S, Jabid T, “Image Recognition with Deep Learning,” in 2018 International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS), 2018; 3:106–110, doi: 10.1109/ICIIBMS.2018.8550021.
- [35] Botelho B, virtual assistant (AI assistant). techtarget, 2017. <https://searchcustomerexperience.techtarget.com/definition/virtual-assistant-AI-assistant> (accessed Aug. 21, 2021).
- [36] Akkara S, and T J, PI Controller Based Switching Reluctance Motor Drives using Smart Bacterial Foraging Algorithm. EAI Endorsed Trans AI Robotics, 2022; 1:1–8.
- [37] Duermyer R, Defenition and examples of virtual assistant: The balance small business. 2021. <https://www.thebalancesmb.com/virtual-assistant-1794441> (accessed Aug. 21, 2021).

- [38] Mourtas S, Katsikis V, and Kasimis C, Feedback Control Systems Stabilization Using a Bio-inspired Neural Network. *EAI Endorsed Trans AI Robotics*, 2022; 1;1–13.
- [39] Nourouzi S, Kolahdooz A, Bakhshi-Juybari M, Hosseinipour SJ, Effect of temperature on the Microstructure of semi-solid casting in cooling slope method. *Aerospace Mechanics Journal*. 2013; 9(3): 55–63.
- [40] Page LC, Gehlbach H, How an Artificially Intelligent Virtual Assistant Helps Students Navigate the Road to College,” *AERA Open*. 2017; 3(4):2332858417749220. Oct., doi: 10.1177/2332858417749220.
- [41] Kepuska V, Bohouta G, Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home),” in 2018 IEEE 8th annual computing and communication workshop and conference (CCWC), 2018; 99–103.
- [42] Iannizzotto G, Lo Bello L, Nucita A, Grasso GM, A vision and speech enabled, customizable, virtual assistant for smart environments. in 2018 11th International Conference on Human System Interaction (HSI), 2018; 50–56.
- [43] Someshwar D, Bhanushali D, Chaudhari V, Nadkarni S, Implementation of Virtual Assistant with Sign Language using Deep Learning and TensorFlow. in 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA), 2020; 595–600.
- [44] Vishnu R, Krishna Prakash N, Mobile Application-Based Virtual Assistant Using Deep Learning. in *Soft Computing and Signal Processing*, 2021; 609–617.
- [45] Carnier M, Albertti R, Gavidia L, Severeyn E, A La Cruz, ToraxIA: Virtual Assistant for Radiologists Based on Deep Learning from Chest X-Ray. in *Artificial Intelligence, Computer and Software Engineering Advances*, 2021; 49–63.
- [46] Shawar BA, Atwell E, Chatbots: are they really useful?. in *Ldv forum*, 2007, vol. 22, no. 1, pp. 29–49.
- [47] Parmar M, Microsoft Bing search is getting its own AI-powered assistant. <https://www.windowlatest.com/>, 2021. <https://www.windowlatest.com/2021/05/31/microsoft-bing-search-is-getting-its-own-ai-powered-assistant/> (accessed Aug. 27, 2021).
- [48] DALE R, The return of the chatbots,” *Nat. Lang. Eng.*, vol. 2016; 22(5):811–817. DOI: DOI: 10.1017/S1351324916000243.
- [49] Bhagwat VA, *Deep Learning for Chatbots*. 2018.
- [50] Weizenbaum J, ELIZA — a Computer Program for the Study of Natural Language Communication between Man and Machine. *Commun. ACM*, 1983; 26(1):23–28, DOI: 10.1145/357980.357991.
- [51] Benoit R, *Making a Clever Intelligent Agent: The Theory behind the Implementation*, vol. 3. 2009.
- [52] Wu W, Yan R, Deep chit-chat: Deep learning for chatbots. in *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2019, 1413–1414.
- [53] Yan R, Song Y, Wu H, Learning to respond with deep neural networks for retrieval-based human-computer conversation system. in *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, 2016, pp. 55–64.
- [54] Yan R, Zhao D, Coupled context modeling for deep chit-chat: towards conversations between human and computer. in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018; 2574–2583.
- [55] Yan R, Chitty-Chitty-Chat Bot’: Deep Learning for Conversational AI. in *IJCAI*, 2018; 18:5520–5526.
- [56] Esteva A, et al., A guide to deep learning in healthcare. *Nat. Med.* 2019; 25(1):24–29. DOI: 10.1038/s41591-018-0316-z.
- [57] Hojjati-Najafabadi, A, et al., Antibacterial and photocatalytic behaviour of green synthesis of Zn_{0.95}Ag_{0.05}O nanoparticles using herbal medicine extract. *Ceramics International*, 2021; 47(22):31617–31624.
- [58] Hojjati-Najafabadi, A, et al, A Tramadol Drug Electrochemical Sensor Amplified by Biosynthesized Au Nanoparticle Using Mentha aquatic Extract and Ionic Liquid. *Topics in Catalysis*, 2021.
- [59] Ardabili S, Mosavi A, Várkonyi-Kóczy AR, Advances in machine learning modeling reviewing hybrid and ensemble methods. in *International Conference on Global Research and Education*, 2019; 215–227.
- [60] Martis RJ, Gurupur VP, Lin H, Islam A, Fernandes SL, Recent advances in big data analytics, internet of things and machine learning. Elsevier, 2018.
- [61] Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A, A primer on deep learning in genomics. *Nat. Genet.* 2019; 51(1):12–18, DOI: 10.1038/s41588-018-0295-5.
- [62] Miotto R, Wang F, Wang S, Jiang X, Dudley JT, Deep learning for healthcare: review, opportunities and challenges,” *Brief. Bioinform.* 2018; 19(6):1236–1246. DOI: 10.1093/bib/bbx044.
- [63] Liu S, Liu S, Cai W, Pujol S, Kikinis R, Feng D, Early diagnosis of Alzheimer’s disease with deep learning. in 2014 IEEE 11th international symposium on biomedical imaging (ISBI), 2014; 1015–1018.
- [64] Brosch T, Tam R, Initiative ADN, Manifold learning of brain MRIs by deep learning. in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2013; 633–640.
- [65] Prasoon A, Petersen K, Igel C, Lauze F, Dam E, Nielsen M, Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network. in *International conference on medical image computing and computer-assisted intervention*, 2013; 246–253.
- [66] Yoo Y, Brosch T, Traboulsee A, Li DKB, Tam R, Deep learning of image features from unlabeled data for multiple sclerosis lesion segmentation in *International workshop on machine learning in medical imaging*, 2014; 117–124.
- [67] Cheng JZ, et al., Computer-aided diagnosis with deep learning architecture: applications to breast lesions in us images and pulmonary nodules in CT scans. *Sci Rep* 6: 24454. 2016.
- [68] Gulshan V, et al., Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*. 2016; 316(22):2402–2410.
- [69] Esteva A, et al., Dermatologist-level classification of skin cancer with deep neural networks. *Nature*. 2017; 542(7639):115–118.
- [70] Cheng Y, Wang F, Zhang P, Hu J, Risk prediction with electronic health records: A deep learning approach. in *Proceedings of the 2016 SIAM International Conference on Data Mining*, 2016; 432–440.
- [71] Lipton ZC, Kale DC, Elkan C, Wetzel R, Learning to diagnose with LSTM recurrent neural networks. *arXiv Prepr. arXiv1511.03677*, 2015.
- [72] Pham T, Tran T, Phung D, Venkatesh S, Deepcare: A deep dynamic memory model for predictive medicine. in *Pacific-Asia conference on knowledge discovery and data mining*, 2016; 30–41.

- [73] Miotto R, Li L, Kidd BA, Dudley JT, Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Sci. Rep.* 2016; 6(1):1–10.
- [74] Miotto R, Li L, Dudley JT, Deep learning to predict patient future diseases from the electronic health records. in *European conference on information retrieval*, 2016; 768–774.
- [75] Liang Z, Zhang G, Huang JX, Hu QV, Deep learning for healthcare decision making with EMRs. in *2014 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2014; 556–559.
- [76] Tran T, Nguyen TD, Phung D, Venkatesh S, Learning vector representation of medical objects via EMR-driven nonnegative restricted Boltzmann machines (eNRBM). *J. Biomed. Inform.*, 2015; 54:96–105.
- [77] Che Z, Kale D, Li W, Bahadori MT, Liu Y, Deep computational phenotyping. in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015; 507–516.
- [78] Lasko TA, Denny JC, Levy MA, Computational phenotype discovery using unsupervised feature learning over noisy, sparse, and irregular clinical data,” *PLoS One.* 2013; 8(6):66341.
- [79] Choi E, Bahadori MT, Schuetz A, Stewart WF, J Sun, Doctor ai: Predicting clinical events via recurrent neural networks. in *Machine learning for healthcare conference*, 2016; 301–318.
- [80] Nguyen P, Tran T, Wickramasinghe N, Venkatesh S, DeepR: a convolutional net for medical records. *IEEE J. Biomed. Heal. informatics.* 2016; 21(1):22–30.
- [81] Razavian N, Marcus J, Sontag D, Multi-task prediction of disease onsets from longitudinal laboratory tests. in *Machine learning for healthcare conference*, 2016; 73–100.
- [82] Dernoncourt F, Lee JY, Uzuner O, Szolovits P, De-identification of patient notes with recurrent neural networks. *J. Am. Med. Informatics Assoc.* 2017; 24(3):596–606.
- [83] Zhou J, Troyanskaya OG, Predicting effects of noncoding variants with deep learning–based sequence model. *Nat. Methods*, 2015; 12(10):931–934.
- [84] Kelley DR, Snoek J, Rinn JL, Basset: learning the regulatory code of the accessible genome with deep convolutional neural networks,” *Genome Res.* 2016; 26(7):990–999.
- [85] Alipanahi B, DeLong MT, Weirauch BJ, Frey, Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning,” *Nat. Biotechnol.*, vol. 33, no. 8, pp. 831–838, 2015.
- [86] Angermueller C, Lee HJ, Reik W, Stegle O, Accurate prediction of single-cell DNA methylation states using deep learning. *BioRxiv.* 2016; 55715.
- [87] Koh PW, Pierson E, Kundaje A, Denoising genome-wide histone ChIP-seq with convolutional neural networks. *Bioinformatics.* 2017; 33(14):i225–i233.
- [88] Fakoor R, Ladhak F, Nazi A, Huber M, Using deep learning to enhance cancer diagnosis and classification. in *Proceedings of the international conference on machine learning*, 2013, vol. 28, pp. 3937–3949.
- [89] Lyons J, et al., Predicting backbone α angles and dihedrals from protein sequences by stacked sparse auto-encoder deep neural network. *J. Comput. Chem.* 2014; 35(28):2040–2046.
- [90] Hammerla NY, Halloran S, Plötz T, Deep, convolutional, and recurrent models for human activity recognition using wearables. *arXiv Prepr. arXiv1604.08880*, 2016.
- [91] Zhu J, Pande A, Mohapatra P, Han JJ, Using deep learning for energy expenditure estimation with wearable sensors. in *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*, 2015; 501–506.
- [92] Jindal V, Birjandtalab J, Pouyan MB, Nourani M, An adaptive deep learning approach for PPG-based identification in *2016 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2016; 6401–6404.
- [93] Nurse E, Mashford BS, Yepes AJ, Kiral-Kornek I, Harrer S, Freestone DR, Decoding EEG and LFP signals using deep learning: heading TrueNorth. in *Proceedings of the ACM international conference on computing frontiers*, 2016; 259–266.
- [94] Sathyanarayana A, et al., Sleep quality prediction from wearable data using deep learning. *JMIR mHealth uHealth*, 2016; 4(4):125.
- [95] Top Use Cases for AI in Media and Entertainment,” *Dataiku.* Sep. 22, 2021, [Online]. Available: <https://www.dataiku.com/stories/ai-in-media-and-entertainment/>.
- [96] Chou CH, Su YS, Hsu CJ, Lee KC, Han PH, Design of Desktop Audiovisual Entertainment System with Deep Learning and Haptic Sensations. *Symmetry (Basel)*, vol. 12, no. 10, 2020, DOI: 10.3390/sym12101718.
- [97] Erhel S, Jamet E, Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Comput. Educ.* 2013; 67:156–167. DOI: <https://doi.org/10.1016/j.compedu.2013.02.019>.
- [98] Justesen N, Bontrager P, Togelius J, Risi S, Deep learning for video game playing. *IEEE Trans. Games.* 2019; 12(1):1–20.
- [99] ParamitaGhosh, The Future of Deep Learning. *DATAVERSITY.* Sep. 22, 2020, [Online]. Available: <https://www.dataversity.net/the-future-of-deep-learning/>.
- [100] Gudmundsson SF, et al., Human-Like Playtesting with Deep Learning. in *2018 IEEE Conference on Computational Intelligence and Games (CIG)*, 2018; 1–8, DOI: 10.1109/CIG.2018.8490442.
- [101] Usukhbayar B, Homer S, Deepfake Videos: The Future of Entertainment. 2020.
- [102] Kulkarni R, Gaikwad R, Sugandhi R, Kulkarni P, Kone S, Survey on deep learning in music using GAN. *Int. J. Eng. Res. Technol.*, vol. 8, no. 9, pp. 646–648, 2019.
- [103] Fessahaye F, et al., T-recsys: A novel music recommendation system using deep learning. in *2019 IEEE international conference on consumer electronics (ICCE)*, 2019; 1–6.
- [104] Nam J, Choi K, Lee J, Chou SY, Yang YH, Deep learning for audio-based music classification and tagging: Teaching computers to distinguish rock from bach. *IEEE Signal Process. Mag.* 2018; 36(1):41–51.
- [105] Sünderhauf N, et al., The limits and potentials of deep learning for robotics. *Int. J. Rob. Res.* 2018; 37(4–5):405–420.
- [106] A. J. Moshayedi, A. S. Roy, and L. Liao, “PID Tuning Method on AGV (Automated Guided Vehicle) Industrial Robot,” vol. 12, no. 4, pp. 53–66, 2020. A
- [107] A. J. Moshayedi and D. C. Gharpure, “Development of position monitoring system for studying performance of wind tracking algorithms,” *7th Ger. Conf. Robot. Robot.* 2012, vol. 32, pp. 161–164, 2012.
- [108] A. J. Moshayedi, S. S. Fard, L. Liao, and S. A. Eftekhari, “Design and development of pipe inspection robot meant for

- resizable pipe lines,” *Int. J. Robot. Control*, vol. 2, no. 1, p. 25, 2019. A
- [109] A. J. Moshayedi, M. Gheibollahi, and L. Liao, “The quadrotor dynamic modeling and study of meta-heuristic algorithms performance on optimization of PID controller index to control angles and tracking the route,” *IAES Int. J. Robot. Autom.*, vol. 9, no. 4, p. 256, 2020, doi: 10.11591/ijra.v9i4.pp256-270.
- [110] Punjani A, Abbeel P, Deep learning helicopter dynamics models. *IEEE Int. Conf. Robot. Autom.*, 2015; 3223–3230.
- [111] Miyajima R, Deep Learning Triggers a New Era in Industrial Robotics. *IEEE Multimed.* 2017; 24(4):91–96. DOI: 10.1109/MMUL.2017.4031311.
- [112] Le TT, Lin CY, Bin-Picking for Planar Objects Based on a Deep Learning Network: A Case Study of USB Packs. *Sensors.* 2019; 19(16) DOI: 10.3390/s19163602.
- [113] Correll N, et al., Analysis and Observations From the First Amazon Picking Challenge. *IEEE Trans. Autom. Sci. Eng.*, 2018; 15(1):172–188, DOI: 10.1109/TASE.2016.2600527.
- [114] McLaughlin E, Charron N, Narasimhan S, Automated defect quantification in concrete bridges using robotics and deep learning. *J. Comput. Civ. Eng.* 2020; 34(5):4020029.
- [115] McLaughlin E, Charron N, Narasimhan S, Combining deep learning and robotics for automated concrete delamination assessment. in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction.* 2019; 36:485–492.
- [116] A. J. Moshayedi, G. Xu, L. Liao, and A. Kolahdooz, “Gentle Survey on MIR Industrial Service Robots: Review & Design,” vol. 10, no. 1, pp. 31–50, 2021.
- [117] A. J. Moshayedi and D. C. Gharpure, “Evaluation of bio inspired Mokhtar: Odor localization system,” 2017, doi: 10.1109/CarpathianCC.2017.7970457.
- [118] A. abbasi Ata Jahangir Moshayedi Liefia Liao, Shuai Li, “Path planning and trajectory tracking of a mobile robot using bio-inspired optimization algorithms and PID control,” in *2019 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 2019, vol. 2, no. 2, p. 60.
- [119] A. J. Moshayedi and D. C. Gharpure, “Design and development of wind tracker,” *ICCAIE 2011 - 2011 IEEE Conf. Comput. Appl. Ind. Electron.*, no. Iccaie, pp. 36–40, 2011, doi: 10.1109/ICCAIE.2011.6162100.
- [120] A. J. Moshayedi, A. S. Roy, and L. Liao, “PID Tuning Method on AGV (Automated Guided Vehicle) Industrial Robot,” vol. 12, no. 4, pp. 53–66, 2020.