International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

https://gssrr.org/index.php/JournalOfBasicAndApplied/index

A Systematic Review of Deep Learning Methods: Classification, Selection, and Scientific Understanding

Zein Samira[a], Denis Kisina[b], Yodit Wondaferew Weldegeorgise[c*]

[a]*Richardson,Texas, USA*
[b]*FortWorth, Texas, USA*
[c]*Texas, USA*
[a]*Email:samirazein43@gmail.com*
[b]*Email:Denis.kisina@gmail.com*
[c]*Email:yoditweldegeorgise@gmail.com*

Abstract

This study, A Systematic Review of Deep Learning Methods: Classification, Selection, and Scientific Understanding, categorizes central deep learning (DL) models—including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Generative Adversarial Networks (GANs), and Autoencoders (AEs)—based on their suitability for specific tasks and data types. While DL has achieved significant success in image recognition, language processing, and anomaly detection, several critical limitations pertain to interpretability, robustness, and scalability. This review summarizes the strengths and weaknesses of each model in a structured manner to guide the choice among DL models. Findings emphasize that theory must be advanced to improve transparency and reliability to better support practitioners and researchers in making informed choices for DL's responsible deployment across sectors.

Keywords: Deep learning; model selection; interpretability; Convolutional Neural Networks (CNN); Recurrent Neural Networks (RNN); Generative Adversarial Networks (GAN); Autoencoders (AE); classification, scalability.

Received: 9/1/2024
Accepted: 11/1/2024
Published: 11/11/2024

---

* Corresponding author.

1. Introduction

Deep learning, a specialized branch of machine learning, has driven innovation onto the fast track in everything from image recognition and language processing to predictive analytics. Equipped with the ability to process enormous amounts of data and self-learn complex patterns, deep learning has brought about industrial transformation by using systems to carry out tasks such as disease diagnosis and self-driving cars. Despite its excellent capabilities, DL remains unfathomably hard to understand. While high-dimensional data processing in DL models supports their performance of specific tasks, the inner mechanisms are usually opaque. It is still an issue to tell why a particular outcome occurs when a given input is applied and what the limitation of their computation is for some DL models; they are also prone to errors when data is slightly perturbed [1, 2]. Because such work does not have theoretical grounds, this, in turn, limits DL's applicability in high-stakes fields where robust, interpretable, and reliable performance is critical [3]. The current study overcomes these challenges by systematically reviewing the DL methods' strengths and limitations, including their suitability for various tasks and data types.

The review classifies DL models from Convolutional Neural Networks to Generative Adversarial Networks use cases and technical requirements, providing an all-encompassing framework that guides model selection. Conclusively, it gives the theoretical gaps in DL and suggests ways to improve interpretability and robustness.The review explores three main questions: (1) What are the strengths, limitations, and appropriate applications for current deep learning (DL) algorithms? (2) How can DL methods be systematically classified to improve model selection for specific tasks and data types? Moreover, (3) What are the theoretical challenges in DL, especially concerning interpretability and robustness?

2. Literature Review

2.1. Overview of Deep Learning and Machine Learning

Deep learning is thus a narrow segment of machine learning, focusing on large volumes of data processing, often with unstructured information represented by images, audio, and text [3]. Contrasting with the traditional models of machine learning, which require explicit feature extraction from experts, these deep learning models learn features through multiple layers of their networks. Thus, they can become more expressive in pattern recognition. This autonomy of feature learning in DL greatly benefits the processing of complicated data types in applications such as image recognition, natural language processing, and autonomous systems. While ML typically performs well on structured data problems and provides more interpretability, DL offers higher accuracy and efficiency in unstructured data, albeit at the cost of being less interpretable [4].

Machine learning depends on explicit feature extraction by humans to extract insights from data. Most of the models are designed for performing either supervised or unsupervised learning. ML models, including decision trees and support vector machines, work well on structured datasets and present precise, traceable results [4]. On the other hand, DL architecture consists of multi-layer neural networks operating on supervised and unsupervised data with algorithms to extract very subtle patterns with no human input. With the potential for analysis of nonlinear data patterns with DL, there has been an expanded use of CNNs and RNNs in various fields. For instance, CNNs process image data through spatial dimensions, while RNNs process sequences in time series and text data [5].

2.2. Historical Development of DL Models

The history of DL can be traced back to the 1950s, with basic models such as the Perceptron, proposed by Frank Rosenblatt, explicitly developed to classify linearly separable data points. The latter improvement on the Perceptron through Backpropagation, created during the 1960s, allowed error-based training across multi-layered networks, a critical development necessary in deeper architectures. This presented the premise for deeper DL architectures that solve nonlinear problems [6].

With the development of DL, major models came forward, such as AE, which was discovered during the 1980s for data dimensionality reduction by keeping only the essential features, and CNNs, which were introduced by Yann LeCun back in 1989 and specialize in recognizing images through their space hierarchies. Since then, CNNs have been at the core of various applications, such as medical imaging and autonomous driving, due to their rigidity in image classification tasks [6]. The 1990s finally witnessed the evolution of RNNs, which are considered very good at sequence-based tasks; hence, their value in speech recognition and language translation [6]. More recent developments include GANs, introduced in 2014 by Ian Goodfellow, which are comprised of two neural network generators and operate in opposition to generating realistic data samples [6]. Applications of GANs range from image synthesis to natural language processing and even extend to healthcare, generating synthetic medical images for research purposes.

These models represent the movement of DL from simple binary classifiers to sophisticated architectures able to manipulate complex, high-dimensional data. On the other hand, this fast advance within DL has also underlined the critical limits of the scientific knowledge about these models on aspects like interpretability, robustness, and computational bounds [1, 2]

2.3. Core Issues in DL Understanding

Notwithstanding the successes of DL, massive gaps remain in basic theoretical understanding regarding model behavior, robustness, and interpretability. Perhaps the most limiting is that they are considered "black-box" models since the processes driving predictions are incompletely understood. Researchers further point out the inability of the field to explain why some models produce a particular output and how to predict the behavior that models will exhibit on unseen datasets [1, 2]. For example, though DL models work well on big datasets, they become brittle once applied to data with slight perturbations or changes in the data distribution.

Another point of incompleteness in understanding DL relates to how these models respond to new input data distributions. In particular, while DL models are pre-trained on specific distributions, they may not generalize well when unseen data are given [7]. This has been an essential issue in applications concerned with autonomous driving or medical diagnosis, wherein such an uninformative input might result in unsafe or incorrect outcomes. Besides, there is also a shortage of approaches for quantifying uncertainty in the predictions of DL, whereas one can measure the confidence intervals using probabilistic ML models. The inability to determine the reliability of DL outputs constrains its utility in high-stakes decision-making [4]

Apart from that, the computational bounds of DL models are not particularly well known. For example, Thompson and his colleagues. raise the question of whether adding more data and increasing computational power result in perpetual improvement in the performance of DL models or whether there is a saturation point at which further addition of these resources does not improve the model anymore [3]. These omissions must be addressed theoretically to enhance safety and reliability when applying DL across industries. Other domains, such as random forests of ML by Biau & Scornet 2015 and high-dimensional sparse statistics, also contain similar gaps between practical performance and theoretical interpretation, thus restating the challenge at large in the machine learning disciplines [8].

2.4. Applications and Limitations in Practice

Applications range from healthcare and manufacturing to financial and natural language processing. However, the field faces a bottleneck of theoretical grounds, hence not being scalable and reliable over a wide range of domains. As expected, CNNs have become very accurate for medical imaging, especially in radiology and pathology, where the extraction of patterns enables early disease detection [9]. This is because CNNs are good at recognizing spatial features and thus work well where image-based applications are concerned. Similarly, most applications of RNN exist for tasks regarding sequence modeling, such as language translation and time series forecasting, where the maintenance of the sequential dependencies is crucial if the accuracy is high [6].

It is used in manufacturing to identify product defects and improve quality control. The model fragility problem typically plagues this application: models developed on particular production conditions misclassify defects when there are slight variations [8]. It also plays a significant role in finance for stock price predictions or fraud detection; however, the well-known vulnerability of DL to adversarial inputs is small data perturbations yielding drastic changes in outputs, which threatens its reliability. Particularly dubious are high-stakes applications where predictions must remain trustworthy and resilient against general manipulation.

The incorporation of DL models such as GANs and CNNs that generate refinements of building models, for example, in architectural design, helps an architect see the model of a building design [9]. This could also enhance energy efficiency. Models might be good at high-dimensional spatial data but are limited in providing transparency for extensive modifications in design due to a lack of interpretability. The black-box nature of the models faces challenges in many fields where decision-making has to be transparent, especially in fields involving public safety. For instance, CNNs have shown tremendous promise in designing structures with optimized layouts and durability; however, due to low interpretability, translating these into valuable insights remains challenging [9]

These limitations in scientific understanding at DL imply practical consequences for its scalability. For instance, a considerable quantity of labeled data must be used to train deep learning models, which is prohibitively expensive and time-consuming to collect, such as in medical diagnostics. The heavy computational requirements for DL necessitate specialized hardware such as GPUs; this makes DL inaccessible in specific settings and significantly smaller organizations. Bound by these factors, DL finds its application, especially in resource-constrained environments, while high energy consumption for training large-scale models raises environmental concerns.

3.Methodology

The primary research questions guiding this review are: (1) What are the strengths, limitations, and appropriate applications for current deep learning (DL) algorithms? (2) How can DL methods be systematically classified to improve model selection for specific tasks and data types? Moreover, (3) What are the theoretical challenges in DL, especially concerning interpretability and robustness?

A systematic review approach was followed to answer these questions, and data collection and synthesis were done according to a well-structured protocol. A structured search will be conducted across major academic databases like Scopus and IEEE Xplore for studies published between 2012 and 2024 since the research area is fast-moving. The search terms covered the primary DL methods, application areas, and challenges determined a priori to cover relevant research.

The screening criteria included English, peer-reviewed articles, conference papers, and proceedings. Title, abstract, and keywords are screened only to select studies that precisely address the applications, strengths, and limitations of DL. The initial screening ensured that only pertinent and high-quality sources would inform the review synthesis.

This systematic approach provided a sound basis for the analysis of the various methods of DL from both the theoretical and applied standpoints. Afterward, the chosen studies were synthesized to create a taxonomy of the methods of DL that could give a better framework for understanding and selecting models based on the task at hand and data types. These recommendations were developed through such an analysis. They should guide practitioners and researchers in effective model selection among DL models, indicating practical applications while underpinning areas where further research is required concerning DL interpretability and robustness.

4. Overview of the Results

The systematic review recognized the significant techniques in DL, including CNN, RNN, AE, and GAN, among others, and logically framed them based on the suitability and strengths of each in target applications. CNN could manipulate images and spatial data; RNN was ideal for sequential data, such as natural language processing and time series analysis applications. The most suitable applications for AEs included unsupervised learning tasks like data compression and anomaly detection. In contrast, GANs helped generate synthetic data and handle complicated picture-generation activities.

It also provides a taxonomy that guides the selection of models based on task type, characteristics of data, and model robustness requirements. Key findings of the work emphasized vital challenges, especially in model interpretability and the capability for generalization under changing conditions, thus probing future research in dealing with the theoretical gaps of DL. Hence, the review extends support to a more fine-tuned approach toward DL method selection that guides practitioners and shows areas where increased interpretability and robustness might enhance the reliability and performance of the models.

Table 1: Summary of the Results

Research Question	**Findings**
Strengths, Limitations, and Applications of DL Algorithms	DL methods excel in tasks like image recognition, language processing, and data generation but suffer from limitations in interpretability and robustness. Sensitive to data variations, DL models may underperform in high-stakes settings requiring reliability.
Classification and Taxonomy of DL Models	DL models were categorized by task and data type: - CNNs for image processing and object detection - RNNs for sequential data (e.g., language, time series) - GANs for data generation - AEs for anomaly detection.
Theoretical Challenges in DL: Interpretability and Robustness	DL's "black-box" nature limits transparency, making it difficult to justify or validate model decisions in critical fields. Unpredictability under data shifts and the high computational demands of DL models remain pressing concerns.
Recommendations for Practitioners and Researchers	Model selection should consider both task requirements and data type. Future research should focus on advancing DL interpretability and developing robust, less resource-intensive DL architectures to expand safe deployment across industries.

5.Discussion

This systematic review presents a clear taxonomy of the DL methods, categorizing them by best task suitability, data type, strengths, and limitations. While this taxonomy gives practical advice for choosing the DL algorithm for a given problem, it also points out several open issues regarding theoretical and practical challenges within the field. While DL has attained arresting milestones across domains, including image processing, natural language processing, and predictive analytics, profound gaps persist in our comprehension of model interpretability, robustness, and scalability, limiting the widespread applicability of DL. This discussion gives a comprehensive view of these themes and how each DL method fits into its application context. Hence, it is confronted by persistent challenges that restrict DL's potential.

5.1. DL Method Suitability and Task-Specific Advantages

The review pointed out that CNNs work exceptionally well in activities involving data with spatial information, such as image processing and object detection [7, 8, 9, 10]. In CNNs, the architecture is usually multilayered, comprising convolutional and pooling layers that extract features from the image data; hence, CNNs are ideal for applications that involve hierarchies of spatial entity recognition. These works have established the efficacy of CNNs in applications ranging from disease detection using medical imaging to autonomous driving, where a CNN endows a vehicle capable of recognizing road signs and obstacles Reference [7]. Although CNNs have achieved high accuracy in these applications, interpretability remains limited; while models can perform well, understanding the feature hierarchies that drive their decisions is difficult.

Recurrent Neural Networks have, till now, shown strengths in handling sequential data and are ideal for tasks such as time series analysis, natural language processing, and speech recognition [1]. Strong points of RNNs lie in the capturing of dependencies within a sequence. However, they are prone to limitations such as vanishing gradients that reduce their effectiveness in handling long-term dependencies. Other works have leveraged Long Short-Term Memory and GRU to mitigate such issues; these have their retrieved information over longer sequences and are applicable in areas such as machine translation and predictive analytics in finance [2]. However, as these networks improved on this weakness, there is still a problem in real-time uses of RNNs where computational efficiency and stability are paramount.

Autoencoders first appeared for anomaly detection, dimensionality reduction, and noise reduction tasks. The general philosophy of AEs is to compress the input data into a latent space and attempt to reconstruct it [8]. It finds its applications in identifying outliers or significant features within unstructured data. Applications range from cybersecurity, which helps detect anomalous patterns, to medical imaging, which reduces noise in diagnostic images [11]. However, AEs are challenged to decide the optimal dimensionality of the latent space, which usually affects the accuracy and reliability of data reconstruction. The second limitation of AEs is that large datasets are required, which constrains application in environments with constrained data [12].The generative adversarial network comprises a generator and a discriminator network, which have proved very helpful in generating synthetic data, considering that their significant applications are found in image synthesis and realistic data generation [13, 14]. GANs have become valuable tools in fields such as fashion and gaming for generating novel images and healthcare and creating synthetic medical images for research and training. Although GANs are notoriously difficult to train, the balance between generator and discriminator networks must be outstanding to get high-quality results [7]. Eventually, the training instability and high computational needs associated with GANs reduce their usability on broader applications.

5.2. Challenges in Interpretability and Robustness

While deep learning methods have had remarkable successes, their black-box nature creates ongoing challenges, especially in high-stakes applications where model interpretability is central. Understanding interpretability means not only what the model predicts but also why it made such predictions [6]. Unlike in more traditional ML models, in which decision boundaries and feature importance can often be traced out, the nature of deep learning operates over many complex layers of transformation that are inherently hard to see into [5]. While CNNs and GANs perform excellently in image-related applications, for instance, their low interpretability constrains their adoption in healthcare, where the rationale behind model decisions must be understandable for clinical validation purposes.

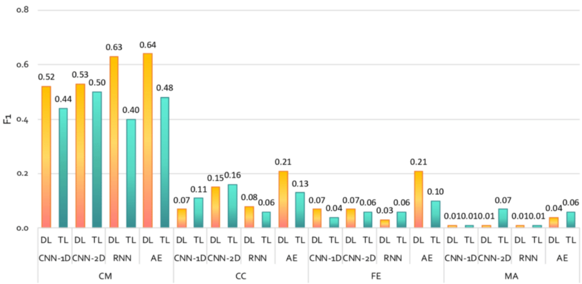

Figure 1: Comparison of Performance of Different DL methods

here are also interpretability implications for model robustness. Since it is difficult to understand how models process the data and make their predictions, there is limited confidence in the DL systems, especially when such models are confronted with novel or slightly different inputs [5]. For example, deep learning research has found evidence that such models are sensitive to changes in the data distribution due to noisiness, perturbations, or data pattern shifts. In one study, slight changes in input images resulted in the misclassification of objects by CNNs; these findings have tended to explain the fragility of DL models even when minor changes present themselves. This is a worrying sensitivity for critical applications that require autonomous vehicles and medical diagnostics to have reliable model performance for safety and efficacy.

Another point is that the methods of quantifying uncertainty within the predictions are missing in DL models [5, 6]. Whereas probabilistic models can estimate confidence in a prediction using probability distributions, most deep-learning models yield deterministic outputs without indicating the reliability of their decisions. This also presents difficulties in assessing when a model may fail in applications where knowing this fact will be beneficial. For example, predicting a degree of uncertainty in forecasting the stock market or climate modeling could help make better decisions.

5.3. Practical Constraints in DL Scalability and Data Requirements

Most DL models require a lot of computational resources and large datasets to be trained appropriately; this usually brings in a scaling inhibitor. For example, CNNs and GANs usually require large datasets to generalize well. They become challenging to implement in scarce data domains [4]. Other models can still work with considerably smaller datasets, such as AEs, but their performance will suffer poorly without adequate data variety and quantity. Demand for labeled data in supervised learning also brings some practical and ethical issues, for example, in healthcare, where data labeling can be extremely time-consuming and costly [8]. A certain offsetting of this burden of data has been done by developing semi-supervised and unsupervised methods. However, these are still very early and require further research.

The scalability is also further complicated by the computational cost associated with DL. Most deep learning models require a GPU during training and inference, making them unreachable for small institutions or resource-constrained environments [9]. All these demands for special hardware bring about financial and environmental implications, as energy-intensive processes contribute to the overall carbon footprint. In response, model optimization techniques like pruning and quantization have been investigated to reduce computational demands; however, these usually come at the cost of model accuracy and, therefore, need further trade-offs [4].

5.4. Theoretical Gaps and Future Research Directions

Besides the practical difficulties, this review has indicated serious theoretical gaps that impede DL improvement. Studies describe some of the scientific open issues concerning the explainability of the behavior of DL models and provide essentially mathematical properties of NNs [1, 2]. Most ML models based on DLs include several matrix transformations whose interactions are far from interpretable, explaining why certain weights have been learned or how these weights relate to final predictions. This lack of theoretical clarity imposes a severe barrier to model validation in that the developers and users cannot find insights on how to verify the performance of DL against different data inputs.

A study investigated whether adding more data or more computation could continue to improve DL models forever or whether, at some point, diminishing returns would kick in and make improvements questionable [3]. These computational limits are essential to consider and are vital for informing the scalability of DL technologies in the future. Besides, other ML techniques, such as random forests and high-dimensional sparse statistics, face similar lapses in theoretical explanation [3, 6, 7, 8]. More significant problems may exist in machine learning areas where empirical results outpace scientific understanding.

These gaps need more research to improve the interpretability of DL and develop metrics for estimating model uncertainty and robustness. Techniques such as XAI aim to enhance the transparency of DL models by attributing the features that contribute most toward specific predictions. Methods such as SHAP and LIME are being explored to provide localized explanations for model decisions. Most existing techniques remain narrow in scope and hardly scalable to handle such complexity from DL models, especially for those applications demanding high accuracy and reliability.

6.Conclusion

This systematic review examined the strengths, limitations, and applications of critical deep learning (DL) methods, aiming to provide a structured framework for model selection and to address gaps in DL's theoretical understanding. By categorizing DL methods such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Autoencoders (AEs), and Generative Adversarial Networks (GANs), this study highlighted each model's specific advantages for tasks involving image processing, sequential data, anomaly detection, and data generation. This taxonomy aids in selecting appropriate DL models based on task type and data characteristics, offering practical guidance to researchers and practitioners alike.

Results have underlined that, despite DL's enormous successes across many tasks and industries, there is a clear need to investigate further the interpretability, robustness, and scalability of deep models. Due to the black-box nature of DL, this severely limits its usability in critical applications, in particular those for which transparency in decision-making is necessary. Besides that, DL models have become sensitive even to minor variations in data and rely heavily on massive datasets, which may be risky in dynamic or data-poor environments. A better solution to these issues is needed with the help of better theoretical frameworks and methods for interpretability so that DL becomes more reliable and accessible.

References

[1] S. Arora, N. Cohen, and E. Hazan, "On the Optimization of Deep Networks: Implicit Acceleration by Overparameterization," pp. 244–253, Jan. 2018.

[2] R. Balestriero and R. G. Baraniuk, "Batch normalization explained," arXiv preprint arXiv:2209.14778, 2022.

[3] N. Thompson, K. Greenewald, K. Lee, and G. F. Manso, "The Computational Limits of Deep Learning," Jul. 2020, doi: 10.48550/arxiv.2007.05558.

[4] Q. Zhang and S. Zhu, "Visual interpretability for deep learning: a survey," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, no. 1, pp. 27–39, Jan. 2018, doi: 10.1631/fitee.1700808.

[5] N. Akhtar and A. Mian, "Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey," *IEEE Access*, vol. 6, pp. 14410–14430, 2018, doi: 10.1109/access.2018.2807385.

[6] B. B. Antal, A. G. Chesebro, H. H. Strey, L. R. Mujica-Parodi, and C. Weistuch, "Achieving Occam's razor: Deep learning for optimal model reduction," *PLoS Computational Biology*, vol. 20, no. 7, pp. e1012283–e1012283, Jul. 2024, doi: 10.1371/journal.pcbi.1012283.

[7] M. Ivanovs, R. Kadikis, and K. Ozols, "Perturbation-based methods for explaining deep neural networks: A survey," *Pattern Recognition Letters*, vol. 150, pp. 228–234, Oct. 2021, doi: 10.1016/j.patrec.2021.06.030.

[8] Y. Kuvvetli, M. Deveci, T. Paksoy, and H. Garg, "A predictive analytics model for COVID-19 pandemic using artificial neural networks," *Decision Analytics Journal*, vol. 1, p. 100007, 2021, doi: 10.1016/j.dajour.2021.100007.

[9] Y.-C. Wang, T. Chen, and M.-C. Chiu, "An explainable deep-learning approach for job cycle time prediction," *Decision Analytics Journal*, vol. 6, p. 100153, 2023, doi: 10.1016/j.dajour.2022.100153.

[10] A. K. Singh and D. Koundal, "A Temporal Convolutional Network for modeling raw 3D sequences and air-writing recognition," *Decision Analytics Journal*, vol. 10, p. 100373, 2024, doi: 10.1016/j.dajour.2023.100373.

[11] F. E. Ayo, L. A. Ogundele, S. Olakunle, J. B. Awotunde, and F. A. Kasali, "A hybrid correlation-based deep learning model for email spam classification using fuzzy inference system," *Decision Analytics Journal*, vol. 10, p. 100390, 2024, doi: 10.1016/j.dajour.2023.100390.

[12] J. Yang, X. L. Zhang, and P. Su, "Deep-learning-based agile teaching framework of software development courses in computer science education," *Procedia Computer Science*, vol. 154, pp. 137–145, 2019, doi: 10.1016/j.procs.2019.06.021.

[13] M. Abadi et al., "TensorFlow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*.

[14] C. Rudin, "Stop explaining black-box machine learning models for high stakes decisions and use interpretable models instead," *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, May 2019, doi: 10.1038/s42256-019-0048-x.