Unveiling the frontiers of deep learning: Innovations shaping diverse domains

Shams Forruque Ahmed[1,2] · Md. Sakib Bin Alam[3] · Maliha Kabir[4] · Shaila Afrin[4] · Sabiha Jannat Rafa[4] · Aanushka Mehjabin[5] · Amir H. Gandomi[6,7,8]

Abstract

Deep learning (DL) allows computer models to learn, visualize, optimize, refine, and predict data. To understand its present state, examining the most recent advancements and applications of deep learning across various domains is essential. However, prior reviews focused on DL applications in only one or two domains. The current review thoroughly investigates the use of DL in four different broad fields due to the plenty of relevant research literature in these domains. This wide range of coverage provides a comprehensive and interconnected understanding of DL's influence and opportunities, which is lacking in other reviews. The study also discusses DL frameworks and addresses the benefits and challenges of utilizing DL in each field, which is only occasionally available in other reviews. DL frameworks like TensorFlow and PyTorch make it easy to develop innovative DL applications across diverse domains by providing model development and deployment platforms. This helps bridge theoretical progress and practical implementation. Deep learning solves complex problems and advances technology in many fields, demonstrating its revolutionary potential and adaptability. CNN-LSTM models with attention mechanisms can forecast traffic with 99% accuracy. Fungal-diseased mango leaves can be classified with 97.13% accuracy by the multi-layer CNN model. However, deep learning requires rigorous data collection to analyze and process large amounts of data because it is independent of training data. Thus, large-scale medical, research, healthcare, and environmental data compilation are challenging, reducing deep learning effectiveness. Future research should address data volume, privacy, domain complexity, and data quality issues in DL datasets.

✉ Shams Forruque Ahmed
    shams.forruque@northsouth.edu;
    shams.f.ahmed@gmail.com

✉ Amir H. Gandomi
    gandomi@uts.edu.au

1   School of Mathematical Sciences, Sunway University, Bandar Sunway, Petaling Jaya 47500, Selangor Darul Ehsan, Malaysia

2   Department of Mathematics & Physics, North South University, Dhaka 1229, Bangladesh

3   Department of Information Technology, University of Information Technology and Sciences (UITS), Chang Wat Pathum Thani, Dhaka 1212, Bangladesh

4   Science and Math Program, Asian University for Women, Chattogram 4000, Bangladesh
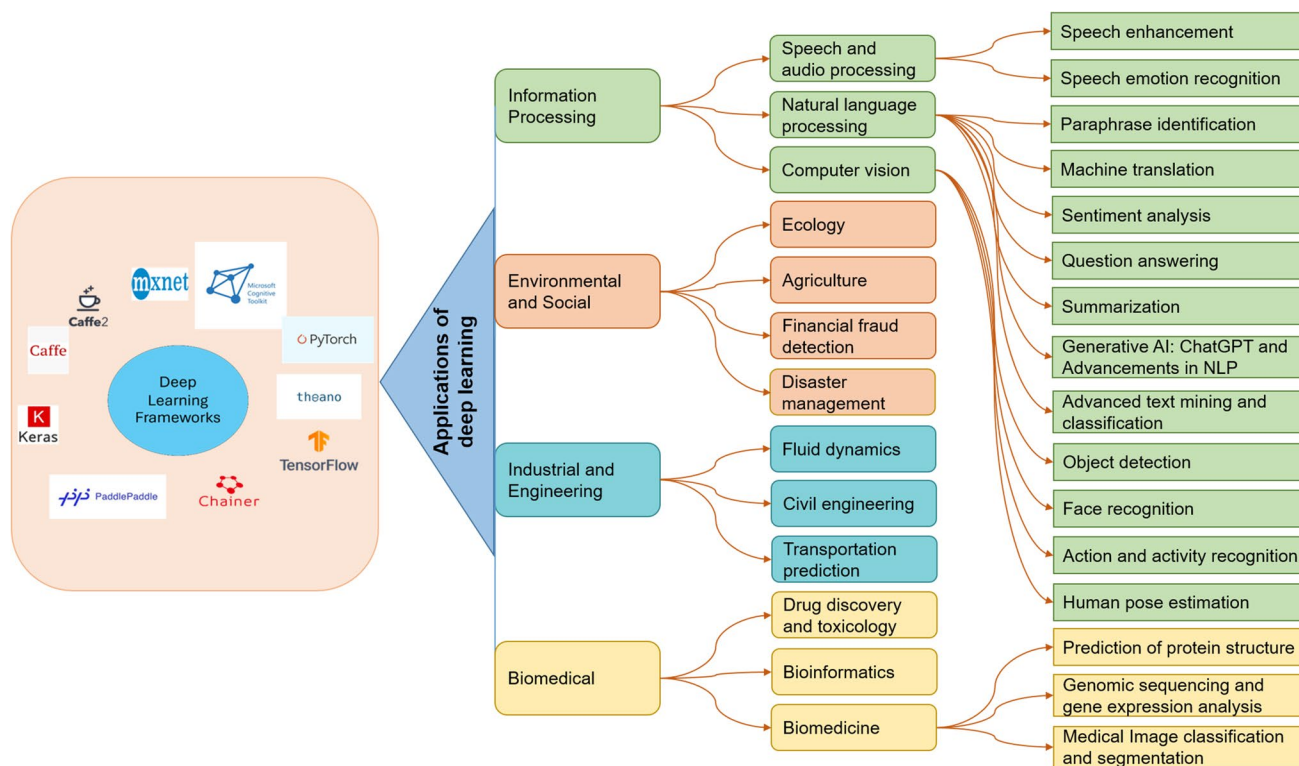
5   School of Biological Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA

6   Faculty of Engineering & Information Technology, University of Technology Sydney, Sydney, NSW 2007, Australia

7   University Research and Innovation Center (EKIK), Óbuda University, 1034 Budapest, Hungary

8   Department of Computer Science, Khazar University, Mahsati 41, Baku, Azerbaijan

Graphical Abstract

Keywords Agriculture · Computer vision · Deep learning · Deep learning framework · Deep neural network · Healthcare · LSTM · Natural language processing

Abbreviations

AI	Artificial intelligence
AMP	Antimicrobial peptide
ANN	Artificial neural network
API	Application programming interface
ATS	Automatic Text Summarization
Bi-LSTM	Bidirectional-LSTM
CAN	Collaborative adversarial network
CDN	Convolutional decoder networks
CEN	Convolutional Experts Network
CFD	Computational fluid dynamics
CNN	Convolutional neural network
CNTK	Computational Network Tool Kit
CPU	Central processing unit
DAM	Domain-attention model
DDI	Drug-drug interaction
DL	Deep learning
DMLC	Distributed Machine Learning Common
ECs	Emerging contaminants
EM	Electron microscopy
EPI	Enhancer-promoter interactions
ER	Estrogen receptor
FBE	Filter bank energies
GMM	Gaussian mixture model
GPU	Graphics processing unit
HAR	Human activity recognition
HDF	Hierarchical Data Format
LMT	Logistic Model Tree
LR	Linear regression
LSTM	Long short-term memory
HDFS	Hadoop Distributed File System
HPE	Human pose estimation
LDA	Linear discriminant analysis
MCA	Multiple correspondence analysis
MDNN	Multimodal deep neural network
MFCC	Mel Frequency Cepstral Coefficients
ML	Machine learning
NLP	Natural language processing
NMT	Neural machine translation
OTE	Opinion target expression
PCA	Principal component analysis
PRM	Pyramid Residual Module
QA	Question answering
RL	Reinforcement learning

RNN	Recurrent neural network
ROM	Reduced-order modeling
SER	Speech emotion recognition
SHBG	Sex-hormone binding globulin
SHM	Structural health monitoring
SNN	Spiking neural network
SNR	Signal-to-noise ratio
SVM	Support vector machine
TEO	Teager Energy Operator
TFX	TensorFlow Extended
TPU	Tensor processing unit
TSS	Text summarization embedding space
WS-MDL	Weakly-supervised multimodal deep learning

1 Introduction

Deep learning (DL) is a method for constructing computational models composed of numerous processing layers in order to investigate and learn the demonstrations of multiple abstraction data [1]. Utilizing these pathways, improvements have been made in various sectors, including audio-visual technology, better recognition and detection, genomics, proteomics, biomedicine, drug discovery, environment, and security. Deep learning unravels and identifies the underlying structures within massive datasets utilizing several algorithms, such as back-propagation, to learn and apply changes to given conditions as a machine would [2]. Deep learning exhibits advantages over earlier machine learning (ML) and artificial intelligence (AI) algorithms that lack the ability to analyze natural raw data. The origins of DL can be traced back to the 1940s, when the notion of artificial neural networks was developed, drawing inspiration from the complicated structure and functioning of the human brain [3]. Nevertheless, the proliferation of backpropagation algorithms and the accessibility of more extensive datasets were crucial factors in the 2000s that propelled deep learning to prominence [4]. The advancements during the 2010s propelled deep learning to the forefront of artificial intelligence research and application [5], such as introducing deep learning frameworks like TensorFlow and PyTorch and utilizing deep neural networks for image classification. Multiple statistical trends demonstrate the incredible expansion of deep learning. The deep learning industry has grown rapidly, with a market size of billions of dollars and a substantial compound annual growth rate. The broad use of deep learning in many sectors, such as healthcare, automotive, finance, and technology, fuels this expansion. The need for high-performance hardware, like graphics processing units (GPUs) and tensor processing units (TPUs), to facilitate the training and deployment of complicated neural networks highlights the continued significance of computational power. Statistical trends like these highlight how dynamic and influential deep

learning is becoming as its applications spread across more and more domains.

By utilizing methods like representation learning, a system is able to take raw data as input and determine the necessary patterns for analysis. Deep learning takes into account many layers of representation data, each of which affects the other. It is essential to examine the recent advancements and applications of deep learning in various fields. Recent progress in deep learning has driven significant advancements in various domains, from healthcare to finance. For instance, deep learning is improving medical diagnostics by utilizing image recognition, and it is transforming fraud detection and portfolio optimization in the field of finance. This demonstrates the wide-ranging influence and adaptability of this technology. The potential fields of DL applications include audio-visual data processing [6–8], agriculture [9–13], transportation prediction [14–16], natural language processing [17–19], biomedicine [20–22], disaster management [23–25], bioinformatics [26–28], healthcare [29–31], drug design [32–36], financial fraud detection [37, 38], computer vision [39–41], ecology [42–44], fluid dynamics [45–47], and civil engineering [48–50], to understand the present state of deep learning. The studies [6–50] reviewed applications of deep learning in a particular domain, while others concentrated on specific topics within that discipline. Most of them offered perspectives on DL application in their respective fields by discussing its advantages and disadvantages in light of future developments. Although various forms of analysis were made in many studies, the specific technical aspects of DL applications were not addressed. For instance, Litjens et al. [35] adopted a broad approach to medical image segmentation and reviewed many different subfields, which did not help to limit the topic. Hesamian et al. [36] shed light on the machine learning and artificial intelligence techniques utilized in recent biomedical image research, focusing on their structure and methodology and evaluating their benefits and limitations. They did not discuss the difficulties of black boxes and the inaccessibility of complex neural networks to human cognition. Getting a complete picture of deep learning's capabilities and limitations is challenging if reviews focus on one study area, as DL models may have issues generalizing across different domains. Additionally, this limited scope fails to account for the more extensive ethical, societal, and interdisciplinary factors vital for a comprehensive assessment of the impacts of deep learning. Therefore, the present investigation is undertaken to address these gaps by taking into account all of the aforementioned potential disciplines.

Several studies [51–54] evaluated features of deep learning-based audio-visual data processing, such as text analysis, activity recognition, and face recognition, without exploring the difficulties of black boxes or data privacy issues. Particularly, in sampling text and picture data, the processing

of personally identifiable information is subject to numerous legal constraints that are not adequately clarified. Sorting and analyzing biomedical data are crucial challenges in the health industry since it is complicated, varied, and dispersed. Health records, picture records, genomes, proteomics, transcriptomics, sensory data, and texts are just a few examples of the many types of data produced by the biomedical sector. To accurately forecast, represent, analyze, and sort this data, deep learning-based data mining approaches have proven to be both effective and fast. The usage of spiking neural networks (SNNs), which are constructed to emulate the information processing methods in biological systems, was reviewed by Pfeiffer et al. [55]. However, the study failed to address the issues posed by these systems, such as the fact that massive volumes of data input cause delays and fail to incorporate uncertainties. Many other reviews concentrating on deep learning techniques for health data overlook the challenges posed by poor data quality and excessive data volume [56]. Privacy problems of health data and genomic data are commonly disregarded in data mining approaches in biomedicine and bioinformatics [28, 57]. Yuan et al. [58] reviewed conventional neural network and deep learning approaches, which pertain to the development of environmental remote sensing procedures and their applications in environmental monitoring. Nevertheless, the authors did not explain how to deal with issues like incomplete or inaccurate data, complex domains, inaccurate models, or the difficulty of using a multidisciplinary approach to merge disciplines.

A search of the scholarly literature was performed to identify available studies on deep learning applications across numerous sectors, deep learning frameworks, and their associated challenges and benefits, utilizing an integrative approach, which comprises a broad collection, careful screening, and intensive assessment of pertinent and high-caliber articles. Searches were carried out utilizing databases sourced from reputable organizations such as Web of Science, Scopus, and Google Scholar, and well-reputed publishers like IEEE, De Gruyter, Elsevier, Nature, Taylor & Francis, Springer, ACM, Wiley, World Scientific, Inderscience, Oxford University Press, Emerald, IOP, Cambridge University Press, and Sage. The most pertinent publications for this study were identified through a search utilizing the following keywords: Deep learning, deep learning framework, deep learning application, use of deep learning, applications of deep learning, challenges/limitations/disadvantages of deep learning, benefits/advantages of deep learning, deep learning in a particular discipline, and similar terms. Then, by collecting and filtering the bibliographies and references of the aforementioned publications, additional relevant papers were identified. Through the search, it has been found that there have been many recent reviews [6, 10, 14, 18, 21, 23, 59–69] of deep learning applications. However, as seen in Table 1, the reviews focused primarily on applications of

deep learning in just one or two fields. In addition, there is currently no available review that investigates all possible applications of deep learning. For the advancement in deep learning research, it is crucial to understand and compare deep learning's performance across different fields. In order to assess the performance, application, progress, and challenges associated with deep learning, this study conducts a comprehensive review of its potential applications. This review took into account four different fields because there is plenty of relevant research literature in these areas where deep learning is widely applied. This report will serve as a valuable resource for academics and industrialists, offering comprehensive insights into the diverse practical applications of deep learning in various industries. Additionally, it will provide a deeper understanding of the potential challenges encountered during the deep learning implementation process.

The paper is structured as follows: Different types of deep learning frameworks are introduced and described in Sect. 2. Section 3 analyzes and explores the applications of deep learning in different sectors. The challenges of different DL application fields are summarized in Sect. 4, along with their benefits. Section 5 outlines the potential future research in order to minimize the identified research gaps in deep learning. A concluding review of the study is presented in Sect. 6.

2 Deep learning frameworks

Deep learning techniques can be referred to as a representation learning method, which has many layers of representation. These representation components can be created by building non-linear elements that can successively change the degree of representation into a subsequent abstract level. Deep learning can perform very complicated functions by combining these modifications. Many deep learning frameworks are used to develop solutions for complex problems, such as Caffe/Caffe2 [70–72], MXNet [73–76], Computational Network Tool Kit (CNTK) [77, 78], Torch/PyTorch [74, 75, 79, 80], Theano [81, 82], Chainer [83, 84], PaddlePaddle [85–87], Keras [75, 88, 89], and TensorFlow [74, 89–91], which are discussed in this section. Figure 1 shows several frameworks of deep learning.
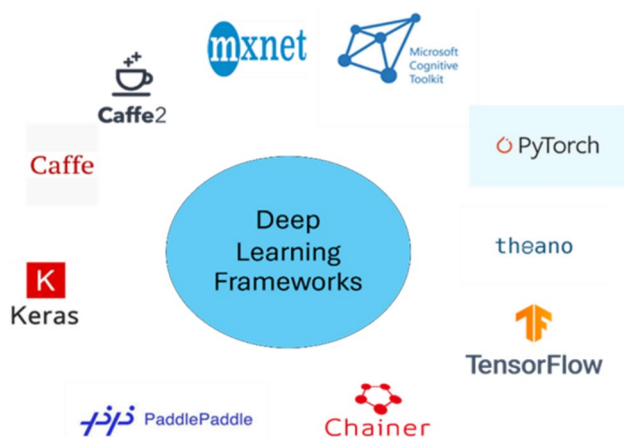
2.1 Caffe/caffe2

Caffe stands for convolutional architecture for rapid feature embedding. It is a free and open-source DL framework that allows users to explore complex structures. This library was created in C + + by the BVLC center and is often used in Matlab and Python [92]. Caffe utilizes declarative configuration for defining neural network designs, and it enables several layer types, such as fully connected, convolutional,

Table 1 Comparison between the present and the most recent review studies on deep learning applications

Review study	Main task (s)	Deep learning frameworks	No. of DL application fields	Challenges and benefits of DL
This study	Investigation of DL applications across all major fields of study, along with their associated benefits and challenges	√	√ (4)	√
Zhu et al. [6]	Examine recent developments in audio-visual learning	×	√ (1)	√
Wang et al. [10]	Explore DL utilization for analyzing hyperspectral images in agriculture	×	√ (1)	×
Wang et al. [14]	Investigate the DL techniques to improve the cognitive capabilities of transportation systems	×	√ (1)	×
Otter et al. [18]	Advances in natural language processing using deep learning	×	√ (1)	×
Haque and Neubert [21]	Implementation of deep learning techniques for biomedical image segmentation	×	√ (1)	×
Sun et al. [23]	Provides an overview of DL applications in disaster management	×	√ (1)	×
Shadmani et al. [59]	Predicting wave energy resources and optimizing the design of wave energy converter with the help of ML and DL	×	√ (2)	×
Kipli et al. [60]	Elucidate the utilization of deep learning techniques and their practical application in the domain of oil palm tree counting	×	√ (1)	√
Huang et al. [61]	Applications of DL in dentistry, with an emphasis on dental imaging	×	√ (1)	×
Erfanian et al. [62]	DL applications in genomics and transcriptomics	×	√ (2)	×
Adam et al. [63]	Use of DL to detect breast cancer based on the basis of magnetic resonance imaging	×	√ (1)	×
Lata and Cenkeramaddi [64]	Investigate the performance of DL methods in medical image cryptography to protect sensitive patient information	×	√ (1)	×
Hederman and Ackerman [65]	Using DL to enhance vaccine development	×	√ (1)	√
Yang et al. [66]	Outline the pathophysiological findings of pathogenic variants in neurological disorders' coding and noncoding transcripts using DL	×	√ (1)	√
Ruszczak et al. [67]	Applying DL to power grid diagnostics	×	√ (1)	×
Wani et al. [68]	Detection of brain tumors through the utilization of deep neural networks	×	√ (1)	√
Alqahtani et al.[69]	The accuracy of dyslexia prediction using deep learning models	×	√ (1)	√

×: Not availbale; √: Available

Fig. 1 Deep learning frameworks

and pooling layers. Both Caffe and Caffe2 are on the basis of feedforward neural networks, which use a non-linear activation function after a linear transformation in each layer [92]. The forward pass can be mathematically expressed as:

$$h^{(l)} = f(W^{(l)}h^{(l-1)} + b^{(l)}) \tag{1}$$

where $h^{(l)}$, $W^{(l)}$, and $b^{(l)}$ represent the activation vectors at the layer l; $f \cdot$ denotes the activation function such as sigmoid, ReLU, and tanh.

Caffe2 is an advanced version of Caffe created by Yangqing Jia, the same person who created Caffe. After Yangqing Jia began operating for Facebook, they collaborated with Facebook and NVIDIA to develop the Caffe2 framework based on Caffe [93]. Caffe can handle over forty million pictures daily with only a Titan GPU or K40 [94]. In Caffe, data are received by data layers. Because of the standard

picture formats of these data layers (.gif,.tiff,.jpg,.jpeg,. png,.pdf), Hierarchical Data Format (HDF5) and efficient databases are acceptable. If Caffe can be incorporated with cuDNN, it will enhance the productivity rate by 36% without using too much memory [95]. Although more recent frameworks have surfaced featuring distinct design philosophies, Caffe continues to be a valuable instrument for DL model development tasks that require simplicity and speed. Caffe2 addresses several of Caffe's shortcomings, including vast dispersed training, deployment in mobile systems, and additional equipment support for quantized processing. NVIDIA is an excellent support system for Caffe2, which includes Python and C++APIs [96]. This enables it to prototype and refine projects quickly.

Caffee and Caffe2 also have some limitations. Caffe's lack of thorough documentation makes it difficult for new users to get going or troubleshoot problems. Because of its more strict design, implementing custom layers or experimental architectures may be difficult. Its poor performance with big neural networks and recurrent neural networks (RNNs) causes complicated models to be handled laboriously and inefficiently. Caffe's development has slowed considerably, restricting its potential and rendering it less appropriate for contemporary applications that demand constant innovation. On the other hand, new users may find compatibility challenging in Caffe2 due to changes from Caffe. Its versatility for some applications is limited by its lack of support for dynamic graph calculations.

2.2 MXNet

MXNet is a multilingual deep learning framework made to achieve speed and flexibility [74]. Pedro Domingos and a team of scientists developed MXNet, which is also an addition of DMLC (Distributed (Deep) Machine Learning Common). This framework is highly scalable with a small memory footprint. The MXNet is constructed using tensor operations and has the capability to handle dynamic computational graphs. The primary function is the convolution, which can be represented as:

$$Y = W * X + b \tag{2}$$

In Eq. (2), X and Y denote the input and output tensors, respectively; W signifies filter tensor, $*$ indicates convolution operation; b denotes bias tensor.

MXNet can operate on a vast range of platforms, including GPU systems and mobile devices, and can perform numerical calculations with a short code of Python and R for dispersed networks and GPUs. It supports imperative and symbolic programming using NDArray API and Symbol API [97]. The MXNet platform offers extensive tools and libraries for constructing and implementing complex neural network models in diverse fields such as natural language

processing, computer vision, and reinforcement learning. The dynamic computation graph of MXNet facilitates convenient experimentation and fast prototyping, rendering it highly favored by both researchers and practitioners. MXNet, with the active participation of a lively community, continues to develop and adapt, ensuring its competitiveness in the ever-changing field of deep learning frameworks.

Despite these strengths, MXNet's limitations include compatibility issues with certain APIs, which might make integrating with other tools and libraries difficult. This can be a major disadvantage in large data pipelines or multi-framework ecosystems where smooth system-to-system communication is essential. Differences in design principles, version mismatches, or inadequate support for more recent features in third-party libraries can all lead to compatibility problems. MXNet provides dynamic graph computing and is compatible with many computer languages; nevertheless, new users may find the learning curve challenging. Furthermore, MXNet might not be as intuitive as other frameworks, making it less suitable for novices.

2.3 Computational network tool kit

Computational network tool kit (CNTK) is a deep learning framework that features Python API and is built on top of C++ code developed by Microsoft [98]. This framework represents neural networks as a sequential computational process. CNTK is renowned for its efficacy in deep neural network training and is specifically engineered to manage complex, extensive tasks. CNTK makes use of computation networks, in which nodes stand for operations and edges for data dependencies [98]. Matrix multiplication following an activation function is a common operation in CNTK:

$$y = \sigma(Wx + b) \tag{3}$$

where x and y represent the input and output vectors, respectively; W denotes the weight matrix; b signifies the bias vector; $\sigma(\cdot)$ defines the activation function.

CNTK's distinctive feature is its seamless integration with the deep learning tools and services offered by Microsoft, which positions it as a favored option among users operating within the Microsoft ecosystem. CNTK can readily combine some of the prominent architectures like Feed-forward deep neural networks (DNNs), convolutional neural networks (CNNs), and recurrent neural networks (RNNs). C# and BrainScript are supported by CNTK, which both offer increased and minimal APIs for simplicity of usage and versatility. By providing a performance-optimized backend and supporting acceleration on both CPU and GPU, CNTK is well-suited for various applications, including natural language processing (NLP) and speech and image recognition. For instance, CNTK was evaluated in a study using a system

of several GPUs in a thoroughly linked four-layer neural network [99] and was shown to outperform Caffe, TensorFlow, Theano, and Torch [100].

Because of its complexity, CNTK poses difficulties for beginners. Compared to more established frameworks, it has less resources and a smaller community because of its relatively recent inception. There can be a high learning curve to grasp CNTK, which limits its accessibility for individuals with little to no prior deep learning knowledge or for short development cycles. Even though CNTK has strong neural network tuning capabilities, reaching optimal performance can be challenging and needs a thorough understanding of the underlying workings of the framework. It may be necessary for users to do extensive performance tuning, which might be intimidating for those without advanced training or experience.

2.4 Torch/PyTorch

Torch is a DL framework written in the Lua programming language. A tensor or an array is used in many of the functions in the torch. These functions include memory exchange, indexing, slicing, and resizing [101]. The torch was designed by Facebook in 2017 and employs dynamic graphics to manage variable-length deliverables. PyTorch is the Python version of Torch [102]. Both PyTorch and Torch depend on dynamic computational graphs and tensors. The fundamental operation entails a linear transformation [103] that is subsequently followed by the activation function:

$$z = Wx + b \tag{4}$$

where x and z represent the input and output tensor, respectively; W denotes the weight tensor; b signifies the bias tensor.

PyTorch has gained considerable acceptance by providing a dynamic computation graph that streamlines imperative programming and permits users to dynamically modify models at runtime. It is written in C, Python, and CUDA and features acceleration libraries from Intel and NVIDIA. This feature has aided PyTorch's rapid growth and adoption among academic communities. By integrating seamlessly with well-known libraries and tools, PyTorch fosters a thriving ecosystem. The machine learning community has witnessed a significant increase in adoption due to its user-friendly API, robust community backing, and seamless transition capabilities from research to production. PyTorch has significantly influenced the development of DL frameworks, thereby bolstering the prevalence of dynamic graph-based methodologies.

Torch has several drawbacks, such as restricted scalability. Thus, it is best suited for smaller applications. PyTorch has come a long way in addressing scalability in production and distributed computing settings. However, it may still be difficult for some applications that demand large-scale deployment and substantial scalability to operate at peak efficiency. The Lua programming language, on which the original Torch framework was based, is becoming less widely used and popular. It is more difficult to locate resources, community support, and knowledge due to Lua's declining popularity in the programming community. This restriction may make development more difficult.

2.5 Theano

Theano is a Python deep learning framework that acts as a compiler for numerical expressions and allows developers and users to evaluate their mathematical process via NumPy's syntax [104]. This package was created at the lab of the University of Montreal. Some packages have been designed to increase the capabilities of Theano, such as Pylearn2 [105], Blocks, Lasagne [106], and Keras [107]. This framework facilitates the optimization, effective definition, and calculation of mathematical expressions that incorporate multi-dimensional arrays. Compiling mathematical expressions into highly optimized code, frequently executed on GPUs to enhance performance, is made possible by Theano's symbolic computation approach. Although Theano's development team actively developed and expanded its infrastructure, its development ended in 2017. Theano is a framework for symbolic computation that uses symbolic definitions of mathematical expressions [108]. Matrix multiplication is a frequently performed operation:

$$Z = dot(W, X) + b \tag{5}$$

In Eq. (5), X and Z represent the input and output tensor, respectively; W denotes the weight matrix; b defines the bias vector.

Since Theano is no longer being actively developed, it is not receiving updates or enhancements to take advantage of new developments in deep learning techniques and technology. Its applicability for cutting-edge research and applications that demand the latest features and optimizations is severely limited by this lack of continuous maintenance and development. Theano operates at a lower abstraction level compared to the modern framework. When building deep learning models, its low-level API might be challenging, making it less desirable than more recent frameworks with higher-level abstractions.

2.6 TensorFlow

TensorFlow is a deep learning framework that employs an individual data flow graph to define all mathematical operations to provide exceptional throughput [109]. It is an open-source framework that was developed by the Google brain

team. TensorFlow is distinguished by its static computation graph, which facilitates the optimization and efficient execution of complicated neural network architectures. Dataflow graphs are used by TensorFlow, with operations as nodes and tensors flowing along their edges [110]. A linear transformation and a non-linear activation are involved in the main operations:

$$y = f(Wx + b) \tag{6}$$

where x and y represent the input and output tensor, respectively; W and b define the weight and bias tensor, respectively; $f(\cdot)$ denotes activation function.

TensorFlow creates huge computational networks wherein every node in the graph refers to a mathematical function, and the edges indicate node interaction. This data flow graph specifically conducts interaction between a large part of the computational system, allowing for concurrent execution of separate calculations or the deployment of numerous devices to run partition operations [111]. TensorFlow Lite (enable low-latency inferences), TensorFlow serving (high performance, open source serving system), TensorFlow Hub (library for the publication, discovery, and consumption), and TensorFlow.JS are closely associated with or variants of TensorFlow (build and train models entirely in the browser) [112]. A general framework of TensorFlow components is illustrated in Fig. 2(a). TensorFlow constructs and executes these graphs using APIs from multiple programming languages, including Python, C++,

and Java. TensorFlow continues to develop, incorporating features such as TensorFlow Extended (TFX) (Fig. 2(b)) to facilitate end-to-end machine learning workflows and TensorFlow Lite to enable model deployment on mobile and edge devices, all in support of a large and active community.

The original low-level API for TensorFlow required users to handle variables, maintain sessions, and create computational networks manually. Creating models can be laborious and time-consuming due to this degree of detail, especially for users who are used to higher-level, more intuitive interfaces. The development process becomes more complicated when computation graphs must be manually created and run, frequently leading to lengthy and complex code. This low-level control might be a major obstacle for novices or those wishing to rapidly prototype models. Furthermore, despite its adaptability, TensorFlow's complexity may be challenging for consumers who want ease of use or quick development.

2.7 Chainer

Chainer is an open-source framework for deep learning distinguished by its adaptability and dynamic construction of computational graphs [83]. Chainer, developed by Preferred Networks, became well-known for its innovative methodology allowing users to dynamically define and alter neural network architectures while the program ran. In contrast to static computation graphs utilized by certain alternative frameworks, this dynamic approach offers practitioners and
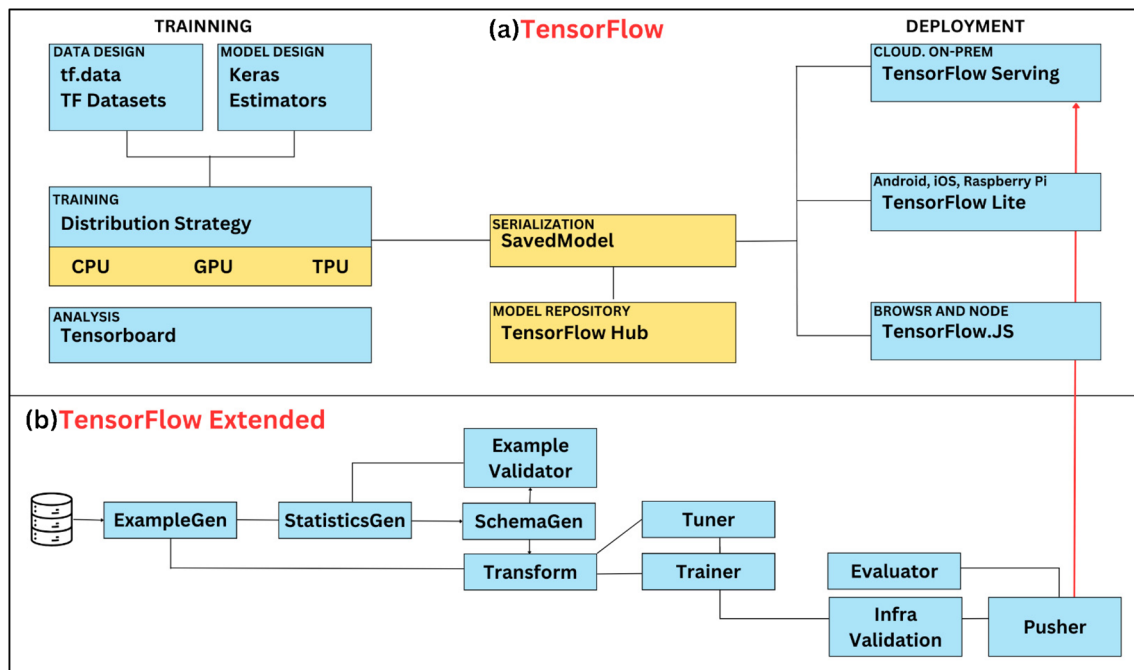

Fig. 2 **a** TensorFlow pipeline and key components; **b** TensorFlow extended pipeline

researchers considerable adaptability to facilitate experimentation and fast prototyping. With Chainer's "define-by-run" technique, computational graphs are built dynamically while the program executes [83]. An activation function is typically implemented subsequent to a linear transformation:

$$y = \sigma(Wx + b) \tag{7}$$

where x and y represent the input and output tensor, respectively; W and b define the weight and bias tensor, respectively; $\sigma(\cdot)$ denotes activation function.

Chainer is applicable to many applications due to its support for both CPU and GPU acceleration. A variety of DL models are offered by Chainer, such as reinforcement learning (RL), variational autoencoders, CNN, and RNN [91]. Chainer, renowned for its intuitive interface and imperative programming compatibility, has garnered extensive implementation across a spectrum of machine learning applications, encompassing natural language processing and image analysis. It has proven to be a valuable instrument for researchers investigating innovative architectures and algorithms within the ever-evolving domain of deep learning due to its user-friendly interface and dynamic graph construction.

The comparatively smaller community and ecosystem of Chainer in comparison to other deep learning frameworks such as TensorFlow or PyTorch is one of its primary limitations. Due to this restriction, users may have less access to help resources, including tutorials, discussion boards, and third-party libraries. Compared to extensively used frameworks, Chainer is rarely used in commercial and industrial environments. Its low industry adoption may lessen its appeal to companies seeking industry-standard tools. Enterprise applications may be better served by frameworks with a longer history and more community support.

2.8 PaddlePaddle

An open-source DL framework developed by Baidu [113], PaddlePaddle is also referred to as Paddle. PaddlePaddle is a highly adaptable framework specifically engineered to encompass various uses, including linguistic recognition, reinforcement learning, computer vision, and more. Its emphasis on supporting static and dynamic computation graphs is one of its distinguishing characteristics; this allows users to optimize and design neural network models with greater flexibility. A non-linear activation followed by a linear transformation is the basic operation [113]:

$$y = ReLU(Wx + b) \tag{8}$$

In Eq. (8), x and y represent the input and output tensor, respectively; W and b define the weight and bias tensor, respectively; $ReLU(\cdot)$ denotes the activation function.

PaddlePaddle architecture comprises four components: core framework, visual interface, scene suites, and algorithm suites [85]. The outstanding suitability of PaddlePaddle for large-scale and distributed training makes it a practical option for industry applications. The framework also provides users with an extensive collection of pre-built models and tools, which facilitate the process of development. PaddlePaddle's dedication to providing an intuitive and readily available interface has significantly improved its prominence, particularly within the Chinese deep learning community. In general, PaddlePaddle acts as a resilient and entire educational and research platform that facilitates the investigation and execution of state-of-the-art deep learning methodologies.

PaddlePaddle's limited acceptance outside of particular locations, especially given its strong correlation with Chinese technology ecosystems. Its use and visibility across the international development community may be limited by this geographical specialization. Because of this, developers and organizations in areas where alternative frameworks like TensorFlow or PyTorch are more popular might not choose it initially. In contrast to other widely used frameworks, this may limit its community support and the availability of pre-built models or tools.

2.9 Keras

Keras is a renowned open-source framework for deep learning due to its user-friendly and simple architecture [91]. Keras, originally designed to facilitate the construction of neural networks from other deep learning frameworks (e.g., TensorFlow), has undergone significant development and is now a vital component of the TensorFlow ecosystem. Its high-level API enables developers to rapidly prototype and experiment with neural network architectures, making it suitable for beginners and professionals [114]. The primary operation of Keras is characterized by layer-based abstractions [115]:

$$y = f(Wx + b) \tag{9}$$

where x and y represent the input and output tensor, respectively; W and b denote the weight and bias tensor, respectively; $f(\cdot)$ denotes activation function.

By removing numerous complications associated with lower-level implementations, Keras offers an efficient and user-friendly interface that facilitates developing, training, and deploying deep learning models. The seamless integration with diverse backends, such as TensorFlow and CNTK, is facilitated by its flexible nature. Keras allows users to concentrate on the design and structure of their neural networks without getting involved in complex implementation details [116], leading to its extensive use in research and

industry applications. This characteristic has enabled Keras to gain extensive acceptance in both academic and industrial sectors.

TensorFlow and Keras are generally interwoven, and although this connection has numerous advantages, there are drawbacks as well. Because of this close relationship, Keras users frequently use TensorFlow for backend processing. Because of this, integrating Keras with other deep learning frameworks may be difficult or time-consuming. For simplicity and quick development, Keras provides a high-level interface by abstracting many of the underlying intricacies of model building and training. For users who want precise control over their models, this abstraction may be a drawback. Due to its higher-level nature, tasks requiring precise tuning, the generation of custom layers, or certain optimization techniques could be difficult to accomplish in Keras.

Summary Contemporary artificial intelligence and machine learning applications are supported by deep learning frameworks, which offer programmers powerful tools for the construction, training, and implementation of neural network models. Caffe2, which is adaptable and mobile-compatible, enhances the efficiency of its predecessor, Caffe, which was established on the foundation of speed. MXNet distinguishes itself through its exceptional scalability and compatibility with multiple languages, whereas CNTK prioritizes peak performance and effective training. PyTorch and Torch, its successors, provide simple interfaces and dynamic computation graphs, supporting a diverse group of developers. Formerly an innovator in symbolic expressions, Theano's popularity has recently declined to be replaced by simpler alternatives. Chainer prioritizes adaptability, PaddlePaddle simplifies model construction, and Keras, now integrated with TensorFlow, emphasizes usability. TensorFlow is a foundation in the deep learning domain due to its extensive usage, scalability, and adaptability. Every framework possesses unique advantages and disadvantages, as outlined in Table 2, that impact its appropriateness for particular use cases and user choices. These frameworks collectively contribute to expanding and progressing deep learning technologies as the field develops. Table 3 explicitly represents the potential advantages of these frameworks.

Deep learning frameworks offer indispensable libraries and tools that enable developers and researchers to efficiently experiment with and implement deep learning algorithms, promoting innovation and breakthroughs in different disciplines. In order to evaluate the performance of deep learning frameworks across a variety of applications, it is necessary to take into account the particular requirements of each field. For instance, renowned for their adaptability and backing from the community, TensorFlow and PyTorch perform admirably in natural language processing, computer vision,

and bioinformatics. Agriculture may benefit from PaddlePaddle's emphasis on usability. Due to their robust libraries, TensorFlow, PyTorch, and Keras dominate drug discovery and financial fraud detection. Fluid dynamics simulations are frequently conducted using TensorFlow or PyTorch, whereas TensorFlow's serving capabilities render it well-suited for real-time endeavors within disaster management.

3 Recent advancements and applications of deep learning

Deep learning has addressed challenges that were considered to be impossible a few years back. Specifically, it has successfully handled issues that conventional machine algorithms failed to address; for example, deep learning has revolutionized image and video analysis, where traditional methods often struggle with variations in lighting, angles, and backgrounds. Because of its exceptional data handling feature, it has piqued the interest of professionals who are bombarded with all kinds of data. The advancement and application of deep learning (Fig. 3) have increased.

Natural language processing is one of the fields that has benefited significantly from deep learning, which is used in the various domains of natural language processing, like the translation of audio and machines [117]. In 2015, Google debuted the word lens identification engine based on deep learning. The best feature of a word lens is that it can read the texts instantaneously and convert them into the target language [118]. Another field where deep learning is being immensely used is transportation. As the world population increases, a smart transportation system is in demand. In the transportation sector, deep learning is being used in Destination Prediction [119], Demand Prediction [117, 118], Traffic Flow Prediction [120–122], Travel Time Estimation [123, 124], Predicting Traffic Accident Severity [125, 126], Predicting the Mode of Transportation [127], Trajectory Clustering [128, 129], Navigation [130, 131], Demand Serving [132, 133], and Traffic Signal Control [134, 135]. Implementation of DL in the bioinformatics domain has increased substantially over the years. In this area, DL has been mostly used to predict the structure of proteins [136, 137], gene regulations and expressions [138, 139], precision medicine[140, 141], biomedical imaging [142, 143], localization of cells [144, 145], clustering [27, 146], classification of proteins [147, 148] and so on. Techniques of DL have become one of the key elements in many multimedia systems [149]. Systems like Convolutional Neural Networks (CNNs) have presented noteworthy outcomes in various tasks from the real world, such as the detection of objects and processing of visual data, including images and videos. There have been numerous advancements, including the development of event detection from sports videos

Table 2 Advantages and disadvantages of the DL frameworks, with brief descriptions of each

DL frameworks	Brief description	Advantages	Disadvantages
Caffe	Caffe represents convolutional architecture for rapid feature embedding. It is a freely available DL framework that may be used to analyze complicated networks	– The code is ideally suited for research due to its flexibility – Caffe nicely functions with CNN architecture – It allows the adjustment of training models without writing perfect code	– Caffe does not have good documentation – It performs poorly and is cumbersome in big neural networks and RNN architecture – Its development has become slow
Caffe2	Caffe2 is an improved variant of Caffe. It fixes a number of issues with Caffe, including scalability, portability, and the lack of hardware support for quantized processing	– Caffe2 is a cross-platform deep learning framework. It can be used in mobile networking systems or edge computing systems – Caffe2 is said to be supported by Amazon, Intel, Qualcomm, and NVIDIA because of its scalability and robustness	– Compatibility with Caffe2 can be challenging for new users – It does not support dynamic graph computations
MXNet	MXNet is a fast and flexible deep learning framework developed for several languages. It has a small memory footprint while still being extremely scalable	– It supports high-level programming languages such as C++, R, Python, Scala, JavaScript, Perl, Go, Julia, Matlab, and Wolfram – It supports dynamic graph computations – Open Neural Network Exchange (ONNX) format also works in MXNet, making switching between other DL frameworks and libraries simple	– It is not compatible with some APIs
CNTK	As a deep learning framework, CNTK is written in C++ and supports a Python API. CNTK simplifies combining popular network architectures, including CNNs, Feed-forward DNNs, and RNNs/LSTMs	– CNTK provides dependable and superior functionality – It's compatible with Azure Cloud, which is supported by Microsoft – The control and use of resources are both efficient	– It is hard for beginners – As it's significantly new, it does not have vast community support
Torch/PyTorch	The Torch framework for deep learning was developed in Lua Many Torch operations depend on either a tensor or an array	– Torch contains a lot of effective extensions and support – It provides GPU support that is both quick and effective – It is research-friendly	– Only the smaller scale projects can be run on this – Lua has become a mid-level programming language; thus, it is not popular anymore
Theano	Theano is a numerical expression compiler for the Python deep learning framework NumPy, enabling both programmers and data scientists to assess the efficacy of their mathematical procedures in the context of the framework	– Theano is a cross-platform framework and open source – It provides support in GPU, and mathematical calculations can be done faster – It makes RNNs efficient because of symbolic API	– No longer under development – Low-level API – Theano is challenging to utilize explicitly for constructing deep learning models
TensorFlow	The TensorFlow deep learning system uses a single data flow graph to express all mathematical operations, resulting in extremely high throughput. The Google Brain team developed this open-source platform to help advance artificial intelligence	– TensorFlow is an open-source, popular, and promptly changing available framework – It has a powerful numerical library as the foundation for deep learning research – Efficiently work with numerical expressions which have a multidimensional array – It has great computational versatility across machines and large datasets	– A low-level API makes generating DL models tough

Table 2 (continued)

DL frameworks	Brief description	Advantages	Disadvantages
Chainer	Chainer is an open-source deep learning framework characterized by its dynamic construction of computational graphs and adaptability	– Chainer's performance for deep learning tasks is improved by its seamless integration with GPUs for rapid computation – It enables flexibility in model development and runtime modification – The intuitive interface of Chainer renders it suitable for users of all skill levels, including beginners	– Although Chainer experiences an active user community, it may lack resources and community support compared to more widely adopted frameworks – In comparison to other DL frameworks, its utilization in industry may be comparatively infrequent
PaddlePaddle	PaddlePaddle stands out for its support for both static and dynamic computation graphs. This gives users more flexibility when optimizing and designing neural network models	– PaddlePaddle is well-suited for managing complex models and large datasets because it supports large-scale and distributed training – It simplifies the development process through its user-friendly interface, comprehensive tools, and pre-built models	– Its adoption may be comparatively limited in scope and recognition beyond specific regions
Kera	Keras is widely recognized as an open-source DL framework due to its user-friendly and simple architecture. It was designed to facilitate the construction of neural networks from other deep learning frameworks, such as TensorFlow	– Keras offers a high-level API that simplifies intricate operations, thereby facilitating the development and training of neural networks by users with minimal coding – Keras's compatibility with multiple backends, including TensorFlow and Microsoft Cognitive Toolkit, provides users with adaptability and the ability to operate in diverse ecosystems	– Although Keras demonstrates adaptability, its integration with alternative deep learning platforms may be hindered by its close association with the TensorFlow ecosystem – It may not provide the same degree of low-level control and flexibility for advanced research as alternative frameworks

Table 3 Comparative analysis of various deep learning frameworks listing potential advantages

Potential Advantages	Caffe	Caffe2	MXNet	CNTK	Torch/PyTorch	Theano	TensorFlow	Chainer	PaddlePaddle	Keras
Flexibility for research	√	√	√	√	√	×	√	√	√	√
Ease of use for beginners	×	×	×	×	√	×	×	√	√	√
Supports dynamic graph computation	×	×	√	√	√	×	√	√	√	√
Good documentation and community support	×	×	√	×	√	×	√	×	×	√
Scalability and distributed training	×	√	√	√	√	×	√	×	√	×
Cross-platform compatibility	√	√	√	√	√	√	√	√	√	√
High-level API availability	×	×	√	×	√	×	×	√	√	√

×: Not available; √: Available

[150], new techniques like recurrent convolution networks (RCNs) for video processing[151], the use of intermediate CNN layers, and Gated Recurrent Units for improving the sparsity and locality in modules.

3.1 Information processing

3.1.1 Speech and audio processing

Speech and audio have been proven to be two important modes of communication in human history. Deep learning technologies have made significant progress in the field of speech processing. Unlike traditional speech enhancement techniques that rely on a statistical model, deep learning models are data-driven. Deep learning is among the most important technologies nowadays and, thus, merits its study. Mapping-based and masking-based are DL approaches used in speech enhancement [152]. Moreover, DL methods for speech emotion recognition have several advantages over conventional machine learning methods, such as identifying complex systems and no requirement for manual extraction of the features [153]. Also, deep learning has the ability to handle a large amount of unlabeled data and has a proclivity for obtaining low-level characteristics from provided raw data. Furthermore, many academics have abandoned traditional signal-processing approaches for sound production due to the emergence of deep learning algorithms. Deep learning techniques have accomplished eloquent voice generation, audio textures, and melodies from simulated instruments [154]. Deep neural networks have shown excellent progress in audio processing.

Speech enhancement Speech enhancement is utilized in video conferencing, hearing aids, microphones, audiovisual screening, and so on. In order to have a good speech enhancement system, pattern mining is needed [155]. Processing background noise is considered a good speech augmentation method. There are different types of conventional machine learning approaches available to filter and eliminate the additional noise from the voice signals. In recent years, DL methods have proven beneficial for speech augmentation. Deep neural networks (DNNs) can learn and approximate complex patterns in the data due to the nonlinearity introduced by activation functions. The model's ability to generalize, manage neuron activation levels, and adjust to the particular demands of various tasks and architectures is further impacted by the activation function selected. Many activation functions may be employed in deep neural networks for speech enhancement; the selection often relies on the particular demands of the required tasks and the nature of the data. Speech enhancement applications frequently employ the exponential linear unit, sigmoid and hyperbolic
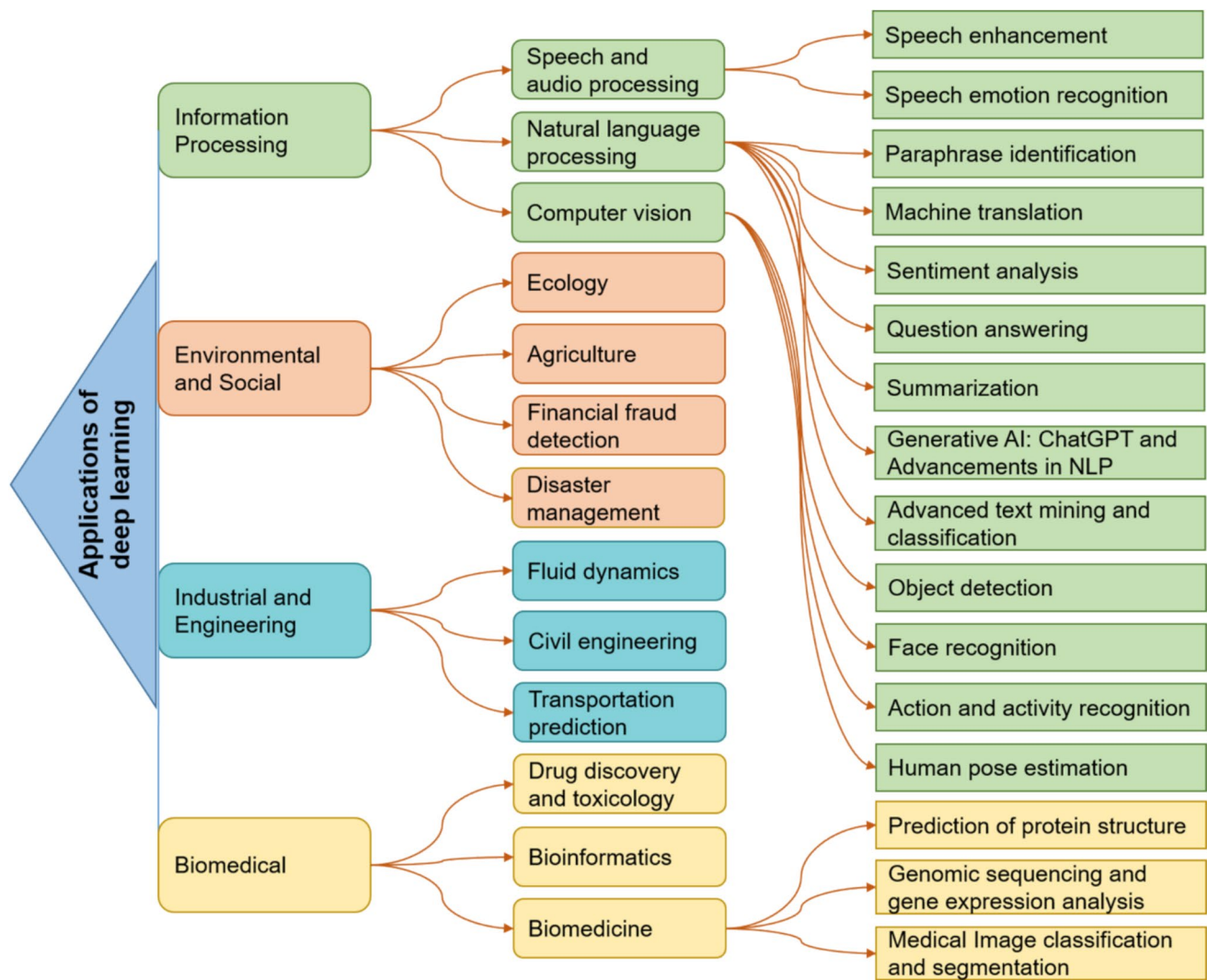
Fig. 3 Applications of deep learning in potential sectors

tangent, GRU and LSTM, and rectified linear unit as activation functions.

The DNN is the most general technique to remove noises from large datasets [156]. However, research shows that DNN struggles to adapt to new voice corpora in low signal-to-noise ratio (SNR) settings [157]. Accordingly, an alternative solution is to utilize a lower frameshift in short-time speech processing that can substantially improve cross-corpus generalization. Noise suppression is a well-developed field of signal processing [158], yet it still heavily relies on the precise tweaking of estimator techniques and variables. In a study by Valin [159], hybrid deep learning technologies were proposed to eliminate noise from the background. In this technique, a great emphasis is placed on minimizing complications while obtaining highly improved speech. This method proves more viable than a standard lowest mean squared error spectrum estimator.

In order to achieve a significant speech enhancement performance, a large DNN is needed, which is computer- and memory-intensive [160]. Therefore, making such a speech enhancement system is challenging because of timing constraints and minimal hardware resources. One study introduced two compression models for DNN-based speech enhancements, mainly including quantization to minimize model size based on clustering, sparse regularization, and repeated pruning [161]. Moreover, experimental results indicate that this method decreases the size of four distinct models by a substantial amount without affecting their enhancing performance.

The DNN model was also used by Mukhutdinov et al. [162] to evaluate the efficiency of speech enhancement by reducing drone ego-noise. Twelve illustrative DNN

models were trained, spanning three different architectures (encoder-decoder, sequential, and generative) and three different operation domains (end-to-end time, time–frequency complex, and time–frequency magnitude). In scenarios with extremely low signal-to-noise ratios (SNRs), ranging from −30 dB to 0 dB, the models' performance was assessed. While competing well on measures of speech enhancement, time–frequency complex domain and UNet encoder-decoder offered a reasonable compromise with respect to other metrics, including model size, context length, and computational complexity. The UNet model performed the best in the time–frequency complex domain and was able to increase ESTOI from 0.1 to 0.4, PESQ from 1.0 to 1.9, and SI-SDR from −15 dB to 3.7 dB, all with an input SNR of −15 dB.

Speech emotion recognition Speech emotion recognition (SER) refers to identifying the underlying emotion in a text or speech regardless of the semantic content. Implementation of deep learning in SER has made it possible to detect emotion in a speech in real-time situations [163]. Emotions can generally be categorized and organized into different groups, such as happiness, anger, sadness, and even a neutral emotional mood. Several researches have aimed to enhance the SER technique using deep learning methods. In a study, a dynamic SER identification system was developed based on aural and visual data processing [164]. In this system, a total of 88 characteristics (Mel Frequency Cepstral Coefficients (MFCC), filter bank energies (FBEs)) are employed to minimize the earlier retrieved features using principal component analysis (PCA) [165]. A comparison of deep learning flow versus conventional machine learning flow techniques for SER is shown in Fig. 4.

Surekha [167] utilized Gaussian mixture model (GMM) classifiers in the SER system to detect five emotions by applying the Berlin Emotional Speech database. An acoustic feature called the MFCC with the Teager Energy Operator (TEO) was used as a prosodic. This approach was employed in another study where 13 MFCC special features derived from the audio data were utilized to identify seven emotions [168]. In this method, the Logistic Model Tree (LMT) method was used, which has a 70% accuracy rate. Nevertheless, there are some challenges related to SER that need to be addressed. For instance, an issue with emotional speech datasets is annotation ambiguity [169]. In a specific task like picture classification, a car will always be labeled as a car, yet in an emotional discourse, asserting something strongly may be categorized as anger. This bias in categorization complicates the work while limiting the ability to combine datasets and generate emotional supersets.

A study by Hema et al. [170] introduced a SER system that demonstrated superior performance compared to an earlier system in terms of data collection and feature selection
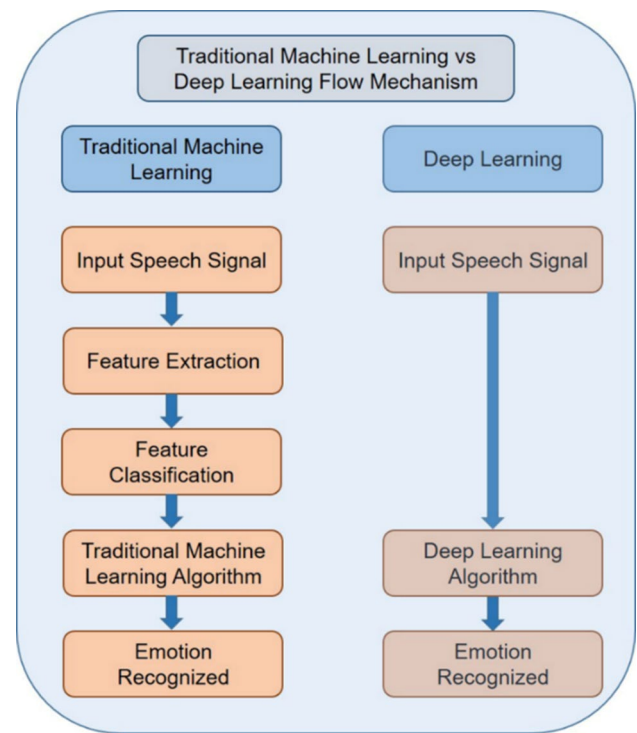

Fig. 4 Traditional machine learning vs. deep learning flow process (modified from[166])

techniques. The study introduced the concept of SER and looked at how emotions play an important role in human communications. As mentioned earlier, this technology analyzes audio signals to extract and predict the emotional state of a speaker. The SER task utilized support vector machine (SVM), radial-basis function (RBF), and back propagation ML models. These sophisticated models outperformed current systems' performance in collecting data and selecting feature techniques, resulting in an accuracy rate of 78% with a reduced number of false positives. Feature normalization can mitigate the impact of speakers' diversity on recognition.

3.1.2 Natural language processing

Deep learning algorithms are rapidly being implemented in natural language processing (NLP) research. Popular ML algorithms, such as SVM and logistic regression, trained on sparse and high-dimensional features have been the foundation of ML approaches for NLP issues for decades. On a range of NLP tasks, neural networks based on dense vector representations have recently surpassed classical models. This trend has been fueled by significant advancements in word embedding and DL approaches [117]. Classical NLP relies on predefined features and protocols, while deep learning NLP trains directly from data representation, which makes it more adaptable and capable of handling complex

tasks. The abstract framework of both classical and DL-based frameworks is demonstrated in Fig. 5.

Word embedding models based on Bidirectional Encoder Representations from Transformers (BERT) have shown effective for several NLP tasks. For instance, three BERT implications, such as regular, ClinicalBERT, and BioBERT, were performed in a study by Turchin et al. [171]. The study evaluated the accuracy of three medical concepts that are linguistically complex. These concepts were i) patients and healthcare providers discussing bariatric surgery, ii) patients not accepting statin treatment recommendations, and iii) documentation of tobacco use status. Compared to the regular BERT trained exclusively on Wikipedia, ClinicalBERT and BioBERT achieved better NLP performance for identifying various complicated medical concepts. This is due to the fact that they are trained on documents from the biomedical domain. The two biomedical BERT implementations tested did not achieve a significantly higher level of accuracy than the other.

Paraphrase identification Identifying whether two statements written in natural language have comparable semantic meanings is known as identification. When two sentences have the same meaning, they are referred to as paraphrases. This methodology is a fundamental approach in various data mining approaches that has tremendous potential in various domains, including plagiarism detection, machine translation, and others. Various techniques of paraphrase identification have been presented, which are distributed into two categories: similarity-based and classification-based methods [172]. Alzubi et al. [173] applied a collaborative adversarial network (CAN) in order to discover the problem of paraphrasing. A common feature extractor is included in CAN to help increase the association between phrases in the recurrent neural network model. Within a word pair, the extractor looks for related features. The integration of adversarial networks with collaborative learning mostly enhances the generator and discriminator working together. The model outperforms the baseline MaLSTM model as well as several baseline approaches. Hunt et al. [174] investigated several ML approaches for modeling the problem of paraphrase identification and alternative input encoding schemes. Since RNN outperformed the other models, the researchers developed a method based on the LSTM, an RNN variation. RNN works by using the output of past time sequences to create the output for next time sequences. This fits in NLP because it can be illustrated as a sequence of tokens, and the appearance of the next token is generally influenced by the preceding tokens.

Machine translation NLP's most well-known application is machine translation, which entails translating texts from one language to another using mathematical and computational procedures. The translation is challenging for humans since it requires knowledge of syntax, semantics, and morphology, as well as a full understanding and evaluation of cultural sensitivity for the target language and its associated communities. Ahmed et al. [175] advocated that attention
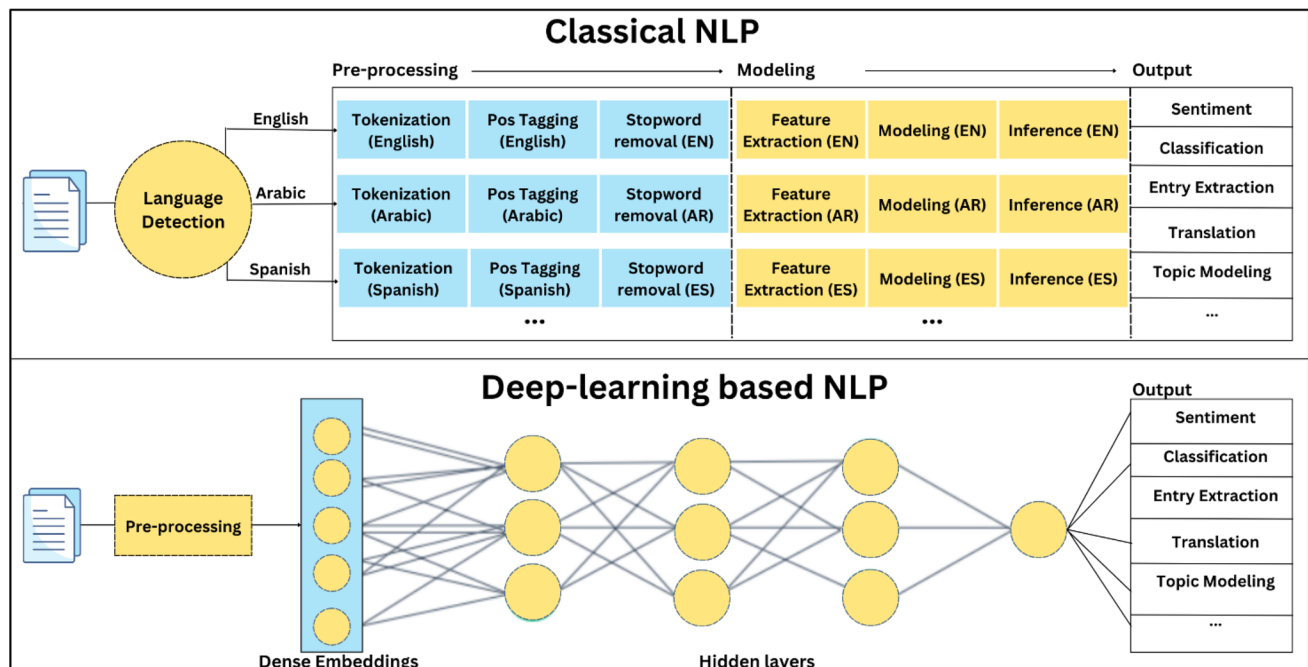
Fig. 5 Framework of classical NLP and DL-based NLP

methods can be used to encode a language from input to output instead of employing a huge number of recurrent and convolutional layers. The following three principles motivate the use of "self-attention" methods over traditional layers: decreasing the complexity of calculations needed for each layer; limiting sequential training steps; and lastly, reducing the length of the path between input and output and its influence on the learning of long-range relationships, which is important in several sequencing tasks.

Johnson et al. [175] reported that a single, basic (but huge) neural network could be applied to translating a number (minimum 12) of several languages into each other. This neural network automatically recognizes the source languages and uses just one input token to determine the output languages. When numerous language tokens are supplied, the method is shown to be capable of interpreting multilingual inputs and producing mixed outputs, at least in part, sometimes even in languages similar to but not identical to those chosen. The performance of such zero-shot translation is frequently insufficient to be practical, as the basic pivoting strategy quickly outperforms it.

A deep-attention technique was introduced by Zhang et al. [176] for a neural machine translation (NMT) system. This model incorporates many stacked attention layers, each of which pays attention to a corresponding encoder layer to advise what should be transmitted or repressed from the encoder layer, resulting in learned distributed representations that are suitable for high-level translation tasks. English-French, NIST-English-Chinese, and WMT14 English-German translation skims were used in this experiment. The technique could be implemented on other tasks like summarization and adapted to more complex attention models. Aside from the work listed above, other academics have suggested a variety of high-performance designs. In the English-German and Chinese-English translation systems, Zhang et al. [177] introduced the Variational NMT approach, which has a novel approach to modeling translation problems. The results demonstrated that it performed
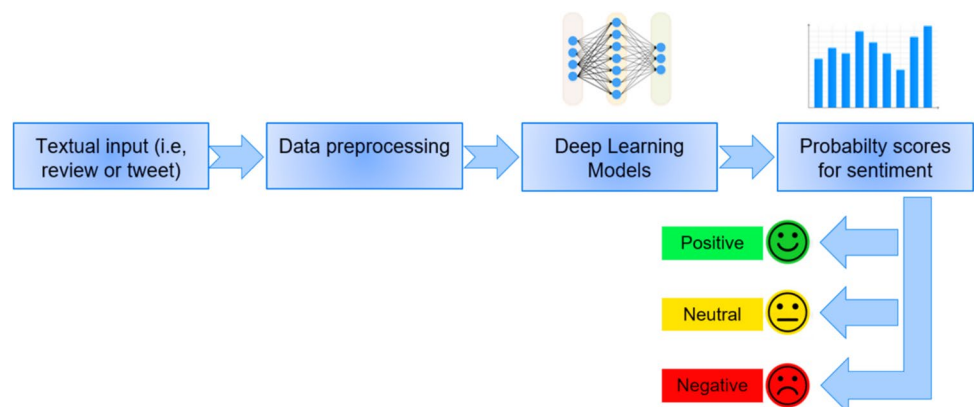
better than the baseline technique. A fast-forward connection for RNN was developed by Zhou et al. [178], which allows for a deeper network in implementation and, hence, greater performance.

Sentiment analysis Users today create massive volumes of data largely and dynamically as the number of internet technologies grows. The number of people who use social media on a regular basis is continuously expanding. In this sense, sentiment analysis appears to be a useful technique for automating the extraction of insights from user-generated data. Deep learning algorithms have recently been presented for several sentiment analysis applications, with the existing results [179]. A deep learning-based sentiment analysis pipeline is depicted in Fig. 6. Text input, such as a review or tweet, is analyzed by a neural network equipped with an attention mechanism. The output is a sentiment distribution displayed as a bar graph, indicating probabilities for emotions such as positive, neutral, and negative. The pipeline demonstrates how deep learning models process textual data to derive meaningful insights.

For extracting aspect opinion target expressions (OTEs), Al-Smadi et al. [180] suggested combining BiLSTM with a CRF method for classifying, as well as for classifying aspect-sentiment polarity. An aspect-based LSTM was used, where the aspect opinion target expressions were applied as attention expressions. A two-stage attention structure was created by Ma et al. [181] that involves paying attention to both the words that make up the target expression and the complete phrase. The author also used extended LSTM to construct a common sense model for target aspect-based emotion detection, which can use external knowledge. However, these methods could not represent several aspects of sentences, and the explicit location contexts of words were not investigated.

A path aggregation relational network for aspect-based sentiment analysis (ABSA) was proposed by Ma et al. [182] that is based on abstract meaning representation (AMR).

Fig. 6 Deep learning-based sentiment analysis pipeline

This model used the AMR semantic structure, which is better in line with the sentiment analysis problem, as opposed to the old ABSA model, which used syntactic structures like dependency trees. The model utilized BERT, a well-known transformer-based model, to encode sentences. Specifically, the authors designed the relation-enhanced self-attention mechanism and the route aggregator, which work well together. Sentence information guides the route aggregator in extracting semantic features from AMRs, and the relation-enhanced self-attention mechanism refines sentence characteristics using this extracted semantic information. When tested against state-of-the-art baselines, experiments on four open data sets showed that AMR-based path aggregation relational network (APARN) improves ABSA by an average of 1.13%. However, the level of accuracy of the AMR parsing results continues to slightly impact the model's performance.

Sentiment analysis has become extensively utilized in microblogging platforms like Twitter in the past decades due to the concise and simple nature of expression, allowing millions of users to share their opinions and thoughts. The BERT exhibits greater comprehension compared to traditional models due to its ability to process all input data through the encoder. Bello et al. [183] proposed to use BERT for text classification in NLP, along with other different variants. After comparing the results with Word2vec and no variant, the experimental results showed that combining BERT with CNN, RNN, and BiLSTM performed better in terms of accuracy, precision, recall, and F1-score. However, the conventional approach of NLP that employs Word2Vec, CNN, RNN, and BiLSTM possesses a few constraints, including the inability to capture the word's deeper context.

By incorporating diverse modalities (e.g., text, images, and audio), multimodal learning approaches improve sentiment analysis by capturing complex expressions and contextual information. This not only expands the informational scope but also enhances the overall performance of DL models by capturing a wider range of media formats and sentiments. Multimodal learning, employing both shared neurons for general activities and specialized neurons for specific tasks, revolutionizes deep learning models by integrating information from various sources like text, images, and audio [184]. The shared neurons capture overarching patterns and general features common across modalities, fostering knowledge transfer. Concurrently, specialized neurons focus on task-specific nuances, refining the model's understanding of particular objectives. This dual neural architecture enhances the overall effectiveness of deep learning models, providing versatility, adaptability, improved robustness, and generalization.

Question answering Question answering (QA) is the process of extracting words, phrases, or sentences from documents that are relevant. In response to a request, QA logically delivers this information. The approaches are similar to the summarizing methods presented in the next section. Wang et al. [185] utilized an attention-based LSTM to associate the questions with answers containing paragraphs. By mapping the full text, a self-matching attention process was employed to develop the machine representations. The position and boundaries of responses were predicted using pointer networks. The networks utilized attention pooling vector representation of passages, and the words were examined to model the crucial tokens or phrases. Modeling more than one aspect simultaneously with the attention mechanism would be an interesting addition to the experiment. Using Wikipedia as the knowledge source, Mozannar et al. [186] investigated the topic of open domain factual Arabic QA. They proposed an approach composed of A BERT document reader and hierarchical TF-IDF document retriever. They also introduced a dataset (ARCD). However, ARCD's questions were created with certain paragraphs in mind; without that context, they might seem ambiguous.

CNNs were compressed with LSTM to speed up processing [187]. For the decomposition of fully connected layers in LSTM and CNN, the researchers suggested applying different decomposition algorithms and regression techniques. Tensor Regression layers replace the Flattening and Fully Connected layers in the final section of the method. The Tensor Contraction layer compresses the flow of features between the layers to compress the parameter further. Determining the rank is a hard NP problem in low-rank decomposition, and their method is still constrained in this area by inserted hyper-parameters. Yu et al. [188] used a tree-LSTM method for capturing a language's linguistic structures. The study created a semantic tree for each question in the dataset, with every node referring to a single LSTM unit and the root node representing the sequence. To boost reasoning capacity, this approach can divide problems into several logical phrases. The representative ability of the network could be improved in the future.

Generative AI: ChatGPT and Advancements in NLP The field of NLP has witnessed transformative developments with the advent of generative AI models such as ChatGPT. Built on the GPT (Generative Pre-trained Transformer) architecture, ChatGPT employs the transformer model's attention mechanisms to understand and generate human-like text. Its training involves two key stages: pretraining on a massive amount of text using autoregressive language modeling and fine-tuning for task-specific applications through reinforcement learning from human feedback [189]. ChatGPT can facilitate cross-lingual communication by providing real-time translation services and addressing language comprehension issues in international interactions [190].

ChatGPT excels in tasks like conversational AI, text summarization, and question answering by leveraging contextual embeddings to maintain coherence over long textual sequences. For instance, compared to earlier architectures such as RNNs and LSTMs, which struggled with long-term dependencies, ChatGPT's self-attention mechanism enables efficient handling of long-range contexts, making it highly effective in dialog systems [189]. However, challenges persist, such as susceptibility to producing plausible but factually incorrect responses, and biases embedded in training data. The widespread adoption of ChatGPT underscores its impact on democratizing NLP tools, yet it also raises ethical concerns related to misinformation (Fig. 7), intellectual property, and data privacy. Future research should focus on addressing these challenges and enhancing the transparency and explainability of such models.

Summarization Automatic Text Summarization (ATS) is a significant topic due to the vast volume of textual data that is rapidly growing nowadays on the internet and other sources. Three ATS approaches are available: abstractive, extractive, and hybrid [191]. To create a summary, the extractive technique chooses and incorporates the most significant sentences from the source text. In contrast, the abstractive approach transforms the input content into an intermediate representation before generating a summary using unique terms. The hybrid approach incorporates both the extractive and abstractive processes. L. Chen and Nguyen [192]

proposed an ATS technique for summarizing sign documents based on a Reinforcement Learning (RL) technique and an encoder-extractor network architecture's RNN model. The key features are selected using a sentence-level encoding approach, after which summary sentences are obtained. However, the method has a chance of suffering from large variances since it uses an approximation in the RL method training objective function.

Clustering, unsupervised NNs, and topic modeling were used to create an Arabic summarization approach in this study [193]. The researchers also presented ensemble learning models and neural network models to combine the information produced by the topic space. Specifically, the ELM-AE method and k-mean approach were used to cluster documents on a large sample of Arabic documents. The LDA technique was implemented to determine the topic space associated with each cluster; then, the discovered subject space was used to generate a numerical document representation. For learning unsupervised features from texts, this format was utilized as the input for various NN and ensemble methods. For the variation of the information in the final summary, the learned characteristics were utilized to rank phrases by following a graph-based model. Key phrases were picked by applying the redundancy removal components. A few more unsupervised NN methods, for instance, stacked-auto-encoder, could be included in future studies to improve the robustness of the suggested technique.

Benefits		Challenges	
✓ Scalability ✓ Innovation catalyst ✓ 24/7 Availability ✓ Cost-efficiency ✓ Transfer learning and fine-tuning ✓ Domain-specific adaptation ✓ Accelerates research ✓ Interactive learning tool ✓ Zero-shot and few-shot learning ✓ Creative writing aid ✓ Personalization ✓ Versatility ✓ Multilingual capabilities ✓ Time-saving	✓ Continuous learning ✓ Language translation ✓ Code assistance ✓ Enhanced learning Experiences ✓ Natural language understanding ✓ Creative collaboration ✓ Mental health support ✓ Multilingual conversational abilities ✓ Context retention ✓ Accessibility ✓ Massive scale data processing ✓ Human-like interactions ✓ Adaptability to domains	✓ Significant computational costs ✓ Misinformation propagation ✓ Ambiguity in responses ✓ Biased outputs ✓ Ethical concerns ✓ No real-time data ✓ Overfitting issues ✓ Lack of emotional intelligence ✓ Over-reliance on training data ✓ Over-reliance by users ✓ Vulnerability to prompt injection attacks ✓ Dependency on prompt phrasing	✓ Limited context understanding ✓ Difficulty with complex queries ✓ Lack of real understanding ✓ Data privacy risks ✓ Security vulnerabilities ✓ Privacy risks ✓ Dependence on human moderation ✓ Limited contextual understanding ✓ Inability to reason causally ✓ Generating misinformation ✓ Dependency on training data

Fig. 7 Benefits and challenges of using ChatGPT

Advanced text mining and classification A methodology named heterogeneous latent topic discovery (HLTD) was proposed by Li et al. [194] that combines word embeddings with topic modeling to make use of both local word sequential relationships and global document topic information. By allocating word distributions according to document topics or word embeddings, HLTD improves the framework's all-encompassing perspective of the text. For parameter and hidden variable inference inside HLTD, a privacy-preserving Monte Carlo expectation–maximization technique is suggested to guarantee data privacy. By integrating differential privacy with conventional Markov Chain Monte Carlo (MCMC) methods, this approach introduced two new sampling algorithms that maintain model performance while safeguarding the privacy of training data. These algorithms can be used in various contexts and incorporated with others that deal with word embedding.

A model agnostic approach, variational continuous label distribution learning (VCLDL), was developed by Zhao et al. [195] to tackle the class imbalance issue in multi-label text classification (MLTC). To mine the data concealed in the observable logical labels, VCLDL creates a suitable link between the feature space and the label space. In particular, VCLDL treats the label distribution as a continuous density function in latent space and develops a versatile variational method to jointly use the feature space and estimate the label density function. When compared to the currently available MLTC models, the experimental results on various benchmark datasets revealed that VCLDL could significantly improve performance. Incorporating VCLDL into MLTC models allows them to focus on the distribution of the entire set of labels instead of individual labels with high response values, effectively resolving the class imbalance issue.

3.1.3 Computer vision

To a broader extent, computer vision is an interdisciplinary branch of computer science study that examines how computers can quickly recognize digital images and movies. The predominant processes include extracting, analyzing, and comprehending relevant information autonomously about a specific picture or a series of images [196]. Computer vision implies the algorithmic and theoretical roots for spontaneous ocular comprehension. The growth of computer vision is reliant on the computer technology system, focusing on image quality enhancement or image classification [197]. Researchers often use the terms interchangeably in the basic approaches because of the overlapping with image processing. However, the core aim of computer vision is to develop models, data extraction, and information from images. On the other hand, image rectification deals with applying computational alterations to pictures, such as sharpening and contrast, among other things [198]. Deep learning

algorithms recently tackled major computer vision tasks like face recognition, object detection, human pose estimation, activity recognition, and fruit defect detection [199].

Object detection Object detection has attracted massive attention as a vital function in computer vision, in which CNN has made substantial headway [200]. Object detection is, in essence, the first step in several computer vision applications, including logo detection, face recognition, pedestrian identification, and video analysis. One-stage detectors (YOLO and its variants) and two-stage detectors (region-based CNN (R-CNN)) are the two primary families of object detection deep learning frameworks [201]. One-stage detectors do not need the cascaded zone classification step to make qualitative predictions of objects on every position of feature maps. Two-stage detectors, on the other hand, have a proposal generator that makes a handful of suggestions before feature extraction. When dealing with complicated models, deep neural architectures are superior to simple ones [202]. Using the weight-sharing principle, CNN is a feed-forward neural network. CNNs are not precise for minor data, but produce high accuracy for vast image datasets [203]. However, to execute computer vision tasks, CNNs require large labeled datasets. It is a fusion of multiple functions and integration that shows how they intersect. The CNN layered structure for object detection is shown in Fig. 8.

For object detection, deep R-CNNs have been widely deployed. Ouyang et al. [205] suggested a deformable deep CNN for generic object detection. This constrained pooling layer in the recommended deep design simulates the deformation of object portions with geometric compensation and limitation. The researchers introduced diverse models by changing the training process, net topologies, and altering some vital components in the detection pipeline. This strategy exceptionally optimized the model's averaging efficacy. The proposed technique elevated the mean averaged precision by RCNN from 31 to 50.3, exceeding GoogLeNet, the winner of the ILSVRC2014, by 6.1%. The extensive experimental evaluation also provides a detailed component analysis. Doulamis and Doulami [206] explored a deep model's representation capacities in a semi-supervised scenario. The method overcame deep learning's crucial drawbacks by utilizing unsupervised data to construct the network and then fine-tuning the data using a gradient descent optimization process. Unsupervised learning appreciates more abstract representations to reduce the input data, thus optimizing the model's stability, accuracy, and reliability. In addition, an adaptive approach allows modifying the model dynamically to the digital photogrammetry conditions.

Face recognition Given the enormous diversity of face images present in the real world, face recognition is amongst
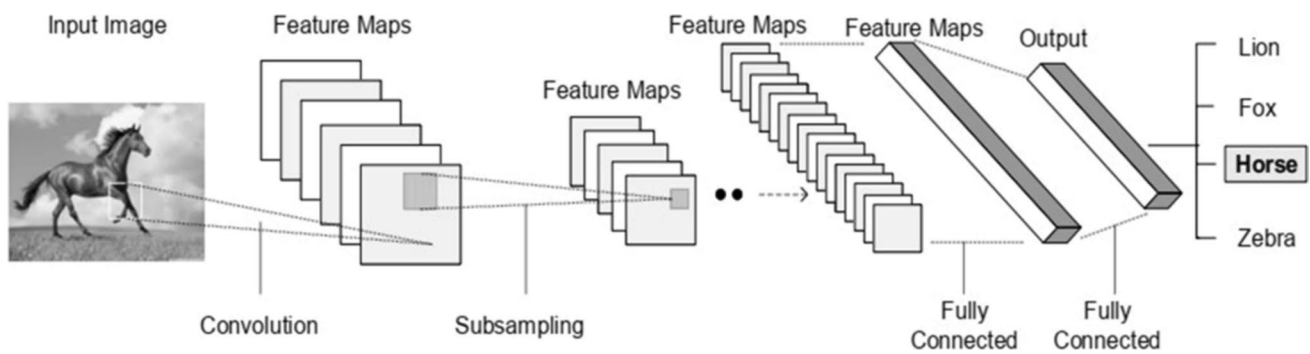
Fig. 8 Layered CNN structure for object detection (reprinted with the permission of Elsevier from [204])

the most complex biometrics attributes to use in open circumstances [207]. Face recognition structures are frequently comprised of several components, such as an image, face detection, face representation, face matching, and face alignment. Researchers have focused on specialized approaches for each face variation owing to the feature's sturdy complexity in unconstrained environments. Many new face recognition architectures have been presented in tandem with deep learning advances, and some even come near to human performance. Extensive research with acceptable outcomes has been conducted on the issues of illumination, position, and expression [208]. However, when dealing with noisy photos, the precision of most methods deteriorates dramatically. Humans can often have a hard time distinguishing an identity from a seriously noisy face.

Zadeh et al. [209] designed a local model with convolutional specialists' constraints for facial landmark detection by adding a set of appearance prototypes of diverse poses and expressions. A pivotal part of the CE-CLMs is the Convolutional Experts Network (CEN), a new local sensor that blends neural structures with expertise in an end-to-end framework. The findings suggest that launching from iteratively-stated CEN network weights while exerting Menpo trained data alone does not yield competent outcomes. However, applying plain datasets of the Curriculum Learning paradigm [210] and switching to Menpo training data gave satisfactory results. Deng et al. [211] suggested UV-generative adversarial networks, a painstakingly built system that unifies global and local adversarial deep CNNs to develop an identity-preserving facial UV closure framework for pose-invariant visual recognition. Combining posture augmentation during training and pose discrepancy depletion throughout testing, the protocol attained a state-of-the-art affirmation accuracy of 94.05%.

To adversarially train the identity-distilled attributes, Y. Liu et al. [212] built an autoencoder system with minimal direction via facial identities. The model was shown to be capable of generating identity-distilled features and extracting concealed identity-dispelled features that can

preserve complementary knowledge, such as backdrop clutters and intra-personal variances. The primary limitation of deep learning approaches is that they must be learned on extremely big datasets with enough variety to generalize to unknown samples. Recently, numerous large-scale face datasets, including images of faces in the wild, have been made available to train CNN models [213, 214]. The studies demonstrate that neural networks may get trained as classifiers and can reduce dimensionality.

An innovative model for detecting and recognizing faces on drones was proposed by Rostami et al. [215] to enhance face recognition performance in situations where query pictures are captured from long distances or high altitudes, resulting in limited facial information of the individuals. The authors also aimed to improve the level of performance by using a deep neural network to carry out these tasks. The proposed method was deconstructed into four phases: preprocessing, detecting faces, extracting features, and recognizing faces, as shown in Fig. 9. A deep CNN model was employed to localize faces in face detection, which was then used in face recognition. The process of feature extraction involved the extraction of deep features from the generated face images for use in the subsequent step. The process of face classification was conducted using CNN, which entails identifying a face from a series of facial images. Experimental testing on the DroneFace dataset compared the suggested framework to the state-of-the-art models, showing that the suggested approach can achieve competitive detection and recognition accuracies. Even when unmanned aerial vehicles (UAVs) take pictures with a wide angle of depression, the suggested method can still identify faces on them within certain angle limits. To overcome this limitation, the developed model can use the grid concept. It primarily models geometric facial components, including the mouth, eyes, and cheeks.

Action and activity recognition Human activity recognition (HAR) plays a vital part in today's society because of its potential to assimilate extensive advanced information
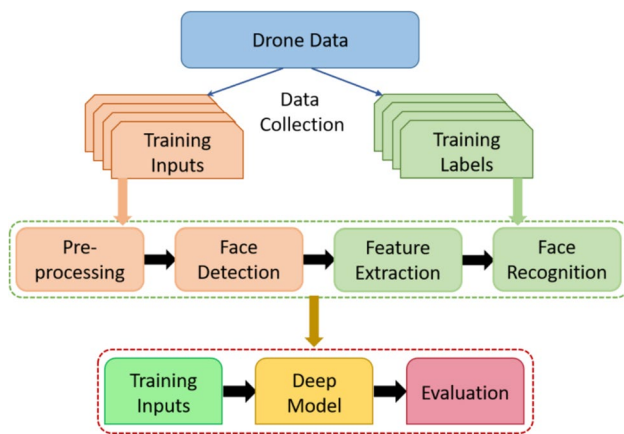
Fig. 9 Flowchart of the proposed framework [215]

from raw sensor data about human actions and activities [54]. To identify sensor acquired data, early studies primarily applied naive bias, decision trees, SVM, and other classic machine learning methodologies [216, 217]. Researchers have recently transitioned from traditional handcrafting to deep learning methods for HAR, especially with the advent and efficient application of deep learning algorithms. Handcrafted model representation cannot tackle complicated cases due to their limitations. However, given the success of DL models in speech recognition, NLP, image classification, and other domains, transferring them to the HAR field is a new research concept for pattern identification [218, 219].

Ordóñez et al. [220] reported a classifier that can classify 5 movements and 27 hand gestures using CNN and LSTM. The methodology surpasses deep non-recurrent networks by an average of 4% and outperformed some previous findings by 9%. Results suggest that this model can be used with homogenous sensor modalities and can fuse multimodal sensors for better execution. To reliably classify subjects or activities, Lin et al. [219] introduced a novel iterative CNN technique using autocorrelation pre-processing rather than the conventional micro-Doppler image pre-processing. An iterative DL framework was used in the proposed method to define and extract features automatically. In addition to outperforming feature-based methods employing micro-Doppler pictures, the proposed iterative CNNs, followed by random forests, also outperform classification methods employing various types of supervised classifiers.

Even though the preceding models may distinguish human activities in general, the entire network structure is quite complex. Furthermore, the large number of parameters has a massive computational cost in these models, which is thus challenging to employ in situations where high real-time performance is required. Agarwal and Alam [221] implemented a lightweight deep learning paradigm for HAR on a Raspberry Pi3. The overall accuracy of this

model on the WISDM dataset was 95.78% using a simplistic RNN incorporating the LSTM technique. Despite its excellent accuracy and simplicity, the conceptual framework was evaluated on a single dataset with only six activities, which does not guarantee that it is generalizable. To address the drawbacks, Xia et al. [222] introduced a unique deep LSTM-CNN for HAR that can extricate activity features and categorize them automatically using only a few factors. The network was tested with three of the most often utilized publicly available datasets. The analysis showed that the network not only has fewer parameters with great precision but also has fast convergence speed with sufficient generalization aptitude. In many computer vision applications, deep learning has outperformed older methodologies due to the capacity of DL algorithms to learn characteristics from raw data, depleting the need for constructed feature detectors and descriptors.

Human pose estimation The goal of the human pose estimation (HPE) problem, which has existed for centuries, is to determine how well sensor inputs can be used to estimate human posture. The HPE system attempts to determine a human's body position from still images or moving video. It is an important research field because it involves various applications, such as action detection, activity recognition, movies, video surveillance, human tracking, sports motion, virtual reality, medical assistance, human–computer communication, and self-driving analysis. HPE is challenging because of the wide variety of body types, self-obscuring stances, and complex environments that might occur from the great degree of freedom of interconnected joints and limbs [223]. Toshev and Szegedy [224] initially tried to instruct an AlexNet-like DNN to determine joint points from complete pictures without utilizing any models or part detectors. A multi-stage refining regressor cascade architecture improved the performance and modified the cropped images from the previous step. Similarly, Pfister et al. [225] used a string of concatenated frames to implement an AlexNet-like network to forecast the human stance in films. The use of joints alone without the context is not robust; however, turning them into numerical-joint positions from heatmap supervision monitoring can benefit both approaches.

Yang et al. [226] developed a Pyramid Residual Module (PRM) to increase Deep CNN (DCNN) stability across levels to swap the Hourglass network's residual module. This method exhibited considerably enhanced performance compared to the earlier state-of-the-art procedures. Papandreou et al. [227] developed a box-free, multi-tasking ResNet-based network for pose evaluation and occurrence classification. In real-time, the ResNet-based network can forecast joint heatmaps of all critical spots of all people, as well as their relative displacements. Then, the most confident detection is grouped, adopting a decoding strategy

focused on a tree-structured kinematic graph. Furthermore, the network acquired a greater average precision and above 5% absolute advancement over the former top-performing approach on a similar dataset. Other studies [228–231] mainly focused on moderately constrained directions, such as 3D HPE, RGB-D-based action identification, body parts-based HPE, and monocular-based HPE model-based HPE.

3.2 Environmental and social

3.2.1 Ecology

The endless data stream offers ecologists a new challenge: finding or developing the analytical models required for data extraction from the massive amounts of video feeds and streaming photos [232]. On the other hand, obtaining usable footage in maritime zones to attain acceptable computing performance provides a unique set of obstacles compared to terrestrial situations. Variable water purity, obstruction due to schooling fish, complicated background formations, and diminished light with increased depth are some environmental complications that might interfere with clear footage in aquatic ecosystems [233]. Though certain elements may affect the image and video quality, deep learning techniques have evinced effectiveness in a variety of marine implementations. Deep learning algorithms are currently pirouetted around marine environments to automate the classification of certain species. Not surprisingly, the most common use of deep learning is identifying species from recordings of sounds, images, or videos. These investigations already cover a wide variety of organisms, from bacteria and protozoa to plants, insects, and vertebrates, both living and extinct, and from microscopic to planetary levels [234–244].

Deep learning's capacity to comprehend features from data is what deems it so powerful. Unsupervised algorithms have no specified output and are frequently used as a systematic appliance to find patterns in data, minimize dimensions, and classify groupings [232]. Ditria et al. [245] emulated the speed and precision of deep learning approaches opposite to human counterparts for verifying fish abundance in underwater video clips and photos to test its adequacy and relevance. They designed three prototypes using Mask R-CNN, an object detection framework to locate the aspired species called luderick (*Girella tricuspidata*). In single image test datasets, the machine topped marine specialists by 7.1% and citizen scientists by 13.4% and outperformed in video datasets by 1.5% and 7.8%, respectively. These results confirm that deep learning is a better tool to evaluate abundance with stable results and is portable across survey locations than humans.

Another vast potential of deep learning is to detect disease symptoms similar to existing applications in the medical sector. For example, CNNs can identify tree defoliation and crop illnesses. Rather than accounting for each feature independently through explicit modeling, Kälin et al. [246] built a joint distribution of techniques, applying fivefold cross-validation for the preliminary investigation to eradicate any train-test split bias. Defoliation frequencies were distributed evenly among all 5 folds. On one of the datasets, this technique performed only 0.9% less than a group of human experts. This protocol can spot malnutrition, scars, and the presence of apparent diseases in wild plants and animals. Although species-specific models might forecast tree stress in a better way, the study does not have enough training data for each species to build a robust algorithm. In consequence, complex situations, including abundantly defoliated trees or a green canopy with trivial defoliation, can produce errors.

Traditional methods for classifying coastal pollution suffer from inefficiency and poor accuracy. A DCNN based on deep learning was proposed by Sathish et al. [247] to address these problems. The image's undesired noise was removed using anisotropic diffusion. To improve the picture's contrast, global histogram equalization (GHE) was used. After the image underwent pre-processing, it was analyzed and classified using a deep learning-based DCNN to identify and assess coastal pollution. The DCNN's weights were fine-tuned for optimal performance in classification by employing a reptile search algorithm. The suggested method was found to be 97.2% accurate, 96% precise, 2.8% error, 95.4% recall, 96.5% f1-score, and 96% specific. When tested against other methods, the developed approach performed better. In order to maintain environmental health and ecological safety, this model aids in the detection of coastal pollution.

Humans can abstract widely between topics and make accurate judgments with minimal information. However, deep learning algorithms have limited abstraction and reasoning capabilities. Only a few researchers have looked into combining different plant organs or perspectives to improve accuracy [248, 249]. Owing to the high effort required to collect and label the datasets, completely automated species identification remains a long way off. The excellence of an automatic detection model is determined not only by the quantity but also by the caliber of the given training data [250]. On the contrary, most of the reviewed research indicates a deficit in the available qualifying data.

3.2.2 Agriculture

The successful application of deep learning has spread to many fields, including agriculture. Some specific problems in this field that can be explored with deep learning methods include land cover categorization, plant classification, fruit counting, and crop type classifications. In a survey, Kamilaris and Prenafeta-Boldú [251] found that

deep learning outperforms frequently-used image processing approaches in terms of accuracy in the agricultural domain. Many methods based on deep learning have been developed recently for identifying leaf stress in plants. Ramcharan et al. [252] offered a cassava vegetable-based transfer learning approach. An Inception V3 model was used on a sample of 15,000 photos, and the approach outperformed more common models based on machine learning, such as K-Nearest Neighbor (KNN) and SVM, with an accuracy of 93%. For certain diseases (CBSD, BLS, GMD), using the leaflet rather than the full leaf increased diagnostic accuracy. However, whole-leaf photos enhanced accuracy for diseases like CMD and RMD.

Plant diseases are responsible for a significant portion of the agricultural produce loss. Reduced agricultural production might worsen the global food crisis and significantly impact national economies. Therefore, preventing the spread of plant diseases through early diagnosis is crucial to maintaining global health. This motivation drives scientists to develop automated systems to diagnose plant diseases early. To improve the accuracy of plant leaf disease classification, Ali et al. [253] utilized an ensemble of several deep learning models, including DenseNet201, inception-resnetV2, efficientNetB0, and efficientNetB3. An innovative method for processing images was proposed by the authors to make deep-learning models more effective. PlantVillage, the largest dataset for plant disease, was used to train and test five different DL models. To find the most accurate DL ensemble, ten randomly selected DL model ensembles were selected to test and compare them. On the dataset, the suggested ensemble model achieved an accuracy of 99.89%. This high accuracy demonstrates the system's generalizability to new datasets or real-world situations.

Instead of evaluating the entire leaf, Arnal Barbedo [254] employed a CNN to categorize specific diseases and spots on plant leaves. This revealed a number of illnesses that all damage the same leaf. In addition, the author used deep learning to recognize specific lesions and spots in 14 plant species. The models in this study were trained using a pre-trained GoogLeNet CNN. The study could not determine the number of pictures of disease symptoms needed for the neural network to learn their characteristics. In all situations, a few hundred photos appeared to be sufficient to produce credible findings, but this quantity must be approached with caution.

Lin et al. [255] applied an RGB sensor along with a UNet-CNN to classify and identify cucumber powdery mildew-affected leaves. Because the loss value of the affected pixels in this study was lower than the non-affected pixels, the experiment presented a loss function (Binary-Cross-Entropy) to tenfold the loss values. The semantic segmentation CNN model segmented the sick powdery mildew on cucumber leaf pictures with a mean pixel accuracy of 96.08%. However, taking images in a controlled setting is necessary rather than in an open field. Furthermore, a lack of appropriate amount and diversity of the datasets, where symptoms produced by other conditions were not included, may hinder the effectiveness of DL methods.

DL approaches have recently been implemented in smart fish farms. A deep learning method was demonstrated to differentiate better variations in traits, classes, and the environment, allowing it to extract target fish attributes from a picture captured in an uncontrolled underwater environment [256]. On the well-known LifeCLEF-14 and LifeCLEF-15 fish datasets, CNN outperformed conventional techniques with classification accuracies of over 90% [257]. General deep structures in the experiment should be fine-tuned to increase the efficacy of identifying vital information in the feature space of interest, thus reducing the requirement for vast volumes of annotated data.

3.2.3 Financial fraud detection

Financial fraud is a type of thievery whereby a business entity or anyone illegally takes money or assets to profit from it. Since the advent of new technologies, there has been a massive rise in financial fraud involving all aspects of the business world. For instance, financial fraud victims in the United States individually lost an average of $1,090, accumulating over $3.2 billion in 2017 [258]. The selection procedure for practitioners is frequently known as safeguards of fraud analysis, assessment, or detection by information security specialists and associations. Some methods for detecting anomalies include decision trees, logistic regression, SVM, and others. However, these methods are constrained as they are supervised algorithms that rely on labels to determine transaction validity [259].

Several studies [260–262] revealed that developing measures based on fraud estimations has been an infuriating procedure that is excessive and futile when tackling composite non-linear data, has inefficient memory utilization, lengthy calculation, and generates abundant false alarms. Financial fraud detection is still strenuous because of the constant changes in fraudulent behavior, the absence of a mechanism to track fraudulent transactions' information, certain limitations of the existing detection techniques, and highly skewed datasets [263]. Therefore, optimization of previous procedures and novel approaches are necessary to increase fraud detection rates [264, 265]. As a result, deep learning, a subset of modern artificial intelligence, is a promising technique that has been recognized in recent years. DL also facilitates computational models with multiple processing layers to adapt data representation to different abstraction levels.

Credit card fraud is one of the most pervasive forms of financial fraud in which a hacker gains unauthorized access

to a legitimate card without the cardholder's consensus. The simultaneous expansion of e-commerce and the internet has resulted in a significant increase in credit card use, resulting in an unusual increase in credit card theft in recent years. For instance, a report by the Boston Consulting Group showed that North American financial institutions lost $3 billion in 2017 due to credit card fraud. Despite industry attempts to combat economic fraud, the Nilson Report site revealed that global financial losses exceeded $24.71 billion in 2016 and $27.69 billion in 2017, attributed to credit card fraud [266]. These fraudulent occurrences may erode consumer confidence, destabilize economies, and elevate people's living costs. As a result, several studies have applied deep learning to detect credit card fraud.

Over 80 million credit card transactions labeled as fraudulent or legitimate were used to assess the efficacy of the deep learning process reported by Roy et al. [267]. This study assessed the performance of various DL algorithms based on their parameters, class imbalance, scalability, and sensitivity analysis. The researchers used LSTM, Gated Recurrent Unit, RNN, and ANN in a distributed cloud computing environment. The findings confirm that the LSTM technique achieved the best performance and reveal that each of the four topologies' execution expanded with the increased network size. However, the study failed to specify the conditions under which the model functions cease to ameliorate, model sensitivity, and limit the network size.

Gated architectures, such as LSTMs and GRUs, excel in sequential learning tasks, making them invaluable for processing large datasets with sequential or temporal dependencies [268]. They address challenges like vanishing and exploding gradients, common in traditional RNNs, ensuring stable and efficient training [220]. They are crucial in unraveling complex data by capturing long-term dependencies, adapting to variable-length sequences, and serving as effective feature extractors. Their ability to automatically learn and emphasize salient features within the data is instrumental for discerning patterns and making informed predictions. As versatile tools, gated architectures contribute significantly to analyzing complicated data structures, allowing for a more nuanced understanding of complex information.

To improve upon existing detection strategies and increase the detection accuracy of complex data sets, Alghofaili et al. [263] introduced an LSTM-based fraud detection method. The system was tested with a unique dataset of credit card frauds. Simultaneously, the outputs are analogized to an auto-encoder model and machine learning technologies in a pre-existing deep learning model to detect suspicious financial activities and alert the appropriate authorities, allowing them to take appropriate action. Although some machine learning techniques have demonstrated promising upshots, they do not detect new patterns or deal with large amounts of data to improve

accuracy after reaching a saturation point. LSTM worked brilliantly in the trials, reaching 99.95% accuracy in less than a minute. The unique applicability of LSTM in prediction is ensured by prior experience and the correlation between prediction outputs and historical input. Many-state memory cells and gates are incorporated into the framework. Due to its vital role in ensuring that data is transmitted unmodified, the cell's state is at the core of the information transfer process [236].

Several studies used other deep learning frameworks productively for a similar cause. To illustrate, Pandey [269] introduced the H2O deep learning framework in credit card forgery detection. The pattern of the model enables multiple algorithms to aggregate as modules, and their outputs can be integrated to improve the final output accuracy. Concurrently, the efficiency of the algorithms rises as the dataset length grows while using this framework; thus, adding more algorithms with equivalent formats and datasets can elevate this model. Another type of financial fraud is tax fraud, the malicious act of falsifying a tax return document to reduce someone's taxation. Low fiscal revenues are weakening governmental investment [270]. When it comes to the property acquisition tax, Lee [271] implemented a reliable sampling technique for tax-paying citizens. The data from 2,228 returns were fed into autoencoder, a well-known unsupervised deep learning technique, to calculate the potential tax shortfalls for each return based on an estimate of the rebuilding faults. The sorted reconstruction ratings are compatible with practical context, implying that the defects can be exploited to identify suspicious payers for auditing in a cost-efficient strategy. Utilizing the recommended strategy in real-world tax administration can reinforce the self-assessment acquisition tax system.

López et al. [272] researched tax fraud detection in the context of individual income tax filings. Neural networks application enabled the taxpayers' segmentation and the possible assessment of a particular taxpayer attempting to avoid taxes. The neural network outcome categorizes whether a taxpayer is deceitful or not as well as reveals a taxpayer's proclivity for unethical activities. In other words, it classifies individuals based on their potential of convicting and also calculates the propensity of tax fraud per taxpayer. The chosen model excelled over other tax fraud detection algorithms with an accuracy of 84.3%. It would be interesting to see this concept applied to additional taxes in the future.

To improve the fraud detection of credit cards, Khalid et al. [273] proposed a new ensemble model combining several classifiers: SVM, KNN, Bagging, RF, and Boosting. The model was tested using a dataset that included credit card transaction records from European customers to ensure a realistic evaluation. With the computational capabilities of Google Colab, the proposed model's methodology included feature engineering, data pre-processing, model choice, and

evaluation. This allowed for effective training and testing of the model. Compared to conventional ML techniques and individual classifiers, the suggested ensemble model has been demonstrated to be more effective in reducing difficulties related to fraud detection of credit cards. The ensemble achieved better results than the preexisting models in recall, accuracy, precision, and F1-score metrics.

3.2.4 Disaster management

In terms of disaster or natural calamity management, uncertainty, lack of resources in the affected areas, and dynamic environmental changes are the major characteristics of natural catastrophes. The inability to predict outcomes indicates that catastrophic effects on persons and property during several disasters cannot be foreseen with a level of accuracy that is reasonable [274]. Big data is a technology paradigm that enables researchers to efficiently analyze enormous amounts of data made accessible by current practices [275]. Making the most of big data is possible using a variety of technological and scientific strategies and equipment. Recent advancements in big data and IoT technologies open up a wide range of opportunities for disaster management systems in order to obtain leading assistance, guidance, as well as better observations and ideas for precise and suitable decision-making.

In order to resolve catastrophic challenges, Anbarasan et al. [276] introduced concepts and strategies for identifying flood catastrophes based on IoT, big data, and CDNN. Big data derived from the flood catastrophe was first used as the input data source. Afterward, the Hadoop Distributed File System (HDFS) was used to decrease the high frequency of data. After excluding high frequency data, the data were preprocessed via missing value approximation and a normalizing technique. A conjunction of attributes was utilized to construct the pattern centered on the pre-processed data. In the final step, the created parameters were sent into the CDNN classifier, which categorized them into two categories: the probabilities and infeasibility of a flood emerging. The performance of the suggested system was analyzed in terms of precision, accuracy, recall, F-score, specificity, and sensitivity in comparison to the existing systems, DNN and artificial neural network (ANN). The comparison findings demonstrate that the CDNN method has a greater level of accuracy than the present approach, with 93.2% accuracy, 92.23% precision, 90.36% recall, and 91.28% F-score with 500 data points utilizing CDNN. The results described above can be improved using DNN as well as ANN models. In conclusion, the detection system performed better than other top methods currently in use. In the future, the work presented here could be improved using IoT devices that have even longer sensor ranges at lower costs and cutting-edge algorithms at each stage of the flood identification process.

Disasters caused by fire do have a negative impact on the environment, society, and economy. Early fire detection and an automated response are vital and essential for disaster management schemes in order to minimize these damages. Fire detection early in disaster management systems while monitoring public spaces, woods, and nuclear power plants can prevent ecological, monetary, and social harm [277]. The movements of entities with a retardant appearance and varied atmospheric factors make early detection hard [278]. Subsequently, an improved accuracy algorithm that reduces the number of false alerts in the aforementioned conditions is required. In order to accomplish this, Khan et al. [279] investigated deep neural networks and developed a prominent architecture for early fire detection through surveillance for efficient disaster management systems. Another preferable criterion was to notify the disaster management system after a successful fire detection and include the representative frames. A system of flexible prioritization was developed for the monitoring system's camera nodes, taking into account the contents they could visualize. The experiment primarily examined two datasets. The initial dataset demonstrated additional progress, as accuracy increased from 93 to 94% and false positives decreased from 0.11% to 0.9%. Despite producing false negatives of around 21%, the approach maintained more robust balance, enabling the method to more accurately achieve fire identification to achieve fire identification more accurately. The findings for the second dataset [280] were gathered using a different set of metrics, including recall, precision, and F-score, in order to properly conduct the performance. Results produced by using deep features and optimizing the fire prediction model are compared and contrasted, showing that the detection approach and model are feasible with a maximum score of 82% for precision, 98% for recall, and 89% for F-score. Despite evaluating the sustainability of the proposed approach utilizing noise intrusions, scaling, and rotations, each image comprised of fire and showed a high accuracy ranging from 89 to 99%. Considering the potential of such occurrences occurring under surveillance, the results indicate that the CNN-based model could detect fires at an early stage in a range of circumstances, even when images are blurry. The relevance of the framework for successful fire disaster management was supported by experimental results, which also confirm the high accuracy of the fire detection technique compared to cutting-edge methods.

Over the past few years, social media networks have contributed to managing natural catastrophes. In order to interpret better messages and extract relevant information from social media, text data mining techniques employing standard machine learning approaches have been constructed. These techniques are distinct and challenging to generalize for different categorizations. Therefore, considering the crisis management efforts related

to hurricanes, Yu et al. [51] investigated the ability of a convolutional neural-based deep learning model to classify the trending catastrophic topics from Twitter. SVM and LR are two conventional machine learning techniques that were evaluated against the logistic regression of the CNN model. The outcomes of the experiment demonstrate that CNN models consistently outperformed SVM and LR models in terms of accuracy for both types of assessment scenarios. Moreover, the CNN classifier surpassed the classification methods of SVM (63–72%) and LR (44–60%), achieving an accuracy of up to 81% among all datasets. In order to categorize tweets posted during a later event, the evaluation was carried out using CNN based on Twitter data from prior events. Results of the experiment demonstrated that while SVM and LR's accuracy dramatically decreased, CNN maintained a steady performance. This suggests that the CNN model could be trained in advance using Twitter data from previous events to categorize new occurrences for situational awareness. However, CNN took longer to learn than SVM and LR because there were more parameters to consider, which makes it difficult to employ CNN for web-based learning.

When a disaster occurs and geotagging data associated with images shared on social media is typically unavailable, the framework proposed by Sathianarayanan et al. [281] can extract geographic locations from images using phone numbers. It consists of three main modules: phone number identification, image collection, and location extraction through Google Maps API. Using the Street View House Numbers (SVHN) and manually interpreted multi-digit phone numbers, the authors trained a CNN-based detection model—specifically, the RetinaNet object recognition algorithm—to identify and extract these numbers from images. The phone number identification model was able to achieve an Average Precision (AP) of all digits greater than 79% (0.79) and a respectable mean AP of 82% (0.8248) using an IoU (Intersection over Union) threshold of 0.5. Utilizing phone numbers retrieved from object detectors, the Google Maps API can render location data with reduced distance distortion. All kinds of disasters, including earthquakes, floods, and wildfires, can benefit from the proposed method's enhanced disaster assessment, situational awareness, and management capabilities.

3.3 Industrial and engineering

3.3.1 Fluid dynamics

Fluid dynamics is an area of applied science dealing with the flow of liquids and gases. Three conservation laws define fluid dynamics: mass conservation, linear momentum conservation, and energy conservation. Computational fluid dynamics (CFD) is a set of numerical methods for providing a rough solution to fluid dynamics and thermal difficulties. CFD is not a science in itself but rather a means of applying numerical analysis principles to heat and mass transfer [282]. It extensively analyzes airflow motion, contaminant transport, and heat transfer in enclosed places. CFD helps with wind flow and contamination dispersal analysis surrounding buildings in metropolitan surroundings but still confronts several difficulties concerning computing cost and accuracy.

The neoteric success of DNNs has been facilitated by a high abundance of computational power that takes advantage of the multi-layer architecture. Not long ago, DNNs gained notoriety in turbulence modeling or, comprehensively, in high-dimensional areas, complex dynamical systems [45]. Fonda et al. [283] scrutinized heat transport characteristics of turbulent Rayleigh–Bénard convection in horizontally expanded systems by using deep-learning frameworks. The researchers applied the deep CNN-trained method to measure the heat transfer's fraction and time variations. The slowly evolving turbulent superstructures received special attention because they are larger than the height of the convection layer. The strategy trains a deep CNN with a U-shaped configuration that has a contraction branch and an expansion branch, simplifying the complicated 3D superstructure in the midplane layer to a temporal planar network. As a result, data compression happens when the maximum Rayleigh number is more than five magnitudes. This shows deep learning's utility in parameterizing convection in global models of stellar and atmospheric convection. The U-specialized net's architecture requires fairly minimal training datasets that proved to be the most effective for ridge extraction, especially for noisier data at higher Rayleigh numbers. Nonetheless, the study did not provide information on the network's efficiency at lower Rayleigh numbers.

Another study conducted by Daw et al. [284] applied extensive fluid flow simulations to climatology and turbulence to forecast turbulent flow by identifying exceptionally nonlinear phenomena from spatiotemporal velocity fields. Specifically, they introduced adaptable spectral filters along with a specific U-net for assertion. The concept exhibits substantial declines in error prediction compared to state-of-the-art baselines. Most crucially, this procedure accurately previses over physical entities that satisfy favorable physical attributes, such as the conservation of mass, and also simulates the acute turbulent kinetic energy field and spectra for precise predictions. Another novel approach in this discipline is model recognition of reduced-order fluid dynamics process applying deep learning, which is a computational simulation-based method to create mathematical patterns of dynamic physical systems.

Reduced-order modeling (ROM) is one of the effective system identification strategies for reducing the complex and excessive dimensional size of discrete dynamical systems. Air pollution modeling, nonlinear large-scale systems, optimal control, ocean modeling, form optimization, neutron difficulties, sensor placement optimization, porous media problems, aerospace, multiscale fracture, and shallow water are some examples of its usage [47, 285–292]. Subsequently, Wang et al. [47] developed a ROM initially by combining a deep learning algorithm with suitable orthogonal decomposition methodologies. The results revealed that the DL ROM (DLROM) can capture sophisticated fluid dynamics with a CPU cost of less than 1/1000. While DLROM offers better potentiality in prediction than earlier ROMs [293], deep learning frequently demands the availability of massive training data that greatly surpass the network's parameters. When the probability distributions of the novel input data deviate from those of the training data, the resulting models are typically acceptable for interpolation but may not be adequate for extrapolation. Nevertheless, the effects of expanding this method to variable parametric obstacles, such as the real-time response to natural disasters, are unknown.

In order to achieve super-resolution of flow-field data in both time and space from a small set of noisy measurements in the absence of high-resolution historical data, physics-informed neural networks (PINNs) were employed in the study of Eivazi et al. [294]. Finding a physically consistent forecast at any location in the solution domain—a continuous solution to the problem—was the primary goal of the study. Three standard examples were used to show that PINNs can be used to super-resolve time and space flow-field data: Burgers' equation, two-dimensional vortex shedding inside a circular cylinder, and minimum turbulent flow through a channel. A real-world experimental dataset, including hot-wire-anemometry measurements, demonstrated that PINNs' capabilities improved resolution and reduced noise. The results also demonstrated that PINNs are capable of handling data augmentation adequately.

3.3.2 Civil engineering

Deep learning is currently exerting a substantial influence in the domain of civil engineering, specifically in the realm of structural health monitoring (SHM) [295]. SHM systems have been extensively used on a variety of civil infrastructures to detect abnormalities or damage to structures and to track their overall structural health. These systems monitor environmental conditions, structural loadings, and responses over an extended period of time [48]. Many sensors are integrated into an SHM system to collect massive monitoring data that can accurately represent the operational state of the target structure. Deep neural networks are capable of real-time assessment of the structural integrity and detection of anomalies in sensor data originating from infrastructure such as dams and bridges. By taking this proactive stance, engineers are able to detect potential problems prior to their escalation, which ultimately results in improved safety and more effective maintenance strategies.

Early detection of damage constitutes the first stage of monitoring structural health. A new damage index was defined by Entezami et al. [296] using some empirical measures and the notion of the lowest distance value. This method was proposed within the theory of empirical ML and operates in an unsupervised learning method. This non-parametric approach successfully distinguishes between damaged and undamaged states with a high degree of damage detectability and negligible false positive/negative errors. The comparative studies also found that this method outperformed the anomaly detection approaches. When compared to the anomaly detection methods examined in the studies, this method also performs better.

On the basis of unsupervised anomaly detection, a probabilistic ML approach for civil structure health monitoring was proposed by Sarmadi and Yuen [297]. Among the several advantages of the suggested approach are its ability to estimate a trustworthy threshold, prepare discriminative novelty ratings for damage detection, handle the challenging problem of environmental and/or functional variability, and integrate the decision-making and threshold-estimating processes into a single structure. The results revealed that the method is a dependable and impactful tool for monitoring the health of civil structures in different environmental and operating circumstances. However, data quality, the interpretability of complicated models, and the necessity to carefully assess model transferability across various time scales are potential obstacles to structural health monitoring.

When evaluating the time and money saved by utilizing high-performance concrete (HPC), it is essential to accurately predict its compressive strength. To reliably forecast the compressive strength of the HPC, Islam et al. [298] proposed four DL models, namely BiLSTM, GRU, CNN, and LSTM. The model was developed by utilizing a substantial database that contained information about fly ash, cement, coarse aggregate, age, water, sand, blast furnace slag (input variable), and compressive strength (output variable). The study employed a training dataset comprising 80% of the total, with a testing dataset comprising the remaining 20%. During training, DL models demonstrated an R-squared value of approximately 0.960; during testing, it was nearly 0.940. However, the GRU model outperforms other models with an R-squared value of 0.990 during training and nearly 0.961 during testing, indicating high accuracy in both operations. Predicting the compressive strength of HPC using this method can be useful for building environmentally friendly infrastructures without resorting to expensive and time-consuming experiments.

3.3.3 Transportation prediction

Advanced transportation operations and services are enabled by traffic prediction, which is a critical component of the Intelligent Transportation System (ITS) that is designed to address the growing traffic congestion issues. The approach used for forecasting traffic has substantially transformed in the past few decades, transitioning from basic statistical models to the complicated integration of various DL models. Previously, analytical or statistical methods were used to solve problems in this domain. These enhancements have aided traffic management and planning, raised transit road safety and security, reduced maintenance costs, improved public transportation and ride-sharing firm performance, and propelled driverless car development to a new level.

Zheng et al. [299] presented a DL model that automatically extracts fundamental properties of traffic flow data using hybrid and multiple-layer architectures. They designed an attention-based ConvLSTM module that extracts spatial and short-term temporal features using a CNN and LSTM network. By automatically employing various weights to flow sequences at different periods, the attention method was correctly constructed to discern the relevance of the sequences at several intervals. In order to investigate long-term temporal features, the researchers proposed a bidirectional-LSTM (Bi-LSTM) architecture that extracts regular (daily and weekly) periodical features in order to grab the variance trend of traffic flow in prior and posterior orientation. However, only relatively small and simple road networks were considered in the study, which are more complicated and large scale. As a result, CNN and Bi-LSTM methods might be unable to completely utilize traffic flow's complex and dynamic properties. A deep learning-based traffic flow prediction system is demonstrated in Fig. 10. Input data from traffic sensors and GPS devices are processed by a CNN-LSTM model. The model analyzes sequential data to predict congestion levels, which are visually represented on a map with color-coded regions indicating traffic density. The neural network architecture is outlined, showing data progression from input to output prediction.

A multistep prediction model built on attention mechanism-based convolutional neural network (CNN)-LSTM was proposed by Vijayalakshmi et al. [300]. To increase model accuracy, the suggested technique employs the spatial and time-based features of data, which are retrieved by CNN and LSTM. This method facilitates the detection of short-term traffic characteristics (e.g., speed), which is essential for calculating future flow values. Weather and other conditions like accidents and road closure data could be taken into account for the reliability of the suggested technique. Abdollahi et al. [301] proposed a method for forecasting travel time using a multi-step process, which begins with removing both temporal and spatial outliers. To increase prediction accuracy, reduce overfitting risks, and achieve a more robust learner, the researchers employed a deep stacked autoencoder to present lower dimensions' features. While the proposed technique could capture traffic dynamics in general, it was not successful when heavy snow was present or another uncommon occurrence that had a substantial influence on travel times prevailed. To overcome this problem, several deep architectures and representation learning algorithms could be implemented, followed by a performance comparison to find an effective method.

The capacity to aggregate massive amounts of data from multiple sources for identifying events is a key feature of ITS. To address this problem, Reddy et al. [302] introduced a distributed fog-cloud ensemble that is IoT-integrated and context-aware. This ensemble intelligently handles context instances across fog nodes, assuring that the ITS has access to them. Employing a hybrid CNN system improved prediction accuracy, in which each vehicle only remembers local information, but nearby fog nodes can access global events through federated learning. The hybrid CNN model outperformed the LeNet by about 3% when tested with identical

Fig. 10 Deep learning-based traffic flow prediction system

RGB input images; this results in an accuracy of over 95% in the experiments. The adoption of lightweight tracking schemes is required for ITS event detection due to extremely high latency requirements, which severely compromises the efficiency of ITS and makes prediction accuracy a challenge.

3.4 Biomedical

3.4.1 Biomedicine

Many biological and medical data, including information about medical imaging, biological sequences, and protein structures, has been accumulated in recent decades due to the advancements in high-throughput technology. Consequently, deep learning has been extensively applied to biomedical data in biomedicine. For instance, CNNs are widely employed in biomedical image processing due to their extraordinary capacity to assess spatial features. CNNs have a lot of potential in omics analysis [57] and the study of biological signals [303], even though sequencing data employing CNNs is not very common. RNN-based architectures, on the other hand, are designed for sequential data and are more frequently utilized for transcriptome analysis [304] and dynamic biomedical signals [305]. Hence, the focus on deep learning is increasing in the field of biomedical information, and each paradigm may soon find new implementations.

Prediction of protein structure Protein structural analysis now relies heavily on cryo-electron microscopy (cryo-EM) [306], making atomic resolution possible. However, on all but the purest density maps with a resolution of less than 2.5 angstroms, estimating the structural trace of a protein remains difficult [307]. A DL model proposed by Si et al. [308] predicted the alpha carbon atoms throughout the core structure of proteins using a series of cascaded convolutional neural networks (C-CNNs). Using a semantic image classifier, the cascaded CNN was trained on a large number of generated concentration maps. With only a suggested threshold value needed per protein concentration map, this procedure was automatic and significant. To create the primary core trace with alpha carbon placements, a customized tabu-search path walking algorithm was applied. The alpha-helix secondary structural elements were further enhanced by a helix-refinement technique. Finally, to create full protein structures, a unique quality assessment-based creative approach was employed to successfully align protein sequences into alpha carbon traces, using 50 trial maps with resolutions between 2.6 and 4.4 angstroms. In terms of the proportion of connecting alpha carbon atoms, the proposed model produced more comprehensive core traces (88.9%), thus outperforming the Phoenix-based structure construction method with an accuracy of around 66.8%. By including additional protein structural details in the C-CNN or

training the networks with experimental data, further study may enhance this research area.

In order to understand sequence patterns and folding structure, Park et al. [309] introduced deepMiRGene, which utilizes RNNs, particularly LSTM networks. The most significant contribution of this suggested approach is that it does not involve rigorous manual feature development. By utilizing end-to-end DL, this approach eliminates the need for extensive domain expertise and instead utilizes simple preprocessing. However, it is challenging to use an LSTM network right away due to the palindromic secondary structure of microRNA. To solve this problem, deepMiRGene employs a novel learning method in which the secondary structure from the input sequence is segmented into front and back streams. Additionally, deepMiRGene performed better than all other options in terms of sensitivity and specificity on testing datasets with higher sensitivity accuracies of 89%, 91%, and 88% in the method's three datasets. Even though there were significant disparities between the features of the various species, deepMiRGene also performed best when using cross-species data, thus demonstrating the potential for identifying inherent traits.

AlphaFold has revolutionized the field of structural biology by accurately predicting protein structures, a breakthrough that has propelled biomedical research into new territories. However, beyond AlphaFold, significant cutting-edge advancements are shaping the biomedical landscape. One notable example is the integration of generative AI models, such as large language models, into drug discovery pipelines. These models aid in predicting molecular interactions, designing novel compounds, and optimizing drug candidates, significantly reducing the time and costs associated with traditional drug development. Additionally, CRISPR-based genome editing technologies continue to expand their scope, enabling precise modifications at the genomic level to address previously untreatable genetic disorders.

AlphaFold by DeepMind revolutionized protein structure prediction by accurately modeling three-dimensional structures from amino acid sequences. This breakthrough addresses the long-standing protein-folding problem using advanced DL techniques. However, other cutting-edge innovations like RoseTTAFold, graph neural networks (GNNs), and multimodal techniques also merit discussion. A complementary approach to AlphaFold, RoseTTAFold utilizes a tri-branch architecture to integrate sequence information and structural predictions effectively. It shows promise in identifying protein complexes and interactions, which remain challenging for AlphaFold [310]. GNNs are increasingly applied in drug discovery to model molecular interactions and predict binding affinities, enabling the identification of novel therapeutics [311]. By incorporating these advancements, the biomedical domain is poised to address critical

challenges, from precision medicine to understanding complex biological systems.

Another area witnessing rapid advancements is the use of personalized medicine, powered by multi-omics data integration and artificial intelligence. By combining genomics, proteomics, and metabolomics data, researchers can now create highly individualized therapeutic strategies tailored to a patient's unique molecular profile. In parallel, advancements in synthetic biology are enabling the development of living therapeutics, such as engineered bacterial or T-cell therapies, which can be programmed to target specific disease pathways with unprecedented precision. Furthermore, advancements in neurotechnology, such as brain-computer interfaces, are opening possibilities for treating neurological disorders and restoring motor functions in paralyzed patients. Combined, these breakthroughs demonstrate that while AlphaFold has laid the groundwork for transforming biomedicine, a broader ecosystem of innovative technologies is driving the field toward a future of more precise, efficient, and personalized healthcare solutions.

With the development of AlphaFold [136], DeepMind has made a major contribution to computational biology by tackling the problem of protein structure prediction from amino acid sequences. Using advanced neural networks, AlphaFold predicts 3D protein structures very accurately. To describe the spatial configurations of amino acids, it combines evolutionary data with attention mechanisms. With this method, AlphaFold can learn patterns from massive volumes of sequence and structural data to predict complicated protein structures, often with near-experimental accuracy. With its ability to shed light on the molecular mechanics of proteins, AlphaFold has great promise for advancing biotechnology, expediting drug development, and improving our comprehension of biological processes.

A significant enhancement in the modeling accuracy was demonstrated in 2021 by the AlphaFold2 method, which led to a milestone in predicting protein structure. The AlphaFold Database made available to the public the predicted structures of proteins in 21 different species after the AlphaFold2

software was released. Aderinwale et al. [312] developed a tool called 3D-AF-Surfer to assist in protein structural analysis and utilization of AlphaFold2 models. This tool enables real-time structure-based searching for the AlphaFold2 models. The structures in 3D-AF-Surfer were modeled using 3D Zernike descriptors (3DZDs), which are mathematical representations of 3D shapes that are invariant to rotation. In the study, a neural network was built to take 3DZDs of proteins as input and achieved more precise retrieval of proteins with the same fold compared to 3DZDs' direct comparison. The 3D-AF-Surfer tool was utilized to classify the structures of AlphaFold2 models, and the relationship between the confidence levels of these models and intrinsically disordered regions was examined. The 3D-AF-Surfer demonstrated high accuracy while enabling real-time structure search, facilitating interactive analysis of Alphafold2 models.

A variety of predictors have been developed for specific interaction types, along with a generalist method, but their deep-learning accuracy often falls short compared to physics-based methods. These methods are typically specialized and cannot predict complex biomolecular structures involving multiple entity types. AlphaFold 3 (AF3) addresses this by accurately predicting complexes with nearly all molecular types found in the Protein Data Bank (PDB)[211]. It outperforms specialized methods in most categories, particularly in protein structure and protein–protein interactions, due to significant advancements in its architecture and training, as illustrated in Fig. 11.

Genomic sequencing and gene expression analysis Large-scale gene expression profiling has been frequently utilized in the characterization of cellular states in response to various illness circumstances, genetic mutations, and so on [314]. Creating a compilation consisting of thousands of samples of gene expression profiles is still quite unreasonable, even though the expense of entire gene expression profiles has been gradually declining. The computational strategy used by the LINCS program, however, currently relies on linear regression, which restricts its accuracy because
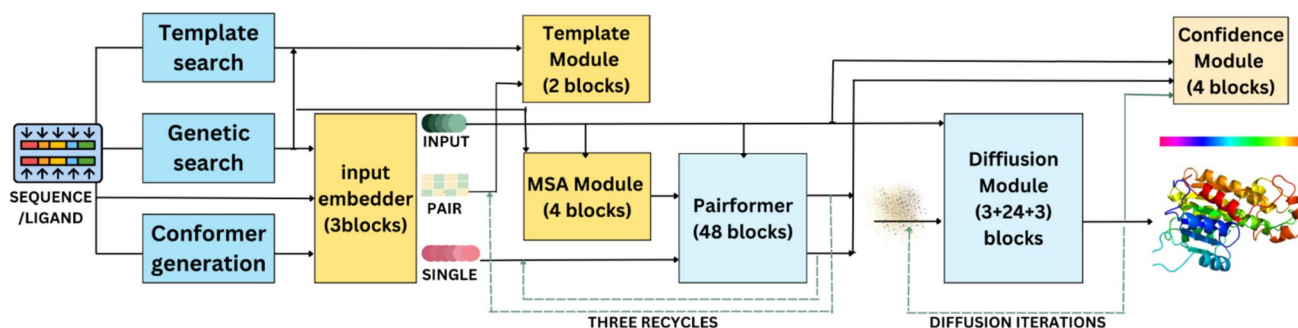

Fig. 11 AlphaFold3 architecture for the inference [313]

it cannot detect complicated complex correlations among gene expressions.

Chen et al. [315] suggested using a DL technique called DGEX to predict the expression of key genes from the landmark gene profile expression. The model was trained using the microarray-based GEO dataset, which contains around 111,100 expression profiles, and its performance was evaluated against other approaches. With a 15.33% proportional improvement, deep learning greatly exceeded linear regression (LR) in terms of the average absolute error of mean through all of the genes. Deep learning outperformed LR in 99.97% of the key genes as per the gene–gene comparison investigation. A separate RNA-Seq-based GTEx dataset with around 20,000 expression profiles was also used to test the performance of the trained model.

Medical image classification and segmentation Deep learning strategies have been utilized in studies segmenting cerebral tumors as a result of their achievement in general image analysis disciplines, including the classification of images [316] and semantic identification [317]. Conditional random fields (CRFs) and fully CNNs (FCNNs) in a single framework were proposed by Zhao et al. [318] to develop a unique deep learning-based identification approach for cerebral malignancy. The model was created to produce a malignancy segment that accurately identifies both appearance and concentration. Instead of employing CRFs as a step after FCNN post-processing, the authors utilized CRF-RNN to construct CRFs, making it simple to train both frameworks as a single deep network. The combined deep learning model was trained in three steps by employing segments and pixel patches. FCNNs were trained in the first step using image patches, while picture slices were used to train the next CRF-RNN in the second stage. In the third stage, the entire network was tuned using image slices. The unification of FCNNs and CRF-RNN, which acquired an accuracy of 88%, could increase segment resilience to model training factors, such as picture patch size and training picture patch count, as per the experimental data. Therefore, using 3D CRF as a step of post-processing could enhance malignancy identification performance.

Gliomas are the most prevalent and deadly type of cerebral tumor, with an extremely low survival probability in the maximum grade [319]. Planning the timeline of treatment is consequently essential for expanding the quality of life for cancer patients. The assessment of these tumors is frequently done using magnetic resonance imaging (MRI), but the volume of data generated by MRI makes it impossible to manually segment the images promptly. Accordingly, this restricts the use of profound quantitative assessments in clinical practice. The enormous spatial and functional heterogeneity among cerebral malignancies makes automatic segmentation a difficult task; hence, dependable and automated segmentation approaches are needed.

An automated segmentation technique based on CNN was suggested by Pereira et al. [320] to explore kernels for glioma segmentation in the images derived from MRI. As a result of the network's smaller number of weights, using small kernels enabled the design of more intricate architectures while also helping to prevent overfitting. Intensity normalization was also utilized as an essential step for pre-processing. While this is considered uncommon in CNN-based segmentation approaches, it proved successful when combined with data augmentation for segmenting brain tumors in images collected from MRI. By simultaneously receiving the top spot for the full, core, and expanding areas in the DSC measure (88%, 83%, and 77%), the proposed methodology demonstrated to be legitimate in the Cerebral Tumor Classification Challenge database.

3.4.2 Bioinformatics

The natural innate immune components known as antimicrobial peptides (AMPs) are commonly the focus of novel therapeutic developments due to the increment of antibiotic-resistant bacteria. Nowadays, wet-lab researchers frequently use machine learning techniques to find promising applicants. For instance, Veltri et al. [321] proposed a DL model to identify antibacterial activity by generating a neural network model that uses a core sequence composition and contains CCN and RNN layers. A comprehensive training and testing dataset incorporating the most recent antibacterial peptide information was utilized. In contrast to the current approaches, the proposed novel deep neural network (DNN) classifier performed better AMP identification. Through the implementation of DNN and RNN layers, the dependency on prior feature generation has decreased. Besides, a reduced-alphabet representation demonstrated that adequate AMP identification can be preserved using 9 different kinds of amino acids based on the embedded parameters. The DNN model performed the best in terms of MCC, auROC, and ACC and achieved a superior accuracy of 91.01% compared to other models, such as AntiBP2 (89.37%) and the CAMP Database RF model (87.57%). Thus, the suggested model eliminates the dependency on domain experts for feature creation by employing a deep network model that automatically extracts expert-free characteristics.

Nowadays, it is possible to measure DNA methylation at a single-cell level due to recent technological advancements. To facilitate genome-wide analysis, strategies to predict unknown methylation patterns are necessary because existing procedures are constrained by insufficient Chg. coverage. Angermueller et al. [322] developed DeepCpG, a computational method focusing on DNN to predict the methylation stages of a single cell and model the origins of DNA methylation dispersion. DeepCpG utilizes relationships

between patterns of DNA sequences, stages of methylation, and nearby CpG sites both within and between cells. The extraction of useful characteristics and model training parts are not separated in the proposed method; rather, DeepCpG is built on a modular architecture and develops predicted sequences of DNA and patterns of methylation through data-based learning. Notably, DeepCpG performed better than RF-trained DNA and CpG characteristics. After training both the RF and DeepCpG models using only DNA sequence information, the highest relative improvements in the accuracy of 80% and 83% were achieved, respectively.

A novel deep learning technique named EPIVAN was introduced by Hong et al. [323] that allows lengthy EPIs to be predicted only from genomic sequences. Pre-trained DNA vectors were utilized to encode regulators and promoters in order to investigate the essential sequential properties. Subsequently, a one-dimensional convolutional model and deep neural units were used to identify both local and global characteristics to make a prediction regarding EPIs in different cell lines. The effectiveness of EPIVAN was compared to that of the neural network models, such as SPEID, EPI-ANN, and SIMCNN. Each model applied the identical test and training set to every cell line. On six cell lines, the proposed model's results and the four predictors were provided in terms of AUPR and AUROC. EPIVAN demonstrated the strongest AUROC of any model, with an accuracy ranging from 96.5% to 98.5%. According to test findings on six cell lines, EPIVAN outperformed the existing models, showed the capability of being utilized as a pre-trained model for further transfer learning, and demonstrated strong transfer ability.

Prediction of the functionality of enzymes is an important step in constructing new enzymes and diagnosing disorders connected to enzymes, which is a major task in bioinformatics [324]. Several studies have generally concentrated on predicting the mechanism of the monofunctional enzyme. However, the amount of multi-functional enzymes is continuously increasing, necessitating the development of novel computational techniques [325]. Zou et al. [326] proposed mlDEEPre, a deep learning network specifically designed to predict the functionality of multi-functional enzymes. Using an auto label issuing threshold and a new transfer function linked with the interaction between several labels, mlDEEPre could predict multi-functional enzymes reliably and quickly. The proposed multi-label model surpassed all other approaches, as it correctly predicted 97.6% of all the observed primary categories in the test dataset with an SD of 0.27. The performance of SVM-NN with an accuracy of 84.7% seemed somewhat better than that of mlDEEPre (82.6%) and GA (80.8%), despite the fact that SVM-NN was worthy of predicting such infrequent class labels generated by unbalanced training samples. Comprehensive tests further demonstrated that mlDEEPre outperformed the other techniques in determining the kind of functional enzyme

utilized, as well as the primary class prediction throughout many parameters. As mlDEEPre and DEEPre are flexible, mlDEEPre could be effortlessly merged into DEEPre, allowing the enhanced DEEPre to handle different functional predictions without human interference.

3.4.3 Drug discovery and toxicology

Emerging contaminants (ECs), which pose a serious risk to human health due to their detrimental effect on the endocrine system, include over eighty thousand endocrine-disrupting chemicals (EDCs). To determine the possible impacts of EDCs, numerous in vitro techniques have been developed, such as signaling pathway studies, ligand-binding protein, reporter gene experiments, and cell proliferation [327]. While in vitro methods are often more economical and quicker than in vivo trials, it is still impractical to analyze thousands of molecules in a reasonable amount of time [328]. This has resulted in the usage of alternative computational methods experiencing exponential growth. To anticipate the toxicological effects of chemicals, the QSAR model is a possible substitute for in vitro techniques.

A DL-QSAR model was evaluated by Heo et al. [329] to give predictions regarding the impacts of EDCs on the endocrine system, particularly estrogen receptor (ER) and sex-hormone binding globulin (SHBG). For the classification and forecasting of probable EDCs, DL-QSAR models were created, and three distinct DL algorithms, namely SAE, DBN, and DNN, were employed. The suggested models' performance was assessed using validation metrics by comparing the classification prediction models with traditional machine learning classifiers, such as LR, SVM, and MLR. Results showed that the DL-QSAR algorithm outperformed traditional machine learning based QSAR models. Accordingly, the DNN-QSAR model adequately described most EDCs' qualitative responses to tests. Compared to the LR and SAE-QSAR models (accuracies of 86.49% and 89.91%, respectively), the DNN-QSAR model achieved an accuracy of 90%. As a result, DNN was more effective for evaluating qualitative responses since it could translate dense chemical identifiers into multidimensional space regions. Additionally, by overcoming multicollinearity and overfitting issues, DNN-QSAR exhibited great performance in computational chemistry. As a result, it was determined that DL can effectively utilize the qualitative characteristics of the EDCs.

Effective methods for assessing potential EDCs are majorly required. An evaluation platform was established by Zhang et al. [328] established an evaluation platform by employing three different machine learning models, such as SVM, linear discriminant analysis (LDA), and classification and regression trees (CART), to identify EDCs through estrogen receptors. The model was modified using 440 compounds, and 109 new compounds were added for

the screening of EDCs. Their predictive capabilities were evaluated by contrasting the screening results with those anticipated by the models derived from classification. The most accurate model was found to use an SVM classifier, which correctly identified agonists at an accuracy level of 76.6% and antagonists with an accuracy of 75% on the test set, including an average predicted accuracy of 75.2%. The overall projected accuracy confirmed by the screening of the EDC assay was 87.57%, illustrating the effectiveness of a synthetic alert for EDCs with ER agonistic or antagonistic actions.

Drugs are often the cause of detrimental results, such as accidents, injuries, and enormous medical expenses. Clinicians can build suitable treatment plans and make effective judgments with the aid of accurate drug-drug interaction (DDI) predictions. Recently, numerous AI-based methods for DDI prediction have been proposed. However, the majority of currently used techniques give less consideration to possible connections between DDI processes and other multidimensional data, such as targets and enzymes. A multimodal DNN for DDI events prediction was suggested by Lyu et al. [330] in order to generate drug multimodal representations in a multimodal deep neural network (MDNN). To examine the complementary aspects of the drug's heterogeneous representations, a multimodal neuronal layer was also developed. Several multi-class classification evaluation metrics, such as F1 score, accuracy, and recall, were used to assess the prediction performance. The proposed model, MDNN, achieved the most stable performance and outperformed DDIMDL by 0.08% on accuracy, 7.1% on F1 score, 1.5% on precision, and 10% on recall. The MDNN model achieved an accuracy of around 99%, according to the comparative study with other models used in different journals, such as DDIMDL [331]. The MDNN model's superior performance can be ascribed to its exploration of both the cross-modality embedding representations of the heterogeneous data and the drug topological integrating expressions in the drug network. This effectively demonstrates how structural information and multimodal features can increase the prediction accuracy of drug-drug interaction irrespective of current drugs or newly-launched FDA-approved drugs and provides a solid, trustworthy basis for research on DDI prediction.

For the prediction of drug-induced liver injury (DILI), Yang et al. [332] constructed 48 classification models using a combination of six distinct molecular fingerprints. These models were developed using a DNN and seven different ML algorithms. In comparing the outcomes of the DNN and ML models, the top-performing model was determined to be the one that utilized DNN with ECFP_6 as its input. The model achieved a balanced accuracy of 0.680, an F1-score of 0.753, and an area within the receiver operating characteristic curve of approximately 0.713. The DNN models built using molecular fingerprints could serve as reliable and effective tools for assessing the risk of DILI in the early stages of developing novel medications.

Drug co-prescription can be safer and more productive if the effects of drug-drug interactions (DDIs) are accurately predicted [333]. There is still potential for improvement in prediction performance despite the many computational methods presented to anticipate the impact of DDIs [334]. These methods attempt to make it easier to find these interactions in vivo or in vitro. A deep learning model was proposed by Lee et al. [335] to more precisely predict the impact of DDIs, as well as to employ autoencoders and a feed-forward deep network to predict the therapeutic effects of DDIs. The experiments utilizing just GSP or TSP or combined GSP and TSP did not produce tests with satisfactory classification accuracy. However, integrating TSP and GSP improved classification accuracy to 97–97.5%. Additionally, the suggested model outperformed standard techniques like SVM (80–83%) and Random Forest (RF) (75–91%). The results showed that TSP and GSP improved prediction accuracy in comparison to SSP alone, and the autoencoder outperformed PCA in reducing the parameters of each profile. Hence, the proposed deep learning model provided a more precise prediction of DDIs and their therapeutic effects.

A summary of the reviewed studies on the uses of deep learning in various sectors is provided in Table 4. Based on our prior discussion, Table 5 represents the relationship between several fields in terms of DL applications and their benefits, specifically how DL techniques on vision, audio, and NLP help other concrete applications such as transportation, agriculture, bioinformatics, and ecology.

Summary Applications of deep learning demonstrate the adaptability and revolutionary power of deep learning in solving complicated problems and advancing technology in many fields. Deep learning has strengths as well as shortcomings when applied to different fields. Although sophisticated audio analysis and precise voice recognition are two of the domain's strongest points, dealing with various accents and ambient noises can be a real challenge. Deep learning excels at optimizing routes and predicting traffic, but its limitations may become apparent in highly urbanized areas or situations where traffic is constantly changing. Nevertheless, these strengths are useful for transportation prediction. For instance, traffic forecasting can be made with 99% accuracy using a CNN-LSTM model based on an attention mechanism. Precision farming has strengths in agriculture, such as crop monitoring and yield prediction. For example, the multi-layer CNN model can reveal 97.13% accuracy in classifying mango leaves affected by fungal disease. However, adapting models to diverse agricultural landscapes is still a challenge. When it comes to cultural differences and complex contexts, natural language processing applications may fall short, but they excel when it comes to sentiment

Table 4 Overview of the surveyed studies conducted on the applications of deep learning

Applications	Algorithms/Models	Objective	Outcome	Remarks	Ref
Transportation prediction	Attention based ConvLSTM, Bi-LSTM	Extracts fundamental properties of traffic flow data using hybrid and multiple layer architectures	The combination of attention ConvLSTM and Bi-LSTM performed better than the existing models	They considered a relatively small and simple road network. Therefore, CNN and Bi-LSTM methods might be unable to completely utilize traffic flow's complex and dynamic properties	[299]
	Attention-based CNN-LSTM	Traffic flow forecasting	The accuracy of the model was found 99%	Weather and other factors such as accidents and road closures can be factored into the model to enhance it	[300]
	Deep stacked autoencoder to present features in a lower dimension	Predict the travel times	Showed better performance than applying Deep NN in the initial training data	Wasn't capable of capturing traffic dynamics during heavy snow	[301]
agriculture	Transfer learning to train Deep Convolution NN	Plant leaf stress detection	Achieved 93% accuracy	For certain diseases (CBSD, BLS, GMD), using the leaflet rather than the full leaf increased diagnostic accuracy. However, using whole leaf photos enhanced accuracies for others (CMD and RMD)	[252]
	Multi-layer CNN	Distinguish between healthy and stressed mango leaves	Achieved 97.13% accuracy	Employing a new activation function in place of Softmax can improve CNN's performance	[336]
	CNN	Plant disease identification	Mildly diseased images were difficult to identify but accuracies were higher for other cases	In all situations, a few hundred photos appeared to be sufficient to produce credible findings, but this quantity must be approached with caution	[254]
	UNet-CNN	Classify and identify cucumber powdery mildew-affected leaves	CNN model segmented the sick powdery mildew on cucumber leaf pictures with a mean pixel accuracy value of 96.08%	Lack of appropriate amount and diversity of the datasets	[255]
	CNN	Fish species classification	Achieved accuracies of over 90%	General deep structures in the experiment should be fine-tuned to increase the efficacy to identify vital information in the feature space of interest, in order to reduce the requirement for vast volumes of annotated data	[257]

Table 4 (continued)

Applications	Algorithms/Models	Objective	Outcome	Remarks	Ref
Natural language processing	Collaborative adversarial network	Paraphrase identification	The model outperforms the baseline MaLSTM model	Shows excellent potential using the CAN for paraphrase identification	[173]
	RNN (LSTM) and several other models	Paraphrase identification	RNN (LSTM) outperforms the others	RNN shows significant performance in NLP	[174]
	Weighted Transformer	Machine Translation	Outperformed state-of-the-art methods	The model ignores the modeling of relations among different modules	[337]
	Single Neural Machine Translation	Machine Translation	The method works reliably on Google scale production setting	The performance of zero-shot translation is frequently insufficient to be practical, as the basic pivoting strategy quickly outperforms it	[175]
	Deep-attention model	Machine Translation	In comparison to the best methods currently available, deep attention performs exceptionally well	The technique could be implemented for other tasks like summarization and could adapt to more complex attention models	[176]
	BiLSTM with CRF	Sentiment analysis (extracting aspect opinion target expression)	Outperformed the existing research	Unable to represent several aspects of sentences and did not investigate the explicit location contexts of words	[180]
	LSTM	Sentiment Analysis	Outperformed the existing research	Unable to represent several aspects of sentences and did not investigate the explicit location contexts of words	[181]
	Domain attention model	Multi-domain sentiment categorization	Evaluated on multiple datasets and showed better performance than the state-of-the-art techniques	Can pull out the most distinctive features from the hidden layers, reducing the number of labeled samples required	[338]
	Weakly-supervised multimodal deep learning (WS-MDL)	Prediction of multimodal attitudes in tweets	Better performance than supervised and other weakly supervised models	The order of the emotion icon levels might be further investigated, which may be included as a constraint to the proposed WS-MDL technique	[339]
	CNN, LSTM	Sentiment classification	Achieved accuracy of 87%	In future work, bag-of-word and word embedding techniques can be combined	[340]
	ConvLstm (merged CNN and LSTM)	Sentiment analysis	Performed better than existing works with less number of parameters	Local information loss may be reduced, and long-term dependencies can be captured using the suggested design	[341]
	Attention-based LSTM	Question answering	Better performance than baseline approaches	Modeling more than one aspect simultaneously with the attention mechanism would be an interesting addition to the experiment	[185]
	Hierarchical TF-IDF document retriever and a BERT document reader	Question answering	Significant improvement of the baseline techniques for the Arabic language	They introduced a dataset (ARCD). However, ARCD's questions were created with certain paragraphs in mind; without that context, they might seem ambiguous	[186]
	CNN, LSTM	Visual question answering	Will be helpful to study the reduction of the network model size	Determining the rank is an NP hard problem in the low-rank decomposition, and their method is still constrained in this area by inserting hyper-parameters	[187]
	Tree-LSTM	Visual question answering	Better performance than existing Hie and Deeper LSTM	The representative ability of the network could be improved in the future	[188]
	CNN, LSTM	Visual question answering	Outperformed the existing studies by 5%	The approach might be used on unstructured information sources, such as online text corpora, in addition to structured knowledge bases	[342]
	Reinforcement Learning technique and an encoder-extractor network architecture's RNN sequence model	Summarization	Performed better than baseline Bi-LSTM	The model may suffer from significant variance since they use an approximation in the Reinforcement Learning method training objective function	[192]

Table 4 (continued)

Applications	Algorithms/Models	Objective	Outcome	Remarks	Ref
Biomedicine	Cascaded-CNN (C-CNN)	Prediction of alpha carbon atoms throughout the core structure of proteins	C-CNN outperformed (88.9%) the Phoenix-based structure construction method (66.8%)	Adding protein structural details for training the networks can enhance the model	[308]
	deepMiRGene (RNN+LSTM)	Prediction of structural characteristics of precursor miRNAs	deepMiRGene performed better, having accuracy between 88%–91%	The most important contribution was the elimination of rigorous manual feature development	[309]
	DNN	Prediction of gene expression interference	Outperformed linear regression in 99.97% of the key genes	Demonstrated higher accuracy than the linear regression model	[315]
	FCNNs, CRF, RNN	Cerebral malignancy identification	The integration of FCNNs and CRF-RNN acquired an accuracy of 88%	2D CNNs lack the essential capabilities to fully utilize 3D information from MR	[318]
	CNN	Identifications of kernels for gliomas segmentation in the images derived from MRI	88%, 83%, 77% accuracy acquired	CNN model demonstrated to be legitimate in the Cerebral Tumor Classification database	[320]
Bioinformatics	DNN, RNN	Identification of AMPs	Best performance, having an accuracy of 91.01%	Eliminated the dependency on domain experts for feature creation by employing a deep network model	[321]
	DeepCpG	Single-cell DNA methylation prediction	Obtained accuracy of 83%	The strength of the DeepCpG extracting predictive sequence from large (<1000 bp) DNA sequence	[322]
	EPIVAN	Identification of enhancer-promoter interactions	Accuracy ranging from 96.5% to 98.5% in different datasets	The model can be utilized in transfer learning	[323]
	mlDEEPre	Prediction of Multiple Enzyme Functions	97.6% accuracy with a standard deviation of 0.27	mlDEEPre can be effortlessly merged into DEEPre to handle functional predictions	[326]
Disaster management systems	CDNN algorithm	Flood catastrophe identification	Higher accuracy (93.2%) level than the current approach of DNN and ANN	The model can be improved with IoT-based devices using cutting-edge algorithms at each stage of flood identification	[276]
	CNN-based architecture	Early fire detection	High accuracy of from 89 to 99%	Capable of early-stage fire detection with high accuracy and in addition to providing an automated response	[279]
	SVM, LR, CNN classifier	CNN-based deep learning algorithm to classify the trending catastrophic topics from social media	SVM (63%–72%), LR (44%–60%), CNN (81%)	CNN took longer to learn than SVM and LR as there were more parameters to consider, making it difficult to employ for web-based learning	[51]
	RNN, CNN, MCA-based model	Disaster information management	Minimal accuracy 73%	MCA based model can include more textual and metadata to optimize the final categorization results	[343]

Table 4 (continued)

Applications	Algorithms/Models	Objective	Outcome	Remarks	Ref
Drug discovery and toxicology	DNN–QSAR model (SAE, DBN, DNN)	Prediction of impacts of EDCs on the endocrine system, particularly (SHBG) and (ER)	The accuracy of DBN-QSAR was 90%,	DNN was more effective for evaluating qualitative responses	[329]
	SVM, LDA, CART	Identifying EDCs through the ER	Overall projected accuracy confirmed by the screening of EDC assay was 87.57%	The estrogenic activity of 109 compounds was predicted using the best model, SVM	[328]
	CNN	Reaction prediction and Retrosynthesis	The model achieved a 95% retrosynthesis accuracy and a 97% response prediction accuracy	The model demonstrated the extreme capability of ranking the real reactions precisely	[344]
	Multimodal deep neural network (MDNN)	To build a connection between DDI and other targets and enzymes	DDIMDL outperformed by achieving around a 99% accuracy rate	Demonstrated the impact of structural information and multimodal features on prediction accuracy	[330]
	Autoencoder (SSP, TSP, GSP)	Predict the therapeutic effects of drug-drug interaction	classification accuracy between 97%–97.5%	Therapeutic implications of the DDIs predicted should be verified	[335]
Partial differential equations	5-layer deep neural network	Building multi-physics/multi-scale modeling	For noise-free training data, the error in predicting unknown parameters was 0.023% and 0.006%, respectively	Capable of accurately identifying the unknown parameters despite variations and the significant time interval between the two training images	[345]
	Deep residual network (ResNet)	Solving SBVPs with high-dimensional uncertainty	DNNs predicted the mean with less than 1.35% relative L_2 error	Trained DNNs were found to transition effectively to inputs from non-distribution data	[346]
	CNN-based encoder-decoder network	Solving SPDE and uncertainty assessment tasks	300 epochs and 200 epochs for PCS and DDS, respectively, along with 8 to 12 images	Surrogate model PCS outperformed data-driven DDS and demonstrated the capability of incorporating prediction uncertainty	[347]
Financial fraud detection	LTSM, RNN, ANN	Evaluating the effectiveness of deep learning techniques in financial fraud detection	LSTM technique achieved the best performance	Did not specify the network size and sensitivity	[267]
	LSTM	Credit card fraud detection	Detected suspicious financial activities and alerted the appropriate authorities with 99.95% accuracy	It can study even complex data structures and adjust to changed fraud trends dynamically	[263]
	H_2O framework	Credit card fraud detection	Enabled multiple algorithms to aggregate as modules, and their outputs can be integrated to improve the final output accuracy	Adding more algorithms with equivalent formats and datasets can elevate this model	[269]

Table 4 (continued)

Applications	Algorithms/Models	Objective	Outcome	Remarks	Ref
Computer vision	CNN	Generic object detection	The proposed method increased the RCNN's mean average precision from 31 to 50.3. It also exceeds GoogLeNet, the winner of the ILSVRC2014, by 6.1%	The addition of a def-pooling layer provides the model with a richer set of options for handling deformations and incorporating deep architectures	[205]
	Semi-Supervised	Overcoming the crucial drawbacks of the object detection model	Translated the input data into more compact and abstract representations, which enhanced the model's convergence, stability, and performance	Allows dynamic modification of the model to the digital photogrammetry conditions	[206]
	Global adversarial deep CNNs	Face recognition model	Protocol attained 94.05% state-of-the-art verification accuracy	Can generate identity-distilled features and also extract concealed identity-dispelled features but lacks generalization	[211]
	CNN and LSTM	Human activity recognition	Outperformed previous models by up to 9%	The framework is useful with similar sensor modalities, and it can combine them for enhanced performance	[220, 268, 269]
	LSTM	Decreasing the parameters and computational cost of the preceding action and activity recognition model	The overall accuracy of this project was 95.78% on the WISDM dataset	It was tested on enough datasets to guarantee its generalizability	[221]
	Autoencoder	Building an autoencoder system with minimal supervision via face identities	The model was able to generate identity-distilled features and also extract concealed identity-dispelled features	Did not specify the exact dependence on range resolution	[271, 272]
	ResNet based network	Human pose recognition	The network could forecast joint heatmaps and relative displacements of all key points for all people in real-time	Notwithstanding its simplicity, it offered state-of-the-art outcomes against the challenging benchmark	[227]
Ecology	Mask R-CNN	To compare the deep learning algorithms for determining fish abundance with human counterparts	The deep learning algorithm outperformed marine specialists by 7.1% and citizen scientists by 13.4%	Evaluated abundance with stable results and is more portable across survey locations than humans	[245]
	CNN	Tree defoliation identification	Identified tree defoliation 0.9% less accurately than a group of human experts	Can produce errors during complex situations due to data deficiency	[246]
	DNN and Transfer learning	To automate characteristic recognition and extraction	The top-one species identification accuracy was 80%, while the top-five accuracy was 90%	It is not possible anymore owing to vast differences in visual appearance	[235]
	CNN	Directly identifying ant genera from the profile, head, and dorsal views of ant photos	Gained over 80% accuracy in top-1 classification and over 90% in top-3 classification while reducing total classification error	Contributes novel understanding of ensembles for multi-view structured data and transfer learning processes for probing commonalities in multi-view CNNs	[249, 268, 269]

Table 4 (continued)

Applications	Algorithms/ Models	Objective	Outcome	Remarks	Ref
Fluid dynamics	Deep CNN	To scrutinize heat transport characteristics of turbulent Rayleigh–Bénard convection	Simplified the complicated 3D superstructure in the midplane layer	Did not provide information on the network's efficiency at lower Rayleigh numbers	[283]
	Physics-guided neural networks (PGNN)	To forecast turbulent flow	Showed significant improvements in error prediction over state-of-the-art benchmarks	Precisely simulate the turbulent kinetic energy field and spectrum that are vital for accurate turbulent flow prediction	[284]
	DLROM and POD	Describing the conservation of mass and momentum in fluids	Comparison with earlier ROMs clarified that the DLROM has better potentiality in the prediction	The consequences of applying this approach to transitioning parametric constraints, such as real-time response to natural disasters, are uncertain	[47]
	ROMs and POD	To conduct a comparative analysis using three ROMs applied to a biological model	A comparison of POD-DEIM and Gappy POD solutions revealed similar levels of accuracy	The model did not converge when the number of MPE points was low	[285]
	POD and ROM	Using an efficient adjoint technique to optimally collect targeted observations	When compared to the high fidelity model, the size of the problem is decreased by a factor of 200	It ensures that the sensors are positioned at the optimum distance from one another	[348]

analysis and machine translation. Genomic analysis and drug discovery are two areas where DL excels in biomedicine and bioinformatics. In 99.97% of targeted genes, deep learning can outperform linear regression in terms of error reduction. For prolonged enhancer-promoter interactions (EPIs) that can be predicted solely from genomic sequences, the EPI-VAN deep learning model achieves an accuracy of 96.5% to 98.5%. However, there are a few obstacles, such as the difficulty in understanding complicated models and a lack of data in specific fields.

While alert systems are helpful for disaster management, they may fail to adequately anticipate truly unique occurrences. Toxicology and drug discovery use deep learning to save time and effort, but it's not always easy to predict complicated biological interactions. Though deep learning is an efficient solution to partial differential equations, it can be difficult to figure out the physical meaning of black-box models. Since adversarial attacks are possible, financial fraud detection depends on deep learning's anomaly detection strengths. With an accuracy of 99.5%, the deep learning model LSTM is capable of identifying fraud-related credit card activities and notifying the relevant authorities. Despite their usefulness, computer vision applications like object detection and image recognition can be influenced by biases and necessitate massive labeled datasets. Deep learning has many applications in ecology, but it can be difficult to generalize results to other ecosystem types. Similarly, fluid dynamics models benefit from deep learning, but they still have a ways to go before they can fully capture the complexities of real-world phenomena. Interpretability, generalizability, and domain adaptability are three areas where deep learning applications still face obstacles despite their many strengths.

4 Challenges and benefits of deep learning applications

Deep learning is an incredibly important computational tool primarily due to its accuracy in data prediction and analysis. Since DL does not require the use of previously processed data, the input of raw data can be computed through branched layers that separately analyze the data and represent it to the next layer for further processing [349]. Deep learning is useful when dealing with enormous amounts of data since it can extract information, eliminating the need for data training and the associated computational expenditures. It is also notable for its expression and optimization capabilities, which make it so effective at processing data without training [350]. As the volume and complexity of available data continue to rise exponentially, deep learning is gaining recognition as a crucial tool for more efficient data collection and analysis [351].

Table 5 How deep learning techniques on vision, audio, and NLP help the other concrete applications

Application area	DL application field	Benefits and applications
Transportation	Object detection	- Enhancing autonomous driving by identifying pedestrians, vehicles, and road signs - Real-time surveillance for safer navigation - Traffic flow analysis and optimization
	Image segmentation	- Identifying road lanes and obstacles for self-driving cars - Accurate mapping and localization using satellite imagery
Agriculture	Crop monitoring	- Identifying crop diseases, pests, and nutrient deficiencies through image analysis - Precision agriculture and yield prediction
	Object recognition	- Differentiating between various plant species and weeds for targeted intervention
Disaster management	Image analysis	- Aerial imagery interpretation to assess disaster extent and damage - Identifying trapped survivors in disaster areas
	Sentiment analysis	- Analyzing social media data for real-time disaster response and understanding public sentiment
Drug discovery	Molecular structure	- Predicting molecular interactions and drug-binding sites for drug design - Accelerating the drug discovery process
	Image analysis	- Analyzing medical images to identify potential drug candidates and understand their effects
Toxicology	Toxicity prediction	- Predicting toxicity of chemicals and compounds, aiding risk assessment - Reducing animal testing through computational models
Bioinformatics	Sequence analysis	- Analyzing genetic data for disease prediction and personalized medicine - Identifying genetic markers for various conditions
	Protein structure	- Predicting protein structures to understand their functions and interactions - Advancing drug discovery and understanding diseases
Ecology	Image classification	- Identifying and tracking wildlife species for biodiversity monitoring - Detecting habitat changes and invasive species
	Sound classification	- Monitoring ecosystem health by analyzing animal calls and environmental sounds
Fluid dynamics	Image analysis	- Analyzing fluid flow patterns in images or simulations - Identifying turbulence and vortices for better predictions
	Data analysis	- Processing large volumes of flow data for insights and predictions
Civil engineering	Image recognition and analysis	- Analyze aerial or satellite images to monitor infrastructure - Detect changes or potential concerns in construction areas - Evaluate the state of bridges and roads
	Noise monitoring	- Assessing the effects of construction activities on neighboring communities - Assuring attention to noise regulations - By examining acoustic signals, structural anomalies can be detected

As demonstrated in the literature, deep learning applications offer an extensive range of advantages across diverse domains. Speech and audio processing techniques demonstrate exceptional proficiency in accomplishing tasks such as voice recognition, facilitating enhanced communication interfaces with superior accuracy. Deep learning models in transportation prediction offer accurate forecasts, optimizing traffic flow and improving overall efficiency. These applications enhance agriculture as they facilitate precision farming by monitoring crops and detecting diseases, resulting in higher crop yields. Deep learning is utilized in natural language processing applications to achieve sophisticated language comprehension, which consecutively enhances machine translation and sentiment analysis by providing more nuanced results. It expedites the examination of intricate genomic data in biomedicine and bioinformatics, facilitating drug discovery and disease prognosis.

Several challenges of deep learning applications are common across various fields. Unfortunately, deep learning models are not always practical in environments with limited computing resources because of how much power they consume. Because of its independence from training data, deep learning demands rigorous data collection for proper analysis and processing of such a large number of data. Hence, medical, research, healthcare, and environmental data are challenging for large-scale data compilation, reducing the efficacy of deep learning [352]. Data quality and structure are particularly concerning, as data from health, research, and environmental studies are highly heterogeneous, full of ambiguity and noise, and often incomplete, which frequently presents problems for the model. Another issue with deep learning is that it typically assumes inputs to be static vectors and cannot readily incorporate time variables as inputs. Due to the complex signal relationship across time, samples

are irregular, which influences the model's performance, particularly when dealing with health data [353–358].

Domain complexity, such as diverse data and insufficient information, and black boxes, the complexity of algorithms that aren't understood, offer challenges for the system and inhibit the growth of data comprehension [349, 359, 360]. In important domains like biomedicine, where comprehending the decision-making approach is essential, the interpretability of these models continues to be an issue. However, the interpretability of complex models can be significantly improved in practice through the implementation of knowledge distillation and attention mechanisms. Knowledge distillation involves the transfer of insights from a complex model to a simpler model, achieving a balance between predictive performance and interpretability. Attention mechanisms, inspired by human cognitive processes, facilitate the focus of models on specific elements of input data, thereby providing practitioners with valuable insights into the process of decision-making [361]. In practical implementation, these solutions are seamlessly integrated into training and evaluation pipelines, bridging the gap between model complexity and human understanding. The real-world impact lies in fostering trust and confidence in AI systems, as interpretable models enhance user comprehension and enable practitioners to identify biases, understand model decisions, and troubleshoot effectively, contributing to the overarching goals of transparent and trustworthy artificial intelligence [362].

The intelligibility of complicated biological interactions might be a problem for drug development applications. When detecting financial fraud, deep learning models are vulnerable to adversarial attacks. Despite these obstacles, deep learning still drives innovation and solves complex problems in many fields. Some studies have found that using multimodal data improves the accuracy of the results, while others report that the heterogeneous nature of the data makes it difficult for the program to implement the necessary mechanisms [363–366]. It is also challenging for deep learning algorithms to overcome the problem of label omission in many datasets [56]. The primary challenges and advantages associated with the integration of deep learning techniques within the investigated fields are presented in Table 6.

5 Identified research gaps and future directions

5.1 Research gaps

Although deep learning has achieved notable progress across numerous fields, there remain several domains that continue to be investigated and developed further or have not yet reached their complete understanding:

i) For effective training, deep learning models frequently require vast quantities of labeled data. In domains where it is challenging to obtain large labeled datasets, such as specific medical disciplines or niche industries, the investigation of deep learning adaptations for regimes with limited data remains an unresolved challenge.

ii) Deep learning models, specifically deep neural networks, are frequently referred to as "black boxes" on account of their complicated architectures. Ongoing efforts are directed towards improving the interpretability and explainability of these models, with the ultimate goal of fostering confidence in their decisions, particularly in important domains such as finance and healthcare.

iii) Transfer learning has demonstrated efficacy in general tasks; however, further investigation is needed regarding the adaptation of pre-trained models to precisely match the characteristics and demands of specialized domains. Customizing pre-trained models to fit particular domains may result in improved performance and broader applicability.

iv) Deep learning models are vulnerable to adversarial attacks, in which even minor perturbations in the input can result in inaccurate predictions. Ongoing attention is being given to enhancing the resilience and security of DL models, particularly in domains where safety is paramount, including healthcare and autonomous vehicles.

v) The efficient management of temporal dynamics and sequential data presents an immense challenge. Time-series data that effectively captures long-term dependencies and patterns requires enhanced architectures and methodologies.

vi) An expanding field of study addresses ethical issues and prevents biases in deep learning models. Ongoing investigation is necessary to guarantee fairness, accountability, and transparency in model predictions.

vii) The seamless and successful integration of information from multiple modalities, including text, images, and audio, is an area of ongoing development known as multimodal integration. Ongoing research struggles to develop models capable of utilizing a variety of data types in order to achieve a complete understanding.

viii) The investigation of how to implement deep learning models on devices with limited resources, such as edge devices or Internet of Things devices, while simultaneously ensuring optimal performance and energy efficiency, is a current and dynamic field of study.

5.2 Potential avenues of future research

The potential of deep learning applications in the future is immense, presenting numerous promising avenues for scientific research and advancement. A summary of future directions for the application of deep learning across the various disciplines investigated in the present review is presented in Table 7.

Table 6 Key challenges and benefits of implementing deep learning in the surveyed disciplines

DL application fields	Challenges	Benefits
Natural Language Processing (NLP)	– Imperfect data can lead to skewed results – Computational resources are a restricted commodity when dealing with large models – Some applications may struggle due to a lack of explainability	– Computes at or near the state-of-the-art on several NLP tasks – Allows for the use of chatbots, sentiment analysis, and the translation of languages – Supports massive amounts of textual information
Computer vision	– Annotating and labeling data can take a lot of time – The processing power required by some deep learning models may be high	– The ability to recognize faces and other objects in pictures is unlocked – Reached the level of performance seen in humans in some situations – It has potential uses in robotics and self-driving cars
Speech and audio processing	– Needs a lot of labeled data to work – Extremely computationally demanding, which can limit its applicability to devices capable of real-time processing	– Increased accuracy in speech recognition – Useful for a wide range of audio applications, including speech synthesis and noise cancellation – Capable of handling massive and intricate data sets
Disaster management	– Depending on the quantity and quality of available data – The stability of the model under severe circumstances is a matter of concern – Issues about the morality of data collection and analysis	– Helpful for early warning systems, estimating damages, and allocating available resources – Uses satellite and sensor data for tracking emergencies – Facilitates quicker responses and better decisions
Agriculture	– Needs knowledge and data unique to the given domain – Models may have limited applicability outside a specific geographical area or crop type	– Facilitates disease diagnosis and crop yield forecasting – Drone footage and sensor information improve precision farming – Be able to make the best use of available resources
Transportation prediction	– The use of location data raises issues of data privacy – Deep learning algorithms may be difficult to interpret – Calls for a lot of processing power	– Capable of simulating and predicting complex traffic patterns – Makes use of current information to improve forecasts – Facilitates the use of various transportation modes
Biomedicine and bioinformatics	– Handling sensitive medical information raises ethical and privacy concerns – Model interpretability is especially important for life-saving medical applications	– Facilitates the study of drugs, proteins, and genomes – Reduces the time it takes to analyze medical images and diagnose diseases – Capable of dealing with complex biological data sets
Drug discovery and toxicology	– Drug discovery with limited interpretability carries some inherent risks – Adopting AI-driven drug discovery faces regulatory obstacles	– Finds potential drug candidates faster and helps predict their safety – Supports substantial chemical databases – Reduces the time and costs needed for experiments
Fluid dynamics	– The computational demands of 3D simulations are very high – Black-box models have limited knowledge of the underlying physical processes – Issues with precision and consistency can arise in a variety of contexts	– Facilitates modeling and prediction of complicated fluid dynamics – Increases the speed at which CFD simulations can be run – Helpful in improving engineering and aerodynamic designs
Partial differential equations	– Computer resources for 3D simulations are particularly demanding – The stability and precision of deep learning PDE solvers are not always easy to achieve	– Specializes in fluid dynamics and solving complex PDEs numerically – Allows for varied simulations and models – Ability to enhance engineering designs
Ecology	– Remote areas present unique challenges for data collection – The ability to transfer models between ecosystems varies	– Assists in the tracking and preservation of biodiversity – Forecasts the future results of current environmental changes – Uses sensor and satellite data for ecological study
Financial fraud detection	– Training on large historical datasets is necessary – There is the potential for model drift and false positives/negatives	– Explores financial data for intricate patterns – Facilitates better transaction tracking in real time – Enables more precise fraud detection

Table 7 Prospects for the future application of deep learning across various fields of study

Disciplines	Future directions
Speech and audio processing	– Advances in deep learning about resilient speech and audio processing might entail the construction of models capable of more efficiently managing a range of accents, dialects, and environmental noises – Investigating unsupervised learning techniques for learning representations in audio signals offers an additional prospective field of study
Transportation prediction	– Construct models that aim to optimize traffic management, dynamic traffic flow adaptation, and real-time data integration – Investigate the potential for integrating deep learning with emerging technologies, including autonomous and connected vehicles
Agriculture	– Address issues caused by a lack of data in specific agricultural settings and continue to build models for precision farming decision support, crop disease identification, and yield prediction – Consider incorporating IoT data and remote sensing for a more thorough examination
Natural language processing	– Enhance understanding of cultural variations, idiomatic expressions, and context by models – Investigate developments in contextual embeddings, improve reasoning capabilities, and construct frameworks to facilitate understanding across different languages and cultures
Biomedicine and Bioinformatics	– Develop models for the integration of multi-omics data, personalized medicine, and the discovery of new drugs – In addition to concentrating on interpretable models for clinical decision support, investigate the application of deep learning to understand complicated biological interactions
Disaster management	– Construct decision-making models in real time that effectively incorporate a wide range of data sources, such as social media and satellite imagery – Improve upon existing models to facilitate comprehensive disaster risk assessment, adaptive resource allocation, and early alert systems
Drug discovery and Toxicology	– Explore the potential of deep learning applications to enhance the precision of drug toxicity predictions, optimize drug design processes, and fully understand complicated molecular interactions – Examine the potential of generative models in the generation of novel pharmaceutical compounds
Partial differential equations	– Enhance the stability of deep learning-based solvers, develop more efficient architectures for numerical simulations, and expand the applicability of these methods to high-dimensional problems – Analyze the application of neural networks based on physics to the solution of complex equations
Financial fraud detection	– Improve the interpretability of the model and its ability to adapt to evolving fraud patterns – Investigate how the integration of explainable AI, anomaly detection, and reinforcement learning can enhance the robustness of financial transaction fraud detection
Computer vision	– Enhance adaptability against adversarial attacks, concentrate on achieving precise image recognition, progress models for comprehending visual relationships, and investigate unsupervised learning techniques – Deep learning's application to 3D vision, scene comprehension, and object tracking should be investigated
Ecology	– Tackle the obstacles associated with the scarcity of labeled data in ecological studies – Construct models comprising capabilities for identifying species, monitoring ecosystems, and assessing biodiversity – Consider the integration of environmental monitoring technologies with deep learning
Fluid dynamics	– Develop neural architectures that capture complicated fluid behaviors, optimize for efficient simulation and turbulence modeling, and scale up applications to real-world scenarios – Explore the application of deep learning in aerodynamics, fluid–structure interaction, and optimization of fluid flow problems
Civil Engineering	– Develop predictive models that scrutinize sensor data derived from structures such as roads, bridges, and deterioration – Developing more sophisticated computer vision systems and algorithms for reinforcement learning for use by autonomous construction equipment may be the subject of future studies

6 Conclusions

To effectively forecast and analyze data at all levels, deep learning uses cutting-edge methods, combining ideas from machine learning, deep neural networks, and artificial intelligence to construct models that utilize data representations at each level. This review investigated numerous deep learning frameworks, deep learning application pathways across diverse fields, and the obstacles and technical issues associated with using deep learning. Research gaps and potential future directions for deep learning applications were also addressed. Research domains might specialize in addressing specific difficulties and issues by analyzing the common obstacles encountered by deep learning applications across numerous fields. Data volume, quality, modeling, domain complexity, and representation are some issues that deep learning applications have in common with other

disciplines. In order to facilitate the efficient experimentation and implementation of deep learning algorithms, deep learning frameworks provide developers and researchers with essential libraries and tools. This, consequently, promotes innovation and breakthroughs across various disciplines. TensorFlow and PyTorch, for example, are widely recognized for their flexibility and support from the community. These frameworks exhibit exceptional performance in the domains of computer vision, bioinformatics, and natural language processing. In the fields of financial fraud detection and drug discovery, TensorFlow, PyTorch, and Keras are dominant due to their robust libraries.

Demonstrating its revolutionary potential and adaptability, deep learning advances technology in many fields while solving complex problems. For instance, in the fields of biomedicine and bioinformatics, deep learning is particularly proficient in the areas of drug discovery and genomic analysis. In terms of error reduction, deep learning can outperform linear regression in 99.97% of the targeted genes. The EPIVAN deep learning model can achieve an accuracy of 96.5% to 98.5% for prolonged enhancer-promoter interactions (EPIs) that can be predicted solely from genomic sequences. However, deep learning models use much power, so they aren't always practical in places with limited computing power. In addition, domain complexity, characterized by a wide range of data and inadequate information, along with black boxes, which refer to algorithms that lack understanding, pose challenges for the system and impede the development of data comprehension.

Deep learning applications can be developed by studying sophisticated architectures, focusing on ethical and interpretability issues, incorporating developing technologies like quantum computing, and expanding deep learning applications. When processing massive amounts of data, it is best to employ a gated architecture, such as GRU units, for extracting persistent information. In a neural network designed for multimodal learning, some neurons are used for all tasks, while others are trained to perform specific tasks. The irregularity problem in temporal sequence similarity measurement is addressed by a suggested approach that uses dynamic time warming. For labeling challenges, data can be implicitly labeled by applying acquired knowledge to fresh datasets for the same task or by utilizing an autoencoder for a variant architecture such that transfer learning can be accomplished. Methods such as knowledge distillation, which condenses the data learned from a complex model to make it simpler and easier to execute, and attention mechanisms, which use an understanding of historical data to predict results, are also being used to enhance interpretability. Recent advances in quick solutions to current difficulties point to a promising future for deep learning techniques in multi-field applications. Future research should focus on better understanding the issues related to the construction of adequate datasets for deep learning models, including but not limited to data quality, volume, domain complexity, and privacy.

Declarations

References

1. Sarker IH (2021) Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. SN Comput Sci 2:420
2. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:463–444. https://doi.org/10.1038/nature14539
3. Darwish A, Hassanien AE, Das S (2020) A survey of swarm and evolutionary computing approaches for deep learning. Artif Intell Rev 53:1767–1812
4. Buduma N, Buduma N, Papa J (2022) Fundamentals of deep learning. O' Reilly Media, Inc., CA, USA
5. Jaakkola H, Henno J, Mäkelä J, Thalheim B (2019) Artificial intelligence yesterday, today and tomorrow. In: 2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO). IEEE, pp 860–867. https://doi.org/10.23919/MIPRO.2019.8756913
6. Zhu H, Luo M-D, Wang R et al (2021) Deep audio-visual learning: A survey. Int J Autom Comput 18:351–376
7. Schoneveld L, Othmani A, Abdelkawy H (2021) Leveraging recent advances in deep learning for audio-Visual emotion recognition. Pattern Recognit Lett 146:1–7. https://doi.org/10.1016/J.PATREC.2021.03.007
8. Michelsanti D, Tan Z-H, Zhang S-X et al (2021) An overview of deep-learning-based audio-visual speech enhancement and separation. IEEE/ACM Trans Audio, Speech, Lang Process 29:1368–1396
9. Saleem MH, Potgieter J, Arif KM (2021) Automation in agriculture by machine and deep learning techniques: A review of recent developments. Precis Agric 22:2053–2091

10. Wang C, Liu B, Liu L et al (2021) A review of deep learning used in the hyperspectral image analysis for agriculture. Artif Intell Rev 54:5205–5253

11. Zhang Q, Liu Y, Gong C et al (2020) Applications of deep learning for dense scenes analysis in agriculture: A review. Sensors 20:1520

12. Wang D, Cao W, Zhang F et al (2022) A review of deep learning in multiscale agricultural sensing. Remote Sens 14:559

13. Kashyap PK, Kumar S, Jaiswal A et al (2021) Towards precision agriculture: IoT-enabled intelligent irrigation systems using deep learning neural network. IEEE Sens J 21(16):17479–17491. https://doi.org/10.1109/JSEN.2021.3069266

14. Wang Y, Zhang D, Liu Y et al (2019) Enhancing transportation systems via deep learning: A survey. Transp Res part C Emerg Technol 99:144–163

15. Yuan H, Li G (2021) A survey of traffic prediction: from spatio-temporal data to intelligent transportation. Data Sci Eng 6:63–85

16. Mozaffari S, Al-Jarrah OY, Dianati M et al (2020) Deep learning-based vehicle behavior prediction for autonomous driving applications: A review. IEEE Trans Intell Transp Syst 23:33–47

17. Sorin V, Barash Y, Konen E, Klang E (2020) Deep learning for natural language processing in radiology—fundamentals and a systematic review. J Am Coll Radiol 17:639–648

18. Otter DW, Medina JR, Kalita JK (2020) A survey of the usages of deep learning for natural language processing. IEEE Trans Neural Networks Learn Syst 32:604–624

19. Wu S, Roberts K, Datta S et al (2020) Deep learning in clinical natural language processing: a methodical review. J Am Med Informatics Assoc 27:457–470

20. Stahlschmidt SR, Ulfenborg B, Synnergren J (2022) Multimodal deep learning for biomedical data fusion: a review. Brief Bioinform 23:bbab569

21. Haque IRI, Neubert J (2020) Deep learning approaches to biomedical image segmentation. Inform Med Unlocked 18:100297

22. Li MM, Huang K, Zitnik M (2022) Graph representation learning in biomedicine and healthcare. Nat Biomed Eng 6(12):1353–69

23. Sun W, Bocchini P, Davison BD (2020) Applications of artificial intelligence for disaster management. Nat Hazards 103:2631–2689

24. Daud SMSM, Yusof MYPM, Heo CC et al (2022) Applications of drone in disaster management: A scoping review. Sci Justice 62:30–42

25. Mohan A, Singh AK, Kumar B, Dwivedi R (2021) Review on remote sensing methods for landslide detection using machine and deep learning. Trans Emerg Telecommun Technol 32:e3998

26. Cao Y, Geddes TA, Yang JYH, Yang P (2020) Ensemble deep learning in bioinformatics. Nat Mach Intell 2:500–508

27. Karim MR, Beyan O, Zappa A et al (2021) Deep learning-based clustering approaches for bioinformatics. Brief Bioinform 22:393–415. https://doi.org/10.1093/BIB/BBZ170

28. Li Y, Huang C, Ding L et al (2019) Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. Methods 166:4–21. https://doi.org/10.1016/j.ymeth.2019.04.008

29. Esteva A, Robicquet A, Ramsundar B et al (2019) (2019) A guide to deep learning in healthcare. Nat Med 251(25):24–29. https://doi.org/10.1038/s41591-018-0316-z

30. Shamshirband S, Fathi M, Dehzangi A et al (2021) A review on deep learning approaches in healthcare systems: Taxonomies, challenges, and open issues. J Biomed Inform 113:103627

31. Qayyum A, Qadir J, Bilal M, Al-Fuqaha A (2020) Secure and robust machine learning for healthcare: A survey. IEEE Rev Biomed Eng 14:156–180

32. Gupta R, Srivastava D, Sahu M et al (2021) Artificial intelligence to deep learning: machine intelligence approach for drug discovery. Mol Divers 25:1315–1360

33. Schneider P, Walters WP, Plowright AT et al (2020) Rethinking drug design in the artificial intelligence era. Nat Rev Drug Discov 19:353–364

34. Lavecchia A (2019) Deep learning in drug discovery: opportunities, challenges and future prospects. Drug Discov Today 24:2017–2032

35. Litjens G, Kooi T, Bejnordi BE et al (2017) A Survey on Deep Learning in Medical Image Analysis. Med Image Anal 42:60–88. https://doi.org/10.1016/j.media.2017.07.005

36. Hesamian MH, Jia W, He X, Kennedy P (2019) Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. J Digit Imaging 32:582–596. https://doi.org/10.1007/s10278-019-00227-x

37. Ozbayoglu AM, Gudelek MU, Sezer OB (2020) Deep learning for financial applications: A survey. Appl Soft Comput 93:106384

38. Pourhabibi T, Ong K-L, Kam BH, Boo YL (2020) Fraud detection: A systematic literature review of graph-based anomaly detection approaches. Decis Support Syst 133:113303

39. Chai J, Zeng H, Li A, Ngai EWT (2021) Deep learning in computer vision: A critical review of emerging techniques and application scenarios. Mach Learn with Appl 6:100134

40. Le N, Rathour VS, Yamazaki K et al (2022) Deep reinforcement learning in computer vision: a comprehensive survey. Artif Intell Rev 55:2733–2819. https://doi.org/10.1007/s10462-021-10061-9

41. Esteva A, Chou K, Yeung S et al (2021) Deep learning-enabled medical computer vision. NPJ Digit Med 4:5

42. Christin S, Hervet É, Lecomte N (2019) Applications for deep learning in ecology. Methods Ecol Evol 10:1632–1644

43. Pichler M, Hartig F (2023) Machine learning and deep learning—A review for ecologists. Methods Ecol Evol 14:994–1016

44. Borowiec ML, Dikow RB, Frandsen PB et al (2022) Deep learning as a tool for ecology and evolution. Methods Ecol Evol 13:1640–1660

45. Kutz JN (2017) Deep learning in fluid dynamics. Journal of Fluid Mechanics. 814:1–4. https://doi.org/10.1017/jfm.2016.803

46. Lye KO, Mishra S, Ray D (2020) Deep learning observables in computational fluid dynamics. J Comput Phys 410:109339

47. Wang Z, Xiao D, Fang F et al (2018) Model identification of reduced order fluid dynamics systems using deep learning. Int J Numer Methods Fluids 86:255–268. https://doi.org/10.1002/FLD.4416

48. Ye XW, Jin T, Yun CB (2019) A review on deep learning-based structural health monitoring of civil infrastructures. Smart Struct Syst 24:567–585

49. Cha Y, Choi W, Büyüköztürk O (2017) Deep learning-based crack damage detection using convolutional neural networks. Comput Civ Infrastruct Eng 32:361–378

50. Avci O, Abdeljaber O, Kiranyaz S et al (2021) A review of vibration-based damage detection in civil structures: From traditional methods to Machine Learning and Deep Learning applications. Mech Syst Signal Process 147:107077

51. Yu M, Huang Q, Qin H et al (2019) Deep learning for real-time social media text classification for situation awareness–using Hurricanes Sandy, Harvey, and Irma as case studies. Int J Digit Earth 12:1230–1247. https://doi.org/10.1080/17538947.2019.1574316

52. Wang M, Deng W (2021) Deep face recognition: A survey. Neurocomputing 429:215–244. https://doi.org/10.1016/j.neucom.2020.10.081

53. Masud M, Muhammad G, Alhumyani H et al (2020) Deep learning-based intelligent face recognition in IoT-cloud environment. Comput Commun 152:215–222. https://doi.org/10.1016/j.comcom.2020.01.050

54. Wang J, Chen Y, Hao S et al (2019) Deep learning for sensor-based activity recognition: A survey. Pattern Recognit Lett 119:3–11. https://doi.org/10.1016/j.patrec.2018.02.010

55. Pfeiffer M, Pfeil T (2018) Deep Learning With Spiking Neurons: Opportunities and Challenges. Front Neurosci 12:. https://doi.org/10.3389/fnins.2018.00774

56. Xiao C, Choi E, Sun J (2018) Review Opportunities and challenges in developing deep learning models using electronic health records data: a systematic review. J Am Med Informatics Assoc 25:1419–1428. https://doi.org/10.1093/jamia/ocy068

57. Cao C, Liu F, Tan H et al (2018) Deep Learning and Its Applications in Biomedicine. Genomics, Proteomics Bioinforma 16:17–32. https://doi.org/10.1016/j.gpb.2017.07.003

58. Yuan Q, Shen H, Li T et al (2020) Deep learning in environmental remote sensing: Achievements and challenges. Remote Sens Environ 241:111716. https://doi.org/10.1016/j.rse.2020.111716

59. Shadmani A, Nikoo MR, Gandomi AH et al (2023) A review of machine learning and deep learning applications in wave energy forecasting and WEC optimization. Energy Strat Rev 49:101180. https://doi.org/10.1016/j.esr.2023.101180

60. Kipli K, Osman S, Joseph A et al (2023) Deep learning applications for oil palm tree detection and counting. Smart Agric Technol 5:100241. https://doi.org/10.1016/j.atech.2023.100241

61. Huang C, Wang J, Wang S, Zhang Y (2023) A review of deep learning in dentistry. Neurocomputing 554:126629. https://doi.org/10.1016/j.neucom.2023.126629

62. Erfanian N, Heydari AA, Feriz AM et al (2023) Deep learning applications in single-cell genomics and transcriptomics data analysis. Biomed Pharmacother 165:115077. https://doi.org/10.1016/j.biopha.2023.115077

63. Adam R, Dell'Aquila K, Hodges L et al (2023) Deep learning applications to breast cancer detection by magnetic resonance imaging: a literature review. Breast Cancer Res 25:87

64. Lata K, Cenkeramaddi LR (2023) Deep Learning for Medical Image Cryptography: A Comprehensive Review. Appl Sci 13:8295

65. Hederman AP, Ackerman ME (2023) Leveraging deep learning to improve vaccine design. Trends Immunol 44(5):333–44

66. Yang S, Kim SH, Kang M, Joo JY (2023) Harnessing deep learning into hidden mutations of neurological disorders for therapeutic challenges. Arch Pharm Res. 46(6):535–49

67. Ruszczak B, Michalski P, Tomaszewski M (2023) Overview of Image Datasets for Deep Learning Applications in Diagnostics of Power Infrastructure. Sensors 23:7171

68. Wani S, Ahuja S, Kumar A (2023) A review on Brain Tumor Detection using Deep Neural Networks. In: 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT). IEEE, pp 363–370. https://doi.org/10.1109/CSNT57126.2023.10134594

69. Alqahtani ND, Alzahrani B, Ramzan MS (2023) Deep Learning Applications for Dyslexia Prediction. Appl Sci 13:2804

70. Cengil E, Çınar A, Özbay E (2017) Image classification with caffe deep learning framework. In: 2017 International Conference on Computer Science and Engineering (UBMK). IEEE, pp 440–444. https://doi.org/10.1109/UBMK.2017.8093433

71. Roy P, Song SL, Krishnamoorthy S et al (2018) Numa-caffe: Numa-aware deep learning neural networks. ACM Trans Archit Code Optim 15:1–26

72. Guo F, Huang H, Liu Y, Xu J (2019) Application of neural network based on Caffe framework for object detection in Hilens. In: 2019 Chinese Automation Congress (CAC). IEEE, pp 4355–4359

73. Hodnett M, Wiley JF (2018) R Deep Learning essentials: A step-by-step guide to building deep learning models using TensorFlow, Keras, and MXNet. Packt Publishing Ltd, Birmingham, UK

74. Jain A, Awan AA, Subramoni H, Panda DK (2019) Scaling TensorFlow, PyTorch, and MXNet using MVAPICH2 for high-performance deep learning on Frontera. Proc DLS 2019 Deep Learn Supercomput - Held conjunction with SC 2019 Int Conf High Perform Comput Networking, Storage Anal 76–83. https://doi.org/10.1109/DLS49591.2019.00015

75. Kim S, Wimmer H, Kim J (2022) Analysis of Deep Learning Libraries: Keras, PyTorch, and MXnet. In: 2022 IEEE/ACIS 20th International Conference on Software Engineering Research, Management and Applications (SERA). IEEE, pp 54–62. https://doi.org/10.1109/SERA54885.2022.9806734

76. Yang H, Fritzsche M, Bartz C, Meinel C (2017) Bmxnet: An open-source binary neural network implementation based on mxnet. In: Proceedings of the 25th ACM international conference on Multimedia. pp 1209–1212. https://doi.org/10.1145/3123266.31293

77. Etaati L, Etaati L (2019) Deep learning tools with cognitive toolkit (CNTK). Mach Learn with Microsoft Technol Sel Right Archit Tools Your Proj (pp 287–302), Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-3658-1_17

78. Banerjee DS, Hamidouche K, Panda DK (2016) Re-designing CNTK deep learning framework on modern GPU enabled clusters. In: 2016 IEEE international conference on cloud computing technology and science (CloudCom). IEEE, pp 144–151

79. Morvan M, Tsiaras A, Nikolaou N, Waldmann IP (2021) PyLightcurve-torch: a transit modeling package for deep learning applications in PyTorch. Publ Astron Soc Pacific 133:34505

80. Ayyadevara VK, Reddy Y (2020) Modern Computer Vision with PyTorch: Explore deep learning concepts and implement over 50 real-world image applications. Packt Publishing Ltd, Birmingham, UK

81. Bourez C (2017) Deep learning with Theano. Packt Publishing Ltd, Birmingham, UK

82. Boufenar C, Batouche M (2017) Investigation on deep learning for off-line handwritten Arabic Character Recognition using Theano research platform. In: 2017 Intelligent Systems and Computer Vision (ISCV). IEEE, pp 1–6. https://doi.org/10.1109/ISACV.2017.8054902

83. Tokui S, Okuta R, Akiba T, et al (2019) Chainer: a deep learning framework for accelerating the research cycle. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp 2002–2011. https://doi.org/10.1145/3292500.3330756

84. Dogaru R, Dogaru I (2020) Fast training of light binary convolutional neural networks using chainer and cupy. In: 2020 12th International Conference on Electronics, Computers and Artificial Intelligence (ECAI). IEEE, pp 1–4. https://doi.org/10.1109/ECAI50035.2020.9223156

85. Bi R, Xu T, Xu M, Chen E (2022) PaddlePaddle: A Production-Oriented Deep Learning Platform Facilitating the Competency of Enterprises. In: 2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys). IEEE, pp 92–99

86. He S, Xue H, Guo L, et al (2021) Intelligent Vehicle Design based on PaddlePaddle and Deep Learning. In: Journal of Physics: Conference Series. IOP Publishing, p 12003. https://doi.org/10.1088/1742-6596/2132/1/012003

87. Hu S, He C, Zhang C, et al (2021) Efficient Scene Text Recognition Model Built with PaddlePaddle Framework. In: 2021 7th International Conference on Big Data and Information Analytics (BigDIA). IEEE, pp 139–142

88. Shorten C, Khoshgoftaar TM (2022) Language models for deep learning programming: a case study with Keras. In: Deep learning applications, vol 4. Springer, pp 135–161. https://doi.org/10.1007/978-981-19-6153-3_6

89. Kapoor A, Gulli A, Pal S, Chollet F (2022) Deep learning with tensorflow and Keras: build and deploy supervised, unsupervised, deep, and reinforcement learning models. Packt Publishing Ltd, Birmingham, UK

90. Elshawi R, Wahab A, Barnawi A, Sakr S (2021) DLBench: a comprehensive experimental evaluation of deep learning frameworks. Cluster Comput 24:2017–2038. https://doi.org/10.1007/s10586-021-03240-4

91. Nguyen G, Dlugolinsky S, Bobák M et al (2019) Machine learning and deep learning frameworks and libraries for large-scale data mining: a survey. Artif Intell Rev 52:77–124

92. Jia Y, Shelhamer E, Donahue J et al (2014) Caffe: convolutional architecture for fast feature embedding. Proc 22nd ACM international conference on Multimedia, pp 675–678. https://doi.org/10.1145/2647868.2654889

93. Deng Y (2019) Deep learning on mobile devices: a review. Mobile Multimed /Image Proc Sec Appl 2019(10993):52–66. https://doi.org/10.1117/12.2518469

94. Jia Y, Shelhamer E, Donahue J, et al (2014) Caffe: Convolutional architecture for fast feature embedding. MM 2014 - Proc 2014 ACM Conf Multimed 675–678. https://doi.org/10.1145/2647868.2654889

95. Shi S, Wang Q, Xu P, Chu X (2017) Benchmarking state-of-the-art deep learning software tools. Proc - 2016 7th Int Conf Cloud Comput Big Data. CCBD 2016:99–104. https://doi.org/10.1109/CCBD.2016.029

96. Xu R, Han F, Ta Q (2019) Deep learning at scale on NVIDIA V100 accelerators. Proc PMBS 2018 Perform Model Benchmarking Simul High Perform Comput Syst Held conjunction with SC 2018 Int Conf High Perform Comput Networking, Storage Anal 23–32. https://doi.org/10.1109/PMBS.2018.8641600

97. Dakkak A, Li C, De Gonzalo SG, et al (2019) TrIMS: Transparent and isolated model sharing for low latency deep learning inference in function-as-a-service. IEEE Int Conf Cloud Comput CLOUD 2019-July:372–382. https://doi.org/10.1109/CLOUD.2019.00067

98. Seide F, Agarwal A (2016) CNTK: Microsoft's open-source deep-learning toolkit. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. p 2135

99. Yu D, Yao K, Zhang Y (2015) The Computational Network Toolkit [Best of the Web]. IEEE Signal Process Mag 32:123–126. https://doi.org/10.1109/MSP.2015.2462371

100. Dai W, Berleant D (2019) Benchmarking contemporary deep learning hardware and frameworks: A survey of qualitative metrics. Proc - 2019 IEEE 1st Int Conf Cogn Mach Intell CogMI 2019 148–155. https://doi.org/10.1109/COGMI48466.2019.00029

101. Liu L, Wu Y, Wei W, et al (2018) Benchmarking deep learning frameworks: Design considerations, metrics and beyond. Proc - Int Conf Distrib Comput Syst 2018-July:1258–1269. https://doi.org/10.1109/ICDCS.2018.00125

102. Wu Y, Cao W, Sahin S, Liu L (2019) Experimental Characterizations and Analysis of Deep Learning Frameworks. Proc - 2018 IEEE Int Conf Big Data. Big Data 2018:372–377. https://doi.org/10.1109/BIGDATA.2018.8621930

103. Paszke A, Gross S, Massa F et al (2019) Pytorch: an imperative style, high-performance deep learning library. Adv Neural Inf Process Syst, p 32

104. Shatnawi A, Al-Bdour G, Al-Qurran R, Al-Ayyoub M (2018) A comparative study of open source deep learning frameworks. 2018 9th Int Conf Inf Commun Syst ICICS 2018 2018-Janua:72–77. https://doi.org/10.1109/IACS.2018.8355444

105. Goodfellow IJ, Warde-Farley D, Lamblin P et al (2013) Pylearn2: a machine learning research library. https://doi.org/10.48550/arXiv.1308.4214

106. Miao X, Zhang W, Shao Y, et al (2021) Lasagne: A Multi-Layer Graph Convolutional Network Framework via Node-aware Deep Architecture. IEEE Trans Knowl Data Eng 1–1. https://doi.org/10.1109/TKDE.2021.3103984

107. Moolayil J (2019) An Introduction to Deep Learning and Keras. Learn Keras Deep Neural Networks 1–16. https://doi.org/10.1007/978-1-4842-4240-7_1

108. Bergstra J, Breuleux O, Bastien F, et al (2010) Theano: A CPU and GPU Math Compiler in Python. In: SciPy. pp 18–24

109. Han J, Shihab E, Wan Z et al (2020) What do Programmers Discuss about Deep Learning Frameworks. Empir Softw Eng 25:2694–2747. https://doi.org/10.1007/S10664-020-09819-6/FIGURES/22

110. Abadi M, Barham P, Chen J, et al (2016) {TensorFlow}: a system for {Large-Scale} machine learning. In: 12th USENIX symposium on operating systems design and implementation (OSDI 16). pp 265–283

111. Pang B, Nijkamp E, Wu YN (2020) Deep Learning With TensorFlow: A Review. J Educat Behav Stat 45(2):227–48. https://doi.org/10.3102/1076998619872761

112. Goh HA, Ho CK, Abas FS (2023) Front-end deep learning web apps development and deployment: a review. Appl Intell 53:15923–15945. https://doi.org/10.1007/s10489-022-04278-6

113. Ma Y, Yu D, Wu T, Wang H (2019) PaddlePaddle: An open-source deep learning platform from industrial practice. Front Data Domputing 1:105–115

114. Gulli A, Pal S (2017) Deep learning with Keras. Packt Publishing Ltd, Birmingham, UK

115. Manaswi NK, Manaswi NK (2018) Understanding and working with Keras. Deep learn with appl using Python Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-3516-4_2

116. Moolayil J, Moolayil J, John S (2019) Learn Keras for deep neural networks. Springer

117. Young T, Hazarika D, Poria S, Cambria E (2018) Recent trends in deep learning based natural language processing [Review Article]. IEEE Comput Intell Mag 13:55–75. https://doi.org/10.1109/MCI.2018.2840738

118. Kapoor A., Gulli A., Pal S, Chollet F (2022) Deep learning with tensorflow and Keras: build and deploy supervised, unsupervised, deep, and reinforcement learning models. Packt Publishing Ltd. Birmingham, UK

119. Xu J, Zhao J, Zhou R et al (2021) Predicting Destinations by a Deep Learning based Approach. IEEE Trans Knowl Data Eng 33:651–666. https://doi.org/10.1109/TKDE.2019.2932984

120. Wang W, Bai Y, Yu C et al (2018) A network traffic flow prediction with deep learning approach for large-scale metropolitan area network. IEEE/IFIP Netw Oper Manag Symp Cogn Manag a Cyber World, NOMS 2018:1–9. https://doi.org/10.1109/NOMS.2018.8406252

121. Wu Y, Tan H, Qin L et al (2018) A hybrid deep learning based traffic flow prediction method and its understanding. Transp Res Part C Emerg Technol 90:166–180. https://doi.org/10.1016/J.TRC.2018.03.001

122. Telikani A, Gandomi AH, Choo K-KR, Shen J (2021) A cost-sensitive deep learning based approach for network traffic classification. IEEE Trans Netw Serv Manag 19(1):661

123. Tran L, Mun MY, Lim M et al (2020) DeepTRANS. Proc VLDB Endow 13:2957–2960. https://doi.org/10.14778/3415478.3415518

124. Xu S, Zhang R, Cheng W, Xu J (2020) MTLM: a multi-task learning model for travel time estimation. Geoinformatica 1–17. https://doi.org/10.1007/S10707-020-00422-X/TABLES/2

125. Ren H, Song Y, Wang J, et al (2018) A Deep Learning Approach to the Citywide Traffic Accident Risk Prediction. IEEE Conf

Intell Transp Syst Proceedings, ITSC 2018-Novem:3346–3351. https://doi.org/10.1109/ITSC.2018.8569437

126. Yang Z, Zhang W, Feng J (2022) Predicting multiple types of traffic accident severity with explanations: A multi-task deep learning framework. Saf Sci 146:105522. https://doi.org/10.1016/J.SSCI.2021.105522

127. Yu JJQ (2021) Travel Mode Identification with GPS Trajectories Using Wavelet Transform and Deep Learning. IEEE Trans Intell Transp Syst 22:1093–1103. https://doi.org/10.1109/TITS.2019.2962741

128. Yi D, Su J, Liu C, Chen WH (2019) Trajectory clustering aided personalized driver intention prediction for intelligent vehicles. IEEE Trans Ind Informatics 15:3693–3702. https://doi.org/10.1109/TII.2018.2890141

129. Wang C, Li J, He Y et al (2020) Segmented Trajectory Clustering-Based Destination Prediction in IoVs. IEEE Access 8:98999–99009. https://doi.org/10.1109/ACCESS.2020.2998063

130. Manogaran G, Alazab M (2021) Ant-Inspired Recurrent Deep Learning Model for Improving the Service Flow of Intelligent Transportation Systems. IEEE Trans Intell Transp Syst 22:3654–3663. https://doi.org/10.1109/TITS.2020.3037902

131. Ibrahim HA, Azar AT, Ibrahim ZF, Ammar HH (2020) A Hybrid Deep Learning Based Autonomous Vehicle Navigation and Obstacles Avoidance. Adv Intell Syst Comput 1153 AISC:296–307. https://doi.org/10.1007/978-3-030-44289-7_28

132. Bhattacharjee A, Chhokra AD, Kang Z, et al (2019) BARISTA: Efficient and scalable serverless serving system for deep learning prediction services. Proc - 2019 IEEE Int Conf Cloud Eng IC2E 2019 23–33. https://doi.org/10.1109/IC2E.2019.00-10

133. Ke J, Zheng H, Yang H, Chen XM (2017) Short-term forecasting of passenger demand under on-demand ride services: A spatio-temporal deep learning approach. Transp Res Part C Emerg Technol 85:591–608. https://doi.org/10.1016/J.TRC.2017.10.016

134. Chu T, Wang J, Codeca L, Li Z (2020) Multi-agent deep reinforcement learning for large-scale traffic signal control. IEEE Trans Intell Transp Syst 21:1086–1095. https://doi.org/10.1109/TITS.2019.2901791

135. Chen C, Wei H, Xu N, et al (2020) Toward a thousand lights: Decentralized deep reinforcement learning for large-scale traffic signal control. AAAI 2020 - 34th AAAI Conf Artif Intell 3414–3421. https://doi.org/10.1609/AAAI.V34I04.5744

136. Jumper J, Evans R, Pritzel A et al (2021) (2021) Highly accurate protein structure prediction with AlphaFold. Nat 5967873(596):583–589. https://doi.org/10.1038/s41586-021-03819-2

137. Torrisi M, Pollastri G, Le Q (2020) Deep learning methods in protein structure prediction. Comput Struct Biotechnol J 18:1301–1310. https://doi.org/10.1016/J.CSBJ.2019.12.011

138. Zrimec J, Börlin CS, Buric F et al (2020) (2020) Deep learning suggests that gene expression is encoded in all parts of a co-evolving interacting gene regulatory structure. Nat Commun 111(11):1–16. https://doi.org/10.1038/s41467-020-19921-4

139. Singh R, Lanchantin J, Robins G, Qi Y (2016) DeepChrome: deep-learning for predicting gene expression from histone modifications. Bioinformatics 32:i639–i648. https://doi.org/10.1093/BIOINFORMATICS/BTW427

140. Grapov D, Fahrmann J, Wanichthanarak K, Khoomrung S (2018) Rise of deep learning for genomic, proteomic, and metabolomic data integration in precision medicine. Omi A J Integr Biol 22:630–636. https://doi.org/10.1089/OMI.2018.0097/ASSET/IMAGES/LARGE/FIGURE4.JPEG

141. Parekh VS, Jacobs MA (2019) Deep learning and radiomics in precision medicine. 4:59–72. https://doi.org/10.1080/23808993.2019.1585805

142. Duncan JS, Insana MF, Ayache N (2020) Biomedical Imaging and Analysis in the Age of Big Data and Deep Learning. Proc IEEE 108:3–10. https://doi.org/10.1109/JPROC.2019.2956422

143. Dai M, Xiao G, Fiondella L et al (2021) Deep learning-enabled resolution-enhancement in mini- and regular microscopy for biomedical imaging. Sensors Actuators A Phys 331:112928. https://doi.org/10.1016/J.SNA.2021.112928

144. Rizk H, Torki M, Youssef M (2019) CellinDeep: Robust and Accurate Cellular-Based Indoor Localization via Deep Learning. IEEE Sens J 19:2305–2312. https://doi.org/10.1109/JSEN.2018.2885958

145. Rizk H, Youssef M (2019) MonoDCell: A ubiquitous and low-overhead deep learning-based indoor localization with limited cellular information. GIS Proc ACM Int Symp Adv Geogr Inf Syst 109–118. https://doi.org/10.1145/3347146.3359065

146. Jia D, Chen C, Chen C, et al (2021) Breast Cancer Case Identification Based on Deep Learning and Bioinformatics Analysis. Front Genet 12:. https://doi.org/10.3389/FGENE.2021.628136

147. Guo L, Wang S, Li M, Cao Z (2019) Accurate classification of membrane protein types based on sequence and evolutionary information using deep learning. BMC Bioinformatics 20:1–17. https://doi.org/10.1186/S12859-019-3275-6/TABLES/12

148. Du N, Shang J, Sun Y (2021) Improving protein domain classification for third-generation sequencing reads using deep learning. BMC Genomics 22:1–13. https://doi.org/10.1186/S12864-021-07468-7/FIGURES/9

149. Greenspan H, Van Ginneken B, Summers RM (2016) Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. IEEE Trans Med Imaging 35:1153–1159. https://doi.org/10.1109/TMI.2016.2553401

150. Tsagkatakis G, Jaber M, Tsakalides P et al (2017) Goal!! Event detection in sports video. Electron Imaging 29:15–20. https://doi.org/10.2352/ISSN.2470-1173.2017.16.CVAS-344

151. Donahue J, Hendricks LA, Rohrbach M, et al (2017) Long-term recurrent convolutional networks for visual recognition and description. In: IEEE Trans Pattern Anal Mach Intell, pp 677–691. https://doi.org/10.1109/TPAMI.2016.2599174

152. Yuliani AR, Amri MF, Suryawati E et al (2021) Speech Enhancement Using Deep Learning Methods: A Review. J Elektron dan Telekomun 21:19. https://doi.org/10.14203/JET.V21.19-26

153. Byrd J, Lipton ZC (2019) What is the effect of importance weighting in deep learning? In: 36th Int Conf Mach Learn ICML, pp 872–881

154. Natsiou A, O'Leary S (2022) Audio representations for deep learning in sound synthesis: A review. 1–8. https://doi.org/10.1109/aiccsa53542.2021.9686838

155. Widodo W, Nugraheni M, Sari IP et al (2019) An idea based on sequential pattern mining and deep learning for text summarization. J Phys Conf Ser 1402:077013. https://doi.org/10.1088/1742-6596/1402/7/077013

156. Kang TG, Shin JW, Kim NS (2018) DNN-based monaural speech enhancement with temporal and spectral variations equalization. Digit Signal Process 74:102–110. https://doi.org/10.1016/J.DSP.2017.12.002

157. Pandey A, Wang D (2020) On Cross-Corpus Generalization of Deep Learning Based Speech Enhancement. IEEE/ACM Trans Audio Speech Lang Process 28:2489–2499. https://doi.org/10.1109/TASLP.2020.3016497

158. Xu Y, Du J, Dai LR, Lee CH (2015) A regression approach to speech enhancement based on deep neural networks. IEEE/ACM Trans Audio Speech Lang Process 23:7–19. https://doi.org/10.1109/TASLP.2014.2364452

159. Valin JM (2018) A hybrid DSP/Deep learning approach to real-time full-band speech enhancement. 2018 IEEE 20th Int Work Multimed Signal Process MMSP 2018. https://doi.org/10.1109/MMSP.2018.8547084

160. Das N, Chakraborty S, Chaki J et al (2021) Fundamentals, present and future perspectives of speech enhancement. Int J

Speech Technol 24:883–901. https://doi.org/10.1007/S10772-020-09674-2/TABLES/6

161. Tan K, Wang D (2021) Towards Model Compression for Deep Learning Based Speech Enhancement. IEEE/ACM Trans Audio Speech Lang Process 29:1785–1794. https://doi.org/10.1109/TASLP.2021.3082282

162. Mukhutdinov D, Alex A, Cavallaro A, Wang L (2023) Deep learning models for single-channel speech enhancement on drones. IEEE Access 11:22993–23007

163. Lech M, Stolar M, Best C, Bolia R (2020) Real-Time Speech Emotion Recognition Using a Pre-trained Image Classification Network: Effects of Bandwidth Reduction and Companding. Front Comput Sci 2:14. https://doi.org/10.3389/FCOMP.2020.00014/BIBTEX

164. Aouani H, Ben AY (2020) Speech Emotion Recognition with deep learning. Procedia Comput Sci 176:251–260. https://doi.org/10.1016/J.PROCS.2020.08.027

165. Noroozi F, Marjanovic M, Njegus A et al (2019) Audio-Visual Emotion Recognition in Video Clips. IEEE Trans Affect Comput 10:60–75. https://doi.org/10.1109/TAFFC.2017.2713783

166. Khalil RA, Jones E, Babar MI et al (2019) Speech Emotion Recognition Using Deep Learning Techniques: A Review. IEEE Access 7:117327–117345. https://doi.org/10.1109/ACCESS.2019.2936124

167. Bandela SR, Kumar TK (2017) Stressed speech emotion recognition using feature fusion of teager energy operator and MFCC. 8th Int Conf Comput Commun Netw Technol ICCCNT 2017. https://doi.org/10.1109/ICCCNT.2017.8204149

168. Zamil AAA, Hasan S, Jannatul Baki SM, et al (2019) Emotion detection from speech signals using voting mechanism on classified frames. 1st Int Conf Robot Electr Signal Process Tech ICREST 2019 281–285. https://doi.org/10.1109/ICREST.2019.8644168

169. Jahangir R, Teh YW, Hanif F, Mujtaba G (2021) Deep learning approaches for speech emotion recognition: state of the art and research challenges. Multimed Tools Appl 80:23745–23812. https://doi.org/10.1007/S11042-020-09874-7/FIGURES/17

170. Hema C, Garcia Marquez FP (2023) Emotional speech Recognition using CNN and Deep learning techniques. Appl Acoust 211:109492. https://doi.org/10.1016/j.apacoust.2023.109492

171. Turchin A, Masharsky S, Zitnik M (2023) Comparison of BERT implementations for natural language processing of narrative medical documents. Informatics Med Unlocked 36:101139. https://doi.org/10.1016/j.imu.2022.101139

172. Altheneyan A, Menai MEB (2020) Evaluation of State-of-the-Art Paraphrase Identification and Its Application to Automatic Plagiarism Detection. Int J Pattern Recognit Artif Intell 34:. https://doi.org/10.1142/S0218001420530043

173. Alzubi JA, Jain R, Kathuria A, et al (2020) Paraphrase identification using collaborative adversarial networks. J Intell Fuzzy Syst 39:. https://doi.org/10.3233/JIFS-191933

174. Hunt E, Janamsetty R, Kinares C et al (2019) Machine learning models for paraphrase identification and its applications on plagiarism detection. In: Proceedings - 10th IEEE Int Conf Big Knowl, ICBK. https://doi.org/10.1109/ICBK.2019.00021

175. Johnson M, Schuster M, Le Q V., et al (2017) Google's Multilingual Neural Machine Translation System: Enabling Zero-Shot Translation. Trans Assoc Comput Linguist 5:. https://doi.org/10.1162/tacl_a_00065

176. Zhang B, Xiong D, Su J (2020) Neural Machine Translation with Deep Attention. IEEE Trans Pattern Anal Mach Intell 42:. https://doi.org/10.1109/TPAMI.2018.2876404

177. Zhang B, Xiong D, Su J, et al (2016) Variational neural machine translation. arXiv preprint arXiv:1605.07869. https://doi.org/10.48550/arXiv.1605.07869

178. Zhou J, Cao Y, Wang X, et al (2016) Deep Recurrent Models with Fast-Forward Connections for Neural Machine Translation. Trans Assoc Comput Linguist 4:. https://doi.org/10.1162/tacl_a_00105

179. Yadav A, Vishwakarma DK (2020) Sentiment analysis using deep learning architectures: a review. Artif Intell Rev 53:. https://doi.org/10.1007/s10462-019-09794-5

180. Al-Smadi M, Talafha B, Al-Ayyoub M, Jararweh Y (2019) Using long short-term memory deep neural networks for aspect-based sentiment analysis of Arabic reviews. Int J Mach Learn Cybern 10:. https://doi.org/10.1007/s13042-018-0799-4

181. Ma Y, Peng H, Khan T, et al (2018) Sentic LSTM: a Hybrid Network for Targeted Aspect-Based Sentiment Analysis. Cognit Comput 10:. https://doi.org/10.1007/s12559-018-9549-x

182. Ma F, Hu X, Liu A, et al (2023) AMR-based Network for Aspect-based Sentiment Analysis. In: Proceedings of the Annual Meeting of the Association for Computational Linguistics

183. Bello A, Ng S-C, Leung M-F (2023) A BERT framework to sentiment analysis of tweets. Sensors 23:506

184. Li X, Zhang L, Wu Z et al (2024) Artificial general intelligence for medical imaging, IEEE Rev Biomed Eng 18:113–129. https://doi.org/10.1109/RBME.2024.3493775

185. Wang Y, Huang M, Zhao L, Zhu X (2016) Attention-based LSTM for aspect-level sentiment classification. In: EMNLP 2016 - Conference on Empirical Methods in Natural Language Processing, Proceedings

186. Mozannar H, El Hajal K, Maamary E, Hajj H (2019) Neural arabic question answering. In: ACL 2019 - 4th Arabic Natural Language Processing Workshop, WANLP 2019 - Proceedings of the Workshop. https://doi.org/10.18653/v1/W19-4612

187. Bai Z, Li Y, Woźniak M, et al (2021) DecomVQANet: Decomposing visual question answering deep network via tensor decomposition and regression. Pattern Recognit 110:. https://doi.org/10.1016/j.patcog.2020.107538

188. Yu D, Gao X, Xiong H (2018) Structured semantic representation for visual question answering. In2018 25th IEEE International Conference on Image Processing (ICIP) (pp 2286–2290). IEEE. https://doi.org/10.1109/ICIP.2018.8451516

189. Brown T, Mann B, Ryder N et al (2020) Language models are few-shot learners. Adv Neural Inf Process Syst 33:1877–1901

190. Ray PP (2023) ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope. Internet Things Cyber-Physical Syst 3:121–154

191. El-Kassas WS, Salama CR, Rafea AA, Mohamed HK (2021) Automatic text summarization: A comprehensive survey. Expert Syst. Appl. 165:113679

192. Chen L, Nguyen M Le (2019) Sentence selective neural extractive summarization with reinforcement learning. In: Proceedings of 2019 11th Int Conf Knowl Sys Eng, KSE. https://doi.org/10.1109/KSE.2019.8919490

193. Alami N, Meknassi M, En-nahnahi N et al (2021) Unsupervised neural networks for automatic Arabic text summarization using document clustering and topic modeling. Expert Syst Appl 172:114652. https://doi.org/10.1016/j.eswa.2021.114652

194. Li Y, Jiang D, Lian R et al (2023) Heterogeneous Latent Topic Discovery for Semantic Text Mining. IEEE Trans Knowl Data Eng 35:533. https://doi.org/10.1109/TKDE.2021.3077025

195. Zhao X, An Y, Xu N, Geng X (2024) Variational Continuous Label Distribution Learning for Multi-Label Text Classification. IEEE Trans Knowl Data Eng 36:. https://doi.org/10.1109/TKDE.2023.3323401

196. Sarker IH (2022) Ai-based modeling: Techniques, applications and research issues towards automation, intelligent and smart systems. SN Comput Sci 3:158

197. O'Mahony N, Campbell S, Carvalho A et al (2020) Deep learning vs. traditional computer vision. In: Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), vol 1 1. Springer, pp 128–144. https://doi.org/10.1007/978-3-030-17795-9_10

198. Wu X, Sahoo D, Hoi SCH (2020) Recent advances in deep learning for object detection. Neurocomputing 396:39–64. https://doi.org/10.1016/J.NEUCOM.2020.01.085

199. Nithya R, Santhi B, Manikandan R et al (2022) Computer Vision System for Mango Fruit Defect Detection Using Deep Convolutional Neural Network. Foods 11:3483

200. Behera A, Wharton Z, Keidel A, Debnath B (2020) Deep cnn, body pose, and body-object interaction features for drivers' activity monitoring. IEEE Trans Intell Transp Syst 23:2874–2881

201. Sumit SS, Watada J, Roy A, Rambli DRA (2020) In object detection deep learning methods, YOLO shows supremum to Mask R-CNN. In: Journal of Physics: Conference Series. IOP Publishing, p 42086

202. Manifold B, Men S, Hu R, Fu D (2021) A versatile deep learning architecture for classification and label-free prediction of hyperspectral images. Nat Mach Intell 3:306–315

203. Latif G, Abdelhamid SE, Mallouhy RE et al (2022) Deep Learning Utilization in Agriculture: Detection of Rice Plant Diseases Using an Improved CNN Model. Plants 11:2230

204. Pathak AR, Pandey M, Rautaray S (2018) Application of Deep Learning for Object Detection. Procedia Comput Sci 132:1706–1717. https://doi.org/10.1016/J.PROCS.2018.05.144

205. Ouyang W, Zeng X, Wang X et al (2017) DeepID-Net: Object Detection with Deformable Part Based Convolutional Neural Networks. IEEE Trans Pattern Anal Mach Intell 39:1320–1334. https://doi.org/10.1109/TPAMI.2016.2587642

206. Doulamis N (2014) Doulamis A (2014) Semi-supervised deep learning for object tracking and classification. IEEE Int Conf Image Process ICIP 2014:848–852. https://doi.org/10.1109/ICIP.2014.7025170

207. Taskiran M, Kahraman N, Erdem CE (2020) Face recognition: Past, present and future (a review). Digit Signal Process 106:102809

208. Guo Y, Zhang L, Hu Y, et al (2016) MS-Celeb-1M: A Dataset and Benchmark for Large-Scale Face Recognition. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 9907 LNCS:87–102. https://doi.org/10.1007/978-3-319-46487-9_6

209. Zadeh A, Lim YC, Baltrušaitis T, Morency LP (2017) Convolutional experts constrained local model for 3D facial landmark detection. Proc - 2017 IEEE Int Conf Comput Vis Work ICCVW 2017 2018-Janua:2519–2528. https://doi.org/10.1109/ICCVW.2017.296

210. Cirik V, Hovy E, Morency L-P (2016) Visualizing and understanding curriculum learning for long short-term memory networks. arXiv preprint arXiv:1611.06204. https://doi.org/10.48550/arXiv.1611.06204

211. Deng J, Cheng S, Xue N, et al (2018) UV-GAN: Adversarial Facial UV Map Completion for Pose-Invariant Face Recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 7093–7102. https://doi.org/10.1109/CVPR.2018.00741

212. Liu Y, Wei F, Shao J, et al (2018) Exploring Disentangled Feature Representation beyond Face Identification. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 2080–2089. https://doi.org/10.1109/CVPR.2018.00222

213. Nech A, Kemelmacher-Shlizerman I (2017) Level Playing Field for Million Scale Face Recognition. Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017 2017-Janua:3406–3415. https://doi.org/10.1109/CVPR.2017.363

214. Bansal A, Nanduri A, Castillo CD, et al (2016) UMDFaces: An Annotated Face Dataset for Training Deep Networks. IEEE Int Jt Conf Biometrics, IJCB 2017 2018-Janua:464–473. https://doi.org/10.1109/BTAS.2017.8272731

215. Painuly K, Bisht Y, Vaidya H et al (2024) Efficient real-time face recognition-based attendance system with deep learning algorithms. In: 2024 Int Conf Intell Innov Technol Comput, Electrical and Electronics (IITCEE). IEEE, pp 1–5. https://doi.org/10.1109/ICOEI48184.2020.9143046

216. Jain A, Kanhangad V (2018) Human Activity Classification in Smartphones Using Accelerometer and Gyroscope Sensors. IEEE Sens J 18:1169–1177. https://doi.org/10.1109/JSEN.2017.2782492

217. Jalloul N, Poree F, Viardot G et al (2018) Activity Recognition Using Complex Network Analysis. IEEE J Biomed Heal Informatics 22:989–1000. https://doi.org/10.1109/JBHI.2017.2762404

218. Mario MO (2019) Human activity recognition based on single sensor square HV acceleration images and convolutional neural networks. IEEE Sens J 19:1487–1498. https://doi.org/10.1109/JSEN.2018.2882943

219. Lin Y, Le Kernec J, Yang S et al (2018) Human Activity Classification with Radar: Optimization and Noise Robustness with Iterative Convolutional Neural Networks Followed with Random Forests. IEEE Sens J 18:9669–9681. https://doi.org/10.1109/JSEN.2018.2872849

220. Ordóñez FJ, Roggen D, Liu Y et al (2016) Deep Convolutional and LSTM Recurrent Neural Networks for Multimodal Wearable Activity Recognition. Sensors 16:115. https://doi.org/10.3390/S16010115

221. Agarwal P, Alam M (2020) A Lightweight Deep Learning Model for Human Activity Recognition on Edge Devices. Procedia Comput Sci 167:2364–2373. https://doi.org/10.1016/J.PROCS.2020.03.289

222. Xia K, Huang J, Wang H (2020) LSTM-CNN Architecture for Human Activity Recognition. IEEE Access 8:56855–56866. https://doi.org/10.1109/ACCESS.2020.2982225

223. Chen Y, Tian Y, He M (2020) Monocular human pose estimation: A survey of deep learning-based methods. undefined 192:. https://doi.org/10.1016/J.CVIU.2019.102897

224. Toshev A, Szegedy C (2013) DeepPose: Human Pose Estimation via Deep Neural Networks. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit 1653–1660. https://doi.org/10.1109/CVPR.2014.214

225. Pfister T, Simonyan K, Charles J, Zisserman A (2015) Deep convolutional neural networks for efficient pose estimation in gesture videos. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 9003:538–552. https://doi.org/10.1007/978-3-319-16865-4_35

226. Yang W, Ouyang W, Wang X, et al (2018) 3D Human Pose Estimation in the Wild by Adversarial Learning. undefined 5255–5264. https://doi.org/10.1109/CVPR.2018.00551

227. Papandreou G, Zhu T, Kanazawa N, et al (2017) Towards Accurate Multi-person Pose Estimation in the Wild. Proc - 30th IEEE Conf Comput Vis Pattern Recognition, CVPR 2017 2017-Janua:3711–3719. https://doi.org/10.1109/CVPR.2017.395

228. Perez-Sala X, Escalera S, Angulo C, Gonzàlez J (2014) A Survey on Model Based Approaches for 2D and 3D Visual Human Pose Recovery. Sensors (Basel) 14:4189. https://doi.org/10.3390/S140304189

229. Wang P, Li W, Ogunbona P et al (2018) RGB-D-based human motion recognition with deep learning: A survey. Comput Vis

Image Underst 171:118–139. https://doi.org/10.1016/J.CVIU.2018.04.007

230. Gong W, Zhang X, Gonzàlez J et al (2016) Human Pose Estimation from Monocular Images: A Comprehensive Survey. Sensors (Basel) 16:1966. https://doi.org/10.3390/S16121966

231. Sarafianos N, Boteanu B, Ionescu B, Kakadiaris IA (2016) 3D Human pose estimation: A review of the literature and analysis of covariates. Comput Vis Image Underst 152:1–20. https://doi.org/10.1016/J.CVIU.2016.09.002

232. Valletta JJ, Torney C, Kings M et al (2017) Applications of machine learning in animal behaviour studies. Anim Behav 124:203–220. https://doi.org/10.1016/J.ANBEHAV.2016.12.005

233. Mandal R, Connolly RM, Schlacher TA, Stantic B (2018) Assessing fish abundance from underwater video using deep neural networks. Proc Int Jt Conf Neural Networks 2018-July: https://doi.org/10.1109/IJCNN.2018.8489482

234. Satoto BD, Utoyo I, Rulaningtyas R, Khoendori EB (2020) An improvement of Gram-negative bacteria identification using convolutional neural network with fine tuning. TELKOMNIKA (Telecommunication Comput Electron Control 18:1397–1405. https://doi.org/10.12928/TELKOMNIKA.V18I3.14890

235. Carranza-Rojas J, Goeau H, Bonnet P et al (2017) Going deeper in the automated identification of Herbarium specimens. BMC Evol Biol 17:181. https://doi.org/10.1186/S12862-017-1014-Z/FIGURES/7

236. Md. Sanzidul I, Sadia Sultana SM, Abujar S, Hossain SA, (2019) Sequence-to-sequence Bangla Sentence Generation with LSTM Recurrent Neural Networks. Procedia Comput Sci 152:51–58. https://doi.org/10.1016/J.PROCS.2019.05.026

237. Schuettpelz E, Frandsen PB, Dikow RB et al (2017) Applications of deep convolutional neural networks to digitized natural history collections. Biodivers Data J 5:e21139. https://doi.org/10.3897/BDJ.5.E21139

238. Valan M, Makonyi K, Maki A et al (2019) Automated Taxonomic Identification of Insects with Expert-Level Accuracy Using Effective Feature Transfer from Convolutional Networks. Syst Biol 68:876–895. https://doi.org/10.1093/SYSBIO/SYZ014

239. Hansen OLP, Svenning JC, Olsen K et al (2020) Species-level image classification with convolutional neural network enables insect identification from habitus images. Ecol Evol 10:737–747. https://doi.org/10.1002/ECE3.5921

240. Villon S, Mouillot D, Chaumont M et al (2018) A Deep learning method for accurate and fast identification of coral reef fishes in underwater images. Ecol Inform 48:238–244. https://doi.org/10.1016/J.ECOINF.2018.09.007

241. Liu X, Song H (2020) Automatic identification of fossils and abiotic grains during carbonate microfacies analysis using deep convolutional neural networks. Sediment Geol 410:105790. https://doi.org/10.1016/J.SEDGEO.2020.105790

242. Younis S, Weiland C, Hoehndorf R, et al (2018) Taxon and trait recognition from digitized herbarium specimens using deep convolutional neural networks. 165:377–383. https://doi.org/10.1080/23818107.2018.1446357

243. Norouzzadeh MS, Nguyen A, Kosmala M et al (2018) Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc Natl Acad Sci U S A 115:E5716–E5725. https://doi.org/10.1073/PNAS.1719367115/-/DCSUPPLEMENTAL

244. de Lima RP, Welch KF, Barrick JE et al (2020) Convolutional neural networks as an aid to biostratigraphy and micropaleontology: a test on late paleozoic microfossils. Palaios 35:391–402. https://doi.org/10.2110/PALO.2019.102

245. Ditria EM, Lopez-Marcano S, Sievers M et al (2020) Automating the Analysis of Fish Abundance Using Object Detection: Optimizing Animal Ecology With Deep Learning. Front Mar Sci 7:429. https://doi.org/10.3389/FMARS.2020.00429/BIBTEX

246. Kälin U, Lang N, Hug C, et al (2018) Defoliation estimation of forest trees from ground-level images. bioRxiv 441733. https://doi.org/10.1101/441733

247. Sathish T, Maheswari SU, Balaji V et al (2023) Coastal pollution analysis for environmental health and ecological safety using deep learning technique. Adv Eng Softw 179:103441. https://doi.org/10.1016/j.advengsoft.2023.103441

248. Lee SH, Chan CS, Remagnino P (2018) Multi-Organ Plant Classification Based on Convolutional and Recurrent Neural Networks. IEEE Trans Image Process 27:4287–4301. https://doi.org/10.1109/TIP.2018.2836321

249. Marques ACR, Raimundo MM, Cavalheiro EMB et al (2018) Ant genera identification using an ensemble of convolutional neural networks. PLoS ONE 13:e0192011. https://doi.org/10.1371/JOURNAL.PONE.0192011

250. Wäldchen J, Mäder P (2018) Machine learning for image based species identification. Methods Ecol Evol 9:2216–2225. https://doi.org/10.1111/2041-210X.13075

251. Kamilaris A, Prenafeta-Boldú FX (2018) Deep learning in agriculture: A survey. Comput. Electron. Agric. 147:70–90

252. Ramcharan A, Baranowski K, McCloskey P et al (2017) Deep learning for image-based cassava disease detection. Front Plant Sci 8:1850. https://doi.org/10.3389/fpls.2017.01852

253. Ali AH, Youssef A, Abdelal M, Raja MA (2024) An ensemble of deep learning architectures for accurate plant disease classification. Ecol Inform 81:102618. https://doi.org/10.1016/j.ecoinf.2024.102618

254. ArnalBarbedo JG (2019) Plant disease identification from individual lesions and spots using deep learning. Biosyst Eng 180:96–107. https://doi.org/10.1016/j.biosystemseng.2019.02.002

255. Lin K, Gong L, Huang Y et al (2019) Deep learning-based segmentation and quantification of cucumber powdery mildew using convolutional neural network. Front Plant Sci 10:155. https://doi.org/10.3389/fpls.2019.00155

256. Yang X, Zhang S, Liu J et al (2021) Deep learning for smart fish farming: applications, opportunities and challenges. Rev. Aquac. 13:66–90

257. Salman A, Jalal A, Shafait F et al (2016) Fish species classification in unconstrained underwater environments based on deep learning. Limnol Oceanogr Methods 14:570–585. https://doi.org/10.1002/lom3.10113

258. Morgan RE, Statistician B (2021) Financial Fraud in the United States, 2017. https://doi.org/10.17226/23492

259. Ounacer S, Ait H, Bour E et al (2018) Using isolation forest in anomaly detection: The case of credit card transactions. Period Eng Nat Sci 6:394–400. https://doi.org/10.21533/PEN.V6I2.533.G312

260. Mubalaike AM, Adali E (2018) Deep Learning Approach for Intelligent Financial Fraud Detection System. UBMK 2018 - 3rd Int Conf Comput Sci Eng 598–603. https://doi.org/10.1109/UBMK.2018.8566574

261. Makki S, Assaghir Z, Taher Y et al (2019) An Experimental Study With Imbalanced Classification Approaches for Credit Card Fraud Detection. IEEE Access 7:93010–93022. https://doi.org/10.1109/ACCESS.2019.2927266

262. Sanober S, Alam I, Pande S, Arslan F, Rane KP, Singh BK, Khamparia A, Shabaz M (2021) An Enhanced Secure Deep Learning Algorithm for Fraud Detection in Wireless Communication. Wireless Commun Mobile Comput 2021(1):6079582. https://doi.org/10.1155/2021/6079582

263. Alghofaili Y, Albattah A, Rassam MA (2020) A Financial Fraud Detection Model Based on LSTM Deep Learning Technique. J Appl Sec Res 15(4):498–516. https://doi.org/10.1080/19361610.2020.1815491

264. Sculley D, Otey ME, Pohl M, et al (2011) . Proc ACM SIG-KDD Int Conf Knowl Discov Data Min 274–282. https://doi.org/10.1145/2020408.2020455

265. Li SH, Yen DC, Lu WH, Wang C (2012) Identifying the signs of fraudulent accounts using data mining techniques. Comput Human Behav 28:1002–1013. https://doi.org/10.1016/J.CHB.2012.01.002

266. Benchaji I, Douzi S, El Ouahidi B (2021) Credit card fraud detection model based on LSTM recurrent neural networks. J Adv Inf Technol 12:113–118. https://doi.org/10.12720/JAIT.12.2.113-118

267. Roy A, Sun J, Mahoney R et al (2018) (2018) Deep learning detecting fraud in credit card transactions. Syst Inf Eng Des Symp SIEDS 2018:129–134. https://doi.org/10.1109/SIEDS.2018.8374722

268. Ahmed SF, Bin AMS, Hassan M et al (2023) Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artif Intell Rev. https://doi.org/10.1007/s10462-023-10466-8

269. Pandey Y (2017) Credit card fraud detection using deep learning. undefined. https://doi.org/10.26483/IJARCS.V8I5.3514

270. De Roux D, Pérez B, Moreno A et al (2018) Tax fraud detection for under-reporting declarations using an unsupervised machine learning approach. Proc ACM SIGKDD Int Conf Knowl Discov Data Min 215–222. https://doi.org/10.1145/3219819.3219878

271. Lee C (2021) Deep learning-based detection of tax frauds: an application to property acquisition tax. Data Technol Appl. https://doi.org/10.1108/DTA-06-2021-0134/FULL/XML

272. López CP, Rodríguez MJD, de Santos S, L, (2019) Tax Fraud Detection through Neural Networks: An Application Using a Sample of Personal Income Taxpayers. Futur Internet 2019(11):86. https://doi.org/10.3390/FI11040086

273. Khalid AR, Owoh N, Uthmani O et al (2024) Enhancing credit card fraud detection: an ensemble machine learning approach. Big Data Cogn Comput 8:6

274. Sutanta H, Rajabifard A, Bishop ID (2010) Integrating spatial planning and disaster risk reduction at the local level in the context of spatially enabled government. Spatially enabling society: Research, emerging trends and critical assessment. pp 56-68

275. Hashem IAT, Yaqoob I, Anuar NB et al (2015) The rise of "big data" on cloud computing: Review and open research issues. Inf Syst 47:98–115. https://doi.org/10.1016/j.is.2014.07.006

276. Anbarasan M, Muthu BA, Sivaparthipan CB et al (2020) Detection of flood disaster system based on IoT, big data and convolutional deep neural network. Comput Commun 150:150–157. https://doi.org/10.1016/j.comcom.2019.11.022

277. Celik T, Demirel H, Ozkaramanli H, Uyguroglu M (2007) Fire detection using statistical color model in video sequences. J Vis Commun Image Represent 18:176–185. https://doi.org/10.1016/j.jvcir.2006.12.003

278. Mehmood I, Sajjad M, Baik SW (2014) Mobile-cloud assisted video summarization framework for efficient management of remote sensing data generated by wireless capsule sensors. Sensors 14(9):17112–17145. https://doi.org/10.3390/s140917112

279. Muhammad K, Ahmad J, Baik SW (2018) Early fire detection using convolutional neural networks during surveillance for effective disaster management. Neurocomputing 288:30–42. https://doi.org/10.1016/j.neucom.2017.04.083

280. Verstockt S, Beji T, De Potter P et al (2013) Video driven fire spread forecasting (f) using multi-modal LWIR and visual flame and smoke data. Pattern Recognit Lett 34:62–69. https://doi.org/10.1016/j.patrec.2012.07.018

281. Sathianarayanan M, Hsu P-H, Chang C-C (2024) Extracting disaster location identification from social media images using deep learning. Int J Disaster Risk Reduct 104:104352. https://doi.org/10.1016/j.ijdrr.2024.104352

282. Zikanov O (2019) Essential computational fluid dynamics, second edition. John Wiley & Sons, Hoboken, NJ, USA

283. Fonda E, Pandey A, Schumacher J, Sreenivasan KR (2019) Deep learning in turbulent convection networks. Proc Natl Acad Sci U S A 116:8667–8672. https://doi.org/10.1073/PNAS.1900358116/-/DCSUPPLEMENTAL

284. Daw A, Karpatne A, Watkins W et al (2022) Physics-guided neural networks (PGNN): an application in lake temperature modeling. In Knowledge guided machine learning (pp 353–372). Chapman and Hall/CRC

285. Dimitriu G, Ştefănescu R, Navon IM (2017) Comparative numerical analysis using reduced-order modeling strategies for nonlinear large-scale systems. J Comput Appl Math 310:32–43. https://doi.org/10.1016/J.CAM.2016.07.002

286. Fang F, Pain CC, Navon IM et al (2013) Non-linear Petrov-Galerkin methods for reduced order hyperbolic equations and discontinuous finite element methods. J Comput Phys 234:540–559. https://doi.org/10.1016/J.JCP.2012.10.011

287. Fang F, Pain CC, Navon IM, Xiao D (2017) An efficient goal-based reduced order model approach for targeted adaptive observations. Int J Numer Methods Fluids 83:263–275. https://doi.org/10.1002/FLD.4265

288. Diez M, Campana EF, Stern F (2015) Design-space dimensionality reduction in shape optimization by Karhunen-Loève expansion. Comput Methods Appl Mech Eng 283:1525–1544. https://doi.org/10.1016/J.CMA.2014.10.042

289. Manzoni A, Salmoiraghi F, Heltai L (2015) Reduced Basis Isogeometric Methods (RB-IGA) for the real-time simulation of potential flows about parametrized NACA airfoils. Comput Methods Appl Mech Eng 284:1147–1180. https://doi.org/10.1016/J.CMA.2014.11.037

290. Alotaibi M, Calo VM, Efendiev Y et al (2015) Global–local nonlinear model reduction for flows in heterogeneous porous media. Comput Methods Appl Mech Eng 292:122–137. https://doi.org/10.1016/J.CMA.2014.10.034

291. Lozovskiy A, Farthing M, Kees C, Gildin E (2016) POD-based model reduction for stabilized finite element approximations of shallow water flows. J Comput Appl Math 302:50–70. https://doi.org/10.1016/J.CAM.2016.01.029

292. Oliver J, Caicedo M, Huespe AE et al (2017) Reduced order modeling strategies for computational multiscale fracture. Comput Methods Appl Mech Eng 313:560–595. https://doi.org/10.1016/J.CMA.2016.09.039

293. Kou J, Zhang W (2021) Data-driven modeling for unsteady aerodynamics and aeroelasticity. Prog Aerosp Sci 125:100725

294. Eivazi H, Wang Y, Vinuesa R (2024) Physics-informed deep-learning applications to experimental fluid mechanics. Meas Sci Technol 35:75303

295. Flah M, Nunez I, Ben Chaabene W, Nehdi ML (2021) Machine learning algorithms in civil structural health monitoring: A systematic review. Arch Comput methods Eng 28:2621–2643

296. Entezami A, Shariatmadar H, De Michele C (2022) Non-parametric empirical machine learning for short-term and long-term structural health monitoring. Struct Heal Monit 21:2700–2718

297. Sarmadi H, Yuen K-V (2022) Structural health monitoring by a novel probabilistic machine learning method based on extreme value theory and mixture quantile modeling. Mech Syst Signal Process 173:109049. https://doi.org/10.1016/j.ymssp.2022.109049

298. Islam N, Kashem A, Das P et al (2024) Prediction of high-performance concrete compressive strength using deep learning techniques. Asian J Civ Eng 25:327–341

299. Zheng H, Lin F, Feng X, Chen Y (2021) A Hybrid Deep Learning Model with Attention-Based Conv-LSTM Networks for Short-Term Traffic Flow Prediction. IEEE Trans Intell Transp Syst 22:6910. https://doi.org/10.1109/TITS.2020.2997352

300. Vijayalakshmi B, Ramar K, Jhanjhi NZ et al (2021) An attention-based deep learning model for traffic flow prediction using spatiotemporal features towards sustainable smart city. Int J Commun Syst 34:e4609. https://doi.org/10.1002/dac.4609

301. Abdollahi M, Khaleghi T, Yang K (2020) An integrated feature learning approach using deep learning for travel time prediction. Expert Syst Appl 139:112864. https://doi.org/10.1016/j.eswa.2019.112864

302. Reddy KHK, Goswami RS, Roy DS (2024) A deep learning-based smart service model for context-aware intelligent transportation system. J Supercomput 80:4477–4499

303. Mirowski P, Madhavan D, LeCun Y, Kuzniecky R (2009) Classification of patterns of EEG synchronization for seizure prediction. Clin Neurophysiol 120:1927–1940. https://doi.org/10.1016/j.clinph.2009.09.002

304. Lee B, Park S, Baek J, Yoon S (2016) DeepTarget: End-to-end learning framework for MicroRNA target prediction using deep recurrent neural networks. ACM-BCB 2016 - 7th ACM Conf Bioinformatics, Comput Biol Heal Informatics 434–442. https://doi.org/10.1145/2975167.2975212

305. Petrosian A, Prokhorov D, Homan R et al (2000) Recurrent neural network based prediction of epileptic seizures in intra- and extracranial EEG. Neurocomputing 30:201–218. https://doi.org/10.1016/S0925-2312(99)00126-5

306. Bai XC, McMullan G, Scheres SH (2015) How cryo-EM is revolutionizing structural biology. Trends Biochem Sci 40(1):49–57. https://doi.org/10.1016/j.tibs.2014.10.005

307. Bartesaghi A, Aguerrebere C, Falconieri V et al (2018) Atomic Resolution Cryo-EM Structure of β-Galactosidase. Structure 26:848-856.e3. https://doi.org/10.1016/j.str.2018.04.004

308. Si D, Moritz SA, Pfab J et al (2020) Deep Learning to Predict Protein Backbone Structure from High-Resolution Cryo-EM Density Maps. Sci Rep 10:1–22. https://doi.org/10.1038/s41598-020-60598-y

309. Park S, Min S, Choi H, Yoon S (2016) deepMiRGene: Deep neural network based precursor microRNA Prediction. arXiv preprint arXiv:1605.00017. https://doi.org/10.48550/arXiv.1605.00017

310. Baek M, DiMaio F, Anishchenko I et al (2021) Accurate prediction of protein structures and interactions using a three-track neural network. Science 373:871–876

311. Gilmer J, Schoenholz SS, Riley PF, et al (2020) Message passing neural networks. Mach Learn meets quantum Phys, pp 199–214. https://doi.org/10.1007/978-3-030-40245-7_10

312. Aderinwale T, Bharadwaj V, Christoffer C et al (2022) Real-time structure search and structure classification for AlphaFold protein models. Commun Biol 5:316

313. Abramson J, Adler J, Dunger J et al (2024) Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630:493–500. https://doi.org/10.1038/s41586-024-07487-w

314. Baldi P, Sadowski P (2013) Understanding dropout. Adv Neural Inf Process Syst 1–9. https://doi.org/10.17744/mehc.25.2.xhyreggxdcd0q4ny

315. Chen Y, Li Y, Narayan R et al (2016) Gene expression inference with deep learning. Bioinformatics 32:1832–1839. https://doi.org/10.1093/bioinformatics/btw074

316. Krizhevsky BA, Sutskever I, Hinton GE (2012) Cnn实际训练的. Commun ACM 60:84–90

317. Liu Z, Li X, Luo P, et al (2015) Semantic image segmentation via deep parsing network. Proc IEEE Int Conf Comput Vis 2015 Inter 1377–1385. https://doi.org/10.1109/ICCV.2015.162

318. Zhao X, Wu Y, Song G et al (2018) A deep learning model integrating FCNNs and CRFs for brain tumor segmentation. Med Image Anal 43:98–111. https://doi.org/10.1016/j.media.2017.10.002

319. Khani P, Nasri F, Khani Chamani F et al (2019) Genetic and epigenetic contribution to astrocytic gliomas pathogenesis. J Neurochem 148:188–203

320. Thaha MM, Kumar KPM, Murugan BS et al (2019) Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images. J Med Syst 43:1240–1251. https://doi.org/10.1007/s10916-019-1416-0

321. Veltri D, Kamath U, Shehu A (2018) Deep learning improves antimicrobial peptide recognition. Bioinformatics 34:2740–2747. https://doi.org/10.1093/bioinformatics/bty179

322. Angermueller C, Lee HJ, Reik W, Stegle O (2017) DeepCpG: Accurate prediction of single-cell DNA methylation states using deep learning. Genome Biol 18:1–14. https://doi.org/10.1186/s13059-017-1189-z

323. Hong Z, Zeng X, Wei L, Liu X (2020) Identifying enhancer-promoter interactions with neural network based on pre-trained DNA vectors and attention mechanism. Bioinformatics 36:1037–1043. https://doi.org/10.1093/bioinformatics/btz694

324. Hoffmann B, Beck M, Sunder-Plassmann G et al (2007) Nature and prevalence of pain in Fabry disease and its response to enzyme replacement therapy - A retrospective analysis from the Fabry outcome survey. Clin J Pain 23:535–542. https://doi.org/10.1097/AJP.0b013e318074c986

325. De Ferrari L, Aitken S, van Hemert J, Goryanin I (2012) EnzML: Multi-label prediction of enzyme classes using Inter-Pro signatures. BMC Bioinformatics 13:1–2. https://doi.org/10.1186/1471-2105-13-61

326. Zou Z, Tian S, Gao X, Li Y (2019) mlDEEPre: Multi-functional enzyme function prediction with hierarchical multi-label deep learning. Front Genet 10:1–10. https://doi.org/10.3389/fgene.2018.00714

327. Zhang Q, Lu M, Dong X et al (2014) Potential estrogenic effects of phosphorus-containing flame retardants. Environ Sci Technol 48:6995–7001. https://doi.org/10.1021/es5007862

328. Zhang Q, Yan L, Wu Y et al (2017) A ternary classification using machine learning methods of distinct estrogen receptor activities within a large collection of environmental chemicals. Sci Total Environ 580:1268–1275. https://doi.org/10.1016/j.scitotenv.2016.12.088

329. Heo SK, Safder U, Yoo CK (2019) Deep learning driven QSAR model for environmental toxicology: Effects of endocrine disrupting chemicals on human health. Environ Pollut 253:29–38. https://doi.org/10.1016/j.envpol.2019.06.081

330. Lyu T, Gao J, Tian L, et al (2021) MDNN: A Multimodal Deep Neural Network for Predicting Drug-Drug Interaction Events. IJCAI Int Jt Conf Artif Intell 3536–3542. https://doi.org/10.24963/ijcai.2021/487

331. Deng Y, Xu X, Qiu Y et al (2020) A multimodal deep learning framework for predicting drug-drug interaction events. Bioinformatics 36:4316–4322. https://doi.org/10.1093/bioinformatics/btaa501

332. Yang Q, Zhang S, Li Y (2024) Deep Learning Algorithm Based on Molecular Fingerprint for Prediction of Drug-Induced Liver Injury. Toxicology 502:153736

333. Gao H, Korn JM, Ferretti S et al (2015) High-throughput screening using patient-derived tumor xenografts to predict clinical trial drug response. Nat Med 21:1318–1325. https://doi.org/10.1038/nm.3954

334. Gottlieb A, Stein GY, Oron Y et al (2012) INDI: A computational framework for inferring drug interactions and their associated recommendations. Mol Syst Biol 8:1–12. https://doi.org/10.1038/msb.2012.26

335. Lee G, Park C, Ahn J (2019) Novel deep learning model for more accurate prediction of drug-drug interaction effects. BMC Bioinformatics 20:1–8. https://doi.org/10.1186/s12859-019-3013-0

336. Singh UP, Chouhan SS, Jain S, Jain S (2019) Multilayer Convolution Neural Network for the Classification of Mango Leaves Infected by Anthracnose Disease. IEEE Access 7:. https://doi.org/10.1109/ACCESS.2019.2907383

337. Karim Ahmed NSKRS Weighted Transformer Network for Machine Translation

338. Yuan Z, Wu S, Wu F et al (2018) Domain attention model for multi-domain sentiment classification. Knowledge-Based Syst 155:1. https://doi.org/10.1016/j.knosys.2018.05.004

339. Chen F, Ji R, Su J et al (2018) Predicting Microblog Sentiments via Weakly Supervised Multimodal Deep Learning. IEEE Trans Multimed 20:997. https://doi.org/10.1109/TMM.2017.2757769

340. Huang Q, Chen R, Zheng X, Dong Z (2017) Deep sentiment representation based on CNN and LSTM. In: Proceedings - 2017 Int Conf Green Inform, ICGI. https://doi.org/10.1109/ICGI.2017.45

341. Hassan A, Mahmood A (2017) Deep Learning approach for sentiment analysis of short texts. In: 2017 3rd International Conference on Control, Automation and Robotics, ICCAR. https://doi.org/10.1109/ICCAR.2017.7942788

342. Narasimhan M, Schwing AG (2018) Straight to the facts: Learning knowledge base retrieval for factual visual question answering. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pp 451–468

343. Pouyanfar S, Tao Y, Tian H et al (2019) Multimodal deep learning based on multiple correspondence analysis for disaster management. World Wide Web 22:1893–1911. https://doi.org/10.1007/s11280-018-0636-4

344. Segler MHS, Waller MP (2017) Neural-Symbolic Machine Learning for Retrosynthesis and Reaction Prediction. Chem - A Eur J 23:5966–5971. https://doi.org/10.1002/chem.201605499

345. Raissi M, Perdikaris P, Karniadakis GE (2019) Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. J Comput Phys 378:686–707. https://doi.org/10.1016/j.jcp.2018.10.045

346. Karumuri S, Tripathy R, Bilionis I, Panchal J (2020) Simulator-free solution of high-dimensional stochastic elliptic partial differential equations using deep neural networks. J Comput Phys 404:1–63. https://doi.org/10.1016/j.jcp.2019.109120

347. Zhu Y, Zabaras N, Koutsourelakis PS, Perdikaris P (2019) Physics-constrained deep learning for high-dimensional surrogate modeling and uncertainty quantification without labeled data. J Comput Phys 394:56–81. https://doi.org/10.1016/j.jcp.2019.05.024

348. Pumsirirat A, Yan L (2018) Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine. Int J Adv Comput Sci Appl 9:18–25. https://doi.org/10.14569/IJACSA.2018.090103

349. Rusk N (2015) Deep learning. Nat Methods 13:35. https://doi.org/10.1038/nmeth.3707

350. Wang T, Wen CK, Wang H et al (2017) Deep learning for wireless physical layer: Opportunities and challenges. China Commun 14:92–111. https://doi.org/10.1109/CC.2017.8233654

351. Shyu M, Chen S, Iyengar SS (2020) Survey on Deep Learning Techniques. Strad Res 7:. https://doi.org/10.37896/sr7.8/037

352. Miotto R, Wang F, Wang S et al (2018) Deep learning for healthcare : review, opportunities and challenges. Brief Bioinform 19:1236–1246. https://doi.org/10.1093/bib/bbx044

353. Rajkomar A, Oren E, Chen K et al (2018) Scalable and accurate deep learning with electronic health records. NPJ Digit Med 1:1–10. https://doi.org/10.1038/s41746-018-0029-1

354. Baytas IM, Xiao C, Zhang X et al (2017) Patient subtyping via time-aware LSTM networks. Proc ACM SIGKDD Int Conf Knowl Discov Data Min Part F1296:65–74. https://doi.org/10.1145/3097983.3097997

355. Jagannatha AN, Yu H (2016) Structured prediction models for RNN based sequence labeling in clinical text. EMNLP 2016 - Conf Empir Methods Nat Lang Process Proc 856–865. https://doi.org/10.18653/v1/d16-1082

356. Pham T, Tran T, Phung D, Venkatesh S (2016) DeepCare: A deep dynamic memory model for predictive medicine. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 9652 LNAI:30–41. https://doi.org/10.1007/978-3-319-31750-2_3

357. Awasthi MK, Sarsaiya S, Wainaina S et al (2021) Techno-economics and life-cycle assessment of biological and thermochemical treatment of bio-waste. Renew Sustain Energy Rev 144:110837. https://doi.org/10.1016/j.rser.2021.110837

358. Choi E, Bahadori MT, Schuetz A et al (2016) Doctor AI: Predicting Clinical Events via Recurrent Neural Networks. JMLR Workshop Conf Proc 56:301–318

359. Lipton ZC (2018) The mythos of model interpretability. Commun ACM 61:35–43. https://doi.org/10.1145/3233231

360. Koh PW, Liang P (2017) Understanding black-box predictions via influence functions. 34th Int Conf Mach Learn ICML 70:1885–1894

361. Gou J, Yu B, Maybank SJ, Tao D (2021) Knowledge Distillation: A Survey. Int J Comput Vis 129:. https://doi.org/10.1007/s11263-021-01453-z

362. Cho JH, Hariharan B (2019) On the efficacy of knowledge distillation. In: Proceedings of the IEEE International Conference on Computer Vision

363. Huang SC, Pareek A, Seyyedi S et al (2020) Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. NPJ Digit Med 3:136. https://doi.org/10.1038/s41746-020-00341-z

364. Alaa AM, Weisz M, van der Schaar M (2017) Deep Counterfactual networks with propensity-dropout. arXiv preprint arXiv:1706.05966. https://doi.org/10.48550/arXiv.1706.05966

365. Razavian N, Marcus J, Sontag D (2016) Multi-task prediction of disease onsets from longitudinal lab tests 56:73–100

366. Henao R, Lu JT, Lucas JE et al (2016) Electronic health record analysis via deep poisson factor models. J Mach Learn Res 17:1–32