Deep Science Publishing
https://doi.org/10.70593/978-81-981271-4-3_1

Chapter 1

Machine learning and deep learning architectures and trends: A review

Nitin Liladhar Rane [1], Suraj Kumar Mallick [2], Ömer Kaya [3], Jayesh Rane [4]

[1] *Vivekanand Education Society's College of Architecture (VESCOA), Mumbai 400074, India*
[2] *Shaheed Bhagat Singh College, University of Delhi, New Delhi 110017, India*
[3] *Engineering and Architecture Faculty, Erzurum Technical University, Erzurum 25050, Turkey*
[4] *Pillai HOC College of Engineering and Technology, Rasayani, India*
[1] nitinrane33@gmail.com

Abstract: The arrival of machine learning (ML) together with deep learning (DL) has been revolutionizing many fields through advances in data-driven decision-making, automation, and predictive analytics. This has formed the keystone for the exploration of the most recent architectures and upcoming trends in said domains as to how they are significantly impacting other sectors. Recent ML designs, such as Transformers or graph neural networks (GNNs) in combination with neural differential equations, have found remarkable performance in tasks such as natural language processing (NLP) or recommendation systems and molecular modeling. The birth of big language models (LLMs)-from GPT-4 to BERT-has furthered the understanding and production of human languages to degrees where chatbots, translation, and content generation are advanced. At the same time, DL structures have evolved with the advent of state-of-the-art advancements, e.g., convolutional neural networks (CNNs) and generative adversarial networks (GANs), that play an essential role in areas such as image and video processing, autonomous driving, and synthetic data generation. This work focuses on how these structures may be combined with the advanced technologies of the Internet of Things with that of blockchain and quantum computing to enhance security, efficiency, and scalability for intelligent systems. Increasing trends show a concern with artificial intelligence (AI) and explainable AI (XAI) to deal with crucial problems of transparency, fairness, and accountability. The study also examines how federated learning is likely to influence privacy-driven data analysis and the surge in edge AI, which involves pushing computing closer to the source of data, reducing latency and improving real-time decision-making. This research will identify the importance of ML and DL, which is crucially important in showing the shape of technology and society in the future.

Keywords: Machine learning, Deep learning, Architectures, Artificial intelligence, Convolutional neural networks, Recurrent neural networks, Natural language processing

Citation: Rane, N. L., Mallick, S. K., Kaya, O., & Rane, J. (2024). Machine learning and deep learning architectures and trends: A review. In *Applied Machine Learning and Deep Learning: Architectures and Techniques* (pp. 1-38). Deep Science Publishing. https://doi.org/10.70593/978-81-981271-4-3_1

1.1 Introduction

The fields of healthcare and finance have seen significant changes as a result of machine learning (ML) and deep learning (DL), which enable machines to learn from data and make decisions with minimal human intervention (Chauhan & Singh, 2018; Shrestha & Mahmood, 2019). These technologies' adoption and expansion have been accelerated by the rapid advancement of processing power and the wealth of available data (Shrestha & Mahmood, 2019; Shinde & Shah 2018; Dargan et al., 2020). The ML and DL architectures, which are the foundation of these technologies, have made significant progress and shown remarkable capabilities in tasks such as natural language processing, autonomous systems, and image and audio recognition. ML models come in a variety of architectures, from basic linear regression models to intricate neural networks, designed for different tasks and types of data (Chauhan & Singh, 2018; Sengupta et al., 2020; Alzubaidi et al., 2021). DL, a branch of ML, utilizes neural networks with multiple layers to capture complex patterns and features in data (Dargan et al., 2020; Alzubaidi et al., 2021; Minar & Naher, 2018). Architectures such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Generative Adversarial Networks (GANs) have expanded the capabilities of machines, resulting in advancements in computer vision, speech synthesis, and generative art.

Research trends in machine learning and deep learning are driven by continuous advancements and growing demands from various applications (Shrestha & Mahmood, 2019; Janiesch et al., 2021; Topuz & Alp, 2023). Better algorithms, integrating ML and DL with other technologies like blockchain and IoT, and emphasising ethical AI and responsible use of these technologies are some recent developments (Bal & Kayaalp, 2021-Thakkar & Lohiya, 2021). To improve the transparency and interpretability of ML and DL models for users, there is also a developing trend towards explainable AI (XAI) (Wang et al., 2020; Mu & Zeng, 2019; Miotto et al., 2018). This research explores the complex structures of ML and DL, analyzing their progress, present developments, and upcoming paths. By conducting a thorough review of the literature, we pinpoint important developments and upcoming trends in these areas. Moreover, we utilize keyword co-occurrence and cluster analysis to reveal the key topics and research groups, offering a detailed insight into the present scenario and possible future advancements (Fig. 1.1).

Machine learning (ML) continues to evolve rapidly along with advances in data-driven technologies. As a sub-discipline of artificial intelligence (AI), this field is revolutionizing many industries with its abilities to learn and make predictions from data. Current ML trends are expanding the application areas of the technology and increasing the effectiveness of existing methods.

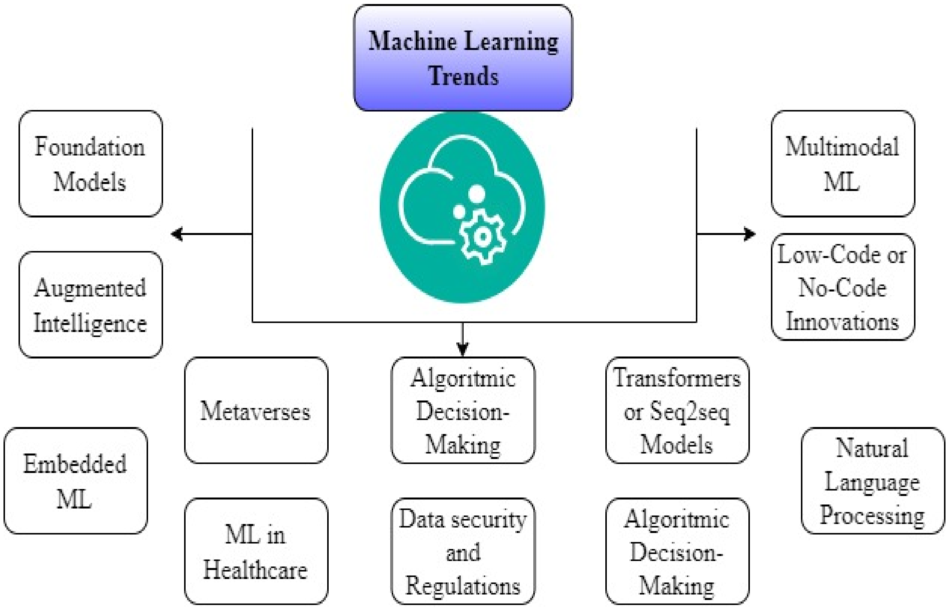

Fig. 1.1 Recent trends in Machine learning and deep learning architectures

Foundation Models are a popular trend; these are models that can be trained on large-scale datasets and are versatile enough to perform a range of tasks. Artificial intelligence (AI) solutions that maximise human-machine cooperation and augment human capabilities are referred to as augmented intelligence. IoT devices in particular use embedded machine learning (ML), which describes models that are housed inside the device and have the ability to interpret data in real time.

Metaverses create new interaction and economy models with the use of ML algorithms in virtual and augmented reality environments. In the healthcare industry, machine learning in healthcare offers innovative solutions in critical areas such as disease diagnosis and patient care. Data Security and Regulations enable the development of reliable AI systems by addressing privacy and ethical issues in the use of ML models.

Algorithmic Decision-Making enables automated and optimized decisions to be made in business processes and daily life. Transformers or Seq2Seq Models are models that show strong performance in language processing and other sequential data tasks. Multi-modal ML deals with models that can perform more comprehensive analyses inter-jointly, considering different styles of information, such as text, image, or audio. Low-Code/No-Code Innovations, on their part, provide tools that will let non-technical users develop models for machine learning. Natural Language Processing does the same but in developing models that understand and produce human language.

These trends are shaping the future direction of machine learning and opening up wider application possibilities across various industries. The increase in data-based decision-making processes and automation increases the importance of ML day by day. These advances are paving the way for smarter, more effective and more accessible AI solutions (Table 1.1).

Table 1.1 Deep learning trends in emerging cloud computing architectures

Deep Learning Technique	Computing Architecture	Devices
Deep Belief Network	Mobile Edge Computing	Medical wearables
Deep Belief Network	Mobile Edge Computing	Smart phones
Deep Belief Network	Mobile Edge Computing	Smart cameras
Deep Belief Network	Mobile Edge Computing	Voice controllers
Deep Reinforcement Learning	Fog Computing	Gateways
Deep Reinforcement Learning	Fog Computing	Computers
Deep Reinforcement Learning	Fog Computing	Switches
Deep Reinforcement Learning	Fog Computing	Automated vehicles
Convolutional Neural Network	Edge Computing	Desktops
Convolutional Neural Network	Fog Computing	Routers
Convolutional Neural Network	Mobile Edge Computing	Smart watches
Convolutional Neural Network	Volunteer Computing	Desktops
Modified Convolutional Neural Network	Edge Computing	IoT sensors
Modified Convolutional Neural Network	Mobile Edge Computing	Smart phones
Modified Convolutional Neural Network	Fog Computing	Gateways
Deep Neural Network	Edge Computing	IoT sensors
Deep Neural Network	Mobile Edge Computing	Smart phones
Deep Neural Network	Fog Computing	Routers
Deep Neural Network	Serverless Computing	Smart speakers
Recurrent Neural Network	Edge Computing	Smart phones

Recurrent Neural Network	Edge Computing	ECG device
Long Short-Term Memory	Edge Computing	Medical devices
Long Short-Term Memory	Edge Computing	Smart wearable devices
Long Short-Term Memory	Fog Computing	Switches
Long Short-Term Memory	Fog Computing	Servers
Long Short-Term Memory	Fog Computing	Routers

Deep learning (DL) is a sub-field of machine learning that uses multi-layer artificial neural networks to learn intricate patterns in data. Assume that deep learning is currently being included into developing cloud computing architectures to offer a theoretical framework across several domains. Then, determining the future course of the technology will require an understanding of how deep learning techniques interact with other cloud computing infrastructures and devices.

Deep learning approaches often improve data processing capacities and optimise device interactions when combined with cloud computing infrastructures. These trends provide notable benefits in applications needing massive data amounts and great computational power by enabling smart systems to function more effectively and efficiently. Future developments of increasingly intelligent and autonomous systems will be facilitated by advances in this dynamic and quickly developing field of deep learning (Balaji et al., 2018; Alzoubi et al., 2024; Deng et al., 2024; Lundberg, et al., 2024).

Significance of the research work:

- Performing an extensive review of literature to outline the progression and present status of ML and DL structures.
- Applying keyword co-occurrence analysis to recognize significant trends and upcoming subjects within the field.
- Conducting cluster analysis to reveal the main research clusters and how they are connected, offering insight into potential future research paths.

1.2 Methodology

In this research, we utilized an extensive literature review, keyword analysis, co-occurrence analysis, and cluster analysis to examine the structures and patterns in the fields of ML and DL. The research was based on a literature review that included a methodical examination of peer-reviewed articles, conference papers, and academic publications on ML and DL architectures and trends. We made use of various academic databases such as IEEE Xplore, Scopus, and Google Scholar to compile a comprehensive selection of pertinent research. The criteria for selection included works published within the last ten years, guaranteeing a contemporary comprehension of the subject. Our attention was directed towards articles covering new ML and DL architectures, their uses, evaluations of how well they perform, and upcoming trends. This thorough examination enabled us to pinpoint key themes and progressions in the literature.

In order to explore more deeply the specific areas of ML and DL research, we carried out a keyword analysis. This included the extraction and analysis of keywords from the chosen literature in order to pinpoint the terms and concepts that appeared most frequently. Automated tools were used in conjunction with manual verification during the keyword extraction process to guarantee accuracy. We proceeded to measure how often each keyword appeared in order to identify the main topics of interest and research focus in the ML and DL community. This examination gave us a deeper understanding of the main subjects and aided in visualizing the research field. Expanding on the keyword analysis, we conducted a co-occurrence analysis to investigate the connections among various keywords and concepts within the field. This technique required forming a co-occurrence matrix to record the frequency of pairs of keywords appearing together in the articles. By utilizing network analysis methods, we depicted these connections through co-occurrence graphs. These charts emphasized groupings of similar keywords, uncovering connections and thematic focuses in the realm of ML and DL research. This was a crucial measure in recognizing connections between different disciplines and merging multiple sub-fields. In the end, we used cluster analysis to group the identified themes and trends into cohesive categories. We used co-occurrence data to apply clustering algorithms like K-means and hierarchical clustering for grouping related keywords and research topics. This examination allowed us to identify separate groups that represent various areas of research, architectural advancements, and trend trends in ML and DL. Through analysis of these groupings, we were able to offer a systematic review of the present condition of the field, emphasizing key research topics and up-and-coming patterns.

1.3 Results and discussions

Co-occurrence and cluster analysis of the trending keywords in ML

The network diagram in Fig. 1.2, illustrating the co-occurrence and clustering of keywords in ML trends, provides important insights into the current trends and connections in the field. This examination investigates important groupings and the connections among main ideas, demonstrating how various aspects of ML study are intertwined and developing. The network diagram shows various separate clusters, each indicating a different thematic area in ML research. Each cluster consists of keywords that often appear together in academic papers, demonstrating a clear thematic relationship. The main clusters that were identified are:

Key Methods and Algorithms in ML (Red Cluster)

The fundamental ideas and techniques of machine learning constitute the heart of the red group. In this group, terms like "classification," "learning systems," "predictive models," "machine learning," and "optimisation" are essential. This team exemplifies the fundamentals of machine learning research, focussing on developing and optimising algorithms that can learn from data and make decisions or predictions. Notable terms in this category are also "decision making," "performance," "forecasting," and "support vector machines." These words designate study fields that aim to improve machine learning models' accuracy and efficacy. Phrases like "internet of things" and "cybersecurity" are included because they imply the application of these techniques in specific domains, demonstrating the interdisciplinary nature of current machine learning research.

Healthcare and Diagnostic Uses (Green Cluster)

The significance of machine learning for the healthcare sector and medical testing is emphasised by the green group. The terms "human," "algorithm," "diagnosis," "diseases," and "controlled study" are essential elements of this category. This demonstrates a focused attempt to analyse medical data using machine learning techniques, identify diseases, and improve patient outcomes. Terms like "sensitivity and specificity," "comparative study," and "major clinical study" refer to the exacting assessment methods used in medical research to ensure the precision and reliability of machine learning models. Demographic terms such as "male," "female," "aged," and "adult" ensure that machine learning algorithms are broadly applicable and effective in healthcare research because they point to a focus on a range of patient populations.

Blue Cluster consists of Neural Networks and DL

The blue cluster is mostly focused on topics related to deep learning and neural networks. Within this category, terms like "deep learning," "artificial neural network," "convolutional neural networks," "training," and "feature extraction" are essential terms. This highlights the significance of neural networks in current machine learning research because of their superior ability to recognise complex patterns in large datasets. Keywords like "image processing," "image segmentation," and "image enhancement" point to specific uses of deep learning in computer vision problems. The persistent emphasis on "precision" and "algorithmic learning" indicates the dedication to improving the efficiency and efficacy of deep learning models. This group demonstrates how important deep learning techniques are to improving machine learning systems' capabilities.

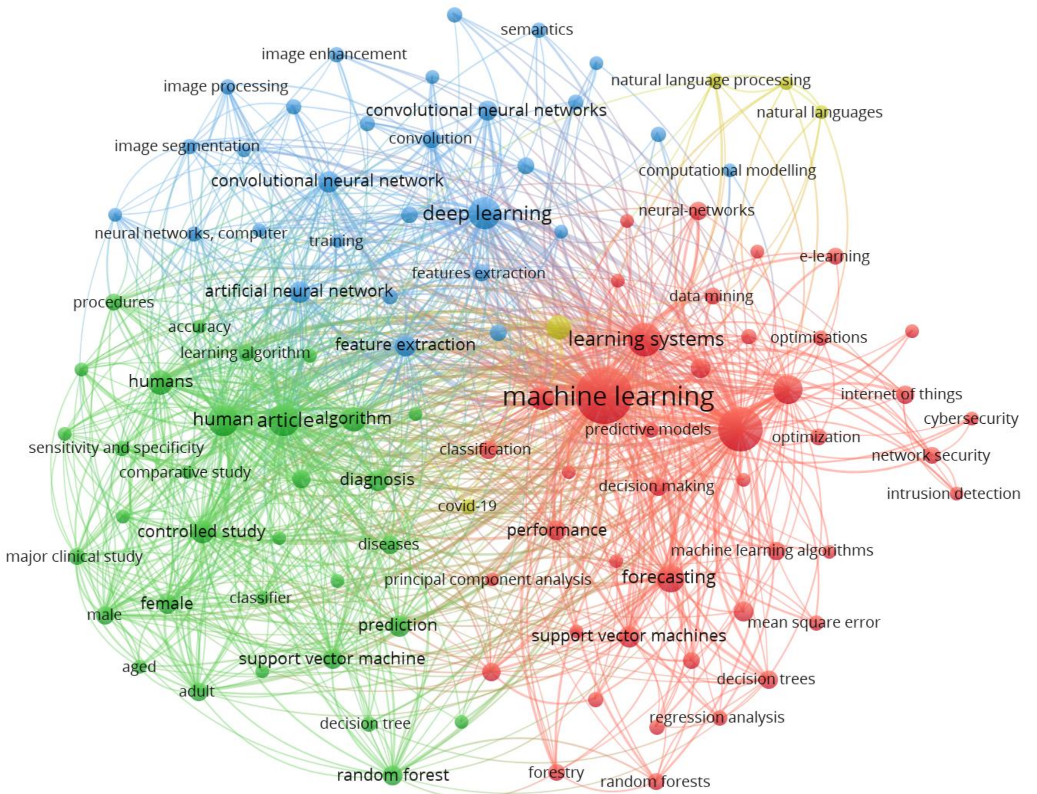

Fig. 1.2 Co-occurrence analysis of the trending keywords in ML

Natural Language Processing and Computational Modeling (Yellow Group)

The yellow cluster is primarily concerned in computational modelling and natural language processing (NLP). Important terminology like "natural language processing," "natural languages," "computational modelling," and "neural networks" are central to this cluster. This indicates that there is a great deal of interest in the study and development of algorithms that can understand, interpret, and produce human language. Words like "e-learning" and "data mining" refer to the application of natural language processing (NLP) techniques to improve educational systems and extract meaningful information from large text datasets. The integration of "semantics" with "feature extraction" highlights the difficulties and developments in comprehending and expressing the meaning of words and phrases in a computationally effective way.

Cybersecurity and Internet of Things (Purple Cluster) focus on protecting connected devices from online threats.

The purple cluster signifies the overlap of machine learning, cybersecurity, and the Internet of Things (IoT). Terms like "cybersecurity," "network security," "intrusion detection," and "internet of things" are fundamental within this group. This shows a strong research emphasis on utilizing machine learning to improve security protocols in connected devices and networks. The use of terms such as "optimization" and "predictive models" indicates the utilization of machine learning methods to forecast and address security risks. This cluster highlights the important role that machine learning plays in creating strong cybersecurity solutions in a world becoming more interconnected.

Connections between different disciplines and collaborative research involving multiple fields.

The network diagram shows both separate clusters and connections between various areas of machine learning research. For example, terms like "optimization," "predictive models," and "performance" are present in numerous groups, showing their wide relevance in different fields. The interconnected clusters demonstrate the interdisciplinary method commonly seen in contemporary machine learning studies. Progress in fundamental methods and algorithms (red group) are utilized in specialized sectors like healthcare (green group) and cybersecurity (purple group), with advancements in deep learning (blue group) and NLP (yellow group) pushing advancements in various areas.

Current patterns and upcoming paths

In machine learning research, the network diagram provides an insight into emerging trends and possible future directions. Deep learning, healthcare, and cybersecurity are highly frequented disciplines, suggesting that these areas will continue to be important areas of study. A step towards developing intelligent systems that can effortlessly connect with both humans and machines is suggested by the combination of machine learning, IoT, and NLP. Furthermore, efforts to improve the efficiency and accuracy of machine learning models are continually demonstrated by the emphasis on optimising performance, using predictive modelling, and extracting features. As machine learning advances, researchers will likely focus on addressing challenges such as interpretability, scalability, and ethical difficulties to ensure that machine learning systems are not only robust but also just and dependable.

Co-occurrence and cluster analysis of the trending keywords in DL

Fig. 1.3 displays an examination of different keywords through co-occurrence and cluster analysis, providing understanding of the connections and importance of various concepts in deep learning studies. Central to the network diagram is the term "deep learning," situated prominently and linked to many other keywords, highlighting its pivotal position in the field. Deep learning, a component of machine learning, employs deep neural networks with multiple layers to examine diverse forms of data. The strong links surrounding "deep learning" demonstrate its essential significance and broad usage in many different fields.

Crimson Group: Educational Platforms and Usage

The red cluster has a high concentration of terms associated with learning systems and how they are used. The close connection between key terms such as "learning algorithms," "reinforcement learning," "learning systems," and "computational modeling" underscores the focus on advancing complex learning methods. Reinforcement learning is well-known for its ability to train models using reward-based learning methods, which are essential for tasks that involve making decisions in situations of uncertainty. This group also includes "neural networks" and "convolutional neural networks," which are important structures in deep learning. The importance of "convolutional neural networks" (CNNs) is highly notable, especially because of their extensive utilization in image and video recognition assignments. CNNs are crucial in handling visual data as they are linked to tasks such as "image enhancement," "object detection," and "object recognition."

Green Cluster: Applications Focused on Humans and Diagnostics

The green group is defined by key terms related to applications centered around humans and diagnostics. Words like "human," "adult," "female," "male," "diagnostic imaging," and "procedures" indicate a clear emphasis on medical and healthcare uses. This grouping suggests that deep learning methods are widely used in the examination of medical images, detecting illnesses, and enhancing healthcare results. Deep learning plays a crucial role in medical imaging, as evidenced by the frequent mention of terms such as "image segmentation," "diagnosis," and "nuclear magnetic resonance imaging" (MRI). Words like "important medical research," "group analysis," and "accuracy in identifying conditions" emphasize the use of deep learning in analyzing large data sets to find patterns and enhance diagnostic precision in clinical studies. The word "algorithm" is also often seen in this group, highlighting the creation of specific algorithms designed for medical use.

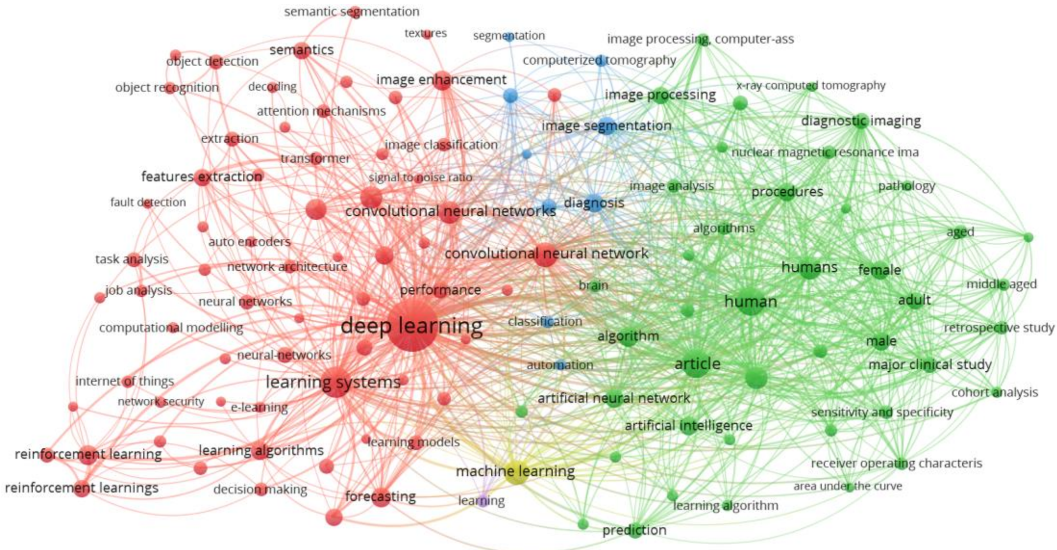

Fig. 1.3 Co-occurrence analysis of the trending keywords in DL

Blue Cluster: Processing and Analysis of Images

Terms like "image processing," "segmentation," "image classification," and "computerised tomography" are concentrated in the blue cluster because they are related to image processing and analysis. In domains like automated surveillance, remote sensing, and medical imaging, technological aspects of managing and interpreting visual data are the focus of this group. To increase the precision and effectiveness of picture analysis, sophisticated methods including "semantic segmentation" and "attention mechanisms" are used. For tasks requiring detailed understanding of visual scenes, semantic segmentation—such as identifying each pixel in an image—is crucial. Attention mechanisms, on the other hand, enable models to focus on significant portions of the data, increasing their efficacy in difficult tasks.

Golden Group: Fundamentals of Machine Learning

The small yellow group is important, emphasizing basic machine learning ideas. The central focus of this cluster is on keywords such as "machine learning," "learning," "prediction," and "algorithm." This indicates a close connection between deep learning and general machine learning principles, emphasizing how progress in deep learning is rooted in core machine learning theories and methods.

The link of this group to "forecasting" and "decision making" suggests the use of deep learning methods in predictive analytics and strategic decision-making processes, crucial in multiple sectors like finance, marketing, and operations management.

Connections and Evolving Patterns

The interconnectivity of various deep learning subfields is further illustrated by the network diagram, which displays multiple noteworthy linkages amongst clusters. One example of this is the association between "image segmentation" in the blue group and "CNNs" in the red group, which highlights the critical role that CNNs play in image processing tasks. Similarly, "diagnostic imaging" in the green cluster and "machine learning" in the yellow cluster are related, indicating that machine learning techniques are being used in medical diagnostics. Emerging trends in deep learning are also depicted in the network diagram. Keywords like "attention mechanisms," "auto encoders," and "transformer" demonstrate the growing importance of sophisticated designs and methods. Transformers, originally made popular in the field of natural language processing, are now being used more and more for different purposes such as image processing and analyzing time series data. Attention mechanisms are increasingly essential for improving model performance in various applications by enabling models to selectively emphasize relevant input components. Autoencoders, which are utilized for unsupervised learning and data compression, demonstrate the continual striving for enhanced model efficiency and effectiveness.

Emerging architectures in machine learning

Models based on transformer architecture

Transformer models, with self-attention mechanisms at their core, heralded a breakthrough in developing ML infrastructures (Sengupta et al., 2020; Bachute & Subhedar, 2021; Penney & Chen 2019). Initially designed for NLP, they have found promisingly effective use in wide applications. Their design based on self-attention mechanisms allows for the efficient handling of dependencies for both long and short distances in sequential data, something that recurrent neural networks (RNNs) had previously struggled with. A good example is the Transformer-based architecture which has enabled the recent models BERT and GPT to set state-of-the-art results for many NLP tasks, including translation, summarization, and question answering (Sengupta et al., 2020; Voulodimos et al., 2018; Jordan & Mitchell, 2015). These models were pre-trained on large enough datasets and fine-tuned on the target tasks. Transformer architectures have also been extrapolated to cater to visual tasks, showcasing scalability and generalization features. Vision Transformers (ViTs) apply the Transformer body on patches within images and perform equally with the CNN structure for tasks like image classification and object detection. The modularity and reusable concept bring to the fore the prospect of change within Transformer architectures in ML.

Graph Neural Networks (GNN)

GNN are powerful tools that have become popular for analyzing data structured in graphs (Shrestha & Mahmood, 2019; Alzubaidi et al., 2021; Deng, 2014; Sarker, 2021). GNNs differ from traditional neural networks in that they can process data in the form of graphs, which are commonly found in social networks, biological networks, and recommendation systems (Shinde & Shah, 2018; Özerol & Arslan Selçuk, 2023; Thakkar & Lohiya, 2021; Miotto et al., 2018). GNNs use message-passing methods, where nodes exchange information with neighboring nodes in a graph, enabling the network to understand the data's underlying structure. The capability to represent connections and interactions among entities makes GNNs highly efficient for tasks such as node classification, link prediction, and graph classification. Recent developments in GNN designs have been concentrating on enhancing scalability and efficiency. Strategies like GraphSAGE, which collect data from a set number of neighbors, and attention mechanisms in Graph Attention Networks (GATs), which give varying importance to different neighbors, have greatly improved the effectiveness and usefulness of GNNs in large-scale environments.

Neural Architecture Search (NAS)

NAS is a revolutionary change in neural network design as it automates the process of finding optimal architectures (Thakkar & Lohiya, 2021; Bashar, 2019; Lee et al., 2017; Moein et al., 2023). In the past, creating successful neural network structures involved a great deal of skill and hands-on trial and error (Dargan et al., 2020; Avci et al., 2021; Nguyen et al., 2018; Yap et al., 2019). NAS uses reinforcement learning, evolutionary algorithms, or gradient-based optimization to find the best architectures designed for specific tasks. One major accomplishment of NAS is the creation of EfficientNet, a series of models that excel in image classification benchmarks by using fewer parameters and requiring less computational resources. EfficientNet models were developed using a compound scaling technique found through NAS, which evenly scales depth, width, and resolution dimensions. NAS has also been expanded to uncover designs for specific tasks like object detection and semantic segmentation, resulting in models that surpass those designed by humans. As NAS techniques progress further, they offer the potential to make the design of neural network structures more widely available, thereby increasing accessibility to advanced ML models.

Architectures for Federated Learning

Federated Learning (FL) is a new approach that tackles privacy and data security issues in ML by allowing joint model training across various devices or organizations without consolidating data (Shrestha & Mahmood, 2019; Alzubaidi et al., 2021; Emmert-Streib et al., 2020; Patil et al., 2020; Kassem et al., 2021). In Florida, a worldwide model is trained by combining updates from local models trained on distributed data sources. Federated learning systems consist of important elements: client devices conduct local training, the server aggregates updates, and communication protocols guarantee secure and efficient data exchange. Recent developments in FL designs prioritize enhancing communication efficiency, managing non-IID data, and guaranteeing resilience against adversarial attacks. Federated learning is being used in different areas such as healthcare for collaborative research while protecting patient confidentiality, and finance for fraud detection without sharing customer data. With the increasing strictness of data privacy regulations, federated learning architectures are expected to have a significant impact on the future of ML.

Capsule Networks

Geoffrey Hinton and his colleagues created Capsule Networks, a novel architecture designed to get beyond the limitations of traditional CNNs. While standard pooling strategies employed in CNNs may not preserve the hierarchical connections between features, Capsule Networks aim to do just that (Dargan et al., 2020; Alzubaidi et al., 2021; Yadav & Vishwakarma, 2020; Angulakshmi & Deepa, 2021; Deng, 2019). The idea behind capsule networks is to represent different aspects of items or parts of objects with accuracy by using capsules, which are collections of neurones. Rather than generating single values, these capsules generate vectors, or matrices, that contain data about the presence and location of features. The capsules have a dynamic routing mechanism that allows higher-level capsules to get input from lower-level capsules that are relevant, which helps in preserving spatial and hierarchical relationships. Despite being in the early stages, Capsule Networks have demonstrated potential in tasks that involve intricate spatial comprehension, such as image recognition and 3D object reconstruction. Further investigation and improvement of Capsule Network structures may result in ML models that are more resilient and easier to interpret.

Self-Supervised Learning (SSL) architectures

The use of SSL is becoming more popular as a useful method for using unlabelled data to train machine learning models (Dargan et al., 2020; Özerol & Arslan Selçuk, 2023; Bashar, 2019). SSL architectures are made to solve automatically produced challenges called pretext tasks, which do not require human labelling, in order to acquire valuable representations from the data itself. Self-supervised learning architectures often follow a two-step process: pre-training on a large dataset with pretext tasks (e.g., guessing the next word in a sentence or missing portions of an image) and fine-tuning on a smaller labelled dataset for the target job. Simple Framework for Contrastive Learning of Visual Representations (SimCLR) and BERT models are two examples of how effective SSL is in identifying significant characteristics from large volumes of unlabelled data. The benefits of self-supervised learning architectures include improved data efficiency, as models can leverage large-scale unlabeled datasets, and enhanced generalization, as the learned representations capture more diverse and robust features. As the volume of unlabelled data continues to grow, self-supervised learning architectures are set to become increasingly important in the ML landscape.

Emerging architectures in deep learning

Transformers and Attention Mechanisms

Transformers have greatly altered the field of NLP (Shrestha & Mahmood, 2019; Dargan et al., 2020; Kassem et al., 2021). Transformers can process sequences in parallel, unlike RNNs and LSTMs, which depend on sequential data processing (Mishra et al., 2021; Sarker, 2021; Lee et al., 2017). Self-attention mechanisms allow the model to assess the importance of various words in a sentence without being limited by their positions. The self-attention mechanism computes attention scores between each word in a sentence, aiding in understanding context and improving accuracy in translations and text generation. Transformers have found success in areas other than NLP. Vision Transformers (ViTs) were introduced to utilize transformer models with image data. ViTs view an image as a series of patches and demonstrate top-notch results on multiple image recognition tests. This change in architecture showcases the adaptability of transformers and their ability to bring together different modalities within one system.

Graph Neural Networks (GNNs)

GNNs are an effective tool for extracting knowledge from data structured as graphs, commonly found in areas like social networks, molecular biology, and recommendation systems (Dargan et al., 2020; Özerol & Arslan Selçuk, 2023; Alzubaidi et al., 2021; Janiesch et al., 2021). Conventional neural networks face challenges when dealing with graph data because of its non-Euclidean characteristics. GNNs tackle this issue by working on the graph structure itself and utilizing message passing to collect data from nearby nodes. Recent developments in GNN structures involve Graph Attention Networks (GATs), which use attention mechanisms to dynamically assess the significance of nearby nodes. Another significant advancement is the introduction of Graph Convolutional Networks (GCNs), which extend the idea of convolution to graph information, allowing for the detection of nearby characteristics. These advancements have greatly enhanced the efficiency and flexibility of GNNs, establishing them as a crucial element of contemporary DL resources (Özerol & Arslan Selçuk, 2023; Alom et al., 2019; Khan & Yairi, 2018; Mu & Zeng, 2019; Miotto et al., 2018).

Neural Architecture Search (NAS)

Creating the best neural network structures typically involves a lot of knowledge and experimentation (Voulodimos et al., 2018; Jordan & Mitchell, 2015; Avci et al., 2021). Neural Architecture Search (NAS) speeds up this process by employing algorithms to seek out the most effective architectures. NAS methods utilize reinforcement learning, evolutionary algorithms, or gradient-based techniques to investigate a wide range of possible architectures. Recent developments in NAS are centered on improving efficiency and scalability. Sharing parameters among different architectures during training in Efficient Neural Architecture Search (ENAS) lowers the computational cost of NAS. DARTS (Differentiable Architecture Search) enhances efficiency by making the architecture search process differentiable, which enables gradient-based optimization. These methods make the design of neural networks accessible to more people, allowing them to find advanced structures without needing to manually adjust them extensively.

Self-Supervised Learning (SSL) Architectures

SSL is becoming a hopeful method for utilizing extensive amounts of data without labels (Shinde & Shah, 2018; Dargan et al., 2020; Bashar, 2019). SSL models produce labels based on the data, typically by performing tasks like predicting missing segments of an input or differentiating between modified versions of the input. This strategy of pretraining enables models to acquire detailed representations that can later be adjusted for particular tasks. New SSL frameworks like BERT and SimCLR have shown impressive results. BERT employs masked language modeling to anticipate the absence of words in a sentence and gather extensive contextual information. SimCLR uses contrastive learning to increase consensus between variously augmented perspectives of the identical image. These methods have greatly decreased the need for labeled data and enhanced the durability and generalizability of DL models.

Capsule Networks

Capsule networks, developed by Hinton et al., seek to overcome the shortcomings of conventional CNN in managing spatial hierarchies and maintaining part-whole relationships (Penney & Chen; 2019; Aziz et al., 2020; Yadav & Vishwakarma, 2020). Capsules are clusters of nerve cells that encode different characteristics of objects, like position, feel, and direction. Capsule networks employ dynamic routing algorithms to guarantee that capsules in lower layers activate the relevant capsules in higher layers. The structure of capsule networks enables them to identify objects even when they are seen in new positions or environments. Capsule networks are well-suited for tasks like object recognition and segmentation that need a thorough comprehension of spatial relationships due to this ability. Although still in its early phases, capsule networks show potential for improving the interpretability and resilience of DL models.

Sparse Neural Networks

Sparse neural networks tackle the inefficiency of dense neural networks by selectively removing connections between neurons. This leads to models with reduced memory and computational requirements, while still achieving or even enhancing performance. Methods for promoting sparsity involve weight pruning, which eliminates unimportant weights, and structured sparsity, which focuses on clusters of weights like whole neurons or filters. Current advancements in sparse neural networks are centered on dynamic sparsity, which involves altering the network's connections as it undergoes training. Techniques such as Sparse Evolutionary Training (SET) and RigL (Rigorous Lottery) adapt the sparsity pattern according to the learning process. These methods have demonstrated that sparse networks can perform just as well as dense networks, making them an appealing choice for environments with limited resources.

Hypernetworks and Meta-Learning

Hypernetworks and meta-learning architectures are designed to improve the flexibility and ability to generalize of neural networks. Hypernetworks create the weights for a main neural network by using the parameters of a secondary network, resulting in models that are more adaptable and concise. This method is especially beneficial for activities that need quick adjustment to new data or surroundings. Meta-learning, also known as "learning to learn," refers to the training of models that have the ability to rapidly adjust to new tasks using very little data. Approaches like Model-Agnostic Meta-Learning (MAML) focus on tuning initial parameters quickly. Meta-learning structures have shown success in scenarios of few-shot learning, a situation where conventional DL models face challenges because of a scarcity of training data.

Energy-Based Models (EBMs)

EBMs depict data as energy landscapes, with lower energy levels indicating higher likelihood of configurations. EBMs provide a structured approach for creating models that generate new samples based on the distribution that has been learned. Recent developments in EBMs are concentrating on enhancing their scalability and efficiency when dealing with high-dimensional data. A significant advancement is the Contrastive Divergence algorithm, which estimates the energy function gradient to train EBMs. Another method, Score-Based Generative Modeling, employs score matching to directly estimate the gradient of the data distribution. These progressions have sparked renewed enthusiasm in EBMs, presenting them as a feasible substitute for conventional generative models such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs).

Architectural innovations and enhancements in ML and DL

Architectures of Transformers

The creation of transformer models is considered one of the most important advancements in DL architecture (Shinde & Shah, 2018; LeCun et al., 2015; Deng, 2019). Transformers have completely transformed the field of NLP (Bashar, 2019; Moein et al., 2023). Transformers, in contrast to traditional RNNs and CNNs, utilize only self-attention mechanisms for processing input data. This enables them to better capture long-range connections and distribute computations in parallel, resulting in significant enhancements in training speed and model effectiveness. Transformers laid the foundation for models such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer). BERT's training in both directions helps it grasp context more effectively, which enhances its performance in tasks like answering questions and analyzing sentiments. However, GPT models like GPT-3 and the newly launched GPT-4 have shown remarkable skills in creating logical and contextually appropriate text, expanding the possibilities of NLP uses. Table 1.2 shows the key architectural innovations and enhancements in ML and DL.

Efficient Neural Networks

The need to utilize ML and DL models on edge devices with restricted computational capabilities has resulted in the creation of effective neural networks (Shrestha & Mahmood, 2019; Mu & Zeng, 2019; Kassem et al., 2021). Methods like model pruning, quantization, and knowledge distillation are utilized to decrease the size and computational complexity of neural networks while maintaining their performance. MobileNet, SqueezeNet, and EfficientNet are well-known architectures created to prioritize efficiency. MobileNet utilizes depthwise separable convolutions in order to decrease the amount of parameters and computations. SqueezeNet utilizes a fire module to reduce the size of the model while still preserving accuracy. EfficientNet, created by Google, effectively adjusts model size (width, depth, and resolution) to maintain a balance, delivering top-notch results using less parameters and computations.

Table 1.2 Key architectural innovations and enhancements in ML and DL

References	Architectural Innovation	Description	Enhancements	Key Applications
(Shrestha & Mahmood, 2019; Shinde	Convolutional Neural	A category of deep neural networks	Noteworthy advancements encompass escalated	Disciplines of interest encompass

& Shah, 2018; Dargan et al., 2020; Mu & Zeng, 2019; Aziz et al., 2020; Deng, 2014)	Networks (CNNs)	predominantly employed for scrutinizing visual stimuli.	precision in discerning and categorizing visual data, alongside refined extraction of features facilitated by convolutional strata.	image and video analysis, in addition to the burgeoning field of medical imaging.
(Alzubaidi et al., 2021; Janiesch et al., 2021; Wu & Xie, 2022)	Recurrent Neural Networks (RNNs)	Neural network architectures characterized by sequential interconnections, adept at processing time-series data.	Distinctive enhancements embrace proficient handling of sequential datasets, coupled with augmented efficacy in language modeling and translation facilitated by variants such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU).	Application domains span language modeling, speech recognition, and the domain of time-series prediction.
(Mishra et al., 2021; Lee et al., 2017; Moein et al., 2023; Avci et al., 2021)	Transformer Architecture	A novel model paradigm predicated upon self-attention mechanisms, supplanting traditional Recurrent Neural Networks (RNNs) in NLP tasks.	Prominent advancements encompass heightened parallelization efficacy and state-of-the-art proficiency in NLP domains, particularly exemplified in tasks such as machine translation and document summarization.	Key areas of application encompass the realm of NLP and text generation.
(Mishra et al., 2021; Wang et al.,	Generative Adversarial	A sophisticated framework comprising two	Preeminent enhancements encompass the	Fields of application include image

2020; Wu & Xie, 2022; Penney & Chen; 2019; Voulodimos et al., 2018)	Networks (GANs)	competing neural networks, namely the generator and discriminator, renowned for their prowess in generating high-fidelity synthetic data.	generation of synthetic data of unparalleled quality and fidelity, thereby engendering realistic images and videos.	and video synthesis, as well as data augmentation endeavors.
(Wu & Xie, 2022; Penney & Chen; 2019; Voulodimos et al., 2018; Yap et al., 2019; Emmert-Streib et al., 2020; Patil et al., 2020)	Autoencoders	Neural network architectures tailored for the acquisition of efficient data encodings, conducive to diverse applications including data compression and anomaly detection.	Salient enhancements comprise the refinement of data compression techniques and the detection of anomalies through effective encoding mechanisms.	Applications span domains such as data compression, anomaly detection, and feature learning endeavors.
(Shinde & Shah, 2018; Özerol & Arslan Selçuk, 2023; Alzubaidi et al., 2021)	Capsule Networks	A progressive augmentation of Convolutional Neural Networks (CNNs) designed to encapsulate spatial hierarchies within data representations.	Pioneering enhancements encompass superior generalization capabilities pertaining to the perception of object orientation and pose.	Noteworthy applications encompass image classification and object detection tasks.
(Voulodimos et al., 2018; Aziz et al., 2020; Jordan & Mitchell, 2015; Bashar, 2019)	Graph Neural Networks (GNNs)	Neural network paradigms tailored for operations on graph structures, demonstrating superior performance in	Notable enhancements include augmented proficiency in tasks necessitating analysis of relational data, such as social networks and	Applications extend to social network analysis, molecular biology, and recommendation systems.

		tasks involving relational data.	molecular structures.	
(Özerol & Arslan Selçuk, 2023; Alzubaidi et al., 2021; Alom et al., 2019)	Self-supervised Learning	A cutting-edge paradigm enabling learning from unlabeled data by predicting unobserved portions from the remainder, thus mitigating dependency on labeled datasets.	Distinctive enhancements encompass reduced reliance on labeled data and expedited pre-training across diverse tasks.	Application domains include NLP and computer vision tasks.
(Dargan et al., 2020; Alzubaidi et al., 2021; Alzubaidi et al., 2021)	Attention Mechanisms	Advanced methodologies facilitating model focus on pertinent segments of input sequences, thereby fostering improved interpretability and performance in sequence-to-sequence tasks.	Notable enhancements encompass heightened interpretability and efficacy in tasks such as language translation and summarization.	Applications span language translation, text summarization, and image captioning domains.
(Shinde & Shah, 2018; LeCun et al., 2015; Deng, 2014)	Neural Architecture Search (NAS)	An automated methodology geared toward the discovery of optimal neural network architectures, fostering enhanced efficiency in designing tailored models.	Noteworthy advancements encompass streamlined architecture discovery and improved task-specific performance.	Domains of application include automated ML and model optimization endeavors.
(Alzubaidi et al., 2021; Janiesch et	Residual Networks (ResNets)	Neural network architectures characterized by	Prominent enhancements include gradient	Applications span image classification

al., 2021; Aziz et al., 2020)		skip connections facilitating seamless gradient flow, thereby mitigating the vanishing gradient phenomenon and enabling training of deep networks.	stabilization and facilitation of training for deeper networks.	and object detection tasks.
(Sharma et al., 2021; Bal & Kayaalp, 2021; Wu & Xie, 2022)	Bidirectional Encoder Representations from Transformers (BERT)	A pre-trained language model leveraging transformer architectures for bidirectional understanding, showcasing state-of-the-art proficiency in NLP tasks.	Salient advancements encompass unparalleled proficiency in NLP domains and heightened context comprehension and representation.	Applications encompass question answering, text classification, and sentiment analysis tasks.
(Bal & Kayaalp, 2021; Thakkar & Lohiya, 2021; Deng, 2014)	Mixture of Experts	A sophisticated network architecture incorporating multiple specialized models and gating mechanisms, thereby facilitating enhanced model capacity and adaptability.	Distinctive enhancements encompass augmented model capacity and adept handling of diverse tasks and data distributions.	Applications span complex decision-making and multi-task learning domains.
(Shinde & Shah, 2018; LeCun et al., 2015; Yap et	Federated Learning	Training models on numerous decentralized devices without exchanging data.	Improved protection of data privacy and security. Enhanced cooperation among various data sources.	Healthcare, finance, IoT

	al., 2019; Deng, 2019)

Neural Architecture Search (NAS)

Automating the design of neural network architectures is the main focus of the emerging field of NAS (Shinde & Shah, 2018; Alzubaidi et al., 2021; Alzubaidi et al., 2021). NAS algorithms search through a large space to find the best architectures for specific tasks, instead of designing them manually. This method has resulted in finding structures that surpass models created by hand. One remarkable instance is the Efficient Net family, which was in part identified through NAS. The utilization of reinforcement learning and evolutionary algorithms in NAS has demonstrated encouraging outcomes, allowing the identification of innovative and effective structures. These structural designs have been used in a range of fields, such as image categorization, identifying objects, and developing language models.

Graph Neural Networks (GNNs)

GNNs are becoming popular because they can effectively handle data that is organized in graph-like structures. GNNs, unlike regular neural networks, can process data in the form of graphs, allowing for more intricate relationships and interactions to be analyzed. This makes them perfect for activities like analyzing social networks, suggesting systems, and studying molecular biology. Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), and GraphSAGE are some of the latest developments in GNN architectures. These models have shown better results in tasks such as node classification, link prediction, and graph classification (Shrestha & Mahmood, 2019; LeCun et al., 2015; Minar & Naher, 2018). GNNs are especially useful in scenarios with abundant relational data and can offer greater understanding of the fundamental structures present.

Attention mechanisms and self-supervised learning.

Attention mechanisms have become an essential part of contemporary DL structures (Sharma et al., 2021; Bal & Kayaalp, 2021; Yap et al., 2019). In addition to transformers, attention mechanisms have been incorporated into different models to improve their effectiveness (Voulodimos et al., 2018; Aziz et al., 2020; Jordan & Mitchell, 2015; Angulakshmi & Deepa, 2021). Self-supervised learning, which involves models learning from unlabeled data through predicting elements within the data, has become increasingly popular. Self-training models such as SimCLR, BYOL, and DINO have demonstrated impressive performance in computer vision assignments. These models use vast quantities of untagged data to acquire valuable representations, which can then be adjusted for particular purposes with minimal tagged data. This method decreases the reliance on annotated data, leading to improved scalability and efficiency in training models.

Federated Learning and Privacy-Preserving Techniques

Federated learning has emerged as a promising solution as data privacy becomes more important. Federated learning enables models to be trained on various decentralized devices without the need to share raw data. Instead of training models globally, they are trained on local devices, with only updates being shared to maintain data privacy and security. Significant progress has been made in federated learning, with uses in healthcare, finance, and IoT (Chauhan & Singh, 2018; Alzubaidi et al., 2021; Minar & Naher, 2018). Privacy and security in federated learning frameworks are being improved by incorporating methods such as differential privacy and secure multi-party computation. These developments allow for collaborative learning while upholding data privacy.

Hyperparameter Optimization

Optimizing hyperparameters is essential for attaining the best performance in ML and DL models. Older methods such as grid search and random search are being complemented by more sophisticated approaches like Bayesian optimization, hyperband, and population-based training. These approaches effectively navigate through the hyperparameter space to uncover setups that result in optimal performance. Automated hyperparameter tuning tools such as Optuna, Ray Tune, and Hyperopt are gaining more traction. These systems offer effective and adaptable solutions for adjusting hyperparameters, simplifying the process of creating top-performing models for users who want to avoid extensive manual testing.

Multi-Modal Learning

Multi-modal learning seeks to combine and interpret data from various sources like text, images, audio, and video (Voulodimos et al., 2018; Aziz et al., 2020; Jordan & Mitchell, 2015). Models such as CLIP and DALL-E have shown the power of multi-modal learning, utilizing both text and image data to produce imaginative and contextually appropriate results. These models use extensive pre-training on a variety of datasets to acquire detailed representations that can be adjusted for different tasks. Multi-modal learning enables new opportunities for use in content generation, virtual assistants, and human-computer interaction by emphasizing the importance of comprehending and analyzing various types of data.

Explainable AI (XAI)

With the increasing complexity of ML and DL models, there is an emerging demand for clarity and explainability. The objective of XAI methods is to ensure that model predictions are both comprehensible and reliable (Chauhan & Singh, 2018; Özerol & Arslan Selçuk, 2023; Alzubaidi et al., 2021). Approaches like SHAP, LIME, and attention-based explanations offer understanding on the decision-making process of models. XAI holds particular importance in sectors such as healthcare, finance, and legal systems, where the impact of model predictions is crucial. Recent developments in XAI concentrate on creating understandable models and improving the clarity of opaque models, guaranteeing that AI systems are both powerful and responsible (Minar & Naher, 2018; Topuz & Alp, 2023).

Meta-Learning

Meta-learning, also known as "learning to learn," is a developing approach that centers on teaching models how to rapidly acquire new tasks with limited data. This method is especially beneficial in situations with limited labeled data or when models must quickly adjust to new surroundings. Approaches like Model-Agnostic Meta-Learning (MAML) and Prototypical Networks have demonstrated encouraging outcomes in few-shot learning assignments. Meta-learning structures are being used in different areas such as robotics, custom recommendations, and healthcare, where the skill to generalize from limited data is essential.

Integration of Internet of Things (IoT), blockchain, and quantum computing with ML and DL

The combination of IoT, blockchain, and quantum computing with ML and DL shows a major merging of cutting-edge technologies, offering revolutionary effects in multiple industries (Sengupta et al., 2020; Alzubaidi et al., 2021; Mishra et al., 2021). This combination utilizes the benefits of each technology to tackle current obstacles and discover fresh opportunities, promoting innovation and improving efficiency, security, and intelligence in data-focused settings (Wu & Xie, 2022; Voulodimos et al., 2018; Dimiduk et al., 2018). The IoT system, marked by a large number of connected devices, produces huge amounts of data instantly. These various data streams, including environmental sensors, smart home devices, and industrial machinery, offer a valuable base for ML and DL algorithms. Through analysis of data generated by the Internet of Things, ML and DL models can extract practical information, improve processes, and facilitate preventative maintenance. In smart cities, combining IoT with ML and DL can improve traffic management, energy efficiency, and public safety through analyzing and learning from live sensor data (Sharma et al., 2021; Bal & Kayaalp, 2021; Patil et al., 2020). Likewise, within the healthcare sector, wearable gadgets and remote monitoring technologies can input patient information into ML models to anticipate health problems, customize treatments, and enhance results. Fig. 1.4 shows the integration of IoT, blockchain, and quantum computing with ML and DL.

Blockchain technology provides an essential level of security and confidence to the merging of IoT and ML/DL (Sharma et al., 2021; Bal & Kayaalp, 2021). One of the main issues with IoT is its susceptibility to cyber threats and security breaches, due to the large quantity of connected devices. The decentralized and unchangeable ledger of Blockchain offers a strong security solution for protecting IoT networks (Sharma et al., 2021; Khan & Yairi, 2018). Blockchain ensures the integrity and authenticity of data by securely recording transactions and data exchanges in an unalterable manner. When integrated with ML and DL, blockchain can enable safe sharing of data and cooperation among devices and systems, improving the dependability and credibility of the conclusions drawn. For instance, in the field of supply chain management, IoT devices are able to monitor products in real-time, with blockchain ensuring the authenticity and traceability of these items, while ML algorithms improve logistics and forecast potential disruptions.

Quantum computing, with its unmatched computation capabilities, has the ability to transform ML and DL (Shinde & Shah, 2018; Sengupta et al., 2020; Alzubaidi et al., 2021). Quantum computers have the capability to handle and examine large amounts of data at a significantly faster rate than traditional computers, which makes them well-suited for complicated ML and DL assignments that require complex data and intricate computations (Özerol & Arslan Selçuk, 2023; Mu & Zeng, 2019; Deng, 2014). Quantum Machine Learning (QML) and Quantum Deep Learning (QDL) have the potential to greatly speed up model training, enhance algorithms, and tackle problems that cannot be solved using traditional computing methods. In drug discovery, quantum computing can analyze large amounts of molecular data from IoT devices and simulate intricate chemical reactions more efficiently, leading to quicker identification of possible drug candidates. Furthermore, QML has the potential to improve cryptographic methods, offering increased security for blockchain networks. Exploring innovative applications already involves merging IoT, blockchain, and quantum computing with ML and DL. In the financial industry, utilizing a combination of these technologies can improve fraud detection, risk management, and trading strategies. IoT devices have the capability to observe transactions and market trends in real-time, while blockchain can protect transaction records, and ML models can anticipate fraudulent activities and enhance investment portfolios. Quantum computing has the potential to improve these abilities even more due to its ability to analyze financial data faster and with greater accuracy. Another area with a lot of potential is maintaining environmental sustainability. IoT sensors have the ability to track environmental conditions like air and water quality, while

ML models can examine the data to forecast pollution levels and pinpoint contamination origins. Blockchain can guarantee the openness and responsibility of environmental data, whereas quantum computing can enhance intricate environmental models and simulations, allowing for more efficient climate change mitigation plans.

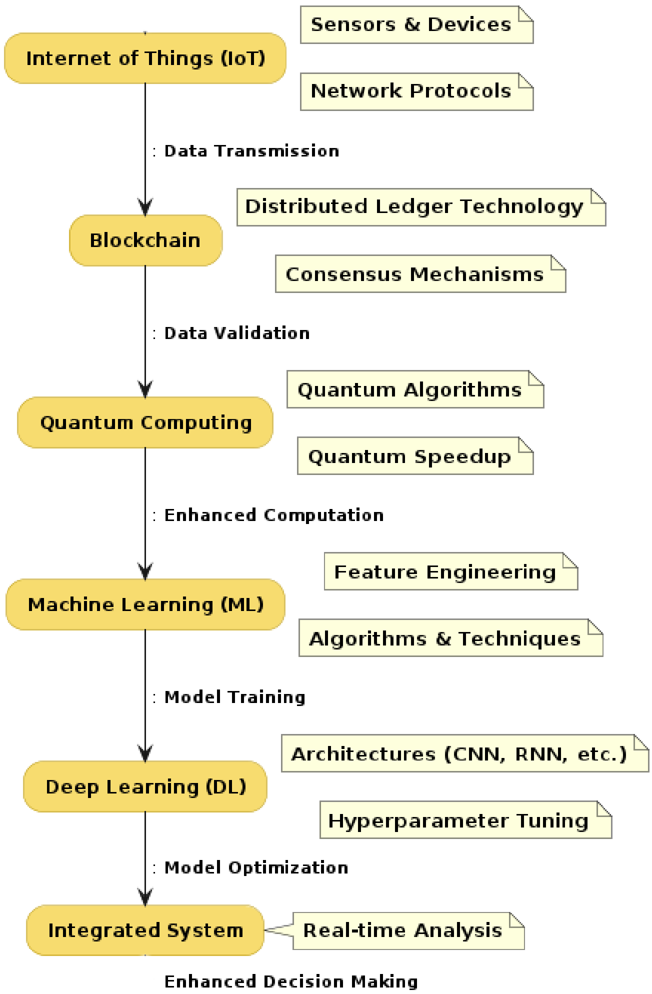

Fig. 1.4 integration of IoT, blockchain, and quantum computing with ML and DL

Even with great possibilities, combining these technologies comes with notable difficulties. Standardization and strong frameworks are necessary for ensuring interoperability among IoT devices, blockchain networks, and quantum computing systems. Making sure data privacy and security are top priorities in a complex ecosystem requires advanced cryptographic methods and following regulatory guidelines.

Furthermore, the full potential of blockchain and quantum computing systems cannot be realized without addressing issues related to scalability and energy consumption. Advancements in research and development in this industry are progressing quickly, powered by partnerships among academia, industry, and government agencies. Trial initiatives and experimental test runs are showcasing the possibility and advantages of combining IoT, blockchain, and quantum computing with ML and DL. For example, the healthcare industry is experiencing efforts that merge these technologies to enhance patient care, streamline operations, and strengthen data security.

ML and DL architectures for specialized applications

BIM is a crucial element of contemporary construction design and planning. ML and DL structures are improving BIM by allowing for more precise forecasts and simulations. Generative Design, fueled by ML algorithms, has the ability to investigate many design options according to set limitations and standards. This method assists architects and engineers in maximizing building designs for aspects like energy efficiency, structural strength, and visual attractiveness. DL models, specifically CNNs, are employed for recognizing and processing 3D images in BIM. These models enable the automatic identification and categorization of various building elements from 3D scans, improving the precision and productivity of design workflows. Integrating DL into BIM enables immediate incorporation of design changes into project plans, ensuring real-time updates and adjustments. The use of ML structures like Gradient Boosting Machines (GBMs) and Random Forests in predictive analytics is being implemented in project scheduling and management. These models use past project data to forecast possible delays and pinpoint important paths. Project managers can prevent delays by recognizing common causes and addressing issues early, which helps keep projects on track and within budget. NLP models are utilized for analyzing project documentation, communications, and reports. These models assist in detecting possible risks and areas of concern at the beginning of the project lifecycle by analyzing textual data. This proactive project management approach improves decision-making and resource allocation, resulting in more effective project implementation.

Safety is of utmost importance in the construction sector, with ML and DL frameworks playing a vital part in improving onsite safety. CNNs, particularly Computer Vision models, are utilized for monitoring safety in real-time. These models examine video feeds from construction sites in order to identify unsafe actions, malfunctioning equipment, and potential dangers. For example, CNNs are able to detect workers who are not wearing correct safety equipment or entering areas they are not allowed to, causing instant notifications to be sent to supervisors on site. DL models are utilized for predicting safety incidents by examining patterns in past safety data as well. RNNs and LSTMs are highly efficient because they can understand patterns and changes in data over time. These models help anticipate when and where accidents may happen, allowing for proactive measures to be taken, thus decreasing the likelihood of injuries and deaths at construction sites. Quality control plays a crucial role in ensuring construction projects adhere to necessary standards and specifications. DL architectures like CNNs are employed for automated detection of defects in construction materials and structures. These models examine construction element images and detect issues like cracks, deformations, or discrepancies instantly. Sophisticated designs such as U-Net and Mask R-CNN are employed for more accurate defect segmentation and identification. These models offer specific data on defects' size, location, and severity, enabling prompt and focused interventions. Automated defect detection enhances construction project quality and decreases costs linked to manual inspections and rework.

ML and DL architectures are having a large influence in the field of predictive maintenance. Through the analysis of sensor data obtained from construction equipment, ML models like Support Vector Machines (SVMs) and Random Forests have the capability to forecast potential machinery malfunctions. This enables prompt maintenance, avoiding unforeseen breakdowns and reducing downtime. DL models, specifically LSTM networks, are employed for more intricate predictive maintenance assignments. These models have the ability to examine time-series data from several sensors, recognizing patterns and irregularities that suggest possible equipment malfunctions. Construction companies can lengthen the lifespan of their equipment, enhance operational efficiency, and lower maintenance costs through the adoption of predictive maintenance strategies. Advancements in ML and DL technologies are leading to a rise in the use of autonomous construction equipment. Autonomous vehicles and drones utilize DL models for activities like conducting site surveys, transporting materials, and monitoring construction sites. CNNs, like Computer Vision models, allow machines to maneuver through intricate surroundings, steer clear of obstacles, and carry out tasks accurately. Reinforcement Learning (RL) structures are used to enhance the performance of self-operating construction machinery. Through their continual learning from the environment, these models can enhance the effectiveness and precision of construction tasks. The utilization of self-driving machinery improves efficiency and decreases the likelihood of incidents and harm at construction locations.

ML and DL structures are being used to evaluate and reduce the environmental effects of building projects. Forecasting the environmental impact of construction activities involves analyzing data on energy consumption, emissions, and waste generation using predictive models. This data is utilized to devise plans for minimizing harmful effects on the environment and advocating for sustainable building methods. DL models excel at analyzing vast amounts of data from different origins like satellite images and IoT sensors, in order to track changes in the environment over time. These models can detect trends and patterns showing environmental harm, allowing for prompt actions to safeguard and conserve natural resources. Companies can help global environmental efforts by integrating sustainability into their construction practices to tackle climate change and conservation.

Healthcare: predictive analytics and personalized medicine

In the field of healthcare, ML and deep learning designs are transforming predictive analysis and individualized healthcare. CNNs and RNNs stand out as notable in the field. CNNs have demonstrated remarkable success in diagnosing diseases like cancer from radiographic images due to their capability to process and analyze medical images. Specialized structures such as U-Net have been created specifically for the segmentation of medical images, allowing for accurate outlining of tumors and other entities. RNNs, along with more recent advancements in Long Short-Term Memory (LSTM) networks, are highly effective in processing sequential data, which makes them well-suited for examining patient records and forecasting the advancement of diseases. These models can predict possible health events by understanding temporal patterns in patient data, enabling timely interventions. Moreover, attention mechanisms and Transformers are more and more used due to their superior performance in handling sequential medical data, resulting in enhanced predictions of patient outcomes and personalized treatment plans.

Finance: fraud detection and algorithmic trading

In the financial industry, utilizing ML and DL structures is essential for duties like spotting fraud and engaging in algorithmic trading. GBMs and Random Forests are well-liked for their strength and ease of understanding when detecting fraudulent transactions. These models have the ability to analyze large amounts of data, picking up on intricate patterns that may signal fraudulent activity, all the while reducing the occurrence of false alarms. DL architectures such as Deep Reinforcement Learning (DRL) are becoming increasingly popular in algorithmic trading. DRL merges reinforcement learning with deep neural networks to make choices using previous trading data and present market conditions. Models like Deep Q-Networks (DQN) and Proximal Policy Optimization (PPO) have shown promising outcomes in creating trading tactics that adjust to market changes instantly, leading to increased profits and risk control.

Manufacturing: predictive maintenance and quality control

In the manufacturing industry, ML and DL structures are improving predictive maintenance and quality control processes. The goal of predictive maintenance is to forecast when equipment failures will happen in order to decrease downtime and maintenance expenses. Support Vector Machines (SVMs) and Random Forests are commonly utilized in this task, using past sensor data to anticipate potential machine failures. CNNs are leading the way in quality control because of their strong performance in tasks related to recognizing images. These models are capable of examining products along production lines and accurately detecting defects. More advanced designs such as ResNet and InceptionNet have enhanced the accuracy and dependability of automated visual inspections by utilizing deeper layers and complex methods like residual connections.

Retail: recommendation systems and demand forecasting

ML and DL architectures play a crucial role in recommendation systems and forecasting demand within the retail sector. Many recommendation systems are built on collaborative filtering and content-based filtering methods, which utilize matrix factorization and nearest neighbor algorithms. More advanced methods include utilizing DL structures such as autoencoders and neural collaborative filtering (NCF) to capture complex connections between users and products, ultimately improving recommendation precision. DL models like LSTMs and TCNs greatly improve demand forecasting. These designs are very good at detecting time-related trends in sales information, allowing retailers to forecast future demand more accurately. This results in improved inventory management, leading to cost reduction and increased customer satisfaction.

Autonomous vehicles: perception and decision-making

Advancements in autonomous vehicle technology depend greatly on ML and DL structures for interpreting data and making choices. Convolutional Neural Networks play a crucial role in the perception phase, as they analyze visual information captured by cameras to recognize objects and comprehend the surroundings of the vehicle. Architectures like YOLO (You Only Look Once) and Faster R-CNN are highly efficient in quickly detecting objects in real time, which is essential for autonomous vehicle operations. DRL is utilized for decision-making to allow vehicles to learn and adjust to different driving conditions. Through ongoing engagement with the surroundings, DRL models like DQN and A3C can acquire tactics for navigating, avoiding obstacles, and engaging with fellow road participants. This results in autonomous driving systems that are both safer and more efficient.

NLP: language understanding and generation

In the realm of NLP, ML and DL structures have made great progress in comprehending and producing language. Transformers like BERT and GPT have achieved recordbreaking results in a range of NLP tasks. BERT's bidirectional processing is particularly adept at grasping the context within text, making it exceptionally useful for tasks such as sentiment analysis, question answering, and named entity recognition. Conversely, GPT models are created specifically for generating text. GPT-4o in ChatGPT, the most recent version, enhances previous models with bigger datasets and more parameters for improved text generation that is coherent and contextually relevant. The uses of these models range from automated content generation to conversational AI, improving user experiences on different platforms.

Environmental monitoring: climate prediction and conservation efforts

Environmental monitoring is a field where ML and DL architectures are having a significant influence. Climate prediction models frequently utilize ensemble learning methods, merging various ML algorithms to enhance predictive precision. Random Forests, Gradient Boosting, and more recently, DL architectures like LSTMs, are employed for modeling intricate climate patterns and predicting weather fluctuations. CNNs and RNNs are crucial in conservation work for analyzing data from diverse sources like satellite imagery and sensor networks. CNNs are utilized for detecting deforestation, monitoring wildlife movement, and tracking changes in habitats, whereas RNNs are employed for forecasting endangered species migration based on time-series data. These findings are essential for creating successful conservation plans and reducing the effects of climate change.

Smart Cities: urban planning and infrastructure management

The idea of smart cities utilizes ML and DL frameworks for city planning and managing infrastructure. Utilizing SVMs, Decision Trees, and DL architectures such as LSTMs, predictive models are utilized to analyze data from IoT sensors for real-time monitoring and management of urban infrastructures like traffic systems, water supply networks, and energy grids. Generative Adversarial Networks (GANs) are growing in popularity for urban planning to model and improve city designs. GANs have the ability to create lifelike city settings using given data, offering planners visual representations and understanding to guide decision-making. This results in urban development that is both more effective and long-lasting.

Scalability and efficiency in ML and DL architectures

The fast advancement of ML and DL requires a strong emphasis on scalability and efficiency (Mishra et al., 2021; Alom et al., 2019). These two elements are essential for implementing ML and DL models in practical situations, where extensive datasets and intricate calculations are typical (Shrestha & Mahmood, 2019; Bal & Kayaalp, 2021;

Mishra et al., 2021). Scalability guarantees that models can manage growing data and computational load, while efficiency ensures that these models can complete their tasks rapidly and with minimum resource usage (Özerol & Arslan Selçuk, 2023; Alzubaidi et al., 2021; Alom et al., 2019). Recent progress in hardware, software frameworks, and algorithmic techniques has had a major influence on the scalability and efficiency of ML and DL architectures.

Advancements in hardware

The development of specialized hardware plays a key role in enhancing the scalability and efficiency of ML and DL architectures. GPUs, TPUs, and specialized accelerators have transformed the industry by offering the essential computing capabilities needed for fast training of large models. GPUs, known for their ability to process in parallel, are now widely used for training DL models, providing much faster speeds compared to traditional CPUs. Google created TPUs for the purpose of enhancing the performance of neural network calculations, particularly for specific DL tasks. Recently, the introduction of AI-dedicated hardware like neural processing units (NPUs) and field-programmable gate arrays (FPGAs) has improved the scalability and effectiveness of ML and DL models. These specialized chips have been specifically designed for the mathematical calculations needed for neural networks, resulting in notable enhancements in both performance and energy efficiency. Consequently, it is now possible to use intricate DL models in real-time tasks like autonomous vehicles and real-time language translation.

Distributed Training and Parallelism

Distributed training and parallelism are now necessary to manage large datasets and complex models. Distributed training splits the training process among multiple machines, enabling data processing and model parameter updates to occur simultaneously. Tools like TensorFlow, PyTorch, and Horovod have simplified the process of carrying out distributed training by utilizing multiple GPUs or TPUs on separate nodes. Model parallelism and data parallelism are the two main approaches used for distributed training. In model parallelism, various sections of the model are spread out over several devices, enabling the concurrent processing of distinct segments of the network. Conversely, data parallelism includes dividing the data among multiple devices, where each device handles a portion of the data and coordinates the updates. These methods greatly decrease the time needed to train big models, allowing for the use of datasets that were previously too large to manage.

Optimized Algorithms and Architectures

Advancements in algorithms have been key in improving the scalability and efficiency of ML and DL structures. Methods like gradient checkpointing, quantization, and pruning have been created to lessen the computational burden and memory usage of models while still maintaining performance. Gradient checkpointing includes the strategic preservation of intermediate outcomes throughout training, leading to decreased memory consumption while requiring some extra computation. This enables the use of bigger models with restricted hardware resources. Quantization decreases the accuracy of the model's parameters by changing them from 32-bit floating-point values to 16-bit or 8-bit integers. This lowers the memory needs and speeds up calculations, since operations with less precision are quicker. Pruning entails eliminating less crucial weights from the model, leading to a more sparse network that demands fewer computations and less memory. Both quantization and pruning are extremely helpful for implementing models on edge devices with restricted computational power and memory capabilities. The emergence of novel structures like transformers and GNN has also played a role in enhancing scalability and efficiency. Transformers like BERT and GPT-3 are known for their use of self-attention mechanisms to manage long-range connections in data, resulting in improved training efficiency and task performance in areas like NLP. In contrast, GNNs are created to analyze data organized in graph formats, which allows for improved handling of intricate connections in various types of data, such as social networks or molecular structures.

Software Frameworks and Libraries

The presence of powerful software frameworks and libraries has played a key role in improving scalability and efficiency. TensorFlow, PyTorch, and Apache MXNet are some of the most widely used frameworks that offer user-friendly APIs to streamline the creation and implementation of ML and DL models. These frameworks are created to make use of the features of current hardware, enabling support for GPU and TPU acceleration, along with distributed training. Specialized libraries like NVIDIA's cuDNN and Intel's MKL-DNN offer optimized versions of common neural network operations, enhancing performance along with these frameworks. These libraries are regularly maintained to utilize the newest hardware improvements, guaranteeing that ML and DL models can adapt effectively to emerging technologies.

Automated ML (AutoML)

AutoML is gaining popularity as a way to increase the efficiency and scalability of ML and DL models. Model selection, hyperparameter tweaking, and feature engineering are made easier by automated machine learning techniques, which also reduce the amount of time and knowledge required to develop successful models. The democratisation of ML and DL allows professionals to work on more complex and creative projects by enabling those without experience to use cutting-edge techniques. Frameworks such as Google's AutoML, Microsoft's Azure AutoML, and open-source libraries like Auto-sklearn and TPOT provide tools for automating various stages of the machine learning process. These technologies significantly reduce the computational resources and time required for model building by automatically identifying the optimal model architectures through techniques like neural architecture search (NAS).

Energy Efficiency and Sustainability

Concerns regarding the energy consumption and environmental impact of larger ML and DL models are growing. Large models like GPT-3,4, which require significant computational resources, lead to considerable energy usage and carbon emissions. Developing more energy-efficient models and training techniques is becoming increasingly important to researchers and professionals. In order to reduce computational requirements while maintaining high performance, techniques such as model distillation use a larger model (teacher) to train a smaller, more efficient model (student). Additionally, advancements in hardware—such as the development of energy-efficient chips and the use of renewable energy sources in data centers—are contributing to the promotion of more environmentally friendly methods in machine learning and deep learning.

1.4 Conclusions

Transformer-based models such as GPT-4 and BERT have caused a significant change in the field of NLP and consistently establish higher levels of performance. These models use self-attention mechanisms to better capture contextual information compared to traditional RNNs and CNNs. Another noticeable development is the merging of ML and DL with other up-and-coming technologies like IoT and blockchain. This coming together is strengthening ML and DL systems, allowing for stronger and safer applications in different industries. In smart cities, IoT devices produce large volumes of data which ML models examine instantly to enhance urban management and infrastructure. Furthermore, there is an increasing emphasis on ethical AI and explainable AI (XAI). With the rise of ML and DL systems, there is a growing need for AI decision-making processes to be more transparent and accountable. Scientists are working on XAI methods to guarantee that these models can be understood and their choices are clear to people, ultimately establishing confidence and enabling broader usage. In sectors like healthcare, finance, and autonomous systems, ML and DL are leading to major advancements in the realm of applications. DL models are utilized in healthcare for the early detection of illnesses, custom treatment strategies, and the development of new medications. In the field of finance, they are improving fraud detection, risk management, and customer service with predictive analytics. In the future, ML and DL are expected to have a transformative impact. Continuing research seeks to tackle present issues like data privacy, model interpretability, and computational efficiency. Moreover, the democratization of ML and DL tools is allowing a wider variety of industries and people to use these technologies, promoting a more inclusive technological environment. As ML and DL continue to develop, they will certainly have a crucial impact on shaping the future of technology and society.

References

Alom, M. Z., Taha, T. M., Yakopcic, C., Westberg, S., Sidike, P., Nasrin, M. S., ... & Asari, V. K. (2019). A state-of-the-art survey on deep learning theory and architectures. electronics, 8(3), 292.

Alzoubi, Y.I., Mishra, A. & Topcu, A.E. (2024). Research trends in deep learning and machine learning for cloud computing security. Artif Intell Rev 57, 132.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8, 1-74.

Angulakshmi, M., & Deepa, M. (2021). A review on deep learning architecture and methods for MRI brain tumour segmentation. Current Medical Imaging, 17(6), 695-706.

Avci, O., Abdeljaber, O., Kiranyaz, S., Hussein, M., Gabbouj, M., & Inman, D. J. (2021). A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. Mechanical systems and signal processing, 147, 107077.

Aziz, L., Salam, M. S. B. H., Sheikh, U. U., & Ayub, S. (2020). Exploring deep learning-based architecture, strategies, applications and current trends in generic object detection: A comprehensive review. Ieee Access, 8, 170461-170495.

Bachute, M. R., & Subhedar, J. M. (2021). Autonomous driving architectures: insights of machine learning and deep learning algorithms. Machine Learning with Applications, 6, 100164.

Bal, F., & Kayaalp, F. (2021). Review of machine learning and deep learning models in agriculture. International Advanced Researches and Engineering Journal, 5(2), 309-323.

Balaji, K., & Lavanya, K. (2018). Recent Trends in Deep Learning with Applications. In: Sangaiah, A., Thangavelu, A., Meenakshi Sundaram, V. (eds) Cognitive Computing for Big Data Systems Over IoT. Lecture Notes on Data Engineering and Communications Technologies, vol 14 . Springer, Cham. https://doi.org/10.1007/978-3-319-70688-7_9

Bashar, A. (2019). Survey on evolving deep learning neural network architectures. Journal of Artificial Intelligence, 1(02), 73-82.

Chauhan, N. K., & Singh, K. (2018). A review on conventional machine learning vs deep learning. In 2018 International conference on computing, power and communication technologies (GUCON) (pp. 347-352). IEEE.

Dargan, S., Kumar, M., Ayyagari, M. R., & Kumar, G. (2020). A survey of deep learning and its applications: a new paradigm to machine learning. Archives of Computational Methods in Engineering, 27, 1071-1092.

Deng, L. (2014). A tutorial survey of architectures, algorithms, and applications for deep learning. APSIPA transactions on Signal and Information Processing, 3, e2.

Deng, Y. (2019). Deep learning on mobile devices: a review. Mobile Multimedia/Image Processing, Security, and Applications 2019, 10993, 52-66.

Deng, Z., Wang, T., Zheng, Y., Zhang, W., & Yun, Y. H. (2024). Deep learning in food authenticity: Recent advances and future trends. Trends in Food Science & Technology, 104344.

Dimiduk, D. M., Holm, E. A., & Niezgoda, S. R. (2018). Perspectives on the impact of machine learning, deep learning, and artificial intelligence on materials, processes, and structures engineering. Integrating Materials and Manufacturing Innovation, 7, 157-172.

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., & Dehmer, M. (2020). An introductory review of deep learning for prediction models with big data. Frontiers in Artificial Intelligence, 3, 4.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets, 31(3), 685-695.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. Science, 349(6245), 255-260.

Kassem, M. A., Hosny, K. M., Damaševičius, R., & Eltoukhy, M. M. (2021). Machine learning and deep learning methods for skin lesion classification and diagnosis: a systematic review. Diagnostics, 11(8), 1390.

Khan, S., & Yairi, T. (2018). A review on the application of deep learning in system health management. Mechanical Systems and Signal Processing, 107, 241-265.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.

Lee, J. G., Jun, S., Cho, Y. W., Lee, H., Kim, G. B., Seo, J. B., & Kim, N. (2017). Deep learning in medical imaging: general overview. Korean journal of radiology, 18(4), 570.

Lundberg, L., Boldt, M., Borg, A., & Grahn, H. (2024). Bibliometric Mining of Research Trends in Machine Learning. AI, 5(1), 208-236.

Minar, M. R., & Naher, J. (2018). Recent advances in deep learning: An overview. arXiv preprint arXiv:1807.08169.

Miotto, R., Wang, F., Wang, S., Jiang, X., & Dudley, J. T. (2018). Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics, 19(6), 1236-1246.

Mishra, R. K., Reddy, G. S., & Pathak, H. (2021). The understanding of deep learning: A comprehensive review. Mathematical Problems in Engineering, 2021(1), 5548884.

Moein, M. M., Saradar, A., Rahmati, K., Mousavinejad, S. H. G., Bristow, J., Aramali, V., & Karakouzian, M. (2023). Predictive models for concrete properties using machine learning and deep learning approaches: A review. Journal of Building Engineering, 63, 105444.

Mu, R., & Zeng, X. (2019). A review of deep learning research. KSII Transactions on Internet and Information Systems (TIIS), 13(4), 1738-1764.

Nguyen, H., Kieu, L. M., Wen, T., & Cai, C. (2018). Deep learning methods in transportation domain: a review. IET Intelligent Transport Systems, 12(9), 998-1004.

Özerol, G., & Arslan Selçuk, S. (2023). Machine learning in the discipline of architecture: A review on the research trends between 2014 and 2020. International Journal of Architectural Computing, 21(1), 23-41.
Patil, T., Pandey, S., & Visrani, K. (2020). A review on basic deep learning technologies and applications. Data Science and Intelligent Applications: Proceedings of ICDSIA 2020, 565-573.
Penney, D. D., & Chen, L. (2019). A survey of machine learning applied to computer architecture design. arXiv preprint arXiv:1909.12373.
Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Computer Science, 2(6), 420.
Sengupta, S., Basak, S., Saikia, P., Paul, S., Tsalavoutis, V., Atiah, F., ... & Peters, A. (2020). A review of deep learning with special emphasis on architectures, applications and recent trends. Knowledge-Based Systems, 194, 105596.
Sharma, N., Sharma, R., & Jindal, N. (2021). Machine learning and deep learning applications-a vision. Global Transitions Proceedings, 2(1), 24-28.
Shinde, P. P., & Shah, S. (2018). A review of machine learning and deep learning applications. In 2018 Fourth international conference on computing communication control and automation (ICCUBEA) (pp. 1-6). IEEE.
Shrestha, A., & Mahmood, A. (2019). Review of deep learning algorithms and architectures. IEEE access, 7, 53040-53065.
Thakkar, A., & Lohiya, R. (2021). A review on machine learning and deep learning perspectives of IDS for IoT: recent updates, security issues, and challenges. Archives of Computational Methods in Engineering, 28(4), 3211-3243.
Topuz, B., & Alp, N. Ç. (2023). Machine learning in architecture. Automation in Construction, 154, 105012.
Voulodimos, A., Doulamis, N., Doulamis, A., & Protopapadakis, E. (2018). Deep learning for computer vision: A brief review. Computational intelligence and neuroscience, 2018(1), 7068349.
Wang, X., Zhao, Y., & Pourpanah, F. (2020). Recent advances in deep learning. International Journal of Machine Learning and Cybernetics, 11, 747-750.
Wu, N., & Xie, Y. (2022). A survey of machine learning for computer architecture and systems. ACM Computing Surveys (CSUR), 55(3), 1-39.
Yadav, A., & Vishwakarma, D. K. (2020). Sentiment analysis using deep learning architectures: a review. Artificial Intelligence Review, 53(6), 4Avci et al., 20215-4385.
Yap M. M., Tekerek, A., & Topaloğlu, N. (2019). Literature review of deep learning research areas. Gazi Mühendislik Bilimleri Dergisi, 5(3), 188-215.