Harnessing Machine Learning: Exploring Novel Techniques and Applications

Parul Chhabra[1], Vipin Babbar[2], Rohit Sharma[3]
Deptt of CSE GJUST, Hisar, Haryana, India[1]
Deptt of CSE GCW, Hisar, Haryana, India[2]
Deptt of CSE CDLU, Hisar, Haryana, India[3]

Abstract: - The current manuscript provides a comprehensive examination of Machine Learning (ML), encompassing its core principles, recent advancements, and prospective future trajectories. The analysis encompasses a broad spectrum of ML theories, including but not limited to the fundamental concepts of supervised, unsupervised, and reinforcement learning, as well as cutting-edge developments such as deep learning and transfer learning. The article additionally examines the pragmatic ramifications of ML in various domains, including healthcare, finance, and environmental science. The objective is to cultivate a more profound comprehension of ML, to facilitate its efficient implementation across diverse sectors, and to lay the groundwork for forthcoming investigations.

Keywords*:* Machine Learning Artificial, Intelligence, Diverse.

1. Introduction

In contemporary times, ML has emerged as a significant paradigm within the domain of artificial intelligence (AI), and has been instrumental in bringing about a profound transformation across various industries and academic fields. ML has its roots in the early days of AI during the mid-20th century and draws inspiration from the human capacity to gain knowledge through experience, as noted by Mnih et al. (2015). In contrast to conventional computational methods, ML algorithms possess the ability to acquire knowledge from data, enhance their understanding, and enhance their efficacy without explicit programming (Krizhevsky et al., 2012). The escalating production of data worldwide, which is further compounded by the widespread adoption of Internet of Things (IoT) devices and the digitization of various industries, has led to an upsurge in the application of ML techniques. These methods facilitate comprehension, integration, and use of the data for operational reasons. Medicine, economics, and ecology are just a few of the domains where ML has gained prominence in recent years. It aids in both patient diagnosis and tailored treatment. In the financial industry, it underpins things like credit rating and automated trading. It is used to model complex climatic systems and optimise the distribution of natural resources. Given its enormous potential, it is crucial to understand the foundational concepts, practical applications, and inherent difficulties of ML (LeCun et al., 2015). Challenges in this context go beyond purely technical ones, such as poor data quality or complicated algorithms. Concerns about data privacy, bias, and the transparency of algorithms fall under this category as well. The purpose of this study is to examine the cornerstones of ML in great detail. It will cover the basic techniques such as supervised, unsupervised, and reinforcement learning, as well as more sophisticated methods like deep learning and transfer learning. The practical applications of ML in diverse sectors will be emphasised, alongside an examination of the ethical and societal issues that are arising in tandem with the development of ML technology. The objective is to facilitate a more profound comprehension of the fundamental principles of ML, foster discourse on its ramifications, and lay the groundwork for forthcoming progressions in this swiftly developing domain.

2. Fundamentals of ML

The field of ML is primarily focused on the development and utilisation of models that are capable of learning from data. Broadly speaking, a model can be defined as any entity that leverages data to generate an outcome. However, in the realm of ML, models are mathematical entities that produce output based on input. According to

Mnih et al. (2015), there exist three primary classifications of ML techniques, which are supervised learning, unsupervised learning, and reinforcement learning. Supervised learning involves the training of an algorithm to acquire a mapping function that can link inputs to outputs. This is achieved through the use of labelled examples, where each instance in the training data comprises input attributes and a corresponding desired output. The objective is for the model to exhibit generalisation capability by accurately predicting outcomes on previously unseen data. The task entails the prediction of a categorical variable for a specific observation. As an illustration, a model could perform email classification into categories of 'spam' or 'non-spam', or make predictions regarding the malignancy or benignity of a tumour by analysing medical imaging data. Forecasts an uninterrupted result. (Krizhevsky et al., 2012). An illustrative instance may involve the prognostication of residential property values, contingent upon attributes such as the quantity of bedrooms, geographical situation, and age of the real estate. Unsupervised learning is a type of ML that differs from supervised learning in that it involves the training of a model using unlabeled data. The model is tasked with autonomously identifying patterns and correlations within the data. Cluster analysis is the process of categorising data points into distinct groups or clusters, which is determined by a specific metric of similarity. As an instance, a model could potentially group customers into distinct segments predicated on their purchasing patterns. The process entails the reduction of input variables within a given dataset. One commonly employed technique for enhancing data visualisation is Principal Component Analysis (PCA). Another ML approach, Reinforcement Learning (RL), involves an agent acquiring knowledge on how to act within an environment by receiving rewards or penalties based on its actions (LeCun et al., 2015). Reinforcement learning (RL) diverges from supervised learning in that it does not undergo training utilising a dataset comprising of accurate responses. The agent engages in environmental exploration, executes actions, and acquires knowledge from the outcomes. Sequential decision making is a prevalent activity in the field of reinforcement learning (RL), and it finds application in diverse domains such as game playing, robot navigation, and resource management.

3. Different Techniques

The identification of optimal ML techniques is heavily reliant on various factors such as the particular task being performed, the characteristics and quantity of the data accessible, computational capabilities, and the ultimate aim of the model. It can be observed that certain techniques have gained widespread recognition owing to their remarkable efficacy across diverse applications. Here are a few:

- Convolutional Neural Networks: CNNs are commonly employed in image processing tasks owing to their capacity to capture spatial hierarchies. CNNs have been employed to attain exceptional performance in tasks such as image classification, object detection, and segmentation (LeCun et al., 2015).
- Recurrent Neural Networks and Transformers: RNNs, including advanced variants such as LSTM networks and Transformer models, have demonstrated remarkable efficacy in tasks that involve sequential data, such as natural language processing or time series prediction (Hochreiter & Schmidhuber, 1997; Vaswani et al., 2017).
- Ensemble Methods: Random Forests and Gradient Boosting Machines (GBMs) are effective methods for achieving strong performance on tabular data. The integration of multiple models results in enhanced predictive performance, surpassing the capabilities of any individual constituent model (Breiman, 2001; Chen & Guestrin, 2016).
- Deep Reinforcement Learning: The integration of deep learning and reinforcement learning has been applied to decision-making tasks with notable success in various fields such as game playing, robotics, and resource management (Mnih et al., 2015).
- Transfer Learning: The utilisation of pre-trained models such as BERT and ResNet, followed by fine-tuning on a particular task, results in a notable reduction in the necessary data and training duration. The utilisation of this methodology has been of utmost significance in natural language processing and computer vision assignments (Devlin et al., 2018; He et al., 2016).
- Generative Adversarial Networks (GANs): GANs have revolutionised the task of generating new data that closely resembles the input data. GANs have found diverse applications in areas such as image synthesis, super-resolution, and other related fields (Goodfellow et al., 2014)

4. . Literature Review on ML Techniques

The field of ML has experienced significant advancements in recent times, with various innovative methodologies gaining traction owing to their remarkable efficacy in diverse domains.

• CNNs :

The models presented by LeCun et al. (2015) are considered to be at the forefront of image processing tasks. CNNs utilise the concept of spatial correlation by imposing a pattern of local connectivity between neurons that are present in adjacent layers. Architectural variations such as Inception networks, ResNets, and EfficientNets have exhibited remarkable efficacy in image-related tasks, as evidenced by studies conducted by Szegedy et al. (2015), He et al. (2016), and Tan & Le (2019). CNNs belong to the category of deep, feed-forward artificial neural networks and are primarily utilised for the analysis of visual imagery. The aforementioned models are adaptations of the multilayer perceptrons introduced by LeCun et al. in 1998, which are designed to run with little preprocessing due to their inspiration from biological phenomena.

One or more convolutional layers constitute the foundation of a standard CNN, with further layers including pooling, fully connected, and normalisation. LeCun et al. (1998) suggested a variant of the standard multilayer perceptron design that has been used here in an effort to reduce the amount of time spent on preprocessing. These neural networks, which have a shared-weights design and translation invariance qualities, are sometimes referred to as shift invariant or space invariant. CNNs have shown their worth in a variety of settings, especially for image and video recognition. In the ImageNet 2012 competition, CNNs achieved a top-5 error rate of 15.3%, according to Krizhevsky et al. (2012), besting the second-best entry's 26.2% mistake rate. Thus, CNNs have become the go-to method for image recognition and classification applications.

• RNNs and Transformers

The transformer architecture, which was first presented in the influential paper "Attention is All You Need" (Vaswani et al., 2017), has significantly transformed the domain of natural language processing. It has become a fundamental component of models such as BERT, GPT, and T5, which have demonstrated exceptional performance in various NLP tasks (Devlin et al., 2018; Radford et al., 2018; Raffel et al., 2019), thereby establishing the transformer architecture as a crucial advancement in the fi Standard RNNs encounter challenges in acquiring long-term dependencies due to the issue of vanishing or exploding gradients. Hochreiter and Schmidhuber (1997) proposed a solution to this problem through the implementation of Long Short-Term Memory (LSTM) networks. Subsequently, Cho et al. (2014) proposed the Gated Recurrent Unit (GRU) as a simplified version of the LSTM. The utilisation of attention mechanisms, as opposed to recurrence, marks a significant departure in the handling of sequential data, as exemplified by Transformers. According to the paper "Attention is All You Need" by Vaswani et al. (2017), the Transformer design is premised on the use of a mechanism called self-attention, or scaled dot-product attention. The success rate of transformers in natural language processing-related activities is astounding. Models built using the Transformer architecture have shown outstanding performance in a variety of natural language processing (NLP) tasks, including BERT (Devlin et al., 2018), GPT-2 (Radford et al., 2019), and T5 (Raffel et al., 2019).

• Ensemble Methods

Due to their robustness and outstanding performance on tabular data, ensemble algorithms like Random Forests and Gradient Boosting Machines have seen widespread implementation since their discovery by Breiman (2001) and Chen & Guestrin (2016). In ML, it is usual procedure to use many models and combine the results to improve prediction accuracy. This method not only reduces the possibility of overfitting, but also increases model stability. Leveraging the ability of several models to achieve improved predictive effectiveness, ensemble approaches have gained substantial momentum in the area of ML. The present literature review furnishes a comprehensive survey of ensemble methods, their diverse forms, and their practical implementations. Breiman (2001) introduced Random Forests as an ensemble learning technique that integrates several decision trees. Individual trees are subjected to training on distinct subsets of the data, with the utilisation of random feature selection. The ultimate forecast is ascertained by consolidating the prognostications of individual trees via either majority voting or averaging. The application of Random Forests has yielded successful outcomes across diverse domains, encompassing classification, regression, and feature selection. Gradient Boosting Machines (GBMs) are a widely used ensemble technique that constructs a robust predictive model through the iterative combination of weak learners, such as decision trees. Sequentially adding weak learners is a common technique employed by boosting algorithms, including AdaBoost and XGBoost (Chen & Guestrin, 2016). These algorithms aim to improve classification accuracy by focusing subsequent learners on samples that previous learners struggled to classify correctly. Gradient Boosting Machines (GBMs) have attained cutting-edge performance in diverse ML contests and are extensively employed in practical applications.

- Deep Reinforcement Learning

The integration of deep learning for representation and reinforcement learning for decision making has exhibited considerable potential in scenarios where data is acquired sequentially and future outcomes are contingent on past events. Silver et al. (2016) and Brown et al. (2019) show that deep reinforcement learning may successfully achieve mastery in games like Chess, Go, and Poker. Exciting research is being done at the intersection of deep learning and reinforcement learning, which is known as Deep Reinforcement Learning (DRL). This synthesis makes it easier for agents to learn optimal behaviours by trying out new approaches and adjusting to their surroundings over time. This literature review delves deeply into the seminal studies and significant developments that have shaped the area of Deep Reinforcement Learning (DRL). The key study by Mnih et al. (2013) used Deep Reinforcement Learning (DRL) to teach itself how to play Atari video games from scratch, using just pixel data. To estimate the Q-function, a novel technique called the Deep Q-Network (DQN) employs a deep convolutional neural network. Several Atari 2600 games were played at a human level using this method. The creators of the DQN framework (Mnih et al., 2016) found that including asynchronous methods increased the efficiency of the training. In this article, we introduced the Asynchronous Advantage Actor-Critic (A3C) approach, which employs several agents to asynchronously update a common network. With this strategy, we were able to achieve convergence more quickly and make better use of our samples. Schulman et al. (2017) presented the Proximal Policy Optimisation (PPO) technique to optimise the efficacy of policies. This approach uses trust region optimisation to promote consistent and efficient learning. Due to its higher performance and sample efficiency compared to earlier approaches, the PPO algorithm has been frequently employed for continuous control applications. The Deep Deterministic Policy Gradient (DDPG) method was introduced by Lillicrap et al. (2015) to bring the policy gradient method to continuous action spaces. The Deep Deterministic Policy Gradient (DDPG) technique uses an actor-critic architecture to estimate the policy through the actor network and the value function via the critic network. Significant successes have been achieved in fields like as robotic manipulation and the learning of complex games using the Deep Deterministic Policy Gradient (DDPG) method.

- Transfer Learning

As a result of its ability to reduce the requirement for vast labelled data, the use of pre-trained models in transfer learning has become more common. Yosinski et al. (2014) and Devlin et al. (2018) show that techniques like the refining of pre-existing models (such BERT and ResNet) have been extensively used in both natural language processing and computer vision applications. Transfer learning is a powerful approach in the area of ML that involves the use of past knowledge from one task or domain to improve learning and performance in a related task or domain. The current literature review provides an exhaustive overview of foundational works and cutting-edge research in the field of transfer learning. The basic goal of domain adaptation methods is to make it easier to move information across domains with different distributions. Co-training for domain adaptation was first proposed by Daume III and Marcu in 2006, and it makes use of unlabeled data from the target domain to improve classification accuracy. The work has been called "seminal" by experts in the subject. Different approaches, such as adversarial domain adaptation, have been developed to address the difficulties associated with this transition. Pre-training on a large dataset followed by fine-tuning on a specific task is an approach that has shown success in the field of transfer learning. The utilisation of pre-trained models like VGG (Simonyan & Zisserman, 2014), Inception (Szegedy et al., 2015), and ResNet (He et al., 2016) has been instrumental in the progression of computer vision, with the ImageNet dataset serving as a pivotal component in this advancement. Studies such as Devlin et al. (2018) and Radford et al. (2018) have illustrated the efficacy of pre-training and fine-tuning in natural language processing tasks, as evidenced by models like BERT and GPT.

- GANs

GANs , as proposed by Goodfellow et al. (2014), have garnered significant attention in research circles owing to their ability to produce data. CNNs have been utilised in various domains, such as image generation, enhancing image resolution, and expanding data sets. The GANs have surfaced as a potent framework within the field of ML for the purpose of generating fresh data that bears a striking resemblance to the distribution of the training data. The present literature review offers a comprehensive survey of seminal works and progressions in the domain of GANs . The GAN framework was introduced by Goodfellow et al. (2014) and comprises of two main components, namely a generator network and a discriminator network. The objective of the generator is to produce plausible instances from stochastic input, whereas the discriminator's objective is to differentiate between authentic and synthesised instances. GANs acquire knowledge through an adversarial learning process, in which the generator and discriminator engage in a competitive game, resulting in the production of progressively authentic instances. Mirza and Osindero (2014) introduced Conditional GANs (cGANs) which integrate supplementary conditioning information to direct the generation procedure. The provision of class labels or supplementary input data enables the regulation of distinct attributes of the produced samples. DCGANs were introduced by Radford et al. (2015), which employ deep CNNs in both the generator and discriminator components. Deep Convolutional GANs (DCGANs) have demonstrated enhanced stability and superior image quality generation in comparison to conventional GAN structures.

Progressive GANs were introduced by Karras et al. (2017), utilising a technique that initiates image generation with low-resolution images and gradually enhances the resolution during the training process. The utilisation of this methodology yields superior image quality and enhanced training stability in contrast to the direct training on high-resolution images. CycleGAN, a variant of GAN, was proposed by Zhu et al. (2017) to facilitate image-to-image translation without the need for paired images. CycleGAN is capable of acquiring the ability to translate images from one domain to another without necessitating the presence of corresponding paired training samples. The aforementioned methodology has been implemented in diverse undertakings, including but not limited to style conversion, object metamorphosis, and domain adjustment. The StyleGAN, as introduced by Karras et al. (2019), enables the manipulation of distinct characteristics of the generated images, including their style and resolution. The StyleGAN model has demonstrated exceptional proficiency in generating images that are both diverse and highly realistic, rendering it a popular choice for creative endeavours and art generation. The authors Brock et al. (2018) introduced BigGAN, a Generative Adversarial Network (GAN) framework that attains the highest level of performance in the domain of image synthesis tasks. The BigGAN model integrates modifications in its architecture and training methodologies, including class-conditional batch normalisation and truncation techniques, to produce superior quality images across diverse datasets.

5. Applications of ML

The utilisation of ML has been observed in diverse domains owing to its proficiency in extracting valuable insights from extensive datasets and generating precise predictions. Here are some notable applications:

Healthcare: The application of ML techniques has the potential to support healthcare practitioners in various aspects including disease diagnosis, patient outcome prediction, treatment plan customization, and drug discovery. One example pertains to the utilisation of ML algorithms for the purpose of pattern recognition in medical imaging data, such as MRI and CT scans, to detect and diagnose tumours, lesions, or other indications of illness. ML has the capability to examine Electronic Health Records (EHRs) for the purpose of forecasting patient risk and providing assistance in preventive care initiatives.

Finance: ML finds extensive applications in the finance industry. Credit scoring models have the ability to evaluate the potential risk associated with extending loans to individuals by analysing their past data. Algorithmic trading models have the capability to execute trades with high velocity and frequency, utilising up-to-the-minute market data to inform decision-making processes. Furthermore, ML has the capability to identify fraudulent transactions through the detection of atypical behavioural patterns.

Environmental Science: Within the field of environmental science, the utilisation of ML techniques has the potential to assist researchers in the modelling of intricate systems, such as global climate models or ecosystems, thereby enabling the formulation of predictions and the provision of guidance for policy-making. ML techniques can be employed to examine satellite imagery for the purpose of deforestation monitoring or identification of illicit fishing practises. ML algorithms have the capability to forecast energy consumption in intelligent power grids and efficiently manage resource allocation.

Transportation: The development of autonomous vehicles heavily relies on the implementation of ML techniques. Visual processing aids in various cognitive tasks such as object detection, traffic sign recognition, prediction of other road-users' behaviour, and decision making related to lane changes or turns. ML is also utilised in the optimisation of logistics, including the optimisation of delivery routes and fleet management.

E-commerce and Marketing: ML algorithms are commonly employed in e-commerce to develop recommendation systems that provide personalised product suggestions to customers, leveraging their browsing and purchasing history. ML is a valuable tool in the field of marketing as it enables the segmentation of customers into distinct groups for the purpose of targeted marketing. Additionally, ML can be utilised to forecast customer churn.

Agriculture: ML has the potential to be utilised in the agricultural industry for various purposes such as forecasting crop yields, resource optimisation, identification of pests and diseases, and automation of tasks such as harvesting and weeding through the use of robotics.

6. Conclusion

It signifies a noteworthy progression in the manner in which we analyse, comprehend, and forecast data. The swift progress and potential of the field of ML is exemplified by its evolution from basic concepts such as supervised, unsupervised, and reinforcement learning to more sophisticated techniques like deep learning, transfer learning, and federated learning. Moreover, the abundance of applications spanning various industries such as healthcare, finance, environmental science, and transportation, highlights its significant influence on society. The field of ML has the potential to enhance efficacy and precision across diverse domains, while also creating prospects for pioneering applications and innovations. Notwithstanding the potential benefits, ML poses a plethora of difficulties. Significant challenges are posed by concerns surrounding data privacy, security, ethical considerations, bias, and algorithmic transparency. Addressing these challenges necessitates a collaborative approach involving extensive research and policy formulation, with a particular emphasis on ensuring that the advantages of ML are achieved in a manner that is fair, morally sound, and upholds the confidentiality of users. As the field of research progresses and technological advancements are made, it is expected that there will be an increase in the utilisation of advanced ML methodologies, leading to the emergence of novel applications and complexities. Hence, it is crucial to sustain a continuous discourse regarding the consequences and regulation of these technological advancements. By means of such endeavours, it is possible to steer the progress of ML towards a direction that optimises its advantages to the community while reducing probable hazards. The potential of ML is highly auspicious, and the trajectory towards its realisation is already in progress.

Refrences

[1] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. In Advances in neural information processing systems (pp. 2672-2680).
[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
[4] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-178
[5] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. Advances in neural information processing systems, 2672-2680.
[6] Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
[7] Radford, A., Metz, L., & Chintala, S. (2015). Unsupervised representation learning with deep convolutional GANs . arXiv preprint arXiv:1511.06434.
[8] Bishop, C. M. (2006). Pattern recognition and ML . Springer.

[9] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.
[10] Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.
[11] Jordan, M. I., & Mitchell, T. M. (2015). ML : Trends, perspectives, and prospects. Science, 349(6245), 255-260.
[12] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436-444.
[13] Russell, S., & Norvig, P. (2016). Artificial Intelligence: A Modern Approach. Malaysia; Pearson Education Limited.
[14] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., ... & Dieleman, S. (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484-489.
[15] Daume III, H., & Marcu, D. (2006). Domain adaptation for statistical classifiers. Journal of Artificial Intelligence Research, 26, 101-126.
[16] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
[17] Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., ... & Rabinovich, A. (2015). Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1-15)
[18] Breiman, L. (2001). Random forests. ML , 45(1), 5-32.
[19] Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785-794).
[20] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing Atari with Deep Reinforcement Learning. arXiv preprint arXiv:1312.5602.
[21] Mnih, V., Badia, A. P., Mirza, M., Graves, A., Lillicrap, T., Harley, T., ... & Kavukcuoglu, K. (2016). Asynchronous Methods for Deep Reinforcement Learning. arXiv preprint arXiv:1602.01783.
[22] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
[23] Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. arXiv preprint arXiv:1406.1078.
[24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).
[25] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.
[26] Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language models are unsupervised multitask learners.
[27] Raffel, C., Shazeer, N., Roberts, A., Lee, K., Narang, S., Matena, M., ... & Liu, P. J. (2019). Exploring the limits of transfer learning with a unified text-to-text transformer. arXiv preprint arXiv:1910.10683.
[28] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.
[29] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep CNNs . Advances in neural information processing systems, 25, 1097-1105.
[30] Breiman, L. (2001). Random forests. ML , 45(1), 5-32.
[31] Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM sigkdd international conference on knowledge discovery and data mining (pp. 785-794).
[32] Hastie, T., Tibshirani, R., & Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer Science & Business Media.
[33] Ng, A. (2012). ML and AI via brain simulations. Stanford University CS231n: CNNs for Visual Recognition.
[34] Bengio, Y., Louradour, J., Collobert, R., & Weston, J. (2009). Curriculum learning. In Proceedings of the 26th annual international conference on ML (pp. 41-48).
[35] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. Nature, 323(6088), 533–536.

[36] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., & Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. Neural computation, 1(4), 541-551.
[37] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
[38] Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.
[39] Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In Advances in neural information processing systems (pp. 3104-3112).
[40] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in neural information processing systems (pp. 5998-6008).
[41] Howard, J., & Ruder, S. (2018). Universal language model fine-tuning for text classification. In Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Vol. 1, pp. 328-339).
[42] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. In Advances in Neural Information Processing Systems.
[43] Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., ... & Chen, Y. (2017). Mastering the game of Go without human knowledge. Nature, 550(7676), 354-359.
[44] Hinton, G., Sabour, S., & Frosst, N. (2018). Matrix capsules with EM routing. In International Conference on Learning Representations (pp. 1-15).
[45] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).
[46] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep CNNs . In Advances in neural information processing systems (pp. 1097-1105).
[47] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 779-788).
[48] Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.
[49] Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.
[50] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).