A Comprehensive Crucial Review of Re-Purposing DNN-Based Systems: Significance, Challenges, and Future Directions

Yaser M Al-Hamzi, Shamsul Bin Sahibuddin
Faculty of Artificial Intelligence (FAI), University Technology Malaysia (UTM), UTMKL, Kuala Lumpur, Malaysia

***Abstract*—The fourth industrial revolution is marked by the significance of artificial intelligence (AI), particularly the remarkable progress in deep neural networks (DNNs). These networks have become crucial in various areas of daily life because of their remarkable pattern-learning capabilities on massive datasets. However, the incompatibility of these systems makes reutilizing them for efficient data analysis and computation highly intricate and challenging due to their fragmentation, internal structure, and complexity. Training in DNNs, a vital essential activity in model development, is often time-consuming and costly intensive computation. More precisely, reusing the entire model during deployment when only a small portion of its required features will result in excessive overhead. On the other hand, reengineering the model without efficient code review could also pose security risks as the model would inherit its defects and weaknesses. This paper comprehensively reviews DNN-based systems, encompassing cutting-edge frameworks, algorithms, and models for complex data and existent limitations. The study, which results from a thorough examination, analysis, and synthesis of observations from 193 recent scholarly papers, provides a wealth of knowledge on the subject, identifying key issues and future research directions by offering novel guidelines to advance the DNN model's repurposing and adaptation, especially in finance, healthcare, and autonomous applications. The demonstrated findings, specifically those related to failure and risk challenges of DNN converters, including factors (n=12), symptoms (n1=4, n2=3), and root causes (n1=4, n2=3), will enrich the ML-DNNs community and guide them toward desirable model development and deployment improvement, with significant practical implications for intelligent industries.**

Keywords*—*DNNs; DNN-based systems; significance and challenges; incompatibility; re-purposing; review

I. Introduction

In the digital age, computer science has increasingly focused on the advancements of artificial intelligence, which has become a thriving area of research [1]. However, despite the urgency of matching the rapid development of the fourth industrial revolution, literature overlooks the scarcity of resources related to AI, ML, and DNN-based systems [2] , [3]. It fails to discuss their importance within ecosystems and their various difficulties [4], [5], [6]. The ability to analyze and forecast potential growth through complex data analytics, leveraging deep learning capabilities, leads to increased business value [7]. More specifically, convolutional neural networks CNN [8] and recurrent neural networks RNN [9], the two primary types of DNN architectures, have been intensively investigated to handle various NLP [10] and computer vision CV [11] Problems. It has been demonstrated that different hyperparameters (e.g. learning rate, number of layers, epochs number, optimizer, hidden size, batch size, and regularization techniques) can cause DNN performance to vary significantly [12]. In addition, optimizing parameters and ensuring compatibility between CNNs and RNNs is critical for their performance [13]. This is because these techniques are relatively new and constantly evolving, with ongoing research exploring how best to leverage their combined strengths [12], [13]. However, while several engineering efforts attempted to overcome the exchange complexities between artificial architectures, specifically machine learning (ML) and deep learning (DL) techniques, the literature indicates that significant limitations still exist, whether in domain knowledge or practical applications [1], [14], [15], [16], [17]. For instance, the inspiration of technology reuse [18], [19], [20], [21], and [22] demonstrated that one of the most significant challenges facing (ML) and (DL) developers, researchers, and end users is the lack of interoperability between their systems [13].

In more detail, reusing DNN-based systems computation is difficult, as with any emerging technology [23], [3]. One of these obstacles is the structural problem of non-interoperability among DNN-based systems, along with a need for more technical skills and engineering methods. [3]. Therefore, this systematic review examines cutting-edge interfaces, models, frameworks, and algorithms tightly associated with machine learning (ML) and deep learning (DL) approaches [24].

Importantly, this study stands out as the first systematic investigation that comprehensively addresses the existing open issues in machine and deep learning systems and related technologies to the best of our knowledge. It explores prominent techniques like "Facebook's Torch-PyTorch and Caffe2" [16], [25], [7], [26], [21], [20], "Montreal University's Theano, TensorFlow founded by Google [27]; [28], Apache's MxNet, Microsoft's CNTK [21], [20], [28], [29], and Hugging-Face [30]" highlighting their contributions to the analysis diverse data types, including complex data as well as the existing limitations. The systematic investigation included a pack of widely used deep neural network models like LeNet-1 [31], LeNet-5 [32], ResNet-18 and ResNet-152 [33], Xception [34], Inception-V3 [35], VGG-16 [36], VGG-19 [37]. The findings of this review illustrated that, in development and deployment processes, the lack of compatibility has three categories: software level, hardware level, and architecture level. The software level includes "type of used programming language" [4], [38], type of "ML and DL framework" [24], type of "ML and DL model" [39], type of" ML and DL algorithm" [12], and type dataset [40], [41]). The hardware level encompasses the type of "computer manufacturer, processor, and type of accelerators (e.g., GPUs, TPUs, FPGAs) [42], [43]. The architecture level comprises structure design and mission misalignment levels [44], [45], [46]. In a relevant context, the findings of this review indicate that a unified model that combines two or more of these methods can significantly enhance deep neural network performance (e.g., memory consumption and inference time) [24], [44].

This review explores the potential of (DNNs) and the challenges they present in the reuse context, enabling more efficient progress in methodologies and processes. It comprehensively addresses their significance and reuse challenges from a unique perspective. The aim is to make a crucial contribution to advancing DNN-based systems. The promising unified approach with its novel techniques can effectively promote compatibility and reuse process, leading to obtaining the desirable accuracy [47], [48], [49] and robustness [50], [49] towards various types of adversarial attacks, reducing computing time and lower computing costs [48], [51], [24].

II. Deep Neural Networks Reuse

A. Overview

The reuse approach involves utilizing existing software engineering and artificial intelligence technologies for different purposes, such as reducing engineering and computation costs and time [23], [52]. Nevertheless, reusability from a deep neural network perspective has received limited attention [53], [54]. One of the main focuses of this paper is the incompatibility problem, challenges, and possible solutions for reusing deep neural networks (DNNs). We do this by highlighting four main reuse paradigms that have been identified through literature review, analysis, and synthesis. These paradigms consider things like the need for computing resources and inference time, hardware configurations, and dataset dependencies [23], [55], [3]. The four include conceptual reuse, development and assessment of the need for reuse, adaptation to reuse, and deployment reuse, as illustrated in Fig. 1.

B. Reuse Paradigms Definition

1) Conceptual reuse: It involves replicating and reengineering algorithms or model architectures from academic literature, often due to licensing or using a specific DL framework. This approach is related to Sommerville's abstraction reuse. [56] and is crucial for scientific reproducibility [57].

2) Model development and reusability assessment: Involve accurately determining the nature of the intended task, followed by data preprocessing, hardware preparation, and algorithm selection for training a model from scratch and evaluating the need and potential for reuse. [58]; [30].

3) Adaptation reuse: Utilizing existing DNN models for different learning tasks, leveraging techniques like transfer learning or knowledge distillation. [59]. This approach is suitable for publicly available pre-trained models PTMs, allowing engineers to customize them for different tasks. [60].

4) Deployment reuse: This method of reusing pre-trained DNN models in various computational environments and frameworks is ideally suited for an engineer's desired task. This approach is similar to Sommerville's "system reuse" and involves fine-tuning, then converting the model from one representation to another, followed by compilation to optimize for hardware. Multiple forms of reuse can be possible in a single engineering project [3].

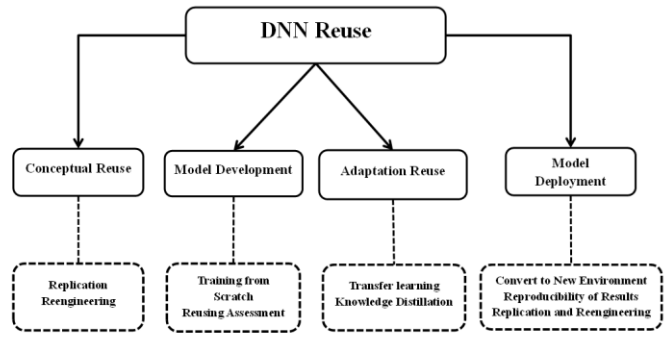

Fig. 1. The four paradigms of deep neural network reuse. The information illustrated in dashed boxes is an example of each paradigm type.

III. Motivation

Machine learning and deep neural networks have benefited numerous aspects of AI, including finance, autonomous applications, and healthcare (e.g. drug discovery, disease predictions, and medical image analysis in intelligent healthcare applications) [61]. However, a critical barrier prevents AI from reaching its full potential. The issue concerns the incompatibility of AI environments, mainly when leveraging previous successful efforts like drawing inspiration from developed codes and pre-trained models [17]; [13]. This incompatibility partially splits the development landscape and prevents code and model sharing, which is an effective way to boost innovations within the field [62]. Reusing existing code and pre-trained models provides significant benefits as follows:

A. Reduced Development Time

For instance, [23] and [54], demonstrated that, by reusing existing code and pre-trained models, experts could focus on possible and innovative solutions, thereby saving significant time that would otherwise be required to construct all components from scratch.in addition, reutilizing eliminates the need to compute similar data multiple times, speeds up development, and reduces costs.

B. Enhanced Reproducibility

Code generation, model sharing, and reusing approaches improve the overall reproducibility of research, allowing for verifying and enhancing existing solutions [13], [55].

C. Accelerated Innovation

By extension, extending previous frameworks and models means that researchers can work on more complicated solutions over a shorter period and advance the field of AI [52]. However, current approaches to reusing and reengineering the previous

ML and DNN productions encounter certain limitations, such as:

1) Performance bottlenecks: Reused models may consume more memory during inference, which may take longer than inference in their native environment [23].

2) Concerns over accuracy and robustness: Fine-tuning necessitates a series of decisions to maintain the model's accuracy in the newly applied context [49].

3) Converting the entire model challenge: Existing methods are ineffective because they require reloading the whole model rather than just the necessary portions, resulting in costly and time-consuming computational resources [23]; [63].

4) Inheritance model defects, when reused, may cause security vulnerabilities: When reusing models, the new system may inherit threats from the previous models, thereby increasing its vulnerability to attacks. These limitations thus advocate for extensive research and development of code and model reuse.

To overcome these challenges, this study suggested the following critical solutions:

- Standardized Frameworks: Many other research areas are challenged to create coherent architectures that facilitate simple integration of off-the-shelf models and built-from-scratch submodules.
- Advanced Transfer Learning Techniques: Studies on techniques for implementing transfer learning, such as domain adaptation, fine-tuning methodologies, and lifelong learning paradigms.
- Developing Novel, Unique Reuse Methods: Involves exploring strategies for reusing only specific aspects of a model, which enables the creation of unique solutions without requiring additional work.
- Development of Security-Aware Reuse Strategies: Creating strategies for distinct discovery and protection of security threats whenever a model is reused to create other new systems. Solving these issues will lead to a healthier ground for code and model sharing in AI. In the long run, this will lead to more innovation, improve the use of AI to its full potential, and bring about changes that will positively impact society.

IV. Scope of Study and Design Architecture

In general, the search strategy for addressing the main research problem and the related challenges in this systematic review is to move in an inverted pyramid, i.e. from the broadest level to the narrowest one, as shown in Fig. 2. The investigation and synthesis then continue to narrow coherently, ultimately reaching the most specific, valid, and critical missing points that previous studies and current methods and approaches have neglected. We examine, investigate, synthesize, analyze, and discuss the most significant open issues and challenges that are closely related to the main research focus of incompatibility, drawing inspiration from the reuse approach and the limitations of current systems reengineering methods. We initially go from the big picture of AI, focusing on computer vision CV and natural language processing NLP, to the more specific view of machine learning with its related techniques (e.g. algorithms, models, and frameworks). Then we move to the much more specific domain of deep neural networks (DNNs), which includes the most computational techniques, such as CNN-based, RNN-based, and hybrid systems, such as DNN-NLP, DNN-CV, NLP-CV, and DNN-NLP-CV [64], [65], [66].

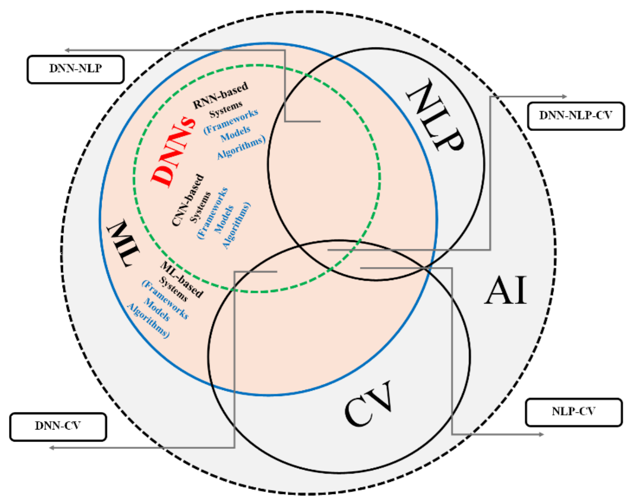

Fig. 2. The architecture of the dependence between AI, ML, computer vision (CV), NLP, and Hybrid DNN-based systems.

V. Concepts

A. The Border Domain of Artificial Intelligence (AI)

Artificial Intelligence (AI) is a “branch of computer science,” and its latest generation is rapidly expanding. This makes it an attractive topic for research that focuses on designing intelligent machines that mimic human thought processes, such as learning and decision-making [1]. This level of automation means that such machines give performances on inference, categorization, and detection for activities formerly involving interfaces [67]. The AI context includes the facilities and technologies used to build intelligent applications and release them into the environment [68].

B. Machine Learning (ML)

ML can be defined as a category of AI involving the implementation of algorithms and methods that enable the computer to learn from data and experience without presenting a set of instructions [69]. DNN-based models can learn more about a particular problem as opposed to other approaches by mirroring the data and identifying patterns and relationships that can be used to construct models [70]. The most frequent categories of ML problems entail classification, regression, and clustering. According to [67], although machine learning models have brought significant improvements to AI applications in fields ranging from finance and healthcare to manufacturing, this progression was predominantly due to their relative accuracy (concerning traditional linear statistics classifiers) when it comes to making predictions. However, this benefit comes with a drawback: the lack of transparency in the design of machine learning models often leads to their characterization as "black boxes". On the other hand [71], demonstrated that the integration of (ML) and (AI) components into public sector applications faces significant limitations due to their fragility and algorithmic mismatches.

C. Deep Neural Networks (DNNs)

Deep learning, also known as "deep neural networks DNN," is a more advanced form of machine learning that requires using an artificial neural network with multiple layers embedded [72]. These deep networks, which optimize data-intensive tasks like image, text, voice, and speech recognition, are based on the structure and functioning of the human brain [73], [74], [16]. With deep learning algorithms, changes are made to the connections between artificial neurons in the network, allowing them to capture progressively more complex features in the data. [75].

Moreover, Recently, DNNs have demonstrated remarkable achievements in the medical field, such as the diagnosis and prediction of diseases such as Alzheimer's disease, heart disease, lung and liver cancer, kidney and brain cancer, and many more. In this context, we will focus on the role of DNN-based algorithms in the diagnosis and prediction of Alzheimer's disease [76], [77], [78]. Furthermore, it has demonstrated remarkable results for low-data drug detection despite the limitation of "out-of-domain generalization" [79]. However, the challenge of incompatibility among ANN-based technologies and the real world is addressed in this study [80].

The fragmentation of models, algorithms, tools, libraries, and frameworks leads to code and model reusability issues. [48], [4], [81]. We also highlight the lack of systematic approaches to enhance interoperability, reduce complexity, and explain mechanisms in artificial neural networks. [82]. Fig. 3 illustrates the basic architecture of the perceptron, as shown in part (a), in DNNs with three input neurons, four neurons in each hidden layer, and two neurons in output layers, as shown in parts (b) and (c) [64]; [65]; [66].

D. The Neuron

Equation 1 describes the calculation of the neural network's output value, or activation, which involves summarizing the activations of all neurons from the "previous layer" connected to the evaluated neuron, adding the neuron's bias, and applying the "activation function" to produce the final neuron activation.

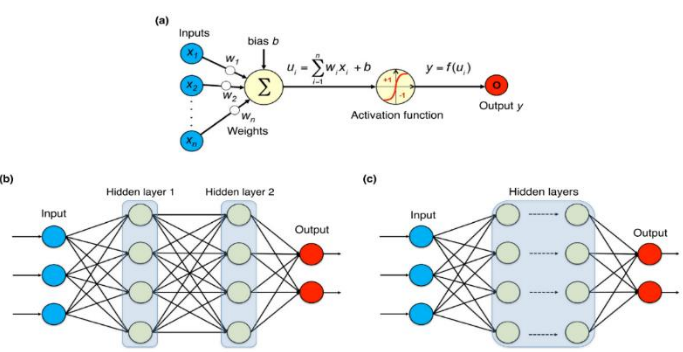

Fig. 3. The basic architecture of Multi-Layer "Perceptron" part (a) and the (DNNs) input and output layers, as shown in parts (b) and (c).

This process introduces non-linearity between inputs and outputs, ensuring the network's outputs are linear combinations of inputs. The process of adjusting weights and biases to achieve desired results is called training, where the weights and biases of the input neurons determine the network's output.

$$U_i = \sigma\left(\left(\sum_j^n U_j * w_{ij}\right) + b_i\right) \tag{1}$$

Where U_i : evaluated neuron activation, σ: activation-function, n: Set of input neurons linked to evaluated neuron, U_j: activation of input neurons from the previous layer, w_{ij}: weight of the connection between neurons, b_i: Bias value of evaluated neuron.

E. Transfer Learning

The transfer learning approach is a popular and effective DL technique that uses pre-trained models PTMs to address challenging issues [83]. However, choosing the best-trained model for a target job remains difficult as most properly training every candidate model is a high computationally cost and long inference time, highlighting the critical necessity for a convenient prediction metric based solely on early training outcomes [81], [3].

F. Model Development and Reusability Assessment

DNNs are novel techniques that have proven to be game-changers in realms such as healthcare, finance, transportation, and technology [2]. Applications include "image and speech recognition" (NLP), recommendation systems, etc. Their ability to learn from data and make predictions is invaluable when dealing with complex problem statements or automating mundane tasks [4], [64]. This section and its connected subsections present the main process of model development and the procedures for reusability assessment shown in Fig. 4.

1) Task selection: Specifying which problem your model will solve, e.g. classification, regression, or clustering.

2) Choosing data types: Depending on the nature of the problem to solve and the availability of sources, one must select the appropriate data types, which could be structured, semi-structured, or unstructured.

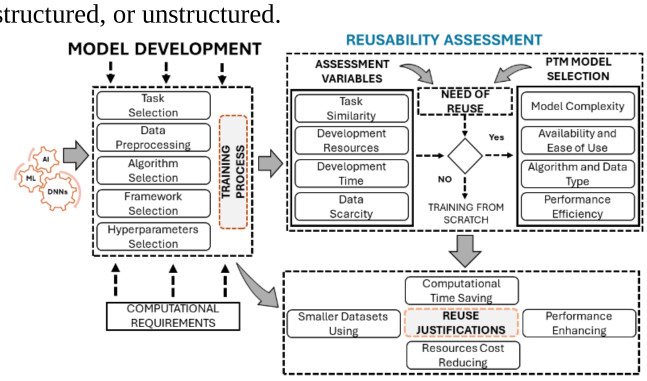

Fig. 4. The development architecture of AI, ML, or DNN-based models, the end-to-end process, and the reusability assessment for the candidate model.

3) Data preprocessing: It includes data cleaning, handling missing values, normalizing (in other words, scaling) features, and encoding them into labels for fitting onto a model.

4) Process steps: The process consists of data collection, preprocessing (sorting), feature selection, model development (training), validation, and testing.

5) Model selection: It involves selecting the appropriate model architecture, such as CNNs for image recognition or RNNs for “sequential data processing”.

6) Choose the framework and algorithms: From TensorFlow, PyTorch, or Keras, select the most suitable framework to meet the project's requirements and incorporate algorithms like backpropagation algorithm and gradient descent.

7) Feature extraction: This involves extracting relevant parts of the data using techniques (e.g. principal component analysis), dimensionality reduction, or feature engineering.

8) Model evaluation: Using metrics such as accuracy, precision-recall, and F1 score to determine how efficient an ML algorithm or model is.

9) Reuse and reengineering: We can reuse pre-trained models by fine-tuning them for new, similar tasks, adjusting the model's parameters, or retraining with more data.

Upon completion of candidate construction, the “cross-entropy loss” ($\mathcal{L}_{ce}$) between the “predictions of the target dataset” and the “actual labels” can be calculated by:

$$\mathcal{L}_{ce} = -\sum_{i=1}^{K} t_i \mathrm{Log}(P_i(\mathcal{M}, \mathcal{H})) \quad (2)$$

“K” represents “class number” for an intended issue, “$\mathcal{M}$” denotes the “mask,” while “$\mathcal{H}$” stands for the “head.” Conversely, “$(P_i(\mathcal{M}, \mathcal{H}))$” denotes the “prediction a candidate formulates for the class of “i-th” utilizing the “$\mathcal{M}$ mask” and the “head $\mathcal{H}$”. Also, the “t_i” signifies the probability of “the class in the single hot representation of the real label,” which can be either *0* or *1*. The “classification accuracy” of the “target dataset” is enhanced when the candidate retains a greater number of weights pertinent to the target problem and exhibits a reduced cross-entropy loss. The mask serves to directly calculate the weight retention rate ($\mathcal{L}_{wr}$) as illustrated below:

$$\mathcal{L}_{wr} = \frac{1}{L}\sum_{i=1}^{L} \mathcal{M}[i] \quad (3)$$

L represents the “number of weights in the original mode (3) “lower weight retention ($\mathcal{L}_{wr}$) rate”, suggests that the “candidate model” keeps “fewer weights,” whereas “objective function” O *is* defined based on $\mathcal{L}_{ce}$ and $\mathcal{L}_{wr}$:

$$\mathcal{O} = \mathcal{L}_{ce} + \alpha \times \mathcal{L}_{wr} \quad (4)$$

Where α is an empirically determined weighting factor of 1.0.

Importantly, to minimize the $\mathcal{O}$, function, some researchers suggest building a search-based technique that initially focus on find the expected “candidate model that maintains only the weights relevant to the target-related problem” to be solved. The idea is that this “candidate model” has the potential to accomplish the desired “classification accuracy” and “robustness” while maintaining the minimum weights.

G. Model Deployment

In the development section, we highlighted the process of developing a candidate model for a new task. We obtained a candidate model by following all the steps and procedures from the previous stage of development, as shown in Fig. 4. We then did a reusability assessment to make sure it was suitable for the new task and met all its requirements. As a result, the candidate model has been approved for reuse and employment in the new task. This will be achieved through reengineering and reproduction processes through the development of the new target environment, as illustrated in Fig. 5, and the subsequent steps:

1) Infrastructure setup: Deploying the model to the target environment and setting up all back-end infrastructure, such as servers, cloud services, or edge devices.

2) Deploy the model: This is where the trained models are integrated into the target environment, such as a mobile application, web service, or embedded system.

3) Test and validate: Testing the deployed model to ensure that it performs adequately in the target environment.

4) Monitoring and maintenance: Deploy techniques to monitor the model, update it when needed, and maintain ongoing performance support.

The search strategy efficiently explores large models with billions of parameters using a gradient-based method, finding new candidates with smaller objective function values each round and updating the mask accordingly. The updated mask, M1, represents a new candidate with a reduced objective function value, and the process of updating it involves dropping the gradient, as seen below: ξ represents the learning rate:

$$\mathcal{M}' = \mathcal{M} - \xi \times \nabla_{\mathcal{M},\mathcal{H}}\mathcal{O} \quad (5)$$

$$\nabla_{\mathcal{M},\mathcal{H}}\mathcal{O} = \nabla_{\mathcal{M},\mathcal{H}}\mathcal{L}_{ce} + \alpha \times \nabla_{\mathcal{M}}\mathcal{L}_{wr} \quad (6)$$

As a summary, this section outlines the steps involved in deploying the candidate model, which includes setting up the infrastructure for the source model and transforming it into an intermediate model using the intermediate representation (IR) technique, as shown in Fig. 4. The deployment also covers testing, validation, and maintenance after deployment to the target environment. The paper emphasizes the importance of sharing code and models to encourage innovation while also considering the potential consequences of incompatibility. The model's performance efficiency, accuracy, and reliability depend on data selection, preprocessing, and training methods. Successful deployment requires an understanding of all decisions made during the development process.

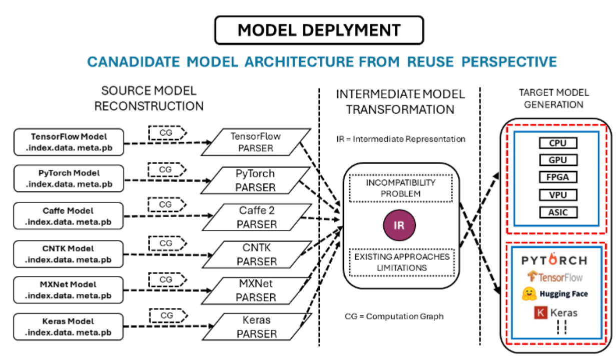

Fig. 5. The deployment architecture of AI, ML, or DNN candidate pre-trained models from source to target environment/s, the end-to-end converting process.

VI. Related Work

In the past decade, specifically the last seven years, there has been a surge in surveys, reviews, and varied studies focused on various applications of machine learning (ML), particularly in complex architectures such as (DNNs) [84], [71], [25]. These systems can potentially enhance multiple business functions and address varied organizational needs [25]. For example, leveraging deep neural networks and related models enables the recommendation of products according to (e.g., previous purchases and audience ratings) [85], [86], image recognition for video surveillance [87], identification of spam and malware emails [88], [89], healthcare applications (e.g. Alzheimer's prediction [78] and cancer prognosis and early detection [90]), among other applications explored by [91], [92]. However, given the novelty of transfer learning, comprehensive survey and review studies have been relatively rare, focusing on providing an overall understanding of this domain [83].

Most notably, relatively little research has been done into the difficulty of compatibility, which can present an insurmountable challenge when attempting to reuse AI systems [55], such as machine learning models and deep neural networks [17]. As far as we know, this systematic review is the first extensive attempt to look at the incompatibility problem through the lens of AI systems reusability. It is an area that has been woefully underexplored, and still, with a deep dive into the nuance of reusability [53] of these cutting-edge technologies, we hope to raise awareness about some of the major issues and factors that come into play [2]. A lack of guidance is available to assist organizations in developing these capabilities [93]. Consequently, our review aims to rectify this gap in literature and serves as a valuable guide to AI system reuse for researchers and practitioners. The integration of ML components into applications faces challenges due to the fragility of algorithms, framework fragmentation, and their susceptibility to changes in data, which can cause prediction shifts over time [94]. Consequently, mismatches between system components hinder the seamless integration of ML capabilities [71]. Thus, [69], [1], [95] Suggested that further research is needed to explore the unique advantages that arise from combining these technologies, particularly considering the increasing availability and complexity of big data characterized by its variety, volume, veracity, volatility, and velocity [96], taking into account understanding the synergy between AI systems and extensive data methods according to [92].

Therefore, this paper aims to fill this gap by comprehensively investigating the state-of-the-art machine and deep learning systems applied in analyzing the bid and complex data [96], including frameworks, models, algorithms, and libraries. It provides an in-depth discussion of current deep neural network technologies, covering their features, categorization, and classification. Additionally, it identifies critical open issues and outlines opportunities for further research in the advancement of big machine deep learning technologies.

VII. The Systematic Review

This review provides a valuable contribution to the research community by offering insights and opportunities for further advancement into the domain of artificial intelligence (AI), particularly advancements in machine learning (ML) technologies and deep neural networks (DNNs). It serves as a guide for researchers and developers aiming to achieve success in evaluating the whole environment of these technologies and networks and related software and hardware components, including the factors that impact the development and deployment process [12], [97]. The first review question (RQ1) seeks to explore the popular AI, ML, and DNN systems and conduct multiple comparisons in various aspects, such as their purposes and goals, tasks, and functions, attributes and characteristics, advantages and disadvantages, similarities and differences, and strengths and weaknesses [98], [99], [100], [101].

The review highlights the role of these systems in preprocessing, training, analyses, storage, and deploying massive and complex data. Multiple comparisons are presented to serve the purpose of this research [98], [39]. The second question (RQ2) specifically aims to identify the open issues and challenges associated with target systems under this study, elaborated in the following sections. The key issues identified include the data and learning complexity, coding from scratch, the dearth of benchmarks, and selecting the proper technology that matches the target task [102], [29], [4], [99], [44], [48], [17]. Through systematic review analysis of the literature, several factors and dimensions affecting these systems were identified, highlighting the significance of the issues closely connected to the primary research problem of incompatibility. [44], [24]. Consequently, understanding these challenges and related factors and dimensions can aid the researchers and the developers in overcoming them, thus achieving the desired results (e.g. prediction, classification, recognition, and detection) [103] (L. Liu et al. 2018b), [99]. The third research question (RQ3) focuses on the main research problem; the findings reveal that interdependencies and compatibility remain open research areas. Hence, further investigation and discussion are needed to find appropriate solutions. The fourth research question (RQ4) addresses the limitations of the existing common empirical approaches that deal with the current interoperability problem and its related challenges [13], [104], [49], [61], [92]. The answer to this question emphasizes the need for a more efficient novel unified method to enhance the compatibility between the source model (e.g., original pre-trained models "PTM") and target model (e.g., the destination model) [102], [48], [105]By addressing these aspects, our authors expect the research community to confidently make strides toward improved compatibility, thus reducing the computational sources' costs and time consumption.

VIII. Methods and Materials

The methods and materials of this study are comprised of three primary stages: the planning of the review, the conducting of the review, the "actual execution of the review," and the reporting of the review. The stages consist of many phases, actions, procedures, measurements, and instruments for achieving the desired findings and outcomes and optimizing the process.

A. The Review Planning

The planning review is a standard step in SLRs that entails two main sets of reviews and procedures. The first set includes formulating research questions, scoping the review, and drafting a review protocol, as shown in Fig. 6. On the other hand, the second set involves identifying the selection of strategy through the development of the (Inclusion/Exclusion) criteria that involve various factors that affect the ML and DNN's performance and reuse process, as presented in Fig. 7. The review process in both sets includes defining objectives, data sources, data extraction, data analysis, and data synthesis.

1) Need of this review: The main goal of this systematic review is to summarize the current evidence on machine learning (ML) techniques and deep neural networks (DNNs), point out gaps in current research on the development and deployment process, and lay the groundwork for new research projects, and explained previously in (Fig. 5 and 6).

2) Research question formulation: The research question formulation stage involves identifying essential elements of domains like AI, ML, and DNNs and addressing methodological aspects such as search, data extraction, and data analysis. The research questions should be specific enough to be feasible within the evaluation scope yet broad enough to make a significant contribution to the field. The research questions include identifying common characteristics and differences, unsolved issues, incompatibility problems, and limitations of existing methods. The goal is to answer these questions while reducing computational processes' high resource costs and time consumption. The review aims to provide a framework for future research and contribute to the fields of machine learning and deep neural networks.

3) Develop the protocol of the systematic review: This study uses a pre-established protocol to avoid researcher bias. It aims to provide a justification for the research, address specific inquiries, and use a systematic approach to identify relevant information. The data extraction strategy identifies variables and methods of interest. The synthesis strategy focuses on addressing primary research inquiries about factors and dimensions impacting the effectiveness of AI, CV, NLP systems, ML techniques, and DNNs. Sub-questions include data synthesis, success definitions, influencing direct factors, indirect factors, and dimension classifications.

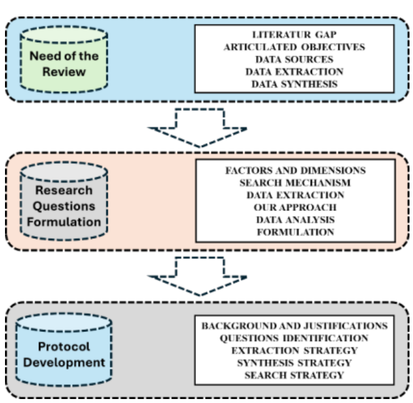

Fig. 6. The first set of review planning includes the need to conduct this review, the research question formulation process, and protocol development.

4) Selection of strategy (inclusion/exclusion criteria): The selection strategy for a review of AI, CV, NLP, ML, and DNN-based systems was based on a series of inclusion and exclusion criteria and derived from research questions and the quality assessment process. The criteria included study focus, empirical studies, evaluation, impact factor, language, theories, publication focus, domain participants, quality control, and replication studies. The emphasis was on English-language studies, focusing on the period from 2015 to 2024 and assessing credible journals.

This review excluded studies that lacked explicit information, publications of low quality, research that focused on artificial intelligence (AI), machine learning (ML), deep-learning/deep neural networks, big and complex data, parallel computing, or technology-based approaches, research that lacked methodology, numerical test findings and analysis, and studies published before 2015 because they were not relevant. Furthermore, we excluded duplicate studies and methods irrelevant to the main research problem.

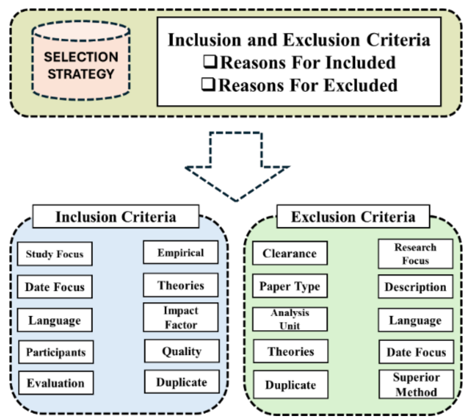

Fig. 7. The second set of review planning involves identifying the strategy selection through the development of inclusion (Inclusion/Exclusion) criteria.

The study focused on addressing the incompatibility observed among AI, CV, NLP, ML, and DNN-based systems, focusing on existing unresolved concerns and constraints. The review included publications from reputable peer-reviewed journals with impact factors and high-quality international scientific conferences.

B. Conducting the Review

The systematic literature review (SLR) is conducted using a well-defined search strategy. Standard electronic databases are the primary source for high-quality primary research, but alternative methods like browsing the internet, seeking advice, and snowballing techniques can also be beneficial. The strategy of search covers selecting the source of data and a search-strings formulating. To determine “search strings”, the authors identify key terms related to machine deep learning software, examine abstracts and titles of selected primary research, and use Boolean operations "OR" and "AND" to construct a comprehensive list of related words. To enrich the search process, we also employed subsequent strings of searches related to the main keywords and terms. Fig. 8 illustrates a sample of the terms and keywords used.

1) *Data sources:* The study systematically gathered data from a variety of sources, including highly indexed international scientific conferences and impact-factor journals. Literature selection was based on relevance to AI, ML, and DL (DNNs), as well as related frameworks, models, and algorithms. We used snowballing and backward search techniques to identify additional relevant studies [106]. From a theoretical perspective, the study examined the main concepts, remarkable successes, and achievements of AI, machine learning, and deep neural networks. It also examined related models, algorithms, and frameworks, as well as the related limitations, open issues, and existing challenges. The used theories included Artificial Intelligence, Machine Learning Theory, Deep Learning and Complexity Theory, Computation Theory, Theory of Programming, Transactive Memory Theory, Structured Process Modelling Theory, Coding Theory, Data Science Theory, and Transfer Learning and Reuse Theory.

2) *Evaluating eligibility and research quality:* It is crucial to evaluate the eligibility and relevance of the primary studies identified in the preceding stage, in addition to the inclusion and exclusion criteria. [107]. As per the instructions outlined in the reference [108], we evaluate the eligibility and quality of each study following the factors, including the significance of the study, the quality of the results and analysis, and potential future research guidelines or discoveries, using the criteria presented in Table I. We evaluate the articles and select studies that exhibit exceptional quality. We have formulated nine questions; each indicated with either (Y)/(Yes) or (P)/(Partly), or (N)/No, to evaluate the eligibility and quality. The questions and their corresponding answers were included. The assigned values for scoring are as follows: Y = 1, P = 0.5, and N = 0. Additionally, each primary study should receive a score ranging from 0 to 13 points. A quality appraisal guarantees that the review will only include the research that is most trustworthy and pertinent to the topic at hand.

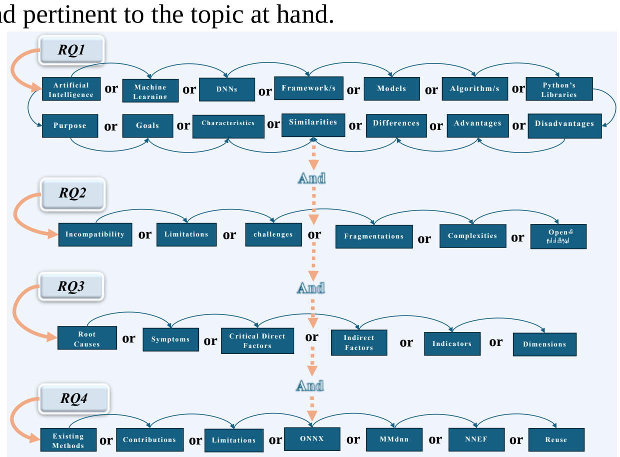

Fig. 8. A list of frequently used terms/strings and keywords on AI, CV, NLP, ML, and DNN-based systems using the Boolean operations "OR" and "AND."

TABLE I. ELIGIBILITY AND QUALITY EVALUATION FORM USING THE "Y-P-N" SCORE

Quality Assessment Questions	Recorded Score
Do the articles provide enough information on "AI, CV, NLP, ML, and DNN" from a theoretical perspective about frameworks, models, and algorithms and their environments?	"Y-P-N"
Does the research study offer a comprehensive comparison between these highlighted techniques?	"Y-P -N"
Does the research paper provide a clear methodology/technique to select proper measurements to evaluate the different aspects of the "AI, CV, NLP, ML, and DNN" domains?	"Y-P -N"
Does the review study present factors, indicators, and dimensions that positively promote the success of "AI, CV, NLP, ML, and DNN" domains?	"Y-P -N"
Do the review papers highlight forward propagation and backpropagation for deep neural networks (DNNs)?	"Y-P -N"
Do the papers efficiently identify the research gaps and well formulate the problem statements?	"Y-P -N"
Do the target studies provide an obvious explanation of the main research problem and causative factors?	"Y-P -N"
Do the target studies provide an explanation and understanding of the related issues challenges and causative factors?	"Y-P -N"
Does the article provide strategies to mitigate or overcome the current DNN issues and challenges?	"Y-P -N"
Does the research article explain and understand the concepts of reuse, transfer learning, and reengineering?	"Y-P -N"
Do the articles provide an explanation, comparison, and understanding of the limitations of the current approaches and techniques?	"Y-P -N"
Are the investigated study outcomes applicable and/or generalizable?	"Y-P -N"
Is the retrieved data appropriately described?	"Y-P -N"
Is the description of inclusion/exclusion/criteria in the study sufficient?	"Y-P -N"

3) *The extraction of data:* Data extraction in a systematic review is a crucial step in evaluating research potential and combining results. As depicted in Table II, manual data extraction can be laborious and expensive, but sophisticated software can help. In this review, the authors developed a data extraction form to collect primary study information and simplify analysis. Two steps were involved: a preliminary study and a second extraction from a random selection of papers. The data extraction was conducted using tools such as Microsoft Excel, REDCap, and Google Sheets. The intake form included columns for each research question and quality assessment items.

The chosen primary studies provided answers to specific or all the issues raised by the investigations. This systematic and standardized procedure ensures consistency and efficiency in data extraction. Among the results of the 1052 studies, we extracted information that was relevant to each of the four study questions and objectives. Following the extraction of the data, we proceeded to examine and discuss it to accomplish the intended objectives of this SLR exploration.

TABLE II. MAPPING THE INTEREST OF SEARCH TO THE RESEARCH QUESTIONS

Interest of Search	Extracted Data Description
RQ1	"Characteristics, fragmentation, common differences, similarities, advantages, disadvantages, systems, models, libraries, frameworks, methods, tools, techniques, applications, practices, operations, algorithms, and complexities"
RQ2	"Limitations, problems, open issues, current challenges, bottlenecks, processing, complexities, mechanisms, software, fragmentation, diversity, lacking, reusing AI, CV, NLP, ML, DNN, RNN, CNN, (BD), and (DS)."
RQ3	"Preprocessing, learning, training, accuracy, adversarial, robustness, quality, inference time, model size, memory consumption, efficiency, and performance"
RQ4	"Existing methods limitations, transfer learning, reuse, reengineering, conversions, compilers, converters, techniques, accelerators, future trends, and opportunities"
Research Domain	"Systems of AI, CV, NLP, ML, DNN in terms of characteristics, opportunities, and existing challenges.
Utilized Techniques	"Ms. Excel/(spreadsheets), Google/Sheets, and Google/Collab"
Employed Methods	"Whether it is a new method, modified, or hybrid approach"
Relevant Data	"Title/Abstract/Keywords/Authors", "scientific databases/venue"
Publication/ Year	"Included papers from "2015 to 2024" substantially"
Publication/ Type	"Journal, conference, or book chapter"
Study Type	"Whether it is an analysis, survey, or a combination of both"

Deep neural networks (DNNs) have made significant progress in large dataset applications, but they have faced criticism for their ambiguity. This systematic review of AI, machine learning, and DNNs covers conceptualization to dissemination, with a detailed research procedure in six phases. The revised study mapping process follows PRISMA and Kitchenham guidelines, with research questions achieving SLR objectives [109] and kitchenham [110].

C. Phase 1: Broad Preliminary Searching

The preliminary study used Google Scholar to search for relevant material related to AI, CV, NLP, ML, and DNN-based systems. It aimed to derive research questions and establish a variety of search strings and keywords. The preliminary scan of many research papers is conducted for general outlook purposes, facilitating the formulation of the research question. The researcher of this study identified search criteria and selected six eminent scientific sources and submission venues to obtain the most associated studies, identifying 1052 research papers presented in detail in Fig. 9.

D. Phase 2: Identifying Research Criteria to Refine the Search

The search strategy involved selecting keywords/terms related to AI, CV, NLP, ML, and DNN-based systems, Fig. 9. The search string was adjusted for each platform to ensure a systematic search. The review encompassed a variety of criteria and parameters, such as the title, background, review methods, studies included and excluded, results, review questions, discussion, authorship and acknowledgments, citations and references, and appendices.

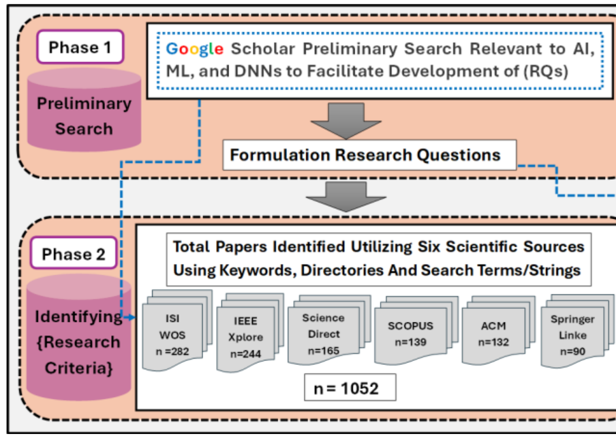

Fig. 9. Phases 1-2. The research question development employed the preliminary search and terms/strings according to the PRISMA and Kitchenham guidelines.

E. Phase 3: Applying the (Inclusion/Exclusion Criteria)

The review phase involved reviewing articles based on inclusion/exclusion criteria, selecting 691 papers related to research questions, while 361 were filtered out, as shown in Fig. 10. The title should be clear and informative, identifying the review's subject and demonstrating its classification as a systematic review.

The review methods should clearly describe the search strategy, the source of data, the study criteria selection, the quality and eligibility assessment methods, and the approach to extracting findings. The discussion section should address the strengths and limitations of evidence, practical implications of findings, research gaps, and areas for future work. The authorship and acknowledgments section should set clear authorship criteria and acknowledge the contributions of individuals or institutions who substantially affected the research but did not meet the author's standards.

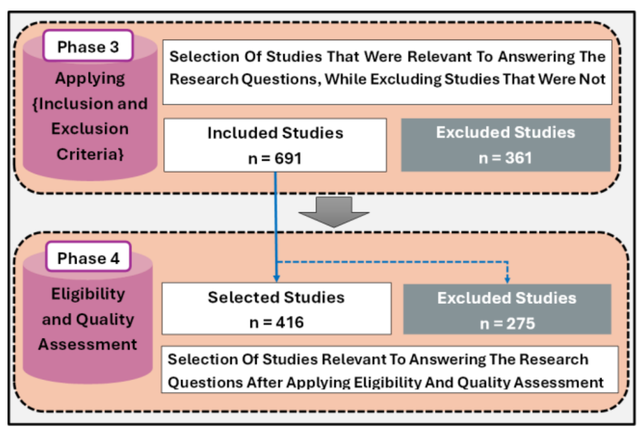

Fig. 10. Phases 3-4. Applying inclusion/exclusion criteria and eligibility/quality assessment according to the PRISMA and Kitchenham guidelines.

F. Phase 4: Evaluating Eligibility and the Quality of Resources

The study critically analyzed 691 selected studies, assessing their suitability and validity. As presented in Fig. 10, we selected 416 studies, ensuring all methodologies were methodologically appropriate and relevant, and omitted 275 papers due to insufficient eligibility and quality standards.

G. *Phase 5: Examining Abstracts and Key Sections*

The review process involved a thorough review of 416 papers, focusing on the abstract, primary findings, and conclusion sections, as shown in Fig. 11. The process involved thoroughly examining the articles and identifying those worth further examination. We considered 297 papers for further examination, eliminating 119 due to their unsuitability. The abstract should be 250 words or less long and written in the true style of creative nonfiction.

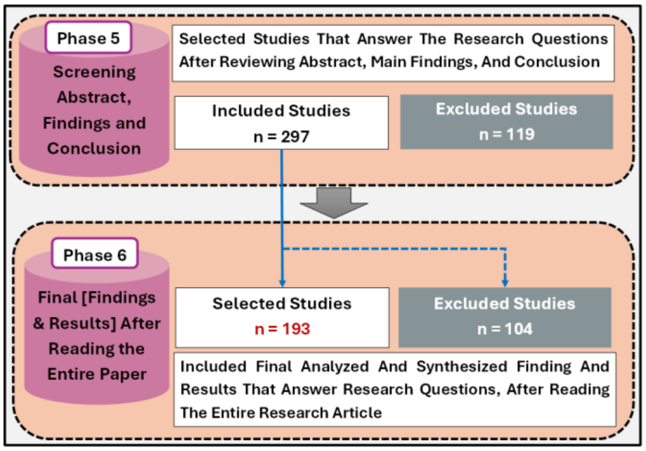

Fig. 11. Phases 5-6. Examining abstracts and key sections and integrating the analyzed selected studies that answer research questions

The findings should include a description of the primary studies, quantitative data, meta-analysis stipulations, a narrative summary of the key findings, and data presentation using tables and diagrams. The conclusion should summarize the main findings, their implications for practice and policy, and suggest suggestions for future research.

H. *Phase 6: Further Analysis and Integration*

The review comprehensively decoded and comprehended the 297 individual studies in Phase 5, focusing on AI, ML, and DNNs. After presenting the advantages, drawbacks, and existing research issues, the review identified 193 qualitative and quantitative studies. Ultimately, after eliminating 104 studies that were not relevant, the final net tally of the papers is 193. The systematic review report should be formatted as shown in Fig. 11.

IX. FINDINGS

This research is a crucial contribution to advancing deep neural networks, aiming to provide insights and guidelines for advancing the repurposing and adaptation of the DNN model. The systematic review used a six-phase robust approach to search for and extract research on domains of artificial intelligence AI, machine learning ML, deep neural network DNN, and their state-of-the-art techniques.

This study used the search string technique to refine the search area and cross-reference the references and citations of the included studies, thereby identifying additional research, as we can understand from the illustrated information in Table III and the chart in Fig. 12.

TABLE III. A SIX-PHASE ROBUST APPROACH TO CARRYING OUT A SYSTEMATIC REVIEW WAS USED TO SEARCH FOR THE REQUIRED TERMS/KEYWORDS

(Search Source)	Phase 1	Phase 2	Phase 3	Phase 4	Phase 5	Phase 6	Percentage (%)
(Google Scholar)	Preliminary Search Relevant to AI, ML, and DNNs to Facilitate Development of Research Questions						
ISI-WOS	X	282	193	120	84	54	28%
IEEE	X	244	165	101	72	47	24%
Science Direct	X	165	108	66	50	32	17%
SCOPUS	X	139	87	51	37	25	13%
ACM		132	81	46	32	21	11%
SpringerLink	X	90	57	32	22	14	7%
Total	X	1052	691	416	297	193	100%

A. *Phase 1: Groundwork and Theory Building Around Research Questions*

Following up on the above essential explanation sections, this phase involved conducting an exploratory Google Scholar search to identify research questions and keywords associated with AI, ML, and DNN and their related applications. The study found that ISI-WoS is the most frequently accessed source of published literature concerning these systems and their associated challenges, followed by IEEEXplore, ScienceDirect, Scopus, ACM, and SpringerLink.

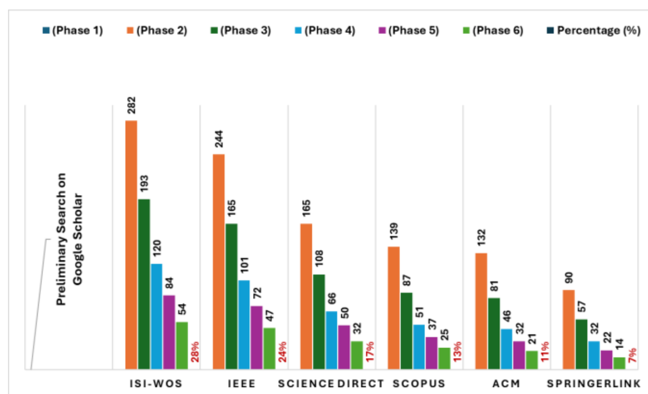

Fig. 12. The intensity of publications for AI, CV, NLP, ML, and DNN-based systems in each scientific database.

B. *Phase 2: Identification of Studies with Search Strings: (Narrowing Down the Search Parameters)*

Then, we refined the search criteria by defining the acceptable characteristics. This helped us to gather some knowledge of the basics of AI, including ML and DNNs. Deep Neural Networks (DNNs) DNNs are a potent technique that learns multiple layers of data representations or features and yields impressive prediction performances.

The findings identified a total of 1052 publications from six selected scientific sources and databases, simulating a deliberate, non-random selection: ISI-WOS (n=282), IEEE Xplore (n=244), Science Direct (n = 165): SCOPUS (n = 139), ACM (n=132 papers) Spring-Link ((n=90). These databases allow researchers to search thousands of leading academic journals, periodicals, and conference proceedings.

C. Period of Search

The period focused on in this research included papers from "2015 to 2024" substantially, as depicted in Table IV and Fig. 13.

D. Phase 3: Inclusion/Exclusion Criteria and References Validation

Among the identified studies (n=1052) from the initial search, our reviewers independently assessed inclusions based on predefined criteria. After removing 361 studies that did not meet the requirements (e.g. research questions answers), we were left with 691 relevant to our research questions. We then assessed all eligible studies' references to discover other potentially appropriate reports.

TABLE IV. THE COVERED (TIME-PERIOD) (FROM 2015 TO 2024)

Publication Year	Number of Publications	Percentage (%)
2015	53	5%
2016	69	7%
2017	74	7%
2018	79	8%
2019	88	8%
2020	96	9%
2021	109	10%
2022	121	12%
2023	187	18%
2024	167	16%
Total	1052	100%

E. Phase 4: Evaluation of Eligibility and Adherence Following Reference Review

This phase involved thoroughly evaluating 691 studies and assessing their relevance and quality. We selected 416 studies for their methodological solid foundation and relevance and excluded 275 due to non-compliance with eligibility and quality criteria.

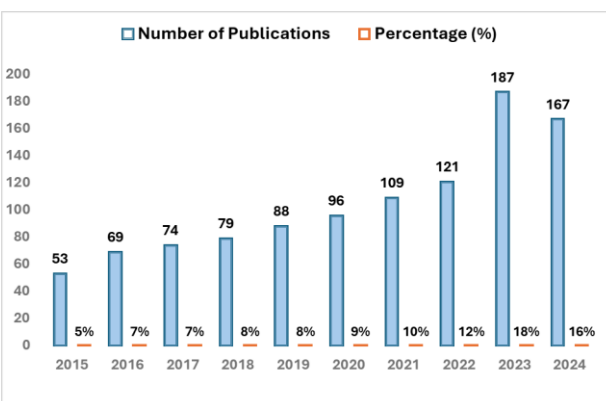

Fig. 13. Overall distribution of AI, CV, NLP, ML, and DNN-based systems from reuse and reengineering perspectives - (Year-Wise).

F. Phase 5: Screening Abstract, Findings, and Conclusion

In this phase, we filtered selection by analyzing abstracts, findings, and conclusions, excluding 119 out of 416 studies for review. This resulted in 297 applicable studies and complete data extraction.

G. Phase 6: Advanced Data Analysis and Answering the RQs

This systematic review concluded with a large set of relevant, high-quality studies that were carefully reviewed, investigated, and synthesized. The study comprehensively analyzed 297 selected studies on AI, ML, and DNNs, excluding 104 irrelevant ones. A structured six-phase process identified and analyzed 193 studies, offering insights to address the study's concerns, as shown in Table V and Fig. 14.

TABLE V. THE ULTIMATE RESULT OF THE REVIEWED PUBLICATION HAS BEEN FOUND AND FILTERED TO BECOME (N = 193)

Publication Year	Number of Publications	Percentage (%)
ISI-WOS	54	28%
IEEE	47	24%
Science Direct	32	17%
SCOPUS	25	13%
ACM	21	11%
SpringerLink	14	7%
Total	193	100%

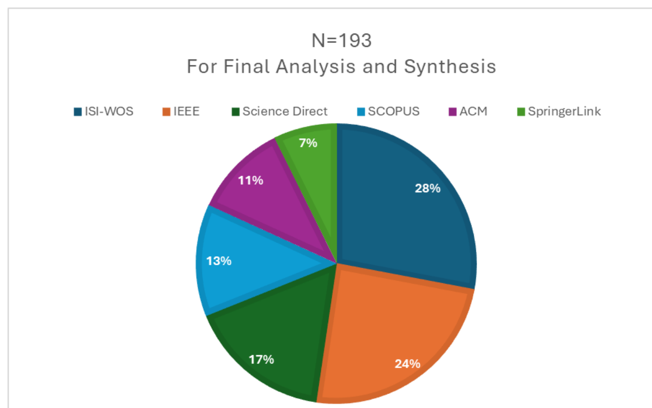

Fig. 14. The scrutinized publication yielded a result of N = 193 after comprehensive analysis and synthesis.

H. Mapping Results and Findings to Research Questions

1) Link each research question to a related dated journal/conference: The systematic review used an iterative design to connect each study to research questions, focusing on RQ1-RQ4. This study developed a new method to classify papers based on the most effectively addressed research questions. Table VI provides details on the mapping process.

TABLE VI. MAPPING PROCESS LISTS THE MOST HIGH-INDEXED REFERENCES PROVIDED ANSWERS TO RESEARCH QUESTIONS

Ref.	Year	Type	Indexing	Cited	Q1	Q2	Q3	Q4
[111]	2019	Journal	ISI/WoS	59	√	√	√	
[99]	2020	Journal	ISI/WoS	58	√	√		√
[6]	2018	Journal	Scie-Dir	56	√		√	
[24]	2020	Journal	Springer	51	√	√	√	
[77]	2021	Journal	Scie-Dir	47	√		√	√
[112]	2020	Journal	Scie-Dir	46	√	√	√	
[113]	2018	Journal	ISI/WoS	40	√	√	√	√
[114]	2019	Journal	ISI/WoS	34		√	√	
[115]	2019	Journal	ISI/WoS	32	√		√	√
[116]	2021	Journal	Scopus	30	√	√	√	√
[117]	2021	Journal	Scie-Dir	30		√	√	
[118]	2018	Conf	IEEE	28	√	√	√	√
[25]	2019	Journal	ISI/WoS	27	√	√		
[119]	2018	Journal	ISI/WoS	26		√		√
[64]	2017	Journal	ISI/WoS	25		√	√	√
[120]	2020	Journal	ISI/WoS	21	√		√	
[73]	2018	Journal	Scie-Dir	20	√	√		√
[121]	2017	Journal	ISI/WoS	19	√	√	√	
[122]	2019	Jr/Cnf	S-D/ACM	18	√		√	√
[123]	2020	Journal	ISI/WoS	17	√	√	√	
[124]	2019	Conf	IEEE	15	√	√		√
[125]	2020	Journal	Scie-Dir	13	√		√	
[126]	2018	Journal	IEEE	12		√	√	√
[127]	2020	Journal	Scie-Dir	11	√	√		
[128]	2019	Conf	IEEE	10	√		√	√
[129]	2020	Journal	ISI/WoS	10	√	√		
[48]	2019	Journal	Springer	9		√	√	√
[130]	2021	Journal	Scie-Dir	9	√	√	√	
[131]	2021	Journal	ISI/WoS	8	√			√
[132]	2018	Conf	IEEE	8	√	√	√	
[133]	2020	Journal	Springer	7	√	√	√	
[134]	2019	Journal	ISI/WoS	7	√		√	√
[135]	2018	Journal	WoS	7	√	√	√	
[136]	2018	Conf	IEEE	6	√	√		√
[137]	2017	Journal	Scie-Dir	6	√	√	√	
[138]	2019	Journal	Scie-Dir	5	√		√	
[139]	2020	Journal	Scie-Dir	5		√	√	√
[44]	2017	Conf	ACM	5	√	√	√	√
[140]	2019	Conf	ACM	5	√		√	
[141]	2018	Conf	WoS/SCIE	5	√	√		√
[142]	2020	Journal/Co	Scopus	4	√	√	√	√
[143]	2017	Conf	IEEE	4	√		√	√
[144]	2020	Journal	Scopus	4	√	√		
[145]	2020	Journal	WoS/SCIE	4	√		√	√
[72]	2019	Journal	WoS/SCIE	4		√	√	√
[146]	2019	Conf	IEEE	4	√	√		
[147]	2020	Journal	Scie-Dir	4		√	√	√
[148]	2018	Journal	IEEE	3	√		√	
[69]	2017	Conf	IEEE	3	√	√		
[149]	2019	Conf	IEEE/ACM	3	√	√	√	√
[13]	2020	Conf	ACM	3	√		√	√
[62]	2021	Conf	IEEE/ACM	3	√	√	√	√
[30]	2023	Conf	IEEE/ACM	2	√	√		√
[49]	2022	Conf	IEEE	2	√		√	
[109]	2022	Journal	WoS	2		√		√
[54]	2023	Symposium	IEEE	2	√		√	
[23]	2023	Conf	IEEE/ACM	2		√	√	√
[150]	2024	Journal	Scopus	2	√		√	
[61]	2024	Journal	WoS	2		√		√
[10]	2024	Journal	Scie-Dir	2	√		√	

2) *The most addressed research questions:* As depicted in Table VII, the most addressed research question in the reviewed studies was RQ3, which focuses on the characteristics, differences, similarities, advantages, and limitations of AI systems, ML techniques, and DNN networks. However, RQ4, which addresses the current interoperability challenge among AI environments, was the least addressed, with only 35 papers providing answers. The research papers that answered RQ4 were average and insufficient, highlighting the need for further research into AI system issues. Despite the fewer papers addressing RQ4, some notable contributions include new methods for distributed training and parallel computing.

TABLE VII. PERCENTAGE (%) OF STUDIES COVERING EACH RESEARCH QUESTION

Question No	Question-Text	Number of Publications	Percentage (%)
"RQ1"	"What are the main characteristics and "common differences, similarities, advantages, and limitations" of AI systems, ML techniques, and DNN networks regarding purpose, tasks, and operations?"	54	28%
"RQ2"	"What are the key unsolved problems and bottlenecks that AI computational environments face?"	45	23%
"RQ3"	"What are the relevant "factors and dimensions" that impact the accuracy, adversarial robustness, memory consumption, and inference time of AI, ML, and DNN tasks?"	59	31%
"RQ4"	"What convenient approaches can be developed to address the existing issues and reduce the computation process's high resource cost and time consumption?"	35	18%
Total	Four questions	193	100%

I. *Characteristics of (AI) Systems, ML Technologies, and DNNs and Answering RQ1*

This systematic literature review comprehensively compares popular machine learning and deep learning frameworks (e.g. TensorFlow, Caffe-Caffe2, Torch/PyTorch, Theano, MXNet, and CNTK) and their connected models and algorithms, as illustrated in Table VIII, towards answering RQ-1. It highlights their relevant characteristics in various aspects and demonstrates the significant disparity in compatibility between frameworks, with excellent compatibility for models from the same ecosystem. However, compatibility diminishes when combining models from different ecosystems, emphasizing the importance of considering framework compatibility during model development.

TABLE VIII. A COMPREHENSIVE COMPARISON OF (AI) SYSTEMS, ML TECHNOLOGIES, AND DNNS AND ANSWERING [RQ1]

Framework Factors	Caffe/Caffe2	TensorFlow	Theano	Torch/PyTorch	CNTK	MXNet
Founder	Learning-Center & Berkeley 2013/caffe 2017/caffe2	Google2015	Montréal-Uni (2007)	(Facebook)2016	Microsoft Research 2016	Apache Foundation 2015
Purpose	Image processing & on both CPU and GPU	High-performance for ML, DL, CPU, GPU, and TPU	Python offers GPU for arrays.	A Python-based for DNNs	For DL, graph representation	A DL for DNNs.
Interface	Python, MATLAB	Python (Keras), C/C++, Java, Go, &R	Python	Python, C++, & Julia	Python, C++,	C++, Python, Julia, Matlab,
Core language	C++	C++	Python	Python, C, C++, CUDA	C++	Small C++ core library
Platform	Linux, Mac OS X, Windows	Linux, Mac OS X, Windows	Cross-platform	Linux, macOS, Windows	Windows, Linux	Linux, Mac OS X,
CUDA Support	Yes	Yes	Yes	Yes	Yes	Yes
Multi-Threaded -CPU	Yes (BLAS)	Yes (Eigen)	Yes	Yes (widely Used)	Yes	Yes (OpenMP)
Pretrained Models	Yes	Yes	Partially	Yes	Yes	Yes
Multi-GPU	Yes (only data parallel)	Yes (Most flexible)	Not perfect	Yes	Yes	Yes
Compatibility (within the same environment)	Yes	Yes	Yes	Yes	Yes	Yes
Compatibility in	No	No	No	No	No	No
Advantages	"Fast, Efficient for CNNs"	"Large Community, Flexible"	"Symbolic, Good for Research"	"Imperative, Pythonic"	"Production-ready"	"Scalable, Efficient"
Disadvantages	"Limited framework support"	"Can be complex."	"Slow execution"	"Limited community"	Win limited	"Learning curve"

J. Challenges and Open Issues and Answering of RQ2

1) Overview: The systematic review highlights several challenges facing AI, ML, and deep neural networks (DNNs) and their related frameworks, models, and algorithms. These issues include system fragmentation, complexity, interpretability, bias, security, privacy, and a lack of benchmarks. System fragmentation affects code and model reusability, while complexity involves data, model size, algorithm complexity, high dimensionality, and characteristics of tensor data used in DNNs. Understanding how machine-learning techniques and DNNs make decisions like black boxes makes interpretability challenging. Bias can lead to unfair and discriminatory outcomes, while security is vulnerable in specific fields like autonomous vehicles and financial institutions.

2) Common Challenges

- ***Interpretability:*** Understanding how machine learning (ML) and deep neural networks (DNNs) make decisions as black boxes can be challenging. Trusting the procedures throughout the learning and training process can be challenging without reliable indicators until each iteration, epoch, and batch size yields the desired production results (e.g. prediction accuracy before and after model deployment) [151].

- ***Bias:*** Machine learning and deep learning models can be biased, which can lead to unfair and discriminatory outcomes [152].

- ***Security:*** Machine learning systems, especially in some fields such as autonomous vehicles and financial institutions, can be vulnerable to security and safety attacks [153], [154], [155].

- ***Privacy:*** The intelligent systems discussed in this review can collect and store large amounts of personal data, which raises privacy concerns [154], [156]. This issue is more challenging when dealing with deep neural (DNN) models, as these models need a huge amount of data that is not readily provided, especially in applications that require lots of sensitive data [156]. Moreover, Mendelson and Avi have demonstrated that privacy has been a hot topic, particularly regarding the need for privacy protection [154]. Since then, that has translated into new legislation affecting our rights regarding how they get data collected on them. Furthermore, the authors have implemented new mechanisms to guard against the potential misuse of this information. However, the current initiatives are neither adequate nor sufficient.

- ***Lack of benchmarks:*** Benchmarks are essential for comparing the performance of different AI environments and identifying areas where improvement is needed. According to [29], there is a lack of standard benchmarks for evaluating the machine-learning perf (ML) and DNNs performance for the diver's data, especially "big data analytics" [157], [158]. The need for more benchmarks within these technologies is a particularly significant challenge due to the need for standard benchmarks [159]. Furthermore, several surveys indicate that few discussions have been held about the community's efforts to implement deep learning benchmarking and the gaps between what has been achieved and what still needs to be added [160]. It would be a valuable contribution to the field of research, and developers would pay more attention to this challenge for industry-standard benchmark development [159], [29].

- ***Difficulty in choosing** between the machine learning techniques:* In general, the selection of the proper machine learning and deep neural methods for a given task can be a challenge [44], [99]. Based on the analyzed studies, the systematic review identified factors that affect the performance of machine learning and deep neural systems, such as the type of selected data, the quality and quantity of the training data, the kind of algorithm, model selection, the computational resources availability, and the appropriate hyperparameters and fine-tuning for the intended task [48]. Moreover, the findings have expounded various implications for researchers and practitioners. Researchers should develop new methods to address the open issues and challenges identified in the analyzed and synthesized SLR. On the other hand, practitioners should be aware of the involved factors that affect the computation performance and factors that should be applied carefully to select the proper techniques (e.g. Framework, pre-trained model, algorithm, and other software components) and hardware requirements (e.g. Computer device, memory size, and accelerators such as CPUs GPUs, and TPUs) for their given specific needs.

3) Learning and Training Challenges in DNNs

- ***Insufficient Training Data:*** Low training data or batch size can hinder stable, generalizable solutions. Ensure that there is sufficient data for the local minimum cost function.
- ***Non-deterministic data:*** Breaking the deterministic relationship between training inputs and target outputs can hinder training convergence.
- ***Wrong Learning Rate:*** Small or large learning rates can slow or make training unstable.
- ***Vanishing Gradients:*** Gradients in particular layers can become small, slowing or stopping training. For other solutions, use different activation functions.
- ***Exploding Gradients:*** Gradients can become large and cause instability, limiting their weight and bias change rate.

K. The Critical Direct and Indirect Factors Impact DNNs and Answering RQ3

This study examined various factors that affect the DNN-based system’s performance and robustness. These factors are critical, direct, and indirect, as explained in Fig. 15.

Direct Factors	Definition	Indirect Factors	Description
Fragmentation	Each system has its own features and characteristics.	NCHW/NHWC Formats	Different layouts of data in the memory hierarchy. The algorithm must be transposed to lead to poor performance.
Structure	Each system has its own structural design.	Data Noise	Mislabeled cases, redundant data, borderlines, or errors in the values of attributes.
Complexity	Model Complexity, Algorithm Complexity, Data/Tensor, and Tensor Layout (Curse of Dimensionality).	Outliers	It is a broader concept that contains discordant data
Purpose	Framework Goal, Model Goal, Algorithm Goal, Task Type (e.g., Prediction, Classification, Detection, or Recognition).	Data Size	Data size: (Not sizable).
Programing Language	API, core language, and Frontends/Interface (C++, Python, Julia, Matlab, JavaScript, Go, R, Scala, or Perl)	Imbalance Data	One or some of the classes greater than others result in biased outcomes.
Accelerators	CPUs, GPUs. TPUs, FPGA, VPU, and ASIC	Missing Features	Variables lack information points, resulting in harming accuracy and model dependability.

Fig. 15. Critical direct and indirect factors impact the DNNs reuse and relevant reengineering challenges (n1=6 direct, n2=6 indirect factors).

The direct factors include data quality, structural design, complexity, creation purpose, programming language, and accelerators. On the other hand, indirect factors such as noise, outliers, and imbalances can impede DNNs' ability to learn program scale and dependencies. By carefully designing and addressing these indirect factors, DNNs can receive training on high-quality data for more accurate and reliable results. We can use deep learning accelerators like GPUs, TPUs, FPGAs, VPUs, and ASIC chips to meet real-time requirements. Therefore, our results demonstrate that researchers and developers should focus on developing more efficient methods to address the highlighted challenges. At the same time, practitioners should be aware of factors affecting computation performance and carefully select the proper techniques and hardware requirements for their specific needs.

L. Existing Limitations on DNNs and Answering RQ4

1) Overview: A potential way for companies and the computer industry to amortize costs is by reusing and re-purposing DNNs. However, reusing DNN-based systems face challenges that affect their reuse and reengineering performance, efficiency, and robustness. The direct and indirect factors behind these challenges have been highlighted in previous sections. The structural problem of non-interoperability among DNN-based systems is one of these obstacles, along with a lack of technical skills and engineering methods [3].

In deep neural cutting-edge systems, interoperability refers to the ability of one software component (with which models are associated) to exchange algorithms and data with another [80]. Fig. 16 illustrates the various paths in developing and deploying a model on hardware or a target environment.

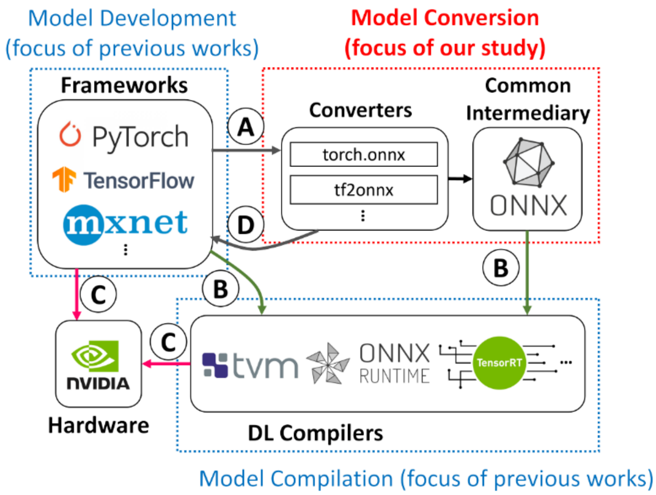

Fig. 16. The paths among model development, hardware deployment, DL Compilers, and model conversion [161].

Interoperability lets models be reused across frameworks and deployment scenarios. While (A) signifies model conversion into a standard intermediary format, letter (B) indicates the compilation process. On the other hand, (C) represents model deployment, and (D) describes transforming a model into a framework according [161].

2) *Failure of conversion:* This software ecosystem mainly includes many dedicated frameworks for deep learning [162], [163] and compilers [164], [105]. The proliferation of these different tools for development and deployment represents a significant obstacle since moving algorithms across frameworks can be difficult. Furthermore, a compiler that understands models from various frameworks is very heterogeneous. Consequently, frameworks such as open neural network exchange (ONNX) are developed to play an essential role in mediating between deep learning software [165].

However, translation errors during conversion can result in distortions and affect the quality of the model, potentially leading to a degradation in prediction, impacting deployment efficiency, and ultimately causing conversion failure.

3) *Failure symptoms and root causes:* Even with the development of modern interoperability solutions such as "ONNX, MMdnn, CoreML, and NNEF," there remains a knowledge gap regarding the failure type of deep learning model conversion errors. This lack of knowledge exposes a significant risk with various interoperability technologies built into the design.

The study findings have divided the failure symptoms into two categories: The main common failure symptoms and relevant practical root causes of reusing and reengineering DNN-based techniques, as elaborated in Fig. 17. The second category is failure symptoms that arise from the structural and procedural root causes when reusing and reengineering DNN systems, as presented in Fig. 18.

- Main and Common Failure Symptoms

Model Size: In memory-constrained environments, pre-trained models (PTM) with large sizes are challenging to deploy and use.

Inference Time: Reusing the entire set of pre-trained models can increase the inference time, affecting real-time applications.

Defect Inheritance: Reusing models without a code review may result in inheriting flaws from the original model, which could pose security risks in the new system. On the other hand, AI models are famously brittle to modest input data perturbations, allowing attackers to create adversarial examples for malevolent assaults.

Poor Performance: Compared to regression testing or hardware performance, the model converter's time and memory costs are substantially higher than developer/user expectations.

Failure Symptom	Definition	Failure Root Cause	Definition
Model Size	In memory-constrained environments, pre-trained models (PTM) with large sizes are a significant challenge to deploy and use.	Reusing the entire Model	The entire model is being reused, even though only a small portion of its functionalities (mainly labels) are necessary.
Inference Time	Reusing only a small portion of the functions provided by the model can lead to overhearing and unwanted computational costs.	Training the entire set of pre-trained models	Reusing the entire set of pre-trained models can increase the inference time, affecting real-time applications.
Defect Inheritance	Inheriting flaws from the original model could pose security and robustness risks in the new system (Weakness Inheriting).	AI models are mostly brittle to modest input data perturbations,	Reusing models without a code review.
Poorly Performance	Compared to regression testing or hardware performance, the model converter's time and memory costs are substantially higher than developer/user expectations.	Algorithmic Error	Optimization: The model converter is experiencing an issue with optimizations, including the elimination of inactive code or incorrect fusing. Tracing: Tracking DNN poses a challenge. This phenomenon is observed in frameworks that employ dynamic models. Dynamic models must be tracked to identify the nodes in the computational graph prior to the conversion.

Fig. 17. The main common failure symptoms and relevant practical root causes of reusing and reengineering DNN techniques (n1=4 symptoms, n2=4 causes).

- Structural failure symptoms

In General, interoperability relies on deep neural network (DNN) conversion methods such as "NNEF, ONNX, MMdnn, and CoreML." While primarily used for deployment and framework conversion, these methods suffer from crashes and performance degradation. This study demonstrated three structural failure types, as shown in Fig. 18:

Crashes and wrong model behavior: because of type of problem and incompatibility issues.

Wrong model selection: because of tensor shape mismatches during transformation, inference, or layout conversion.

Build failures: These occur because of testing issues like missing or broken tests. These issues primarily impact converters' graph translation and optimization stages. This study's findings align with existing research, and further analysis revealed model characteristics linked to converter failures.

Failure Symptom	Definition	Failure Root Cause	Description
Crash	Typically, a report of error is generated when the model converter terminates unpredictably during the conversion process.	Incompatibility	Internal: The API evolution has resulted in API compatibility issues within a model converter. External: There are API compatibility issues between a model converter and third-party libraries (e.g., ONNX, MMdnn, NNEF, TensorFlow, ONNX Run-time).
		Problem Type	Node: A model converter employs a computational network, with nodes representing DL operators such as convolution and pooling and edges representing tensors. Every node receives a variable number of tensors and outputs one. These challenges involve node transformation. Tensor: A tensor is a matrix with several dimensions that consists of elements of the same data type. This applies to the various classifications of tensors. Conventional: Conventional software systems also use customary factors. This subcategory addresses traditional variables in various forms.
Wrong Model	model converter exhibits unanticipated behavior without crashing, resulting in an incorrect—intermediate or final result.	Tensor Shape	This affects the shape of a node's input and output tensors during shape matching, transformation, inference, layout, etc.
Build Failure	Installation of model converter or dependencies fails.	Testing	Problems associated with software tests, including unit tests. This encompasses "flaky" tests, missing tests, broken tests, or novel tests.

Fig. 18. Structural failure symptoms arise from the structural and procedural root causes of DNN reuse (n1 =3 structural symptoms, n2 =4 structural causes).

M. Opportunities

Merging innovation with interoperability of deep neural models:

This study suggests that DNN applications need additional interoperability research to understand how pre-trained models make decisions to improve transparency and accountability. A deeper understanding of model converter failure modes may also lead to improved solutions. Scaling these new technologies may ease existing approaches' limits, enabling scale-up for any emerging model and reducing resource and time costs. Our review study (RQs) and relevant critical direct/indirect factors, failure symptoms, and root causes can help researchers and developers overcome limitations in ML and deep neural network model development and deployment from reuse inspiration. Currently, the study team is developing an innovative and promising approach that could overcome the incompatibility and related reuse issues to enhance the re-purposing capabilities of the DNN's domain and practice.

X. Conclusion and Future Trends

Deep neural networks (DNNs) have revolutionized the fields of machine learning (ML) and artificial intelligence (AI), enabling the creation of various models and algorithms capable of autonomously solving complex tasks. However, the development of these techniques faces significant challenges. This paper comprehensively reviews deep neural models that address the characteristics, challenges, open issues, and critical factors influencing ML and DNN technologies. It guides researchers and developers in achieving successful DNN models from reuse inspiration by evaluating factors impacting the reuse and related learning pipeline process. It carefully examines the symptoms of failure in existing reuse methods and the causes that usually lead to this failure. The SLR examined four research questions to comprehend their characteristics, challenges, influencing factors, and strategies for resolving limitations, particularly incompatibility and associated issues. By analyzing the first 193 research papers, the study identified significant challenges related to software incompatibility within DNNs. Other open issues include performance degradation, complexity, vulnerability to adversarial attacks, difficulty in code reengineering and conversion, a lack of standardized benchmarks, and system selection. The review reveals how these issues hinder parallel computing efficiency by increasing development, deployment, and computing time, complicating task organization, and raising costs. Additionally, goal mismatches can decrease computing accuracy, making training and performance complex. This study contributes to understanding DNN as a black box and benefits various fields, including AI, ML methods, and DNN-based systems in performance (e.g., accuracy, inference time, memory consumption) and robustness towards multiple adversarial attacks. The study suggests that DNN applications require further research on interoperability, model converter failure modes, and scaling technologies to improve solutions and reduce costs.

Future work focuses on reducing model size computational cost without sacrificing performance efficiency, whether for basic DNN models (e.g., pre-trained models) or LLMs using unique techniques (e.g. "Low-Rank-Adaptation "Lora" and Quantized Lora "QLora," model compression, and pruning) for optimal converted model size and inference time). Furthermore, during future work procedures, the team ensures that various adversarial attacks, especially in finance, healthcare, and autonomous applications, are avoided, including data security, privacy, veracity, and benchmarking challenges.

Acknowledgment

The authors gratefully acknowledge all contributors whose valuable insights and suggestions significantly improved this work. We confirm that no funding was received for this work and declare no conflicts of interest.

References

[1] Y. Duan, J. S. Edwards, and Y. K. Dwivedi, "Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda," Int J Inf Manage, vol. 48, no. February, pp. 63–71, 2019, doi: 10.1016/j.ijinfomgt.2019.01.021.

[2] J. Liu et al., "Usability Study of Distributed Deep Learning Frameworks For Convolutional Neural Networks," 2019. [Online]. Available: https://caffe2.ai/

[3] J. C. Davis, P. Jajal, W. Jiang, T. R. Schorlemmer, N. Synovic, and G. K. Thiruvathukal, "Reusing Deep Learning Models: Challenges and Directions in Software Engineering," Apr. 2024, [Online]. Available: http://arxiv.org/abs/2404.16688

[4] Z. Wang, K. Liu, J. Li, Y. Zhu, and Y. Zhang, "Various Frameworks and Libraries of Machine Learning and Deep Learning : A Survey," Archives of Computational Methods in Engineering, no. 0123456789, 2019, doi: 10.1007/s11831-018-09312-w.

[5] J. Liu et al., "Usability Study of Distributed Deep Learning Frameworks For Convolutional Neural Networks," ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. In Proceedings of the Deep Learning Day at SIGKDD Conference on Knowledge Discovery and Data Mining (KDD'18)., no. 1, p. 9, 2018, [Online]. Available: https://caffe2.ai/

[6] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," Applied Soft Computing Journal; Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX, vol. 70, pp. 41–65, 2018, doi: 10.1016/j.asoc.2018.05.018.

[7] M. Hilbert, "Big Data for Development: A Review of Promises and Challenges," Development Policy Review peer-reviewed academic journal Impact factor:1.093 2019 SJR, vol. 34, no. 1, pp. 135–174, 2016, doi: 10.1111/dpr.12142.

[8] S. Dobilas, "Transposed Convolutional Neural Networks — How to Increase the Resolution of Your Image," pp. 1–24, 2022.

[9] I. Tecnol, "Performance Comparison of Deep Learning Frameworks in Image Classification Problems using Convolutional and Recurrent Networks," In2017 IEEE Colombian Conference on Communications and Computing (COLCOM) 2017 Aug 16 (pp. 1-6). IEEE. 978-1-5386-1060-2/17/$31.00 ©2017 IEEE, no. In2017 IEEE Colombian Conference on Communications and Computing (COLCOM) 2017 Aug 16 (pp. 1-6). IEEE., p. 6, 2017.

[10] J. Just, "Natural language processing for innovation search – Reviewing an emerging non-human innovation intermediary," Technovation, vol. 129, Jan. 2024, doi: 10.1016/j.technovation.2023.102883.

[11] A. Khang, V. Abdullayev, O. Hrybiuk, and A. K. Shukla, "COMPUTER VISION AND Al-INTEGRATED IOT TECHNOLOGIES IN THE MEDICAL ECOSYSTEM EDITED BY," 2024.

[12] J. Rezaeenour, M. Ahmadi, H. Jelodar, and R. Shahrooei, "Systematic review of content analysis algorithms based on deep neural networks," Multimed Tools Appl, pp. 17879–17903, 2022, doi: 10.1007/s11042-022-14043-z.

[13] Y. Liu et al., "Enhancing the interoperability between deep learning frameworks by model conversion," Proceedings of the 28th ACM Joint Meeting on European Software Engineering Conference and Symposium on the Foundations of Software Engineering, p. 10, 2020, doi: 10.1145/3368089.3417051.

[14] B. Jan et al., "Deep learning in big data Analytics: A comparative study," Journal of Computers and Electrical Engineering Science Direct, vol. 75, no. September 2018, pp. 275–287, 2019, doi: 10.1016/j.compeleceng.2017.12.009.

[15] A. Shatnawi, G. Al-Bdour, R. Al-Qurran, and M. Al-Ayyoub, "A comparative study of open source deep learning frameworks," 2018 9th International Conference on Information and Communication Systems, ICICS 2018, vol. 2018-Janua, no. April, pp. 72–77, 2018, doi: 10.1109/IACS.2018.8355444.

[16] M. P. Reyes et al., "A Survey on Deep Learning," ACM Comput Surv, vol. 51, no. 5, pp. 1–36, 2018, doi: 10.1145/3234150.

[17] J. Nilsson, System of Systems Interoperability Machine Learning Model System of Systems Interoperability Machine Learning Model. Dept. of Computer Science and Electrical Engineering Lule˚a University of Technology Lule˚a, Sweden: Printed by Luleå University of Technology, Graphic Production 2019 ISSN 1402-1757 ISBN 978-91-7790-458-8 (print) ISBN 978-91-7790-459-5 (pdf) Luleå 2019 www.ltu.se, 2019. [Online]. Available: www.ltu.se

[18] Jason . Brownlee, "A Gentle Introduction to Tensors for Machine Learning with NumPy," pp. 1–17, 2018.

[19] Facebook Open Source, "Caffe2 Open Source Brings Cross Platform Machine Learning Tools to Developers," Https://Caffe2.Ai/, pp. 3–6, 2017, [Online]. Available: https://caffe2.ai/blog/2017/04/18/caffe2-open-source-announcement.html

[20] J. Li et al., "Feature selection: A data perspective," ACM Computing Surveys Journal ISI:Science Citation Index Expanded, vol. 50, no. 6, 2017, doi: 10.1145/3136625.

[21] Y. E. Wang, C.-J. Wu, X. Wang, K. Hazelwood, and D. Brooks, "Exploiting Parallelism Opportunities with Deep Learning Frameworks," ACM Digtal librar ACM Trans. Archit. Code Optim., Vol. 18, No. 1, Article 9, Publication date: December 2020., vol. abs/1908.0, 2020.

[22] Y. You, J. Demmel, K. Keutzer, Z. Zhang, and C. Hsieh, "Fast Deep Neural Network Training on Distributed Systems and Cloud TPUs," IEEE Transactions on Parallel and Distributed Systems, vol. 30, no. 11, 2019, doi: 10.1109/TPDS.2019.2913833.

[23] B. Qi, H. Sun, X. Gao, H. Zhang, Z. Li, and X. Liu, "Reusing Deep Neural Network Models through Model Re-engineering," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.00245

[24] Z. Wang, K. Liu, J. Li, Y. Zhu, and Y. Zhang, "Various Frameworks and Libraries of Machine Learning and Deep Learning: A Survey," Archives of Computational Methods in Engineering, no. 0123456789, p. 24, 2019, doi: 10.1007/s11831-018-09312-w.

[25] Y. Wu et al., "A Comparative Measurement Study of Deep Learning as a Service Framework," IEEE Trans Serv Comput, no. Dl, p. 15, 2019, doi: 10.1109/TSC.2019.2928551.

[26] L. Ruthotto, S. J. Osher, W. Li, L. Nurbekyan, and S. Wu, "A machine learning framework for solving high-dimensional mean field game and mean field control problems," Proceedings of the National Academy of Sciences, PNAS April 28, 2020 117 (17) 9183-9193; first published April 9, 2020, vol. 117, no. 17, p. 11, 2020, doi: 10.1073/pnas.1922204117.

[27] W. Jiang, G. Zeng, S. Wang, X. Wu, and C. Xu, "Application of Deep Learning in Lung Cancer Imaging Diagnosis," J Healthc Eng, vol. 2022, 2022, doi: 10.1155/2022/6107940.

[28] P. Skryjomski, B. Krawczyk, and A. Cano, "Speeding up k-Nearest Neighbors classifier for large-scale multi-label learning on GPUs," Neurocomputing Journal Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX; ISI: Science Citation Index Expanded, vol. 354, 2019, doi: 10.1016/j.neucom.2018.06.095.

[29] F. Bajaber, S. Sakr, O. Batarfi, A. Altalhi, and A. Barnawi, "Benchmarking big data systems: A survey," Comput Commun, vol. 149, pp. 241–251, 2020, doi: 10.1016/j.comcom.2019.10.002.

[30] W. Jiang et al., "An Empirical Study of Pre-Trained Model Reuse in the Hugging Face Deep Learning Model Registry," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.02552

[31] R. A. Hazarika, A. Abraham, D. Kandar, and A. K. Maji, "An Improved LeNet-Deep Neural Network Model for Alzheimer's Disease Classification Using Brain Magnetic Resonance Images," IEEE Access, vol. 9, pp. 161194–161207, 2021, doi: 10.1109/ACCESS.2021.3131741.

[32] J. Zhang, X. Yu, X. Lei, and C. Wu, "A Novel Deep LeNet-5 Convolutional Neural Network Model for Image Recognition," Computer Science and Information Systems, vol. 19, no. 3, pp. 1463–1480, Sep. 2022, doi: 10.2298/CSIS220120036Z.

[33] R. U. Khan, X. Zhang, R. Kumar, and E. O. Aboagye, "Evaluating the performance of ResNet model based on image recognition," in ACM International Conference Proceeding Series, Association for Computing Machinery, Mar. 2018, pp. 86–90. doi: 10.1145/3194452.3194461.

[34] F. Chollet, "Xception: Deep Learning with Depthwise Separable Convolutions," Oct. 2017, [Online]. Available: http://arxiv.org/abs/1610.02357

[35] C. Wang et al., "Pulmonary image classification based on inception-v3 transfer learning model," IEEE Access, vol. 7, pp. 146533–146541, 2019, doi: 10.1109/ACCESS.2019.2946000.

[36] D. F. and A. V. Hussam Qassim, "VGG16 - Residual Squeeze," 2018.

[37] [37] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," Sep. 2017, [Online]. Available: http://arxiv.org/abs/1409.1556

[38] S. Sztwiertnia, M. Grübel, A. Chouchane, D. Sokolowski, K. Narasimhan, and M. Mezini, Impact of programming languages on machine learning bugs, vol. 1, no. 1. Association for Computing Machinery, 2021. doi: 10.1145/3464968.3468408.

[39] C. Luo, X. He, J. Zhan, L. Wang, W. Gao, and J. Dai, "Comparison and Benchmarking of AI Models and Frameworks on Mobile Devices," arXiv preprint arXiv:2005.05085. Cornell University Journal, vol. abs/2005.0, 2020, [Online]. Available: https://www.semanticscholar.org/paper/3858ccebd64b63e27a8cbc2d4d4498e063c05688

[40] A. Chapman et al., "Dataset search: a survey," in VLDB Journal, Springer, Jan. 2020, pp. 251–272. doi: 10.1007/s00778-019-00564-x.

[41] N. Sharma, V. Jain, and A. Mishra, "An Analysis of Convolutional Neural Networks for Image Classification," in Procedia Computer Science, Elsevier B.V., 2018, pp. 377–384. doi: 10.1016/j.procs.2018.05.198.

[42] Y. Wang et al., "Benchmarking the Performance and Energy Efficiency of AI Accelerators for AI Training," Proceedings - 20th IEEE/ACM International Symposium on Cluster, Cloud and Internet Computing, CCGRID 2020, pp. 744–751, 2020, doi: 10.1109/CCGrid49817.2020.00-15.

[43] Y. Wang, G. Y. Wei, and D. Brooks, "Benchmarking TPU, GPU, and CPU platforms for deep learning," arXiv preprint arXiv:1907.10701. Coenell University, vol. 4, 2019.

[44] L. Nguyen, P. Yu, and M. Chowdhury, "No!: Not Another Deep Learning Framework," Proceedings of the Workshop on Hot Topics in Operating Systems - HOTOS '17, Whistler, BC, Canada, May 08-10, 2017, https://doi.org/10.1145/3102980.3102995. ACM, vol. Part F1293, no.

May 08-10, 2017, pp. 88-93 (6 Pages), 2017, doi: 10.1145/3102980.3102995.

[45] T. Zhao and X. Huang, "Design and implementation of DeepDSL: A DSL for deep learning," Comput Lang Syst Struct, vol. 54, pp. 39–70, 2018, doi: 10.1016/j.cl.2018.04.004.

[46] W. Xiang, "APPLICATIONS OF DEEP LEARNING IN LARGE-SCALE OBJECT DETECTION AND SEMANTIC SEGMENTATION," ProQuest Number:13855813, no. March, p. 142, 2018.

[47] V. Kovalev, A. Kalinovsky, and S. Kovalev, "Deep Learning with Theano, Torch, Caffe, TensorFlow, and Deeplearning4J: Which One Is the Best in Speed and Accuracy? Recognition of multi-drug resistant tubercolusis based on CT and X-ray image analysis View project UAV: back to base problem View project," XIII International Conference on Pattern Recognition and Information Processing; SPRINGER, pp. 99–103, 2019, [Online]. Available: http://imlab.grid.by/

[48] G. Nguyen et al., "Machine Learning and Deep Learning frameworks and libraries for large-scale data mining: a survey," Artificial Intelligence Review (2019) 52:77–124 Springer, vol. 52, no. 1, pp. 77–124, 2019, doi: 10.1007/s10462-018-09679-z.

[49] M. Openja, A. Nikanjam, A. H. Yahmed, F. Khomh, Z. Ming, and Jiang, "An Empirical Study of Challenges in Converting Deep Learning Models," Jun. 2022, [Online]. Available: http://arxiv.org/abs/2206.14322

[50] [50] K. Zhou, T. P. Michalak, and Y. Vorobeychik, "Adversarial robustness of similarity-based link prediction," Proceedings - IEEE International Conference on Data Mining, ICDM, vol. 2019-Novem, pp. 926–935, 2019, doi: 10.1109/ICDM.2019.00103.

[51] R. . ShyamChari, "IMAGE RECOGNITION AND STUDY OF HYPERPARAMETER OPTIMIZATION OF CONVOLUTIONAL NEURAL NETWORKS USING TENSORFLOW AND KERAS FRAMEWORKS Deep," PhD diss., 2018 Lyles College of Engineering California State University, Fresno May 2018, no. May 2018, p. 400, 2018.

[52] B. Qi, H. Sun, H. Zhang, and X. Gao, "Reusing Convolutional Neural Network Models through Modularization and Composition," Nov. 2023, [Online]. Available: http://arxiv.org/abs/2311.04438

[53] H. Kwon, P. Chatarasi, M. Pellauer, A. Parashar, V. Sarkar, and T. Krishna, "Understanding reuse, performance, and hardware cost of DNN dataflows: A data-centric approach," in Proceedings of the Annual International Symposium on Microarchitecture, MICRO, IEEE Computer Society, Oct. 2020, pp. 754–768. doi: 10.1145/3352460.3358252.

[54] J. C. Davis, P. Jajal, W. Jiang, T. R. Schorlemmer, N. Synovic, and G. K. Thiruvathukal, "Reusing Deep Learning Models: Challenges and Directions in Software Engineering," 2023.

[55] J. Ghofrani, E. Kozegar, A. Bozorgmehr, and M. D. Soorati, "Reusability in Artificial Neural Networks: An Empirical Study," in Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI), Gesellschaft fur Informatik (GI), 2021, pp. 37–38. doi: 10.18420/SE2020_07.

[56] M. Niazi, A. M. Saeed, M. Alshayeb, S. Mahmood, and S. Zafar, "A maturity model for secure requirements engineering," Comput Secur, vol. 95, Aug. 2020, doi: 10.1016/j.cose.2020.101852.

[57] C. W. Krueger, "Software Reuse."

[58] H. Hussain, P. S. Tamizharasan, and C. S. Rahul, "Design possibilities and challenges of DNN models: a review on the perspective of end devices," Artif Intell Rev, vol. 55, no. 7, pp. 5109–5167, Oct. 2022, doi: 10.1007/s10462-022-10138-z.

[59] F. Authors, F. Wu, T. Li, F. Luo, S. Wu, and C. Xiao, "A Survey on Deep Transfer Learning and Edge Computing for Mitigating the COVID-19 Pandemic," Journal of Systems Architecture, p. 32, 2020, doi: https://doi.org/10.1016/j.sysarc.2020.101830.

[60] X. Han et al., "Pre-trained models: Past, present and future," AI Open, vol. 2, pp. 225–250, Jan. 2021, doi: 10.1016/j.aiopen.2021.08.002.

[61] A. Rahman et al., "Machine learning and deep learning-based approach in smart healthcare: Recent advances, applications, challenges and opportunities," 2024, American Institute of Mathematical Sciences. doi: 10.3934/publichealth.2024004.

[62] H. V. Pham et al., "Problems and Opportunities in Training Deep Learning Software Systems: An Analysis of Variance," in Proceedings - 2020 35th IEEE/ACM International Conference on Automated Software Engineering, ASE 2020, Institute of Electrical and Electronics Engineers Inc., Sep. 2021, pp. 771–783. doi: 10.1145/3324884.3416545.

[63] P. Jajal et al., "Analysis of Failures and Risks in Deep Learning Model Converters: A Case Study in the ONNX Ecosystem," Mar. 2024, [Online]. Available: http://arxiv.org/abs/2303.17708

[64] V. Sze, Y. Chen, T. Yang, and J. Emer, "Efficient Processing of Deep Neural Networks: A Tutorial and Survey," Proceedings of the IEEE, vol. 105, no. 12, 2017, doi: 10.1109/JPROC.2017.2761740.

[65] S. Sengupta et al., "A review of deep learning with special emphasis on architectures, applications and recent trends," Knowl Based Syst, vol. 194, p. 105596, 2020, doi: 10.1016/j.knosys.2020.105596.

[66] Z. J. Yao, J. Bi, and Y. X. Chen, "Applying Deep Learning to Individual and Community Health Monitoring Data: A Survey," International Journal of Automation and Computing, vol. 15, no. 6, pp. 643–655, 2018, doi: 10.1007/s11633-018-1136-9.

[67] P. Giudici, M. Centurelli, and S. Turchetta, "Artificial Intelligence risk measurement," Expert Syst Appl, vol. 235, Jan. 2024, doi: 10.1016/j.eswa.2023.121220.

[68] N. Soni, E. K. Sharma, N. Singh, and A. Kapoor, "Artificial Intelligence in Business: From Research and Innovation to Market Deployment," Procedia Computer Science Journal / International Conference on Computational Intelligence and Data Science (ICCIDS 2019)., vol. 167, no. 2019, pp. 2200–2210, 2020, doi: 10.1016/j.procs.2020.03.272.

[69] A. Vinothini and S. B. Priya, "Survey of machine learning methods for big data applications," ICCIDS 2017 - International Conference on Computational Intelligence in Data Science, Proceedings, vol. 2018-Janua, pp. 1–5, 2018, doi: 10.1109/ICCIDS.2017.8272638.

[70] P. Singhal and S. Pareek, "Artificial neural network for prediction of breast cancer," Proceedings of the International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2018 IEEE Xplore, pp. 464–468, 2019, doi: 10.1109/I-SMAC.2018.8653700.

[71] G. A. Lewis, S. Bellomo, and A. Galyardt, "Component Mismatches Are a Critical Bottleneck to Fielding AI-Enabled Systems in the Public Sector," Carnegie Mellon Software Engineering Institute Pittsburgh, PA USA arXiv:1910.06136v1.AAAI, vol. 1, pp. 2–5, 2019.

[72] A. Sheth, M. Gaur, U. Kursuncu, and R. Wickramarachchi, "Shades of Knowledge-Infused Learning for Enhancing Deep Learning," IEEE Internet Comput, vol. 23, no. 6, pp. 54–63, 2019, doi: 10.1109/MIC.2019.2960071.

[73] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," Information Fusion, vol. 42, no. August 2017, pp. 146–157, 2018, doi: 10.1016/j.inffus.2017.10.006.

[74] A. Koutsoukas, K. Monaghan, X. Li, and J. Huan, "Deep-learning: Investigating deep neural networks hyper-parameters and comparison of performance to shallow methods for modeling bioactivity data," Journal of Cheminformatics Springer, vol. 9, no. 1, 2017, doi: 10.1186/s13321-017-0226-y.

[75] L. Liu et al., "Deep Learning for Generic Object Detection : A Survey," International Journal of Computer Vision (2020) 128:261–318 Springer, vol. 128, no. 2, pp. 261–318, 2021, doi: 10.1007/s11263-019-01247-4.

[76] A. Abrol, M. Bhattarai, A. Fedorov, Y. Du, S. Plis, and V. Calhoun, "Deep residual learning for neuroimaging: An application to predict progression to Alzheimer's disease," J Neurosci Methods, vol. 339, no. September 2019, p. 108701, 2021, doi: 10.1016/j.jneumeth.2020.108701.

[77] M. A. Ebrahimighahnavieh, S. Luo, and R. Chiong, "Deep learning to detect Alzheimer's disease from neuroimaging: A systematic literature review," Comput Methods Programs Biomed, vol. 187, p. 105242, 2021, doi: 10.1016/j.cmpb.2019.105242.

[78] N. Rahim, S. El-Sappagh, S. Ali, K. Muhammad, J. Del Ser, and T. Abuhmed, "Prediction of Alzheimer's progression based on multimodal Deep-Learning-based fusion and visual Explainability of time-series data," Information Fusion, vol. 92, no. June 2022, pp. 363–388, 2023, doi: 10.1016/j.inffus.2022.11.028.

[79] D. van Tilborg, H. Brinkmann, E. Criscuolo, L. Rossen, R. Özçelik, and F. Grisoni, "Deep learning for low-data drug discovery: Hurdles and opportunities," Jun. 01, 2024, Elsevier Ltd. doi: 10.1016/j.sbi.2024.102818.

[80] Y. Liu, C. Chen, R. Zhang, M. Research, H. Lin, and M. Yang, "MMdnn: Enhancing the Interoperability between Deep Learning Frameworks by

Model Conversion," pp. 6–8, 2020, [Online]. Available: https://en.wikipedia.org/wiki/Interoperability.

[81] C. Jiang, Z. Huang, T. Pedapati, P. Y. Chen, Y. Sun, and J. Gao, "Network properties determine neural network performance," Nat Commun, vol. 15, no. 1, Dec. 2024, doi: 10.1038/s41467-024-48069-8.

[82] Y. Cheng, D. Wang, P. Zhou, and T. Zhang, "A Survey of Model Compression and Acceleration for Deep Neural Networks," IEEE SIGNAL PROCESSING MAGAZINE, SPECIAL ISSUE ON DEEP LEARNING FOR IMAGE UNDERSTANDING, no. SPECIAL ISSUE, Oct. 2020, [Online]. Available: http://arxiv.org/abs/1710.09282

[83] F. Zhuang et al., "A Comprehensive Survey on Transfer Learning," Nov. 2021, [Online]. Available: http://arxiv.org/abs/1911.02685

[84] L. Liu, Y. Wu, W. Wei, W. Cao, S. Sahin, and Q. Zhang, "Benchmarking deep learning frameworks: Design considerations, metrics and beyond," Proceedings - International Conference on Distributed Computing Systems; IEEE Computer Society, vol. 2018-July, pp. 1258–1269, 2018, doi: 10.1109/ICDCS.2018.00125.

[85] S. S. Khanal, P. W. C. Prasad, A. Alsadoon, and A. Maag, "A systematic review: machine learning based recommendation systems for e-learning," Educ Inf Technol (Dordr), vol. 25, no. 4, pp. 2635–2664, Jul. 2020, doi: 10.1007/s10639-019-10063-9.

[86] M. Perano, G. L. Casali, Y. Liu, and T. Abbate, "Professional reviews as service: A mix method approach to assess the value of recommender systems in the entertainment industry," Technol Forecast Soc Change, vol. 169, no. April, p. 120800, 2021, doi: 10.1016/j.techfore.2021.120800.

[87] B. Zoph and J. Shlens, "Learning Transferable Architectures for Scalable Image Recognition," 2018.

[88] J. Li et al., "Networked human motion capture system based on quaternion navigation," in BodyNets International Conference on Body Area Networks, 2018. doi: 10.1145/0000000.0000000.

[89] A. Goyal et al., "Automatic border surveillance using machine learning in remote video surveillance systems," in Lecture Notes in Electrical Engineering, vol. 569, Springer Verlag, 2020, pp. 751–760. doi: 10.1007/978-981-13-8942-9_64.

[90] K. Kourou, T. P. Exarchos, K. P. Exarchos, M. V. Karamouzis, and D. I. Fotiadis, "Machine learning applications in cancer prognosis and prediction," Comput Struct Biotechnol J, vol. 13, pp. 8–17, 2017, doi: 10.1016/j.csbj.2014.11.005.

[91] et al Noureen Talpur, "Deep Neuro Fuzzy System application trends, challenges, and future perspectives; A systematic Review.pdf," 2023.

[92] O. Ali, W. Abdelbaki, A. Shrestha, E. Elbasi, M. A. A. Alryalat, and Y. K. Dwivedi, "A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities," Journal of Innovation and Knowledge, vol. 8, no. 1, 2023, doi: 10.1016/j.jik.2023.100333.

[93] A. Darwish, A. E. Hassanien, and S. Das, "A survey of swarm and evolutionary computing approaches for deep learning," Artif Intell Rev, vol. 53, no. 3, 2020, doi: 10.1007/s10462-019-09719-2.

[94] G. A. Lewis, S. Bellomo, and I. Ozkaya, "Characterizing and Detecting Mismatch in Machine-Learning-Enabled Systems," Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.14101

[95] J. R. Ragini, P. M. R. Anand, and V. Bhaskar, "Big data analytics for disaster response and recovery through sentiment analysis," Int J Inf Manage, vol. 42, no. May, pp. 13–24, 2018, doi: 10.1016/j.ijinfomgt.2018.05.004.

[96] B. Kumar Padhi, S. S. Nayak, and B. N. Biswal, "Machine Learning for Big Data Processing: A Literature Review," INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY; IJIRT 140001 December 2018 | IJIRT | Volume 5 Issue 7 | ISSN: 2349-6002 samples, vol. 5, no. 7, p. 10, 2018, [Online]. Available: https://www.researchgate.net/publication/331497335

[97] D. A. Neu, J. Lahann, and P. Fettke, "A systematic literature review on state-of-the-art deep learning methods for process prediction," 2022. doi: 10.1007/s10462-021-09960-8.

[98] B. Krishna, N. Rekulapelli, and B. P. Kauda, "Comparison of different deep learning frameworks," Mater Today Proc, no. xxxx, p. 5, 2020, doi: 10.1016/j.matpr.2020.09.608.

[99] A. Darwish, A. Ella, and H. Swagatam, "A survey of swarm and evolutionary computing approaches for deep learning," Artif Intell Rev, 2020, doi: 10.1007/s10462-019-09719-2.

[100] M. Modasshir, A. Quattrini Li, and I. Rekleitis, "Deep neural networks: A comparison on different computing platforms," Proceedings - 2018 15th Conference on Computer and Robot Vision, CRV 2018 978-1-5386-6481-0/18/$31.00 ©2018 IEEE, pp. 383–389, 2018, doi: 10.1109/CRV.2018.00060.

[101] J. Dai and A. Gibson, "Comparison of deep-learning software originally," https://en.wikipedia.org/wiki/Comparison_of_deep-learning_software, pp. 1–5, 2020.

[102] Y. Liu, C. Chen, R. Zhang, M. Research, H. Lin, and M. Yang, "Enhancing the Interoperability between Deep Learning Frameworks by Model Conversion Tingting Qin," 2020.

[103] S. Hassan, "Deep context of citations using machine learning models.pdf," 2018.

[104] Khronos G, "Neural Network Exchange Format (NNEF)," 2019.

[105] W. F. Lin et al., "ONNC: A Compilation Framework Connecting ONNX to Proprietary Deep Learning Accelerators," Proceedings 2019 IEEE International Conference on Artificial Intelligence Circuits and Systems, AICAS 2019, pp. 214–218, 2019, doi: 10.1109/AICAS.2019.8771510.

[106] S. Sanchez-Gordon and S. Luján-Mora, "Technological Innovations in Large-Scale Teaching: Five Roots of Massive Open Online Courses," Journal of Educational Computing Research, vol. 56, no. 5, pp. 623–644, 2018, doi: 10.1177/0735633117727597.

[107] H. N. F, Y. S. S, and S. S. M, "A Systematic Literature Review on Features of Deep Learning in Big Data Analytics," Int. J. Advance Soft Compu. Appl, vol. 9, no. 1, 2017.

[108] F. Hujainah, R. B. A. Bakar, M. A. Abdulgabber, and K. Z. Zamli, "Software requirements prioritisation: a systematic literature review on significance, stakeholders, techniques and challenges," IEEE Access, vol. 6, pp. 71497–71523, 2018.

[109] L. A. Kahale et al., "PRISMA flow diagrams for living systematic reviews: a methodological survey and a proposal," F1000Res, vol. 10, p. 192, Mar. 2022, doi: 10.12688/f1000research.51723.1.

[110] Kitchenham, "Guidelines for performing Systematic Literature Reviews in Software Engineering," p. 65, 2007, doi: 10.1145/1134285.1134500.

[111] S. Mittal and S. Vaishay, "A survey of techniques for optimizing deep learning on GPUs," Journal of Systems Architecture; Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX, vol. 99, p. 45, 2019, doi: 10.1016/j.sysarc.2019.101635.

[112] S. Fatima, K. C. Desouza, and G. S. Dawson, "National strategic artificial intelligence plans: A multi-dimensional analysis," Journal of Economic Analysis and Policy; Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX, vol. 67, pp. 178–194, 2020, doi: 10.1016/j.eap.2020.07.008.

[113] Q. Mao, F. Hu, and Q. Hao, "Deep learning for intelligent wireless networks: A comprehensive survey," IEEE Communications Surveys and Tutorials, vol. 20, no. 4, pp. 2595–2621, 2018, doi: 10.1109/COMST.2018.2846401.

[114] A. Lavecchia, "Deep learning in drug discovery: opportunities, challenges and future prospects," Drug Discovery Today Journal 1359-6446/ã 2019 Elsevier Ltd. All rights reserved SCI, vol. 24, no. 10, p. 16, 2019, doi: 10.1016/j.drudis.2019.07.006.

[115] T. R. Rao, P. Mitra, R. Bhatt, and A. Goswami, "The big data system, components, tools, and technologies: a survey," Knowledge and Information Systems ISI : Science Citation Index Expanded, vol. 60, no. 3, pp. 1165–1245, 2019, doi: 10.1007/s10115-018-1248-0.

[116] Y. Zahidi, Y. El Younoussi, and Y. Al-amrani, "A powerful comparison of deep learning frameworks for Arabic sentiment analysis," International Journal of Electrical and Computer Engineering (IJECE), vol. 11, no. 1, pp. 745–752, 2021, doi: 10.11591/ijece.v11i1.pp745-752.

[117] M. Salvi, U. R. Acharya, F. Molinari, and K. M. Meiburger, "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis," Comput Biol Med, vol. 128, p. 104129, 2021, doi: 10.1016/j.compbiomed.2020.104129.

[118]A. Shatnawi, G. Al-Bdour, R. Al-Qurran, and M. Al-Ayyoub, "A comparative study of open source deep learning frameworks," 2018 9th International Conference on Information and Communication Systems, ICICS 2018; IEEE, vol. 2018-Janua, no. April, pp. 72–77, 2018, doi: 10.1109/IACS.2018.8355444.

[119]A. Oussous, F. Z. Benjelloun, A. Ait Lahcen, and S. Belfkih, "Big Data technologies: A survey," Journal of King Saud University - Computer and Information Sciences ISI: Science Citation Index Expanded, vol. 30, no. 4, pp. 431–448, 2018, doi: 10.1016/j.jksuci.2017.06.001.

[120]N. Mahendran, P. M. D. R. Vincent, and K. Srinivasan, "Machine Learning Based Computational Gene Selection Models : A Survey , Performance Evaluation , Open Issues , and Future Research Directions," Frontiers in genetics Journal SCIE (web of science), vol. 11, no. December, pp. 1–25, 2020, doi: 10.3389/fgene.2020.603808.

[121]R. Vinayakumar, K. Soman, and P. Poornachandran, "Evaluation of recurrent neural network and its variants for intrusion detection system (IDS)," International Journal of Information System Modeling and Design, vol. 8, no. 3, 2017, doi: 10.4018/IJISMD.2017070103.

[122]A. Awan, K. Manian, C. Chu, H. Subramoni, and D. Panda, "Optimized large-message broadcast for deep learning workloads: MPI, MPI+NCCL, or NCCL2?," Parallel Comput, vol. 85, 2019, doi: 10.1016/j.parco.2019.03.005.

[123]D. W. Bates, A. Auerbach, P. Schulam, A. Wright, and S. Saria, "Reporting and Implementing Interventions Involving Machine Learning and Artificial Intelligence," Annals of internal medicine Jour, vol. 172, no. 11, 2020, doi: 10.7326/M19-0872.

[124]W. Dai and D. Berleant, "Benchmarking contemporary deep learning hardware and frameworks: A survey of qualitative metrics," Proceedings - 2019 IEEE 1st International Conference on Cognitive Machine Intelligence, CogMI 2019, pp. 148–155, 2019, doi: 10.1109/CogMI48466.2019.00029.

[125]Y. Chen, Y. Tian, and M. He, "Monocular human pose estimation: A survey of deep learning-based methods," Computer Vision and Image Understanding, vol. 192, no. December 2019, p. 102897, 2020, doi: 10.1016/j.cviu.2019.102897.

[126]X. Lu, H. Shi, R. Biswas, M. Javed, and D. Panda, "DLoBD: A Comprehensive Study of Deep Learning over Big Data Stacks on HPC Clusters," IEEE Transactions on Multi-Scale Computing Systems, vol. 4, no. 4, 2018, doi: 10.1109/TMSCS.2018.2845886.

[127]U. Naseem, I. Razzak, and P. W. Eklund, "A survey of pre-processing techniques to improve short-text quality: a case study on hate speech detection on twitter," Multimedia Tools and Applications Journal; Publisher: SPRINGER , VAN GODEWIJCKSTRAAT 30, DORDRECHT, NETHERLANDS, 3311 GZ, p. 28, 2020, doi: 10.1007/s11042-020-10082-6.

[128]A. Jain, A. A. Awan, H. Subramoni, and D. K. Panda, "Scaling TensorFlow, PyTorch, and MXNet using MVAPICH2 for high-performance deep learning on Frontera," Proceedings of DLS 2019: Deep Learning on Supercomputers - Held in conjunction with SC 2019: The International Conference for High Performance Computing, Networking, Storage and Analysis, 2019, doi: 10.1109/DLS49591.2019.00015.

[129]T. Meng, X. Jing, Z. Yan, and W. Pedrycz, "A survey on machine learning for data fusion," Information Fusion, vol. 57, no. 2, pp. 115–129, 2020, doi: 10.1016/j.inffus.2019.12.001.

[130]F. Zaki, A. E. Mohamed, and S. G. Sayed, "CtuNet : A Deep Learning-based Framework for Fast CTU Partitioning of H265 / HEVC Intra-coding," Ain Shams Engineering Journal, no. xxxx, pp. 1–8, 2021, doi: 10.1016/j.asej.2021.01.001.

[131]M. Li et al., "The Deep Learning Compiler: A Comprehensive Survey," IEEE Transactions on Parallel and Distributed Systems; Publisher: IEEE COMPUTER SOC , 10662 LOS VAQUEROS CIRCLE, PO BOX 3014, LOS ALAMITOS, USA, CA, 90720-1314, vol. 32, no. 3, pp. 708–727, 2021, doi: 10.1109/TPDS.2020.3030548.

[132]Y. Oyama, S. Matsuoka, T. Ben-Nun, and T. Hoefler, "Accelerating Deep Learning Frameworks with Micro-Batches," Proceedings - IEEE International Conference on Cluster Computing, ICCC, vol. 2018-, no. 11, 2018, doi: 10.1109/CLUSTER.2018.00058.

[133]T. Doleck, D. J. Lemay, R. B. Basnet, and P. Bazelais, "Predictive analytics in education: a comparison of deep learning frameworks," Educ Inf Technol (Dordr), vol. 25, no. 3, pp. 1951–1963, 2020, doi: 10.1007/s10639-019-10068-4.

[134]E. M. Dogo, N. I. Nwulu, B. Twala, and C. Aigbavboa, "A survey of machine learning methods applied to anomaly detection on drinking-water quality data," Urban Water J, vol. 16, no. 3, pp. 235–248, 2019, doi: 10.1080/1573062X.2019.1637002.

[135]J. H. Tao et al., "BenchIP: Benchmarking Intelligence Processors," J Comput Sci Technol, vol. 33, no. 1, 2018, doi: 10.1007/s11390-018-1805-8.

[136]C. L. Hung, Y. Y. Lin, C. Y. Tang, C. Wang, and M. C. Chen, "Performance of convolution neural network based on multiple GPUs with different data communication models," Proceedings - 2018 IEEE/ACIS 19th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2018, pp. 87–92, 2018, doi: 10.1109/SNPD.2018.8441056.

[137]L. Zhou, S. Pan, J. Wang, and A. V Vasilakos, "Neurocomputing Machine learning on big data : Opportunities and challenges," ELSEVIER JOURNAL Science Direct, vol. 237, no. September 2016, pp. 350–361, 2017, doi: 10.1016/j.neucom.2017.01.026.

[138]S. Mao, B. Wang, Y. Tang, and F. Qian, "Opportunities and Challenges of Artificial Intelligence for Green Manufacturing in the Process Industry," Engineering Journal; Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX, vol. 5, no. 6, pp. 995–1002, 2019, doi: 10.1016/j.eng.2019.08.013.

[139]L. E. Lwakatare, A. Raj, I. Crnkovic, J. Bosch, and H. H. Olsson, "Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions," Information and Software Technology 127 (2020) 106368 Contents ELSEVIER, vol. 127, no. June, p. 17, 2020, doi: 10.1016/j.infsof.2020.106368.

[140]E. Breck, N. Polyzotis, S. Roy, S. E. Whang, and M. Zinkevich, "DATA VALIDATION FOR MACHINE LEARNING," Proceedings of the 2 nd SysML Conference, Palo Alto, CA, USA, 2019, pp. 1--14, 2019.

[141]Y. Roh, G. Heo, and S. E. Whang, "A survey on data collection for machine learning: A big data - AI integration perspective," IEEE Transactions on Knowledge and Data Engineering. Publisher: IEEE COMPUTER SOC , 10662 LOS VAQUEROS CIRCLE, PO BOX 3014, LOS ALAMITOS, USA, CA, 90720-1314, vol. 2, pp. 1–20, 2018, doi: 10.1109/tkde.2019.2946162.

[142]G. Al-bdour, R. Al-qurran, M. Al-ayyoub, A. Shatnawi, and M. Al-ayyoub, "Benchmarking open source deep learning frameworks," International Journal of Electrical and Computer Engineering (IJECE), vol. 10, no. 5, pp. 5479–5486, 2020, doi: 10.11591/ijece.v10i5.pp5479-5486.

[143]P. Xu, S. Shi, and X. Chu, "Performance Evaluation of Deep Learning Tools in Docker Containers," Proceedings - 2017 3rd International Conference on Big Data Computing and Communications, BigCom 2017, pp. 395–403, 2017, doi: 10.1109/BIGCOM.2017.32.

[144]M. Štufi and B. Baˇ, "Big Data Analytics and Processing Platform in Czech Republic Healthcare," Applied Science Journal . Appl. Sci. 2020, 10, 1705; doi:10.3390/app10051705 www.mdpi.com/journal/applsci10, 1705. MDPI, no. Appl. Sci. 2020, 10, 1705; doi:10.3390/app10051705, p. 23, 2020.

[145]F. Cappa, R. Oriani, E. Peruffo, and I. McCarthy, "Big Data for Creating and Capturing Value in the Digitalized Environment: Unpacking the Effects of Volume, Variety, and Veracity on Firm Performance*," Journal of Product Innovation Management Publisher: WILEY , 111 RIVER ST, HOBOKEN, USA, NJ, 07030-5774 ISI: Science Citation Index Expanded | Social Sciences Citation Index, vol. 0, no. 0, pp. 1–19, 2020, doi: 10.1111/jpim.12545.

[146]E. Kusmenko, S. Nickels, S. Pavlitskaya, B. Rumpe, and T. Timmermanns, "Modeling and Training of Neural Processing Systems," In: Conference on Model Driven Engineering Languages and Systems (MODELS'19), pp. 283--293, IEEE, Munich, Sep. 2019. www.se-rwth.de/publications/, p. 11, 2019.

[147]L. F. Sánchez-Peralta, L. Bote-Curiel, A. Picón, F. M. Sánchez-Margallo, and J. B. Pagador, "Deep learning to find colorectal polyps in colonoscopy: A systematic literature review," Artificial Intelligence in Medicine; Publisher: ELSEVIER , RADARWEG 29, AMSTERDAM, NETHERLANDS, 1043 NX, vol. 108, no. March, p. 101923, 2020, doi: 10.1016/j.artmed.2020.101923.

[148]S. Bianco, R. Cadène, L. Celona, and P. Napoletano, "Benchmark Analysis of Representative Deep Neural Network Architectures," IEEE Access IEEE TRANSACTIONS and JOURNALS, vol. 6, pp. 64270–64277, 2018, doi: 10.1109/ACCESS.2018.2877890.

[149]Q. Guo et al., "An Empirical Study towards Characterizing Deep Learning Development and Deployment across Different Frameworks and Platforms," Sep. 2019, [Online]. Available: http://arxiv.org/abs/1909.06727

[150]L. Messeri and M. J. Crockett, "Artificial intelligence and illusions of understanding in scientific research," Nature, vol. 627, no. 8002, pp. 49–58, Mar. 2024, doi: 10.1038/s41586-024-07146-0.

[151]X. Li et al., "Interpretable Deep Learning: Interpretation, Interpretability, Trustworthiness, and Beyond," Mar. 2021, [Online]. Available: http://arxiv.org/abs/2103.10689

[152]J. J. Pfeiffer, J. Neville, and P. N. Bennett, "Overcoming relational learning biases to accurately predict preferences in large scale networks," WWW 2015 - Proceedings of the 24th International Conference on World Wide Web; ACM 978-1-4503-3469-3/15/05, pp. 853–863, 2015, doi: 10.1145/2736277.2741668.

[153]D. Jeong, "Artificial Intelligence Security Threat, Crime, and Forensics: Taxonomy and Open Issues," IEEE Access, vol. 8, 2020, doi: 10.1109/access.2020.3029280.

[154]A. Mendelson, "Security and Privacy in the Age of Big Data and Machine Learning," Computer Journal; Publisher: IEEE COMPUTER SOC , 10662 LOS VAQUEROS CIRCLE, PO BOX 3014, LOS ALAMITOS, USA, CA, 90720-1314 IEEE, vol. 52, pp. 65–70, 2019, doi: 10.1109/MC.2019.2943137.

[155]C. Issues, A. Projects, and W. Security, "Caffe to TensorFlow," 2017, [Online]. Available: https://github.com/ethereon/caffe-tensorflow

[156]H. Park, J. H. Lee, Y. Oh, S. Ha, and S. Lee, "Privacy attacks against deep learning models and their countermeasures," Journal of Systems Architecture journal, vol. 9, no. 114 (2021) 101940, 2021, [Online]. Available: www.elsevier.com/locate/sysarc%0APrivacy

[157]C. Ning and F. You, "Optimization under Uncertainty in the Era of Big Data and Deep Learning : When Machine Learning Meets Mathematical Programming," Computers & Chemical Engineering ; Publisher: PERGAMON-ELSEVIER SCIENCE LTD , THE BOULEVARD, LANGFORD LANE, KIDLINGTON, OXFORD, ENGLAND, OX5 1GB, no. 607, pp. 1–42, 2020.

[158]L. Liu, Y. Wu, W. Wei, W. Cao, S. Sahin, and Q. Zhang, "Benchmarking Deep Learning Frameworks : Design Considerations , Metrics and Beyond," 2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS) IEEE computer Society, no. July 2019, pp. 1258–1269, 2018, doi: 10.1109/ICDCS.2018.00125.

[159]P. Mattson et al., "MLPerf: An Industry Standard Benchmark Suite for Machine Learning Performance," IEEE MICRO, Published by the IEEE Computer Society, no. 0272-1732 2020 IEEE, p. 9, 2020.

[160]Q. Zhang et al., "A Survey on Deep Learning Benchmarks: Do We Still Need New Ones?," in Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Springer, 2019, pp. 36–49. doi: 10.1007/978-3-030-32813-9_5.

[161]P. Jajal et al., "Analysis of Failures and Risks in Deep Learning Model Converters: A Case Study in the ONNX Ecosystem," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2303.17708

[162]A. Paszke et al., "PyTorch: An Imperative Style, High-Performance Deep Learning Library," 2020.

[163]S. Ghemawat et al., "TensorFlow: A system for large-scale machine learning," Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016, p. 21, 2016.

[164]T. Chen et al., Open access to the Proceedings of the 13th USENIX Symposium on Operating Systems Design and Implementation is sponsored by USENIX. TVM: An Automated End-to-End Optimizing Compiler for Deep Learning TVM: An Automated End-to-End Optimizing Compiler for Deep Learning. 2018. [Online]. Available: https://www.usenix.org/conference/osdi18/presentation/chen

[165]I. Soliman, "Build ONNX Runtime from Source on Windows 10," pp. 1–15, 2020.