

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/351603970>

# Machine Learning in Computer Vision: A Review

Article in EAI Endorsed Transactions on Scalable Information Systems · April 2021

DOI: 10.4108/eai.21-4-2021.169418

CITATIONS

142

READS

3,819

3 authors:



**Dr-Abdullah Ayub Khan Ph.D**

Bahria University Karachi Campus

130 PUBLICATIONS 2,811 CITATIONS

SEE PROFILE



**Asif Ali Laghari**

Sindh Madressatul Islam University

246 PUBLICATIONS 6,085 CITATIONS

SEE PROFILE



**Shafique Ahmed Awan**

Benazir Bhutto Shaheed University Lyari (BBSUL) Karachi

18 PUBLICATIONS 277 CITATIONS

SEE PROFILE

## Machine Learning in Computer Vision: A Review

Abdullah Ayub Khan<sup>1,2</sup>, Asif Ali Laghari<sup>1,\*</sup>, and Shafique Ahmed Awan<sup>2</sup>

<sup>1</sup>Faculty of Computer Science, Sindh Madressatul Islam University, Karachi, Sindh, Pakistan.

<sup>2</sup>Faculty of Computing Science and Information Technology, Benazir Bhutto Shaheed University Lyari, Karachi, Pakistan.

### Abstract

**INTRODUCTION:** Due to the advancement in the field of Artificial Intelligence (AI), the ability to tackle entire problems of machine intelligence. Nowadays, Machine learning (ML) is becoming a hot topic due to the direct training of machines with less interaction with a human. The scenario of manual feeding of the machine is changed in the modern era, it will learn automatically. Supervised and unsupervised ML techniques are used as a distinct purpose like feature extraction, pattern recognition, object detection, and classification.

**OBJECTIVES:** In Computer Vision (CV), ML performs a significant role to extract crucial information from images. CV successfully contributes to multiple domains, surveillance system, optical character recognition, robotics, suspect detection, and many more. The direction of CV research is going toward healthcare realm, medical imaging (MI) is the emerging technology, play a vital role to enhance image quality and recognized critical features of binary medical image, covert original image into grayscale and set the threshold values for segmentation.

**CONTRIBUTION:** This paper will address the importance of machine learning, state-of-the-art, and how ML is utilized in computer vision and image processing. This survey will provide details about the type of tools and applications, datasets, and techniques. Limitations of previous work and challenges of future work also discussed. Further, we identify and discuss a set of open issues yet to be addressed, for efficiently applying of ML in Computer vision and image process.

**METHODS, RESULTS, AND CONCLUSION:** In this review paper, we have discussed the techniques and various types of supervised and unsupervised algorithms of ML, general overview of image processing and the results based on the impact; neural network enabled models, limitations, tools and application of CV, moreover, highlight the critical open research areas of ML in CV.

**Keywords:** Machine Learning, Computer Vision, Supervised and Unsupervised Learning, Medical Imaging, Pattern Recognition, Feature Extraction, Neural Network.

Received on 19 January 2021, accepted on 17 April 2021, published on 21 April 2021.

Copyright © 2021 Abdullah Ayub Khan *et al.*, licensed to EAI. This is an open access article distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/3.0/>), which permits unlimited use, distribution and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.21-4-2021.169418

### 1. Introduction

Nowadays, artificial intelligence is an emerging field throughout the world, a vast area with diverse definitions, few researchers elaborate AI as “a system that thinks like humans” [1] while others indicate “a system that acts like humans” [2]. There is no proper formal rationale of AI; it may vary from a distinct environment [3]. It plays a significant role in many aspects of computing technology, but there is a slight difference between weak and strong AI. According to Keng Siau et.al, machine intelligence

can measure two main aspects of AI, “Weak only focus on the narrow task and strong targets more than one area” [4]. Several applications running live in various environments and Google are the popular example of an artificial intelligence crawler search (content-based search) mechanism. In gaming technology, an artificial program that defeats the world champion chess player (Garry Kasparov) in a row of the six-game matches (Deep Blue Machine) [5]. Similarly, machine speaks and understand what human writes, retrieve crucial information by perceiving things through the camera, control highly reactive radio frequency, medical assisted

\*Corresponding author. Email: asifalilaghari@gmail.com

(expert system: MYCIN), self-driving car, and humanoids.

Whereas machine learning is the subfield of AI [6]. The machine can learn automatically (train) from a given dataset and it will be able to take a proper decision (test) by itself [7]. In ML, the dataset is the main key checkpoint contains millions of unique records [8]. Data categories into two main formats, either it may qualitative or quantitative. Qualitative data collected through interviews, observation, and written documents [9], while quantitative have numerical records which can be measured for statically analysis [10]. Qualitative means machine can recognize the shape of different objects, color, size, weight, models, etc.

ML can classify into three most important learning techniques: (i) supervised (SL), (ii) unsupervised (USL), and (iii) semi-supervised learning (SSL). SL, function to labelled training data, used to prediction. Classification and regression come under supervised learning; popular algorithms are Support Vector Machine (SVM), KNN, logistic regression, Decision tree, and random forest [11]. On another end, USL has unlabelled training data, used to extract features from the input data [12].

Clustering means a group of similar data or records and it is the part of unsupervised. If we talk about the clustering algorithm then only one thing popup in our mind, which is K-mean, but there are varied more like hierarchical clustering and DBSCAN, etc. At last, SSL is the collaboration of both SL and USL, it has a small amount of labelled and large amount of unlabelled data, mainly used in information retrieval, image processing, and bio-information [12]. In this process, first, we apply to cluster on an unlabelled dataset for group same data after labelling those data with the help of classification. Preferred algorithms are continuity, cluster, and manifold assumptions.

At last, SSL is the collaboration of both SL and USL, it has a small amount of labelled and large amount of unlabelled data, mainly used in information retrieval, image processing, and bio-information [13]. In this process, first, we apply to cluster on an unlabelled dataset for group same data after labelling those data with the help of classification. Preferred algorithms are continuity, cluster, and manifold assumptions.

The journey of computer vision (CV) started in 1960, where father of computer vision Larry Roberts proposed 3D geometrical information extraction from 2D perspective polyhedral concepts in his Ph.D. thesis at MIT [13]. At the initial stage, CV illustrates, extraction of crucial information through digital images using computational models. Elaborating dual goals, vision is used as an autonomous system for the engineering point of views just like a human can perform a visual task, while computational models applied in the human biological system for detecting symptoms of diseases in the body [14].

From an engineering perspective, CV has successfully applied to an educational environment where attendance monitoring systems used to get students' attendance

automatically by face detection and recognition through camera [15]. Unmanned Aerial Vehicle (UAV) is a vision-based agent using surveillance of unwanted situation without a human pilot, Kanellakis et.al wrote a survey article based on current development and latest application on UAV systems [16]. In a biological context, medical imaging (MI) is an emerging field, and the number of research is going to increase day-by-day for the betterment of society.

The purpose of MI is to extract essential information from the medical image, but at a certain stage lack of quality of images can create another challenging aspect for all researchers. Kesner et.al highlights the concepts of digitizing and application of MI data in the innovation age [17]. MI is the costly process, automatic extraction of the labelled image with NLP from the radiology report for maintaining expert annotation [18]. In a recent report of McKinsey and Company, New York USA, author Alan Alexander et.al set the future direction of MI in 2021, where technology cluster growth rate, cutting edge practice, and the implication is the key aspect of future blockchain innovations [19]. At the same time, technology has little limitations in distinct areas which are discussed below:

- In an agricultural viewpoint, farming through robots using CV and ML techniques for increasing productivity with quality. Encounter countless agricultural constraints: object color, size, texture, and reflectance properties [20].
- Several applications and research available on handwritten and documented script text recognition but detecting real-time image e.g., street view image and recognize each character properly of the image is still the rising problem for machine learning [21].
- A smaller number of image dataset is one of the limitations in ML where training and testing can only be performed accurately when providing a huge amount of labelled or unlabelled data.

This paper will address the importance of machine learning and deep learning, state of art, and how ML is utilized in computer vision and image processing. This survey will provide details about the type of tools and applications, datasets, and techniques. Limitations of previous work and challenges of future work also discussed. Further, we identify and discuss a set of open issues yet to be addressed, for efficiently applying of ML in Computer vision and image process.

The paper organized in 8 sections and section 2 is based on Machine learning techniques for image processing. In section 3 we provide details of Neural network models, section 4 is based on the tools and applications and section 5 provides information about datasets for computer vision and image processing. Similarly, section 6 based on the challenges and limitations of machine learning research and section 7

provides open research issues for future work. Finally, in section 8 we conclude this review paper.

## 2. Machine Learning Techniques for Image Processing

Image processing (IP) means process image through a digital computer. Utilization of algorithms for enhanced quality of an image, on the bases of this system, can extract crucial information from the image [22, 23]. There are two main types namely: analogous and digital image processing (DIP) [24, 25]. In IP, get input as an image and after process system retrieves output in the form of an image. Mentioning the three important key factors of IP which are discussed below:

- Start with the image acquisition
- Manipulating and analyzing the image
- Output in the form of an image which is based on analyzing

In a digital computer, everything stores in the form of binary, storage of an image in pixels (raster), each pixel represents some numeric values of the two-dimensional array (i.e. x means row and y mean column). In analogue IP, processing can be applying on printout image, photographs mean any hard copies, utilize fundamental visual techniques to analyze the image. While at the other end, DIP manipulates digital images by simply follow three basic rules: (i) pre-processing, (ii) enhancement, and (iii) extraction of the information displayed through digital computer [26].

Phases of IP starts with image acquisition, it is the first step of IP were applying to scale on load image (convert RGB into gray-scale) [27, 28]. Image enhancement technique facilitates to extract hidden detail from the input image. Image restoration deals with the degradation on the bases of the mathematical or probabilistic model. Color image processing is another phase of IP where it handles the color model (pseudo-color and full color) of IP. Wavelets and multi-resolution processing used to check images from various levels of degree [29]. After completion angle rotation, shrink size and resolution of the image, this is known as compression.

Morphological processing is one of the pivotal steps of IP and excerpt information of components and sends it to segmentation phases were partitioning the image into objects (the difficult task is to segment an autonomous picture) [30, 102]. These objects represent the processed data transform from the solution space of segmentation [30, 103]. At last, object detection and recognition assign a label for those objects which use in future processing. In DIP, digitize analogue image or video signal, digitization is the process of converting continuous into a digital image, it can only be done when image function (x, y) digitized both spatial and in amplitude. Sampling refers to take a sample and determine the spatial resolution of an

image, there are two main parts of sampling (i) up-sampling and (ii) down-sampling [31]. Variation occurs in the signal randomly due to noise. In sampling, the process of reducing noise is to get more samples [32]. The process of quantization is opposite from sampling (digitize the amplitudes) and the level of gray in the scale determines the level of quantization [31]. There are three main types of images use to process digitally: (i) binary image, (ii) gray-scale, and (iii) color image, further explain in the next paragraphs.

ML is the imperative factor of IP, unsupervised used for feature extraction, and supervised used to label objects for detection and recognition [33, 34]. In unsupervised ML, researchers share various research publicly related to the elicitation of critical features in an image. Segmentation applied through the process of histogram then clustering. In clustering, color-based segmentation only done by Fuzzy C, where similar color pixels group together means cluster, similarly, segment texture image through the K-mean algorithm [35]. Expectation Maximize algorithm's parameters based on unsupervised operation in model-based segmentation [35].

Several real-time IP applications running on unsupervised ML techniques one work provide by Zhong et.al successfully build computational intelligence in optical remote sensing using IP clustering [36]. Ghosh et.al elaborate unsupervised changes detection of remote sensing images and compare two images of the same geographical area using a fuzzy clustering algorithm [37, 38]. Moreover, it is widely used in the CV medical imaging field where it helps the system to diagnose disease; manual assessment of diagnosing malaria is the long step process and prone to human error. Purwa et.al developed unsupervised automatic detection of sensitive malaria screening techniques [39].

It is not the stopping point of unsupervised learning, but there are more areas cover with the real-time application. Supervised Learning is learning where we train or test machines using the labelled dataset, the label means already tagged data for classification purposes. In supervised learning, Support Vector Machine (SVM) widely use in NLP for recognizing handwritten and scan text images [40]. Classify cell biology to recognize phenotypes in medical imaging [41]. In robotics, supervised machine learning plays a vital role in the visual perception of forest trail for mobile robots [42]. Erickson et.al labels "benign" and "malignant" to detect brain tumour medical image [43]. Tuia et.al emphasis supervised learning in remote sensing image classification [44].

## 3. Neural Network (NN) Based Models

Deep learning (DL) is the newest innovation of researchers, but the fact is DL is the subfield of ML; it is just the model name of neural networks [45, 104]. The idea of NN is popup by the concept of biological neurons of an individual. McCulloch and Walter Pitts derived

human neurons into artificial neurons from the very first time in 1943 [46], the notion of perception in NN given by Frank Rosenblatt, a probabilistic model for information storage, funded by US Naval and successfully implemented at Cornell Aeronautical Research Lab in 1958 [47]. The basic building block of NN; computational model working on parallel with no centralized control unit, weight is the primary mean of storing information in neurons for the long term and amend the weight then updating information storage of NN as well. The architecture of NN is based on three main parts: (i) number of neurons (ii) number of layers and (iii) connections between layers [48].

Feedforward neural networks (FFNN) is an acyclic network and input values move forward from input nodes to hidden, then hidden to output neurons. Feedforward has a single and multilayer perceptron, single layer feedforward perceptron (SFFP) or linear threshold is typically based on a single output layer [49]. It contains an aggregate of input values with weights and activation function of artificial neurons only fire when the threshold value is zero (0). If the value of threshold is less than zero (-1) then add parameter known as bias for stabilizing the output otherwise model goes deactivated. Bias is the constant intercept ( $\text{output} = \text{sum}(\text{input} * \text{weight}) + \text{bias}$ ) [1] add to the product of aggregate input and weight for setting the threshold value of activation function. In a multilayer feedforward perceptron (MFFP), nodes of each layer are fully connected to sub layers of others to make a network; means there is an input layer, one or more hidden layers and a single output layer [50].

Each layer has a different number of neurons and connected with the same as an acyclic graph. The backpropagation learning algorithm, the essential task is to reduce the error rate by tuning the weight of NN [25]. In the phase of feedforward, the model cannot work properly and produce error then backpropagation goes back in the previous iteration for retaining the probabilistic connecting weight for minimizing error value. For optimization algorithms, gradient descent is the iterative process that minimizes loss function by updating the parameter of the model (parameter refers to the weight of neural nets) [51]. In gradient descent, the learning rate is the size of those steps, which we take to reduce the rate of error. If we take a high learning rate, more steps cover to reach our desired result, but risk increases. If we take a

low learning rate (local minima), the overall process is time-consuming but efficient, it takes fewer steps to move toward the goal with low risk [1]. Cost (also known as a loss) function tells how well our model or how our model performs accurately. It plots its own curve and gradient in the graph.

This is the complete and concise overview of NN, advancement occurs when introducing innovation in the field of artificial NN is a convolutional neural network (CNN) in DL [52, 53]. The main agenda behind the concept is to build strong AI which acts and think same as a human [54], CNN is an active learning algorithm [1], the objective is to recognized visual features of input image, adjust weight and bias (if needed) for differentiate object of one from others [55]. Moreover, researchers applied CNN mostly on a visual dataset to retrieve hidden information and make system knowledgeable for automatically recognized different objects from a different facet of an image. If we compare CNN with classification algorithms, CNN takes the low amount in the pre-processing phase while taking the high amount in the filtering process for enough training (ability to learn more characteristics), as compared to this, classification algorithms are totally opposite of CNN. The contribution of CNN is not limited in CV but most of its features used for classifying sentences in Natural Language Processing (NLP) [56]. The architecture of CNN is inspired from the human visual cortex, connective patterns of brain neurons [57]. The receptive field in architecture, individual neuron responds only in a restricted region of visual field simultaneously [58, 105]. In this context, defining the basic difference of DL and neural nets, more than two hidden layers in NN model is probably called DL, which Highlighting the main aspect of gradient descent, learning rate, loss function, convolutional NN, and architecture of CNN.

## 4. Tools and Applications

In this review paper, list of eleven most popular tools of machine learning described in Table 1 with its features, platform capability, commercial and non-commercial products, and availability of supporting programming language domain.

Table 1. Tools and Application of Machine Learning

Name of Tool	Platform Support	Paid/Non-paid	Language	Features
Scikit Learn [59, 60]	Windows Linux Mac OS	Non-paid	C C++ Python	Clustering Classification Regression Pre-processing Model Selection
TensorFlow [61, 62]	Windows	Non-paid	Python	Library for dataflow programming



	Linux Mac OS		CUDA C++	
<b>PyTorch [63, 64]</b>	Windows Linux Mac OS	Non-paid	Python CUDA C++	Autograd Module Optim Module nn Module
<b>KNIME [65]</b>	Windows Linux Mac OS	Non-paid	Java	Text Mining Image Mining Using Plugins
<b>Weka [66, 67, 68]</b>	Windows Linux Mac OS	Non-paid	Java	Clustering Classification Regression Association Data-preparation Visualization
<b>Keras.io [69, 70]</b>	Cross Platform	Non-paid	Python	API for Artificial Neural Networks
<b>Rapid Miner [71]</b>	Cross Platform	Paid	Java	Data Pre-processing Data Loading Transformation Visualization
<b>Shogun [72]</b>	Windows Linux Unix Mac OS	Non-paid	C++	Dimensionality-Reduction Classification Clustering Regression Support Vector Machine Online Learning
<b>Accors.Net [73]</b>	Cross Platform	Non-paid	C#	Hypothesis Test and Kernel Methods Image, Audio, Vision and Signals Distribution Classification and Regression Clustering
<b>Colab [74]</b>	Cloud Service	Non-paid	-	Support libraries of Keras, OpenCV, PyTorch, and TensorFlow
<b>Apache Mahout [75, 76]</b>	Cross Platform	Non-paid	Java Scala	Distributed Linear-Algebra Recommender Pre-processor Clustering Regression

As mention above tools of ML, distinct applications cover different areas of artificial intelligence (AI):

- Video surveillance in CV is the monitoring app, let's consider a single person can monitor multiple cameras but in a certain period of time, impossible to monitor 24/7 in a row. Track unusual criminal activities, generate an alert to security forces before it happens, it is all about CV but at the backend, ML train machine ad handles all the situations.
- In the recent era, everyone probably using the biometric system in an organization for daily attendance, every person has its unique fingerprints but if someone punch attendance through biometric, the machine can easily detect how the person is, this is only possible because of ML.
- Pinterest is the CV application where identity objects in the images and suggest similar pins accordingly.
- Collaborate image processing with machine learning can easily extract features of a person in a picture and recommend how a person looks like in the next ten years. Similarly, extracting an old image and suggest how a person look like in recent year according to the features of the image. This application mostly helps for investigation purpose use by the military.
- Out of the boundary of CV, the biggest example trending nowadays is Netflix, movies recommend on the bases of early watching, recommendation engine running on ML.
- People you may know on Facebook, continuously notice activities like profile visits, recent connection, liking pages, ads clicking, workplace, schooling, etc.

continuously learning, Facebook suggests a list of friends, pages, ads you probably want to connect. One more thing, when we upload picture and person on the picture can automatically tag (face-recognition) on Facebook because of ML unsupervised.

- Email spamming, rule-based spam filter restricts spammer to track. The system security programs powered by ML to detect malware code patterns.
- Google search engine adopts ML techniques to handle top results clicking by users, different algorithms run on the backend for managing results as per visiting. Google also use ML for translation and Gmail spam filtering purpose.

## 5. Online Available Dataset for Computer Vision/Image Processing

There are many datasets were developed for the research purpose, which varies according to the number of images, categories, and types of images such as medical, hyperspectral, satellites, video streams, grayscale and RGB natural images. In this paper, we enlist the most popular image dataset, which is used for image processing/computer vision.

**Labelled Faces in the Wild:** this dataset available on the web site of the University of Massachusetts [77], this dataset contains 13,000 labelled images of human faces, for use in developing applications that involve facial recognition.

**Stanford Dogs Dataset:** this dataset provides by computer vision research group of Stanford University, which Contains 20,580 images and 120 different dog breed categories, with about 150 images per class [78].

**Places:** this dataset developed MIT Computer Science and Artificial Intelligence Laboratory and available on the website, which is based on a Scene-centric database with 205 scene categories and 2.5 million images with a category label [79].

**CelebFaces:** this dataset is designed by Multimedia laboratory, The Chinese University of Hong Kong and this Face dataset contains more than 200,000 celebrity images, each with 40 attribute annotations [80].

**Plant Image Analysis:** this dataset based on the collection of plant images and collection based on over 1 million images of plants. It can choose from 11 species of plants [81].

**Indoor Scene Recognition:** this dataset based on the 67 indoor categories and contains 15620 images. This dataset is freely available for research purpose and all images are in the format of jpg and every category at least 100 images [82].

**MedPix:** this dataset is based on the medical images and freely available for research purposes. This dataset contains 59000 images related to medical [83].

## 6. Challenges, Issues, and Limitations

In computer vision, limitations emerge due to training machine accordingly dataset, the concept rises called underfit and overfit, means not to train a system with less amount of data nor with the huge amount of data. In this paper, mention the five main aspects of raising challenges in the domain of image processing discuss below:

### 6.1. Algorithms Required Huge Amount of Data for Better Performance:

In this context, a machine requires a large amount of data for training and without this machine can't perform well enough as expected [84, 85]. Underfitting occurs when the algorithm does not fit the data and increase bias and decrease variance. Another scenario is overfitting it occurs in training phases, where model overload because of the ration of training data, decrease bias and increase variance [86]. The collection of valid data, which is useful for further process, is another critical issue rising by gathering data from different origins but ensure data is in a particular domain. The problem arises when conduct questionnaires, people can't respond 100% as we expected results received according to the filling of questionnaires by people and training and testing the model through vague data generates error ratio [87].

### 6.2. Require Lengthy Offline Labelling of Training Data:

Right now, AI is facing difficulty to label data, almost 80% of real-world data are gathered, organizing, and label [1], data-labelling requires lots of efforts to train, validate, and tune model, this isn't an easy task [88]. The vast gap between data-increment and data-recognition. Reason to require lengthy offline labelling: low quality, scale, inefficient or costly, quality assurance. Estimate steps to convert unlabelled to labelled training data: (i) data-identification, (ii) data-aggregation, (iii) data-cleansing, (iv) data-augmentation and (v) data-labelling, after that prepare algorithm for further process (vi) ML-optimization, (vii) ML-model tuning, (viii) ML-model training, and (ix) ML-algorithm development. These are the main process of labelling data to execute the ML algorithm.

### 6.3. High Processing Power:

Now, this is another challenging task to handle for the field researchers. ML requires more computational power

to process the input image dataset [89]. High processing consumes more time to validate the massive amount of data if found any error, it takes a cycle to check either code running perfectly or not. In short, the overall scenario is the time deplete with lots of processing power needed to compute several records.

#### 6.4. Bogus Data:

Algorithmic data emerge, fortune model, this is the beginning of synthetic or fake data become the part of the real-world [90, 91]. Types of bogus are online websites or small companies, particularly start-up to create a mimic dataset for model training, creating the perfect dataset for research publication, offline survey collection, or non-existent data. Identification of actual data from different sources and utilizes these collected datasets to find the unique outcome is one of the challenging tasks nowadays [92].

#### 6.5. ML Model/Algorithm does not Collaborate:

Algorithms are categorized into two parts; one is learning style while other is a similarity. Learning style algorithms separated in the form of supervised, unsupervised, and semi-supervised learning aspects [93, 94]. The similarity is used to grouping the same data together, instance-based, regression, regularization, clustering, decision-tree, Bayesian algorithms are the main examples. Limitation rises when the hybrid system means to collaborate with two algorithms to utilize both functionalities and make the machine more powerful than other run-time systems.

### 7. Open Research Areas

A lot of work has been done on computer vision by using machine learning but still issues in the computer vision field.

**Object/ Vehicle Detection:** Vehicle and object detection is still required efficient algorithms to detect required object or vehicle in the war situations to react quickly, so it required further development to fast processing to detect to react and prevent from damage [95, 105]. Compressed and large-scale images are the problems in the dataset to accurately detect objects, this will help to easily detect vehicle and monitor their actives [96]. Poor visual and Low resolution of the satellite image is also an issue in object/vehicle detection, but development advance ML algorithms and model will help to use of Maximally Stable Extremal Regions for vehicle detection in complex situations low lightening conditions or under shadow regions [97, 98].

**Activity Recognition:** Activity recognition in images/videos still required ML-based efficient algorithms and models, which measure indexing,

recognition rate, accuracy, robustness, efficiency, single activity and multiple type and level of activity recognition [99].

**Human pose estimation:** This is also a major research area of computer vision, which based on the analysis of the pose of humans in an image [100]. This is the process of calculating the actual position of human joints in video or image. The research of computer vision is still facing the problem of the location of human joints in images [104]. It is also important to note that pose estimation has various sub-tasks such as single pose estimation, estimating poses in an image with many people, estimating poses in crowded places, and estimating poses in videos [101].

### 8. Conclusion

In this paper, the importance of ML in the image processing (IP) domain is highlighted and address the processes of digital image and how difficult is to feed the computer system. For this processes, the acquisition is the first step of IP where image can load and prepare for further process (briefly elaborate in image processing section). In the end, we get label image objects in the detection and recognition phases. Neural Networks (NN) is one of the ML algorithms, which aims to optimize solutions of the given problem to provide better solution and predicate output by providing input values. Well-known companies adopted NN for maintaining entire applications and one of them is Amazon the world-famous organization utilizes NN for its power recommendation engine, Microsoft for translation purpose, Facebook for facial detection and recognition, and Google use NN for Gmail spam filter, etc. Also, we have discussed the concept of deep learning, the emerging technology of AI, which is the subfield of ML, the advent of the technology it grabs the attention of researchers nowadays. We conclude as machine learning spread all the corners of computer vision and well perform in all those areas, but unfortunately, there are still some open research areas, where researchers need to focus on it and fill these areas, which are mention in open research areas section.

#### Declarations.

##### Availability of data and materials

- No associated data available related to this paper

##### Competing interests

- Authors did not have any conflict of interest.

##### Funding

- There is no funding available for this research

##### Authors' contributions

- First author proposes and write draft and 2nd author is supervise and revised the paper. 3rd authors review the paper for language and typing issues.

### References



- [1] Sadhayo, Intesab Hussain, Muhammad Ibrahim Channa, and Asif Ali Laghari. "Automated Medication System for Rural and War Affected Areas." *Transactions on Machine Learning and Artificial Intelligence* 3, no. 1 (2015): 08.
- [2] Russell, Stuart J., and Peter Norvig. *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited, 2016.
- [3] Dobrev, Dimitar. "A Definition of Artificial Intelligence." *arXiv preprint arXiv:1210.1568* (2012).
- [4] Siau, K., and Yin Yang. "Impact of artificial intelligence, robotics, and machine learning on sales and marketing." In *Twelve Annual Midwest Association for Information Systems Conference (MWAIS 2017)*, pp. 18-19. 2017.
- [5] Campbell, Murray, A. Joseph Hoane Jr, and Feng-hsiung Hsu. "Deep blue." *Artificial intelligence* 134, no. 1-2 (2002): 57-83.
- [6] Laghari, Rashid Ali, Janguang Li, Asif Ali Laghari, and Shu-qi Wang. "A Review on Application of Soft Computing Techniques in Machining of Particle Reinforcement Metal Matrix Composites." *Archives of Computational Methods in Engineering* (2019): 1-15.
- [7] Zhao, Ying, and Arjuna Flenner. "Deep Models, Machine Learning, and Artificial Intelligence Applications in National and International Security." *AI Magazine* 40, no. 1 (2019): 35-36.
- [8] Nichols, James A., Hsien W. Herbert Chan, and Matthew AB Baker. "Machine learning: applications of artificial intelligence to imaging and diagnosis." *Biophysical reviews* 11, no. 1 (2019): 111-118.
- [9] Patton, Michael Quinn. "Qualitative research." *Encyclopedia of statistics in behavioral science* (2005).
- [10] Goertzen, Melissa J. ". Introduction to Quantitative Research and Data." *Library Technology Reports* 53, no. 4 (2017): 12-18.
- [11] Bzdok, Danilo, Martin Krzywinski, and Naomi Altman. "Points of significance: Machine learning: supervised methods." (2018): 5.
- [12] Camacho, Diogo M., Katherine M. Collins, Rani K. Powers, James C. Costello, and James J. Collins. "Next-generation machine learning for biological networks." *Cell* 173, no. 7 (2018): 1581-1592.
- [13] Chapelle, Olivier, Bernhard Scholkopf, and Alexander Zien. "Semi-supervised learning (chapelle, o. et al., eds.; 2006) [book reviews]." *IEEE Transactions on Neural Networks* 20, no. 3 (2009): 542-542.
- [14] Huang, T. "Computer vision: Evolution and promise." (1996).
- [15] Salman, Hasan, Md Nasir Uddin, Samuel Acheampong, and He Xu. "Design and Implementation of IoT Based Class Attendance Monitoring System Using Computer Vision and Embedded Linux Platform." In *Workshops of the International Conference on Advanced Information Networking and Applications*, pp. 25-34. Springer, Cham, 2019.
- [16] Ibrar, Muhammad, Jianing Mi, Shahid Karim, Asif Ali Laghari, Shakir Muhammad Shaikh, and Vishal Kumar. "Improvement of large-vehicle detection and monitoring on CPEC route." *3D Research* 9, no. 3 (2018): 1-7.
- [17] Kesner, Adam, Richard Laforest, Ricardo Otazo, Kwak Jennifer, and Tinsu Pan. "Medical imaging data in the digital innovation age." *Medical physics* 45, no. 4 (2018): e40-e52.
- [18] Olatunji, Tobi, Li Yao, Ben Covington, and Anthony Upton. "Caveats in Generating Medical Imaging Labels from Radiology Reports with Natural Language Processing." (2019).
- [19] Alexander, Alan, Megan McGill, Anna Tarasova, Cara Ferreira, and Delphine Zurkiya. "Scanning the Future of Medical Imaging." *Journal of the American College of Radiology* 16, no. 4 (2019): 501-507.
- [20] Kapach, Keren, Ehud Barnea, Rotem Mairon, Yael Edan, and Ohad Ben-Shahar. "Computer vision for fruit harvesting robots—state of the art and challenges ahead." *International Journal of Computational Vision and Robotics* 3, no. 1/2 (2012): 4-34.
- [21] Netzer, Yuval, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y. Ng. "Reading digits in natural images with unsupervised feature learning." (2011).
- [22] Kussel, Marcel. "Method for enhancing vehicle camera image quality." U.S. Patent 10,257,432, issued April 9, 2019.
- [23] Karim, Sajida, Hui He, Asif Ali Laghari, Arif Hussain Magsi, and Rashid Ali Laghari. "Quality of service (QoS): measurements of image formats in social cloud computing." *Multimedia Tools and Applications* (2020): 1-26.
- [24] Natarajan, Balakrishnan, and T. R. Ramesh. "Noise Removal Using Histogram Equalized Based Contrast Masking For Image Quality Assessment." *International Journal of Applied Engineering Research* 13, no. 10 (2018): 8194-8199.
- [25] Kamble, Vipin Milind, Mayur Rajaram Parate, and Kishor M. Bhurchandi. "No reference noise estimation in digital images using random conditional selection and sampling theory." *The Visual Computer* 35, no. 1 (2019): 5-21.
- [26] Russ, John C. *The image processing handbook*. CRC press, 2016.
- [27] Xing, Guanyu, Yanli Liu, Wanfa Zhang, and Haibin Ling. "Light mixture intrinsic image decomposition based on a single RGB-D image." *The Visual Computer* 32, no. 6-8 (2016): 1013-1023.
- [28] Zhang, Xiaohong, Min Tang, and Ruofeng Tong. "Robust super resolution of compressed video." *The Visual Computer* 28, no. 12 (2012): 1167-1180.
- [29] Lidong, Huang, Zhao Wei, Wang Jun, and Sun Zebin. "Combination of contrast limited adaptive histogram equalisation and discrete wavelet transform for image enhancement." *IET Image Processing* 9, no. 10 (2015): 908-915.
- [30] Kaur, Pahulpreet, and Bikrampal Kaur. "2-D geometric shape recognition using canny edge detection technique." In *2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom)*, pp. 3161-3164. IEEE, 2016.
- [31] Sonka, Milan, Vaclav Hlavac, and Roger Boyle. *Image processing, analysis, and machine vision*. Cengage Learning, 2014.
- [32] Boyat, Ajay Kumar, and Brijendra Kumar Joshi. "A review paper: noise models in digital image processing." *arXiv preprint arXiv:1505.03489* (2015).
- [33] Sérot, Jocelyn, François Berry, and Cédric Bourrasset. "High-level dataflow programming for real-time

- image processing on smart cameras." *Journal of Real-Time Image Processing* 12, no. 4 (2016): 635-647.
- [34] Laghari, Asif Ali, Hui He, Muhammad Shafiq, and Asiya Khan. "Assessment of quality of experience (QoE) of image compression in social cloud computing." *Multiagent and Grid Systems* 14, no. 2 (2018): 125-143.
- [35] Yogamangalam, R., and B. Karthikeyan. "Segmentation techniques comparison in image processing." *International Journal of Engineering and Technology (IJET)* 5, no. 1 (2013): 307-313.
- [36] Zhong, Yanfei, Ailong Ma, Yew soon Ong, Zexuan Zhu, and Liangpei Zhang. "Computational intelligence in optical remote sensing image processing." *Applied Soft Computing* 64 (2018): 75-93.
- [37] Ghosh, Ashish, Niladri Shekhar Mishra, and Susmita Ghosh. "Fuzzy clustering algorithms for unsupervised change detection in remote sensing images." *Information Sciences* 181, no. 4 (2011): 699-715.
- [38] Patra, Swarnajyoti, Susmita Ghosh, and Ashish Ghosh. "Histogram thresholding for unsupervised change detection of remote sensing images." *International journal of remote sensing* 32, no. 21 (2011): 6071-6089.
- [39] Purwar, Yashasvi, Sirish L. Shah, Gwen Clarke, Areej Almugairi, and Atis Muehlenbachs. "Automated and unsupervised detection of malarial parasites in microscopic images." *Malaria journal* 10, no. 1 (2011): 364.
- [40] Pradhan, Ashis. "Support vector machine-A survey." *International Journal of Emerging Technology and Advanced Engineering* 2, no. 8 (2012): 82-85.
- [41] Sommer, Christoph, and Daniel W. Gerlich. "Machine learning in cell biology-teaching computers to recognize phenotypes." *J Cell Sci* 126, no. 24 (2013): 5529-5539.
- [42] Giusti, Alessandro, Jérôme Guzzi, Dan C. Cireşan, Fang-Lin He, Juan P. Rodríguez, Flavio Fontana, Matthias Faessler et al. "A machine learning approach to visual perception of forest trails for mobile robots." *IEEE Robotics and Automation Letters* 1, no. 2 (2015): 661-667.
- [43] Erickson, Bradley J., Panagiotis Korfiatis, Zeynettin Akkus, and Timothy L. Kline. "Machine learning for medical imaging." *Radiographics* 37, no. 2 (2017): 505-515.
- [44] Tuia, Devis, Michele Volpi, Loris Copa, Mikhail Kanevski, and Jordi Munoz-Mari. "A survey of active learning algorithms for supervised remote sensing image classification." *IEEE Journal of Selected Topics in Signal Processing* 5, no. 3 (2011): 606-617.
- [45] Nikolenko, Sergey, A. Kadurin, and E. Arkhangelskaya. "Deep learning." Peter, St. Petersburg (2018).
- [46] Abraham, Tara H. "(Physio) logical circuits: The intellectual origins of the McCulloch-Pitts neural networks." *Journal of the History of the Behavioral Sciences* 38, no. 1 (2002): 3-25.
- [47] Huang, Guang-Bin. "What are extreme learning machines? Filling the gap between Frank Rosenblatt's dream and John von Neumann's puzzle." *Cognitive Computation* 7, no. 3 (2015): 263-278.
- [48] Al-Molegi, Abdulrahman, Mohammed Jabreel, and Baraq Ghaleb. "STF-RNN: Space Time Features-based Recurrent Neural Network for predicting people next location." In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1-7. IEEE, 2016.
- [49] Irsoy, Ozan, and Claire Cardie. "Deep recursive neural networks for compositionality in language." In *Advances in neural information processing systems*, pp. 2096-2104. 2014.
- [50] Maren, Alianna J., Craig T. Harston, and Robert M. Pap. *Handbook of neural computing applications*. Academic Press, 2014.
- [51] Ding, Shifei, Hui Li, Chunyang Su, Junzhao Yu, and Fengxiang Jin. "Evolutionary artificial neural networks: a review." *Artificial Intelligence Review* 39, no. 3 (2013): 251-260.
- [52] Banerjee, Imon, Yuan Ling, Matthew C. Chen, Sadid A. Hasan, Curtis P. Langlotz, Nathaniel Moradzadeh, Brian Chapman et al. "Comparative effectiveness of convolutional neural network (CNN) and recurrent neural network (RNN) architectures for radiology text report classification." *Artificial intelligence in medicine* 97 (2019): 79-88.
- [53] Vardhana, M., N. Arunkumar, Sunitha Lasrado, Enas Abdulhay, and Gustavo Ramirez-Gonzalez. "Convolutional neural network for bio-medical image segmentation with hardware acceleration." *Cognitive Systems Research* 50 (2018): 10-14.
- [54] Banerjee, Soham, Pradeep Kumar Singh, and Jaya Bajpai. "A comparative study on decision-making capability between human and artificial intelligence." In *Nature Inspired Computing*, pp. 203-210. Springer, Singapore, 2018.
- [55] Rawat, Waseem, and Zenghui Wang. "Deep convolutional neural networks for image classification: A comprehensive review." *Neural computation* 29, no. 9 (2017): 2352-2449.
- [56] Kim, Yoon. "Convolutional neural networks for sentence classification." *arXiv preprint arXiv:1408.5882* (2014).
- [57] Miranda, Vladimiro, Pedro A. Cardoso, Ricardo J. Bessa, and Ildemar Decker. "Through the looking glass: Seeing events in power systems dynamics." *International Journal of Electrical Power & Energy Systems* 106 (2019): 411-419.
- [58] Ahmad, Mobeen, Jooyeon Joe, and Dongil Han. "CortexNet: Convolutional Neural Network with Visual Cortex in human brain." In *2018 IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, pp. 206-212. IEEE, 2018.
- [59] <https://scikit-learn.org/> accessed [11/06/19]
- [60] Hao, Jiangang, and Tin Kam Ho. "Machine Learning Made Easy: A Review of Scikit-learn Package in Python Programming Language." *Journal of Educational and Behavioral Statistics* 44, no. 3 (2019): 348-361.
- [61] Shukla, Nishant. *Machine learning with TensorFlow*. Manning Publications Co., 2018.
- [62] <https://www.tensorflow.org/> accessed [11/08/19]
- [63] Mishra, Pradeepta. "Introduction to Neural Networks Using PyTorch." In *PyTorch Recipes*, pp. 111-126. Apress, Berkeley, CA, 2019.
- [64] <https://pytorch.org/> accessed [11/07/19]
- [65] <https://www.knime.com/> accessed [11/08/19]
- [66] <https://www.cs.waikato.ac.nz/ml/weka/> accessed [11/08/19]
- [67] Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. "The WEKA data mining software: an update." *ACM*

- SIGKDD explorations newsletter 11, no. 1 (2009): 10-18.
- [68] Jabez, J., S. Gowri, S. Vigneshwari, J. Albert Mayan, and Senduru Srinivasulu. "Anomaly Detection by Using CFS Subset and Neural Network with WEKA Tools." In *Information and Communication Technology for Intelligent Systems*, pp. 675-682. Springer, Singapore, 2019.
- [69] <https://keras.io/> accessed [11/10/19]
- [70] Si, Minxing, Tyler J. Tarnoczi, Brett M. Wiens, and Ke Du. "Development of Predictive Emissions Monitoring System Using Open Source Machine Learning Library-Keras: A Case Study on a Cogeneration Unit." *IEEE Access* 7 (2019): 113463-113475.
- [71] <https://rapidminer.com/> accessed [11/12/19]
- [72] Sonnenburg, S. Č., Sebastian Henschel, Christian Widmer, Jonas Behr, Alexander Zien, Fabio de Bona, Alexander Binder, Christian Gehl, and Vojtěch Franc. "The SHOGUN machine learning toolbox." *Journal of Machine Learning Research* 11, no. Jun (2010): 1799-1802.
- [73] <http://accord-framework.net/> accessed [11/13/19]
- [74] <https://colab.research.google.com/notebooks/welcome.ipynb> accessed [11/15/19]
- [75] Zheng, Yong. "A Comparison of Tools for Teaching and Learning Data Analytics." In *Proceedings of the 20th Annual SIG Conference on Information Technology Education*, pp. 160-160. ACM, 2019.
- [76] <https://mahout.apache.org/overview.html> accessed [11/18/19]
- [77] Deng, Jiankang, Yuxiang Zhou, and Stefanos Zafeiriou. "Marginal loss for deep face recognition." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 60-68. 2017.
- [78] Hsu, David. "Using convolutional neural networks to classify dog breeds." *CS231n: Convolutional Neural Networks for Visual Recognition [course webpage]* 2 (2015).
- [79] Zhou, Bolei, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. "Places: A 10 million image database for scene recognition." *IEEE transactions on pattern analysis and machine intelligence* 40, no. 6 (2017): 1452-1464.
- [80] Cao, Jiajiong, Yingming Li, and Zhongfei Zhang. "Celeb-500K: A Large Training Dataset for Face Recognition." In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pp. 2406-2410. IEEE, 2018.
- [81] Lobet, Guillaume. "Image analysis in plant sciences: publish then perish." *Trends in plant science* 22, no. 7 (2017): 559-566.
- [82] Quattoni, Ariadna, and Antonio Torralba. "Recognizing indoor scenes." In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 413-420. IEEE, 2009.
- [83] Lau, Jason J., Soumya Gayen, Asma Ben Abacha, and Dina Demner-Fushman. "A dataset of clinically generated visual questions and answers about radiology images." *Scientific data* 5 (2018): 180251.
- [84] Chang, Zheng, Lei, Zhenyu Zhou, Shiwen Mao, and Tapani Ristaniemi. "Learn to cache: Machine learning for network edge caching in the big data era." *IEEE Wireless Communications* 25, no. 3 (2018): 28-35.
- [85] Mahdaviinejad, Mohammad Saeid, Mohammadreza Rezvan, Mohammadamin Barekatain, Peyman Adibi, Payam Barnaghi, and Amit P. Sheth. "Machine learning for Internet of Things data analysis: A survey." *Digital Communications and Networks* 4, no. 3 (2018): 161-175.
- [86] Santos, Miriam Seoane, Jastin Pompeu Soares, Pedro Henriques Abreu, Helder Araujo, and Joao Santos. "Cross-validation for imbalanced datasets: Avoiding overoptimistic and overfitting approaches [research frontier]." *IEEE Computational Intelligence Magazine* 13, no. 4 (2018): 59-76.
- [87] Anderson, Hyrum S., and Phil Roth. "Ember: an open dataset for training static PE malware machine learning models." *arXiv preprint arXiv:1804.04637* (2018).
- [88] Gadepally, Vijay, Justin Goodwin, Jeremy Kepner, Albert Reuther, Hayley Reynolds, Siddharth Samsi, Jonathan Su, and David Martinez. "AI Enabling Technologies: A Survey." *arXiv preprint arXiv:1905.03592* (2019).
- [89] Pouyanfar, Samira, Saad Sadiq, Yilin Yan, Haiman Tian, Yudong Tao, Maria Presa Reyes, Mei-Ling Shyu, Shu-Ching Chen, and S. S. Iyengar. "A survey on deep learning: Algorithms, techniques, and applications." *ACM Computing Surveys (CSUR)* 51, no. 5 (2019): 92.
- [90] Chen, Yuantao, Jiajun Tao, Jin Wang, Xi Chen, Jingbo Xie, Jie Xiong, and Kai Yang. "The novel sensor network structure for classification processing based on the machine learning method of the ACGAN." *Sensors* 19, no. 14 (2019): 3145.
- [91] Alshinina, Remah, and Khaled Elleithy. "A highly accurate machine learning approach for developing wireless sensor network middleware." In *2018 Wireless Telecommunications Symposium (WTS)*, pp. 1-7. IEEE, 2018.
- [92] Shu, Kai, Deepak Mahudeswaran, Suhang Wang, Dongwon Lee, and Huan Liu. "Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media." *arXiv preprint arXiv:1809.01286* (2018).
- [93] Kiran, B., Dilip Thomas, and Ranjith Parakkal. "An overview of deep learning based methods for unsupervised and semi-supervised anomaly detection in videos." *Journal of Imaging* 4, no. 2 (2018): 36.
- [94] Chen, Chong, Ying Liu, Maneesh Kumar, Jian Qin, and Yunxia Ren. "Energy Consumption Modelling Using Deep Learning Embedded Semi-supervised Learning." *Computers & Industrial Engineering* (2019).
- [95] Karim, Shahid, Ye Zhang, Asif Ali Laghari, and Muhammad Rizwan Asif. "Image processing based proposed drone for detecting and controlling street crimes." In *2017 IEEE 17th International Conference on Communication Technology (ICCT)*, pp. 1725-1730. IEEE, 2017.
- [96] Memon, Muhammad Sulleman, and Asif Ali Laghari. "Proposal for Wireless Controlled Threat Monitoring and Targeting Vehicles." *International Journal of Computer and Communication Engineering* 2, no. 5 (2013): 580.
- [97] Karim, Shahid, Ye Zhang, Shoulin Yin, Asif Ali Laghari, and Ali Anwar Brohi. "Impact of compressed and down-scaled training images on vehicle detection

- in remote sensing imagery." *Multimedia Tools and Applications* 78, no. 22 (2019): 32565-32583.
- [98] Zhang, Shugang, Zhiqiang Wei, Jie Nie, Lei Huang, Shuang Wang, and Zhen Li. "A review on human activity recognition using vision-based method." *Journal of healthcare engineering* 2017 (2017).
- [99] Karim, Shahid, Imtiaz Ali Halepoto, Adnan Manzoor, Nazar Hussain Phulpoto, and Asif Ali Laghari. "Vehicle detection in satellite imagery using maximally stable extremal regions." *IJCSNS* 18, no. 4 (2018): 75.
- [100] Varadarajan, Jagannadan, Ramanathan Subramanian, Samuel Rota Bulò, Narendra Ahuja, Oswald Lanz, and Elisa Ricci. "Joint estimation of human pose and conversational groups from social scenes." *International Journal of Computer Vision* 126, no. 2-4 (2018): 410-429.
- [101] Ma, Miao, Naresh Marturi, Yibin Li, Ales Leonardis, and Rustam Stolkin. "Region-sequence based six-stream CNN features for general and fine-grained human action recognition in videos." *Pattern Recognition* 76 (2018): 506-521.
- [102] Pandey, Dinesh, Xiaoxia Yin, Hua Wang, and Yanchun Zhang. "Accurate vessel segmentation using maximum entropy incorporating line detection and phase-preserving denoising." *Computer Vision and Image Understanding* 155 (2017): 162-172.
- [103] Pandey, Dinesh, Xiaoxia Yin, Hua Wang, Min-Ying Su, Jeon-Hor Chen, Jianlin Wu, and Yanchun Zhang. "Automatic and fast segmentation of breast region-of-interest (ROI) and density in MRIs." *Heliyon* 4, no. 12 (2018): e01042.
- [104] Sarki, Rubina, Khandakar Ahmed, Hua Wang, and Yanchun Zhang. "Automated detection of mild and multi-class diabetic eye diseases using deep learning." *Health Information Science and Systems* 8, no. 1 (2020): 1-9.
- [105] Du, Jiahua, Sandra Michalska, Sudha Subramani, Hua Wang, and Yanchun Zhang. "Neural attention with character embeddings for hay fever detection from twitter." *Health information science and systems* 7, no. 1 (2019): 1-7.