JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 11(3), SEPTEMBER, 2023
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.net

Systematic Literature Review of Deep Learning models for Computer Vision Applications: Deployment Challenges in Nigeria

Ibrahim Mohammed Abdullahi[1*], Hassan Opotu Siyaka[2], Gambo Asmau Alhaji[3], Danlami Maliki[1], Idris Aji Dauda[1]
[1]Department of Computer Engineering, Federal University of Technology Minna, Nigeria
[,2]Department of Computer Science, Federal Polytechnic Bida, Niger State
[3]Department of Computer Science, Niger State polytechnic Zungeru Niger State

ABSTRACT

Deep learning has gained attention recently. Since its adoption, deep learning has provided state-of-the-art solutions to lots of standing computational problems. One of the areas it has gained unequaled success is computer vision. The success of deep learning is not limited to computer vision only, it has also recorded unmatchable success in areas like natural language processing and speech recognition. With the advent of big data, the use and importance of deep learning can only continue to grow. One downside of this algorithm is its computational requirements: large datasets and high-end computing devices. In this paper, we provide an overview of recent deep learning models for computer vision, and we also highlighted the challenges faced by developing countries in adopting these technologies. No review has covered the challenges faced by Nigeria in deploying this technology. Some of the challenges highlighted include manual data collection and lack of adequate cloud storage services. Inadequate infrastructures such as power and network facilities, and finally, lack of adequate funding of the sector. It was recommended that local cloud services be established to encourage local data storage and reduce storage cost. Also, adequate investment for power and network availability should be made. Finally, there should be enough budget allocation to IT sector that will encourage technocrat and experts to develop and fully harness the benefit of the technology.

ARTICLE INFO
Article History
Received: June 2023
Received in revised form: July, 2023
Accepted: September, 2023
Published online: September, 2023

KEYWORDS
Deep learning, computer vision, Systematic Literature Review, Developing countries

INTRODUCTION

Computer Vision (CV), a scientific field, which started with the desire to automate the process of image analysis and to have computers mimic the vision systems of human, has seen a great deal of advancements and developments over the years. CV is a field which through the development of various methods and techniques, seeks to equip computers with the ability to understand and independently carry out activities which the human visual system can [1]. In the early developmental stages of CV, researchers focused on "low-level" tasks such as edge detection and segmentation. In the process of the development, various algorithms, such as feature extraction algorithms and low-level image processing algorithms were developed to solve the challenges been faced in various computer vision applications [1].However, due to the challenges that comes with the surge in the source, amount and form of data which CV systems have experienced recently, the need for faster performance CV techniques arose, as such there has been an evolution past just the development of programming algorithms, to the adoption of a technology like Deep Learning (DL) in the implementation of the various computer vision systems [2].

DL is a field of machine learning, which makes use of several collective layers of neurons, allowing artificial neural networks to imitate the functioning of the human brain, in the creation of patterns, decision making, as well as in processing data. It is a form of representation learning that is capable of learning and gathering experience from a given set of data (structured or unstructured), without the need for feature extraction by a human being [1]. The adoption of DL in computer vision, provides cutting-edge results in challenging CV applications like: object recognition, image classification and facial recognition [3]. DL provides various schemes, known as models, which are adopted for varying CV applications [4]. Some of the DL schemesapplied in CV include; Restricted Boltzmann Machines (RBM), Recurrent Neural Network (RNN), Deep Belief Networks (DBN), Autoencoders and Convolutional Neural Network (CNN) [5]. The scheme adopted in a computer vision system, depends on the task which such

Corresponding author: Ibrahim Mohammed Abdullahi
✉ amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 11(3), SEPTEMBER, 2023
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.net

a system seeks to carry out.

Over the years, CV has been applied in various fields such as; in agriculture for food quality inspection [6], grading and sorting of fruits and vegetables [7], in the construction industry for structure safety and health monitoring [8] as well as in medicine for medical imaging analysis [1]. There has no doubt, been a reasonable advancement in the application of deep learning models in CV systems. Some researchers as in the case of [9], have even gone ahead to propose the combination of multiple deep learning schemes, to achieve robustness and higher performance in the various computer vision applications. Several works and reviews have been carried out, to define and describe the various deep learning models and to recommend improvements and adjustments that can be made on the various models, in order to obtain better results in terms of performance, for specific CV applications. In recent years, critical reviews like [1][9][10], have focused on outlining and giving an overview, of the various DL schemes that are mostly used in CV systems and applications. There has however, been little to no work done to examine, outline and review the challenges been faced in the adoption of deep learning models for CV systems. To the best of our knowledge, there has also not been a systematic review work that has a wide scope focused analysis of deep learning schemes for CV systems as a whole, as existing works have rather focused on a particular field of computer vision. Consequently, this systematic review will offer a comprehensive overview of the various DL schemes, explaining their areas of applications, at the same time, pay attention to the challenges and issues being faced in the process of their adoption in CV applications, thereby serving as a blueprint to guide upcoming researchers, on deep learning algorithms suitable for particular CV applications and as well, give insight into research opportunities for scholars who would like to conduct research, that would proffer solutions to the various challenges being encountered in the adoption of deep learning schemes for CV systems.

METHODOLOGY

This Systematic Literature Review (SLR) was conducted in accordance with the specifications of PRISMA framework. A search strategy was developed to identify relevant search keywords, search phrases, resources and literature used for the study. Figure 1 shows the entire study methodology.

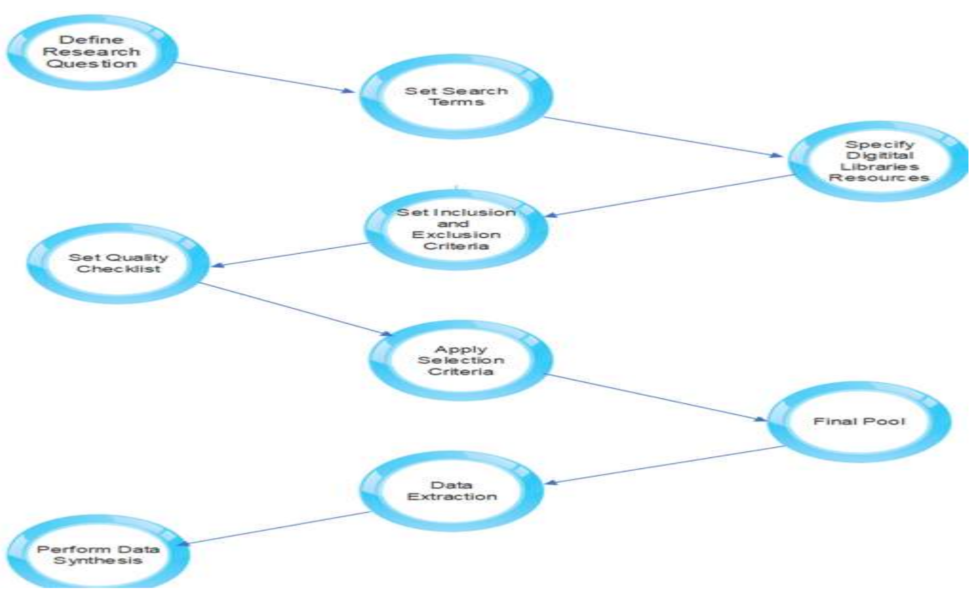

Figure 1: Study methodology

This search process was tailored to Google Scholar, and Web of Science (WoS) databases. The

Corresponding author: Ibrahim Mohammed Abdullahi
amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 11(3), SEPTEMBER, 2023
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.net

search mainly focused on mapping existing literature in the field of DL and then narrowed down the subject area to CV, DL models and the problems of applying these DL models in Nigeria. In Subsection 2.1, the paper outlines the review questions needed to be addressed by this research. Subsection 2.2 outlines the search process, while, Subsection 2.3, explained the primary study selection process and the inclusion and exclusion criteria that was followed in grading the selected papers. In Subsection 2.4, the quality assessment rules were applied. Finally, in Subsection 2.5, the data extracted from each research was presented, and the synthesis of extracted data.

Research Questions

This section presents the research questions needed to be addressed in this review:

Q1. Which DL models are used in CV?

Q2. What are the challenges faced by developing countries such as Nigeria in adopting DL models for CV?

Search Process

The search process that was used in this study is outlined in the following sections:

Resources

The search process is manual and the resources were gotten from the internet. It includes journal, and articles. Two databases were used to get resources, they include: WoS and google scholar.

Search Keywords

The search keywords used in the research process are as follows: “Deep learning”, “Computer vision”, “deep learning models”, “challenges of deep leaning in developing countries”, Convolutional Neural Network, “Deep Belief Networks, Variational Autoencoders, Deep Boltzmann machine, Boltzmann machine and GANS.

Search Terms

We followed the systematic literature review process to construct the search terms of this review. The main terms are gotten from the research questions which are related to Challenges of deep learning for computer vision in developing countries. The search terms consist of the combination of the selected keywords using Boolean operator AND, and the joining of the terms and synonyms using Boolean operator OR as the following:

“deep learning” OR “deep learning models” OR “deep learning architecture” AND “computer vision” AND “developing countries”

(“CNN” OR “GANS” OR “Deep Belief Networks” OR “Boltzmann Machine “OR “Deep Boltzmann Machine” OR “Variational Autoencoders”) AND “developing countries” AND “Computer vision”

(“deep learning” OR “deep structured learning” OR “hierarchical learning”) AND “deep learning challenges” AND (“CNN” OR “GANS” 0R “Deep Belief Networks” OR “Boltzmann Machine “OR “Deep Boltzmann Machine” OR “Variational Autoencoders”) AND “developing countries” AND “Computer Vision”.

Search Phases

The identified search terms were searched through the chosen digital libraries to retrieve the primary related literature. The Boolean operators AND/OR were applied to construct search terms to collect as many results as possible. The search included articles published on Google scholar, WoS. The search took place in February 2021.

Primary Study Selection Process

Figure 2 shows the flowchart of the sequential steps of the search and selection process we followed to obtain relevant and qualified studies as illustrated below: The first step of the search process produced 425 results.

In the next step, we used another filtration process to remove the duplicates and exclude all publications that did not provide the necessary information to answer the research questions.

After that, the inclusion and exclusion criteria and quality assessment rules were applied to the papers to exclude irrelevant publications.

Finally, a quick scan was done on the selected publications to obtain publications that were more related to the primary research topic.

RESULTS

The articles reviewed are in consonant with the current reality in the field of computer vision, acknowledging the recent spike (since 2016) in interest in computer vision applications based on deep learning. In this section we present the summary of our results in two sub sections: descriptive analysis and literature classification.

Descriptive Analysis

Here we show the distribution of past literatures using publication based on: year, journal,

Corresponding author: Ibrahim Mohammed Abdullahi
✉ amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 11(3), SEPTEMBER, 2023
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.net

citation, country. Figure 3 shows that the past few years have witnessed a surge of implementation of deep learning models for computer vision systems. It is noticed that the number of publications has increased since 2016 from a mere 8% to 35% in 2020 and 29% in 2021.

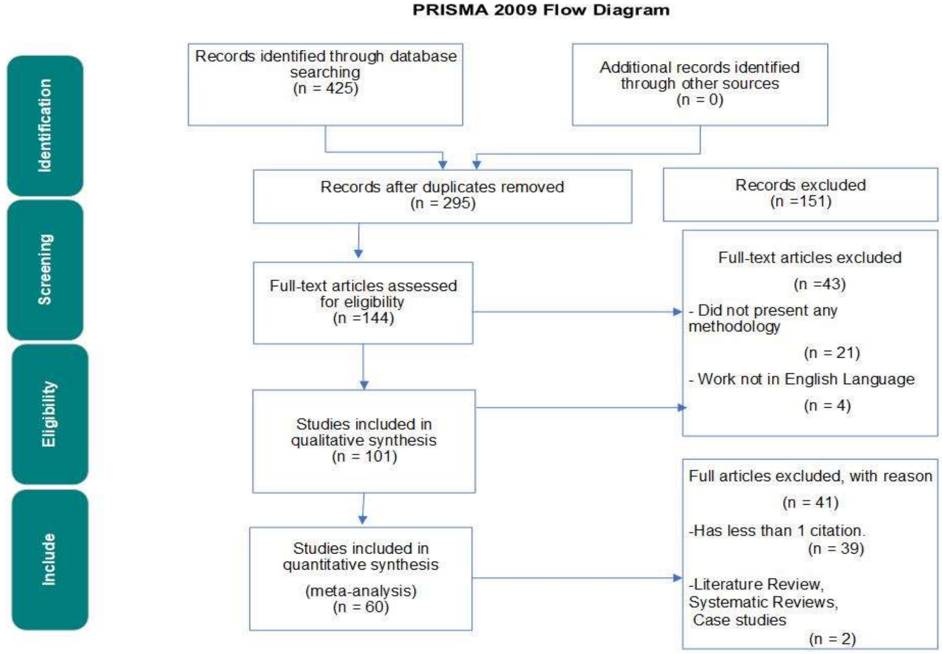

Figure 2: PRISMA Search and Selection process

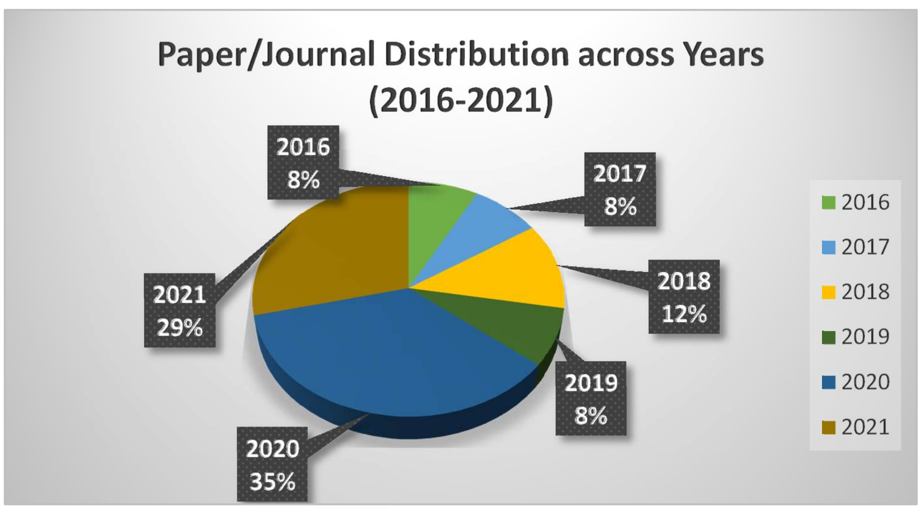

Figure 3: Distribution of Journals across years

LITERATURE CLASSIFICATION

Deep learning models are broadly classified based on the task they are used for: models used for image classification problems (discriminative models) and generative models. In this section we highlight recent deep learning models used in computer vision, their areas of

Corresponding author: Ibrahim Mohammed Abdullahi
✉ amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 11(3), SEPTEMBER, 2023
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.net

application, advantages and disadvantages.

Generative Models

They are unsupervised models (learn hidden and underlying structure of data) that can take input training samples from some distribution and learn a model that represent that distribution. They are used for two main goals: density estimation and sample generation [16].

Variational Autoencoder (VAE)

Like classical autoencoders, VAEs are unsupervised models [11], and are able create a latent space representation of a problem domain. This results in a compressed data (lower-dimensional representation of the high-resolution input data). The ultimate goal of VAE is to reveal the density distribution underlying some data.

VAE is composed of two structures: Encoder, and Decoder.

a. *Encoder:* The input data is mapped to the internal representation of the network. The encoder may have several hidden layers, like convolutional layers, in general.
b. *Decoder:* Tries to reconstruct the input from the network's internal data representation. The decoder may also have a complicated structure that resembles that of the encoder [16].

The benefits of using this model come from its ability to compact high-resolution input data into a lower-dimensional representation and expose the frequency distribution underlying a dataset of interest. The drawback of the algorithm is that it will lose information related to the input data during the encoding and reconstruction processes if it is not properly equipped.

Deep Belief Networks (DBN)

Deep Belief Network (DBN) is a deep neural network model that is made of multiple layers of Restricted Boltzmann Machines (RBM) neural networks [12]. Proposed in 2006 by Geoffrey Hinton, each layer of a DBN is trained independently with each layer encoding the representation of the statistical dependencies from the previous layer [13]. It is classified as a generative statistical model. The DBN is a stack of several Restricted Boltzmann Machines (RBM), It learns through a greedy and stepwise approach. The first RBM in the stack gets trained with a dataset, and through this, gets optimized to reconstruct the original distribution using an algorithm known as a Contrastive Divergence algorithm (CD-k). The output from the reconstruction of the dataset by the first RBM, gets passed on to serve as input for the next RBM. The next RBM gets trained and also produces a reconstruction of its input as output. This output is passed on to the next RBM, and this process continues till all the blocks of RBM in the Deep belief Network have been trained. The contrastive divergence algorithm is a frequently used approach adopted for training energy-based latent variable models which are widely used by Deep Belief Networks and Restricted Boltzmann Machines [13]. Deep belief Networks are able to create different intermediate internal representation of the dataset, these various representations can then be adopted for a variety of purposes. DBNs over time became an attractive choice which is adopted for computer vision applications like; recognition, generation and clustering of images, video sequences and motion-capture data.

Figure 4 shows the graphical representation of the structure of the Deep Belief Network (DBN). It is made by stacking several Restricted Boltzmann Machines (RBM) together. However, unlike the RBM, the nodes of the DBN do not have any form of communication with each other [14]. A node in each layer is connected to all other nodes in the previous and the subsequent layer but not with other nodes in the same layer with it. Also, in the DBN, only the 2 topmost layers have a bi-directional connection, the other layers have a Uni-directional connection with symmetrical weights between them.

Corresponding author: Ibrahim Mohammed Abdullahi
amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 11(3), SEPTEMBER, 2023
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.net

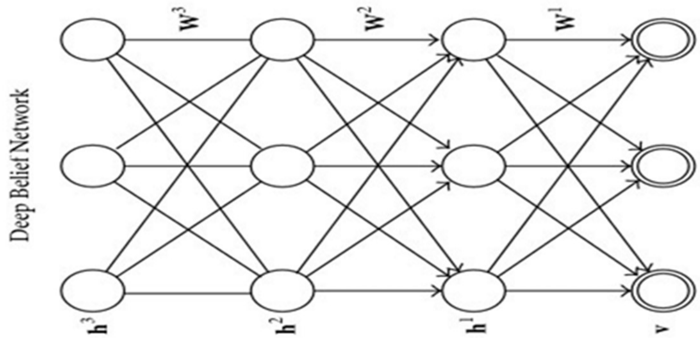

Figure 4: Structure of Deep Belief Network [19]

The main advantage of Deep Belief Network stems from the fact that it is resistant to Vanishing Gradient problem. Vanishing gradient problem arose after the introduction of increase in number of hidden layers in a neural model. Vanishing gradient problem is one in which a neural model learns nothing of the learning process, which is due to a decrease in gradient as it moves from a layer to another, so much that the gradient almost vanishes thus, preventing a realization of an optimal value. The development of Deep Belief Network (DBN) neural model however provides a solution to this. Other advantages of Deep Belief Networks include; small dataset required for learning, shorter training time on GPU powered machines, more accurate than the shallow nets. One major challenge that is however faced during adoption of Deep Belief Networks is, deciding the right number of RBM units needed. The number of RBM units adopted is a crucial factor that affects the success of the training process [14].

Generative Adversarial Networks (GANs)

GANs are generative models generated by pitting two neural networks against each other. The GANs' learning process involves concurrently training a discriminator D and a generator G [15]. G is a generator that creates images using features learned from the training dataset. D, the discriminator, predicts whether G's image is real. This state-of-the-art generative model was recently implemented in [15].

A. GAN Architecture

The GAN architecture is composed of two components:

i. Generator: tries to convert random noise into observations that look as if they have been sampled from the original dataset.
ii. Discriminator: tries to predict whether an observation comes from the original dataset or is one of the generator's forgeries.

From the architecture shown in Figure 5, the steps GAN takes are:

i. The generator takes in random numbers and returns an image.
ii. This generated image is fed into the discriminator alongside a stream of images taken from the actual, ground-truth dataset.
iii. The discriminator considers both true and false images and returns probabilities, which range from 0 to 1, with 1 indicating authenticity and 0 indicating fake.

B. Training the GAN

The discriminator is being trained to improve as a classifier in order to increase the likelihood of correctly labeling both training examples (real) and images produced by the generator (fake). On the other hand, the generator is being conditioned to become a better forger in order to improve its chances of fooling the discriminator. Thanks to their adversarial existence, both networks are improving at what they do [16, 17].

Training the GAN involves the following as shown in Figure 6:

i. Train the discriminator: This is a simple supervised training procedure. When given labeled images from the generator (false) and training data (real), the network learns to differentiate between real and fake images using a sigmoid prediction output.
ii. Train the generator: this is a tricky operation. Unlike the discriminator, the generator model cannot be learned on its own. It requires the discriminator model to tell it whether or not it did a good job of

Corresponding author: Ibrahim Mohammed Abdullahi
amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.

faking its picture. As a result, a hybrid network made up of both G and D models is used to train the generator [17].

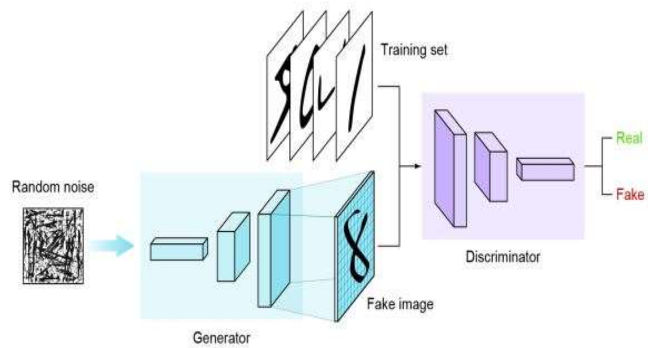

Figure 5: GAN architecture [17]

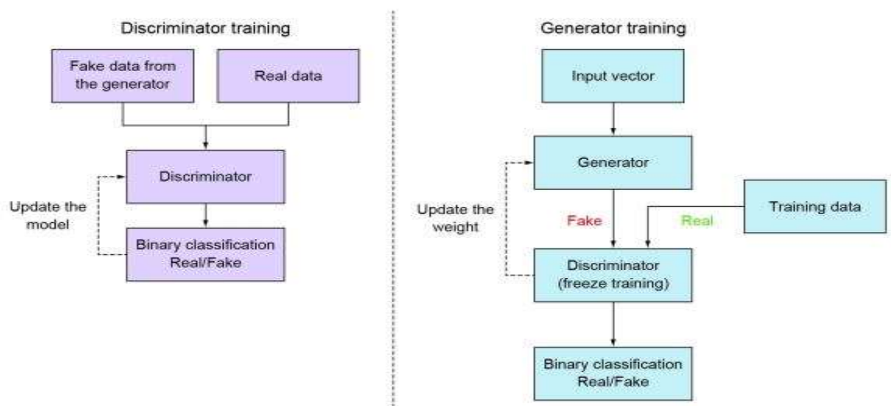

Figure 6: Training Generative Adversarial Network [17]

C. GAN Minimax Function

It's more of a zero-sum game than an optimization problem when it comes to GAN training. The total utility score is divided among the players in zero-sum games. A rise in one player's score causes a decrease in the score of another player [16]. This is known as a minimax game theory in AI. Minimax is a decision-making algorithm that is commonly used in two-player turn-based games. The algorithm's aim is to find the best next move. In which one player is known as the maximizer, who strives for the highest possible score, and the other is known as the minimizer, who strives for the lowest possible score by attempting to counter shift the maximizer. GANs plays a minmax game, where the entire network attempts to optimize the function V (D, G) [17].

Discriminative Models

A discriminative model tries to deduce conclusion from observed data. A classification model will attempt to predict the value of one or more outcomes given one or more inputs. Labels that can be applied to a dataset are called outcomes [18]. Unlike the generative models presented so far, discriminative models do not generate output, given input, they rather predict one of many output probabilities.

Corresponding author: Ibrahim Mohammed Abdullahi
✉ amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 11(3), SEPTEMBER, 2023
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.net

Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNN) are one of the most well-known deep learning techniques, in which several layers are robustly trained [19]. CNN has proven to be extremely powerful and it's the most widely used in a variety of computer vision applications. Below are CNN properties that makes it standout:

i. They connect neurons that only correspond to pixels in the image's immediate vicinity. The neurons are "forced" to only take feedback from neurons that are spatially similar to them in this way. Since not all neurons are related, this decreases the number of weights.

ii. Parameter sharing is used by a CNN. In other words, all neurons in a layer share a small number of weights. This further reduces the number of weights and helps fight over fitting.

A forward stage and a backward stage are both used to train the network. The forward stage's primary objective is to represent the input image in each layer using the current parameters (weights and bias). The loss cost is then measured using the forecast output and the ground truth labels. Second, the backward stage uses chain rules to calculate the gradients of each parameter based on the loss cost. The gradients are used to update all of the parameters, which are then prepared for the next forward computation [20]. Figure 7 shows the CNN Architecture.

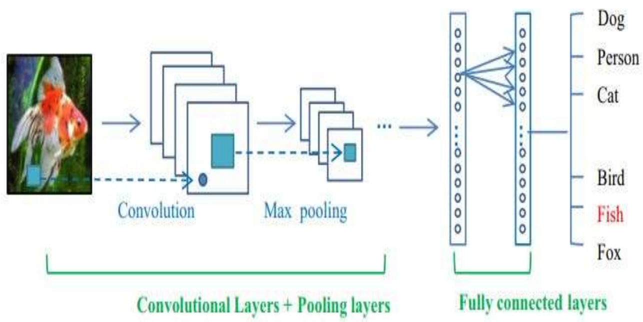

Figure 7: CNN architecture [20]

A. Convolutional Layer

The most critical component of a CNN is the convolutional layer. It is made up of a series of filters (also known as kernels or feature detectors), each of which is applied to all areas of the input data. A filter is defined by a set of weights that can be learned. The following picture, which is a nod to the topic at hand, exemplifies this very well:

Convolution is a mathematical procedure that combines two functions to create a third modified function. The first function is the input image, and the second function is the CONV filter, to illustrate this concept in the sense of CNNs. To create a changed image with new pixel values, we'll use some mathematical operations [21]. Kernel is another name for the convolution filter. It's a weighted matrix that slides over an image to extract information. The CONV filter's dimensions are referred to as the Kernel size (width x height).

B. Pooling Layer

A pooling layer splits the input slice into a grid, where each grid cell represents a receptive area of a number of neurons (just as a convolutional layer does). Then, a pooling operation is applied over each cell of the grid. Pooling layers exits in number of shapes and sizes. Since the pooling process is conducted separately, the volume depth is not affected by the pooling layers on any slice. The most common method of pooling is *max pooling*.

The max pooling operation selects the neuron in each local receptive field (grid cell) with the highest activation value and propagates only that value forward. With a receptive field of 2x2, we can see an example of max pooling in the diagram below [19, 22].

The convolutional layers' dimensionality is reduced by the pooling layers. The importance of reducing dimensionality is due to the fact that in complex projects, CNNs can have several CONV layers, each with tens or hundreds of convolutional filters (kernels). As a result, adding POOL layers would allow us to retain the important features and move them on to the next layer while simultaneously reducing the image dimensionality.

C. Fully Connected Layers

After going through the feature learning process with the picture, the extracted features from this image are placed together in a long tube of features using

Corresponding author: Ibrahim Mohammed Abdullahi
amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 11(3), SEPTEMBER, 2023
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.net

the CONV+POOL layers. The image is categorized by the classification layer based on the existence of certain features. MLPs work great in classification problems.

The key benefit of CNN over its predecessors is that it identifies essential features without the need for human interaction. Given a large number of pictures of cats and dogs, it can learn distinct features for each class on its own. In addition, CNN is computationally efficient. It performs parameter sharing and uses special convolution and pooling operations. CNN models will now run on any computer, making them universally appealing [23].

CHALLENGES IN IMPLEMENTATION OF DEEP LEARNING MODELS IN COMPUTER VISION SYSTEMS

Several works have been carried out to ensure that processes, procedures, machines and systems have some component of intelligence. This includes the ability to learn and gather knowledge from experience and as well apply the acquired knowledge to improve and probably automate the same process, procedure and system. Several approaches, frameworks and methodologies have been proposed, all in a bid to achieve the goal of 'intelligence'. A trend that can be noticed from the systematic review of the various works that have been carried out is that, each work seeks to address a challenge, limitation or restraint of existing methods. This has led to the development of domain-generic and domain-specific frameworks and methodologies.

However, in some regions of the world, such as in a developing nation like Nigeria, a lot of challenges are being faced in a bid to adopt these existing deep learning frameworks for computer vision applications. These challenges can be placed in one of three classes, which are; infrastructural, budgetary and management, and finally, dataset and model training challenges.

a. Infrastructure Challenges:

i. Network Bandwidth, Latency and Coverage

In the implementation of any deep learning model, there comes a point where a decision is made on whether to adopt an on-premise or off-premise infrastructure approach [24]. On-premise deployment is an approach to deployment of deep learning that involves setting up and running of the deep learning system within the confines of your own organization, meaning that such organization has total control of the infrastructural setup [25]. On the other hand, is the Off-premise approach, which involves adopting cloud computing strategies, whereby a third-party provider hosts all of the hardware and infrastructure needed, all that is required of the organization deploying the deep learning-based application is to pay on 'as-needed' basis [25]. It is this approach to deployment of deep learning models that is mostly affected by the challenges of Network Bandwidth, Latency and Coverage. Assume there is a deep learning-based application that involves streaming of video feed as data to the cloud-based model, there is need for reliable and stable internet connection which provides needed bandwidth for the application, with as little latency as possible. However, in a country like Nigeria, there are several places that do not have any form of Internet connectivity, while some places that do have the Internet connectivity, do not have access to the required network bandwidth with required latency for their deep learning-based application.

ii. Power Supply

This is a challenge that affects mostly, the on-premise infrastructure deployment of deep learning-based applications. This is because there is the need to provide stable power supply for whatever hardware infrastructures have been adopted for the deep learning-based application. However, according to [26], average per capita electricity consumption per person in Nigeria is 144kWh, which is a miserly value when compared the 14,612kWh of Canada. The value shows that the challenge of power supply is a big one and can frustrate and inhibit the successful implementation of deep learning-based applications.

b. Budgetary and Management Challenges:

i. Management Complexity and Limitations in skillset

The adoption and implementation of deep learning by an organization brings about the need to put in place procedures (that define the use), strategies (that guide the further development of the application) and most importantly skillsets (that actually implement the procedures and application). However, in a country like Nigeria, there is the issue of brain drain where individuals with skillsets needed for particular deep learning-based applications are not readily available in the country, brought about by the fact that individuals with the required skillset have gone out of the country in search for more favorable professional opportunities[27].

ii. Financial and Budgetary Limitations.

As stated earlier on, there is the need for adoption of some form of hardware or infrastructures for

Corresponding author: Ibrahim Mohammed Abdullahi
✉ amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 11(3), SEPTEMBER, 2023
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.net

whatever deep learning-based application. Also, there is the need to employ and pay individuals with the skillsets to implement the deep learning-based application. However, in a country like Nigeria, there is so little budgetary allocation made for the Information Technology (IT) sector, as such, it becomes increasingly difficult to obtain the necessary infrastructures and skillset.

c. Dataset and Model Training Challenges

i. Data Capturing

In any deep learning-based application, the adopted deep learning model needs to obtain experience from a set of data through a process known as training [28, 29]. The dataset with which the deep learning model is trained is expected to be similar in nature to the type of dataset which the model is expected to work on when finally deployed. However, in developing countries, Nigeria for instance, obtaining the dataset needed for training a deep learning model is often very difficult. This is caused by the fact that there are few and sometimes no central and reliable systems in place for data capturing, saving and retrieving. The state of lack of accurate data in Nigeria is such that there are only several estimates of the population size, gender distribution and several other statistics. The last census in the country was conducted in 2006, with plans for another one, stalling several times due to lack of funding [30]. As such it becomes hard to obtain some required dataset needed to train a deep learning model for a particular application.

ii. Manual Data Collection

Most data collection process in Nigeria is manual in nature, cumbersome and in-accurate. This has affected the integrity of the dataset collected in most organizations in charge of data collection.

iii. Lack of Local Cloud Services

In Nigeria, the unavailability of reliable cloud data storage facilities hinders the storage of data and its accessibility. International cloud services are often very expensive.

iv. Data Security and Data Privacy.

Data security and Privacy are terms that are used interchangeably these days caused by the increasing attention being paid to protection of generated user data. Data security is concerned with providing protection for data from any form of external compromises that could be caused by external attackers or even malicious insiders. On the other hand, Data privacy is concerned with defining guidelines for how data is collected, shared and used [31] [32]. In a country like Nigeria, several citizens do not readily entrust their data to third-party for security and privacy reasons. As such it becomes increasingly difficult to obtain data needed for training of models that are resistant to the issue of overfitting and poor end results.

CONCLUSION

The main objective of this papers is to examine the relevance and adoption of deep learning models for computer vision applications as well as its inherent challenges for developing countries such as Nigeria. To achieve this objective, systematic literature review was carried out in accordance with the specifications of PRISMA framework. A search strategy was developed to identify relevant search keywords, search phrases, resources and literature used for the study. This search process was tailored to Google Scholar, and web of science databases. The search mainly focused on mapping existing literature in the field of deep learning. In this paper, an overview of recent deep learning models for computer vision and the challenges faced by developing countries in adopting in the reviewed models were highlighted. It was discovered that the key challenges hindering the adaptation of DL models for CV applications in Nigeria include; Data collection and privacy, hardware and infrastructure, budget and management challenges.

To fully harness the benefit of this technology in developing countries such as Nigeria, the following are recommended:

i. Data capturing and archiving should be taking given high priority. Organizations and institutions should see data collection as a must if the technology will be harnessed. Also, organizations like National Information Technology Development Agency (NITDA) and its subsidiaries must provide local cloud services for reliable data storage. Furthermore, the security and integrity of the data must be guaranteed.

ii. Digitizing data collection will play a long role in easing data collection process and improving the integrity of the data.

iii. Adequate network and power infrastructures must be put in place to provide the required need for adequate deployment of the technology.

Adequate budget and financial provision must be

Corresponding author: Ibrahim Mohammed Abdullahi
✉ amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 11(3), SEPTEMBER, 2023
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.net

made for IT sector to encourage both manpower and technical experts to develop the sector and fully exploit the benefits of the technology

REFERENCE

[1] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, 'Deep Learning for Computer Vision: A Brief Review', *Comput. Intell. Neurosci.*, vol. 2018, 2018, doi: 10.1155/2018/7068349.

[2] J. Brownlee, *Deep Learning for Computer Vision Image Classification , Object Detection , and Face Recognition in Python UNLOCK Computer Vision With Deep Learning*. 2019.

[3] S. K. Shetty and A. Siddiqa, 'Deep Learning Algorithms and Applications in Computer Vision', *Int. J. Comput. Sci. Eng.*, vol. 7, no. 7, pp. 195–201, Jul. 2019, doi: 10.26438/ijcse/v7i7.195201.

[4] J. F. S. Gomes and F. R. Leta, 'Applications of computer vision techniques in the agriculture and food industry: A review', *European Food Research and Technology*, vol. 235, no. 6. pp. 989–1000, Dec. 2012, doi: 10.1007/s00217-012-1844-2.

[5] M. Radhakrishnan and A. El, 'Application of Computer Vision Technique on Sorting and Grading of Fruits and Vegetables Development of atmospheric pressure plasma system for food processing application View project Development of Non thermal plasma system for liquid food sterilisation ', *J Food Process Technol*, p. 5, 2012, doi: 10.4172/2157-7110.S1-001.

[6] J. Seo, S. Han, S. Lee, H. K.-A. E. Informatics, and undefined 2015, 'Computer vision techniques for construction safety and health monitoring', *Elsevier*.

[7] D. Khemasuwan, J. S. Sorensen, and H. G. Colt, 'Artificial intelligence in pulmonary medicine: computer vision, predictive model and COVID-19', *Eur Respir. Soc*, doi: 10.1183/16000617.0181-2020.

[8] Y. Sun, X. Wang, and X. Tang, 'Hybrid Deep Learning for Face Verification', *openaccess.thecvf.com*, 2013, doi: 10.1109/ICCV.2013.188.

[9] R. K. Sinha, R. Pandey, and R. Pattnaik, 'Deep learning for computer vision tasks: A review', 2018.

[10] J. Guo *et al.*, 'GluonCV and GluonNLP: Deep learning in computer vision and natural language processing', 2019.

[11] X. Hou, L. Shen, K. Sun, G. Q.-2017 I. W. C. on, and undefined 2017, 'Deep feature consistent variational autoencoder', *ieeexplore.ieee.org*, Accessed: Mar. 28, 2021. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/7926714/.

[12] A. Khan, A. Zameer, T. Jamal, and A. Raza, 'Deep belief networks based feature generation and regression for predicting wind power', *arXiv*. arXiv, Jul. 2018.

[13] N. Lopes and B. Ribeiro, 'Deep Belief Networks (DBNs)', 2015, pp. 155–186.

[14] Y. Qiu, L. Zhang, and X. Wang, 'Unbiased contrastive divergence algorithm for training energy-based latent variable models'.

[15] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, 'Generative Adversarial Networks: An Overview', 2018. doi: 10.1109/MSP.2017.2765202.

[16] M. Y. Liu and O. Tuzel, 'Coupled generative adversarial networks', 2016. Accessed: Mar. 27, 2021. [Online]. Available: https://arxiv.org/abs/1611.02163.

[17] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, 'Spectral normalization for generative adversarial networks', 2018. Accessed: Mar. 27, 2021. [Online]. Available: https://github.com/pfnet-research/sngan_.

[18] Analytics India Magazine. (2020, September 8). 7 Types of Classification Algorithms. Retrieved April 2, 2021, from https://analyticsindiamag.com/7-types-classification-algorithms/.

[19] Y. LeCun, L. Bottou, Y. Bengio, et al., Gradient-based learning applied to document recognition, Proc. IEEE 86 (11) (1998) 2278–2324. https://analyticsindiamag.com/7-types-classification-algorithms/ (accessed Apr. 02, 2021).

[20] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, vol. 187, pp. 27–48, 2016, doi: 10.1016/j.neucom.2015.09.116.

[21] M. A. Ranzato, "Deep Learning for Vision," no. October, p. 458, 2013.

[22] R. Patel and S. Patel, "A comprehensive study of applying convolutional neural network for computer vision," *Int. J. Adv. Sci. Technol.*, vol. 29, no. 6 Special Issue, pp. 2161–2174, 2020, Accessed: Mar. 12, 2021. [Online]. Available: https://www.researchgate.net/publication/344121826.

[23] "Applied Deep Learning - Part 4: Convolutional Neural Networks | by Arden Dertat | Towards Data Science." https://towardsdatascience.com/applied-deep-learning-part-4-convolutional-neural-networks-584bc134c1e2 (accessed Apr. 02, 2021).

[24] J. Villa, "Choosing Your Deep Learning

Corresponding author: Ibrahim Mohammed Abdullahi
✉ amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.

JOURNAL OF SCIENCE TECHNOLOGY AND EDUCATION 11(3), SEPTEMBER, 2023
ISSN: 2277-0011; Journal homepage: www.atbuftejoste.net

Infrastructure: The Cloud vs. On-Prem Debate," 2018. https://determined.ai/blog/cloud-v-onprem/ (accessed Mar. 28, 2021).

[25] M. Zwicker, "SaaS vs On-Premise vs Off-Premise," 2018. https://info.iointegration.com/blog/saas-vs-on-premise-vs-off-premise (accessed Mar. 28, 2021).

[26] Wikipedia, "List of Countries by GDP," 2016. https://en.wikipedia.org/wiki/List_of_countries_by_GDP_(PPP)_per_capita (accessed Mar. 28, 2021).

[27] M. Leslie, "Brain drain.," *Science of aging knowledge environment : SAGE KE*, 2005. https://www.investopedia.com/terms/b/brain_drain.asp (accessed Mar. 28, 2021).

[28] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, no. 7, pp. 1527–1554, 2006, doi: 10.1162/neco.2006.18.7.1527.

[29] T. Amaratunga and T. Amaratunga, "What Is Deep Learning?," *Deep Learning on Windows*, 2021. https://machinelearningmastery.com/what-is-deep-learning/ (accessed Mar. 28, 2021).

[30] K. Yomi, "Nigeria struggles with citizen data, Buhari criticizes World Bank — Quartz Africa," 2019. https://qz.com/africa/1725537/africa-has-a-data-problem-says-mo-ibrahim-governance-report/ (accessed Mar. 28, 2021).

[31] Jeff Patters, "Data Privacy Guide: Definitions, Explanations and Legislation | Varonis," *Varonis*, 2020. https://www.varonis.com/blog/data-privacy/ (accessed Mar. 28, 2021).

[32] A. Wilson, "A Brief Introduction to Supervised Learning | by Aidan Wilson | Towards Data Science," *Towards Data Science*, 2019. https://towardsdatascience.com/a-brief-introduction-to-supervised-learning-54a3e3932590 (accessed Mar. 28, 2021).

Corresponding author: Ibrahim Mohammed Abdullahi
✉ amibrahim@futminna.edu.ng
Department of Computer Science Federal Polytechnic, Bida, Niger State.