IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 26 August 2022, accepted 16 September 2022, date of publication 26 September 2022, date of current version 13 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3209825

RESEARCH ARTICLE

A Review on Machine Learning Styles in Computer Vision—Techniques and Future Directions

SUPRIYA V. MAHADEVKAR[1], BHARTI KHEMANI[1], SHRUTI PATIL[2], KETAN KOTECHA[2], DEEPALI R. VORA[1], AJITH ABRAHAM[3], (Senior Member, IEEE), AND LUBNA ABDELKAREIM GABRALLA[4]

[1]Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India
[2]Symbiosis Centre for Applied Artificial Intelligence, Symbiosis Institute of Technology, Symbiosis International (Deemed University), Pune 412115, India
[3]Machine Intelligence Research Laboratories (MIR Laboratories), Auburn, WA 98071, USA
[4]Department of Computer Science and Information Technology, College of Applied, Princess Nourah Bint Abdulrahman University, Riyadh 11671, Saudi Arabia

Corresponding author: Shruti Patil (shruti.patil@sitpune.edu.in)

This research work was supported by the Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R178), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

ABSTRACT Computer applications have considerably shifted from single data processing to machine learning in recent years due to the accessibility and availability of massive volumes of data obtained through the internet and various sources. Machine learning is automating human assistance by training an algorithm on relevant data. Supervised, Unsupervised, and Reinforcement Learning are the three fundamental categories of machine learning techniques. In this paper, we have discussed the different learning styles used in the field of Computer vision, Deep Learning, Neural networks, and machine learning. Some of the most recent applications of machine learning in computer vision include object identification, object classification, and extracting usable information from images, graphic documents, and videos. Some machine learning techniques frequently include zero-shot learning, active learning, contrastive learning, self-supervised learning, life-long learning, semi-supervised learning, ensemble learning, sequential learning, and multi-view learning used in computer vision until now. There is a lack of systematic reviews about all learning styles. This paper presents literature analysis of how different machine learning styles evolved in the field of Artificial Intelligence (AI) for computer vision. This research examines and evaluates machine learning applications in computer vision and future forecasting. This paper will be helpful for researchers working with learning styles as it gives a deep insight into future directions.

INDEX TERMS Machine learning techniques, computer vision, supervised learning, multi-task learning, object detection, artificial intelligence, image categorization, zero-shot learning.

NOMENCLATURE

AI	Artificial intelligence.
CV	Computer vision.
ML	Machine learning.
SVM	Support vector machine.
CNN	Convolution neural network.
SLR	Systematic literature review.
KNN	K-Nearest neighbor.
MIL	Multiple instance learning.
AL	Active learning.
MTL	Multitask learning.
RCNN	Region based convolution neural network.
PCA	Principal component analysis.
FL	Federated learning.

The associate editor coordinating the review of this manuscript and approving it for publication was Zhongyi Guo.

I. INTRODUCTION

Machine learning is a type of artificial intelligence (AI) that trains computers to think like humans by learning from and expanding upon previous experiences. It employs minimal

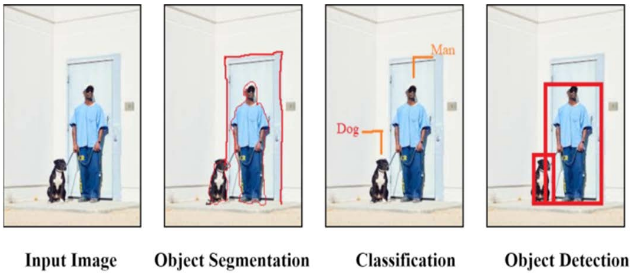

FIGURE 1. Segmentation, classification & Object detection.

human intervention to analyze data and spot trends. Machine learning has a wide range of effects on society, including production lines, healthcare, education, transportation, and food [1]. Machine learning is transforming our lives and industries in housing and apps, cars, retail, the food industry, etc.

The goal of machine learning and computer vision is to impart to computers the ability to gather data, understand it, and make decisions based on previous and present results. Computer vision is important for the Internet of Things, Industrial Internet of Things, and human cognitive interfaces. Computer vision and machine learning techniques are used to identify and track complex human actions in multimedia streams. For the prediction and analysis task of data, there are three types of learning: supervised, unsupervised, and semi-supervised [2].

The ability of computers to gather data, interpret it and make decisions based on past and present results is the aim of machine learning and computer vision. Computer vision is essential for the Internet of Things, Industrial Internet of Things, and human cognitive interfaces. Computer vision and machine learning techniques are used to identify and track complex human actions in multimedia streams. The image segmentation, localization & classification, and object detection are shown in the figure1. The authors have outlined the application areas of computer vision in the figure 4. They have listed which machine learning techniques and Python Libraries have been employed in each field.

An essential method of image processing that examines the contents of the image is segmentation. Image segmentation can be used for pattern recognition, feature extraction, content-based image retrieval, etc. Image segmentation is an important process in most medical image analysis. K-means is a widely used clustering algorithm to partition data into k clusters. The K-means and fuzzy K-means clustering algorithms can be used to identify tumor cells in MR images that may show the characteristics of the tumor's severity, facilitating the necessary diagnosis and therapy. There are numbers of well-established algorithms for prediction and analysis such as supervised learning, un-supervised learning, and semi supervised learning. These methods use the machine learning algorithms such as support vector machine, KNN etc. Scipy, Scikit, OpenCV, Matplotlib and Keras are the popular are libraries used for image segmentation.

For object detection previously sliding window, selective search and Kadane's algorithms were used but now most of the application areas uses deep learning algorithms like RCNN,YOLO,SSD for object detection in CV. The software libraries utilized in object detection for computer vision are Tensor Flow, ImageAI, GluonCV and YOLOv7. Convolution neural networks, recurrent neural networks, long short-term memories, gated recurrent units, and Bayesian networks are all used in traffic detection models. Sensors in intelligent settings collect data that is later analyzed and forecasted. One of the tasks the convolution neural network (CNN) successfully completes for successful object detection is feature extraction [3]. With a big collection of face photos, a deep convolution neural network can recognize faces through supervised learning. Data annotation and labeling is the only problem in computer vision and machine learning applications. The support vector machines, neural networks, KNN and probabilistic graphical models machine learning paradigms for computer vision. A common classification technique is support vector machines (SVMs), a subfield of supervised machine learning techniques. With a maximum margin separating two significant classifed classes, SVMs seek to locate a hyperplane [4]. Layered networks of connected processing nodes make up a neural network. A class of neural networks called convolution neural networks (CNNs) is utilized for image recognition and categorization. It has neurons that are wide, large, and deep. Due to widely available datasets, GPUs, and regularization approaches, CNN has become more and more popular in recent years.

The paper looks at a variety of machine learning applications in computer vision. For instance, biological sciences include segmentation, feature extraction, pattern matching, visual model optimization, form representation, surface reconstruction, and modeling. Computer vision uses machine learning to evaluate data from images that detect cars and pedestrians, using images to analyze remote sensing data for geographic information systems, diagnose faults in railroad ties automatically, identify different varieties of mango fruit based on size attributes, and extract graphical and textual information from document images [5]. Detecting curb ramps in Google Street View, automatically detecting and identifying faces, machine vision, handwriting recognition, enhanced driving assistance systems, and behavioral measures are some of the other techniques. Computer vision and machine learning are also used in the medical field, especially in nuclear medicine, endoscopy, angiography, magnetic resonance, ultrasound, and microscopy, according to studies in this field. Engineering, health, agriculture, astronomy, sports, cyber security, and education are among the many fields where machine learning and computer vision are used [6].

A. MOTIVATION

There is no systematic literature review (SLR) for machine learning styles used in computer vision focuses on methodology, datasets, application areas, comparative analysis, and

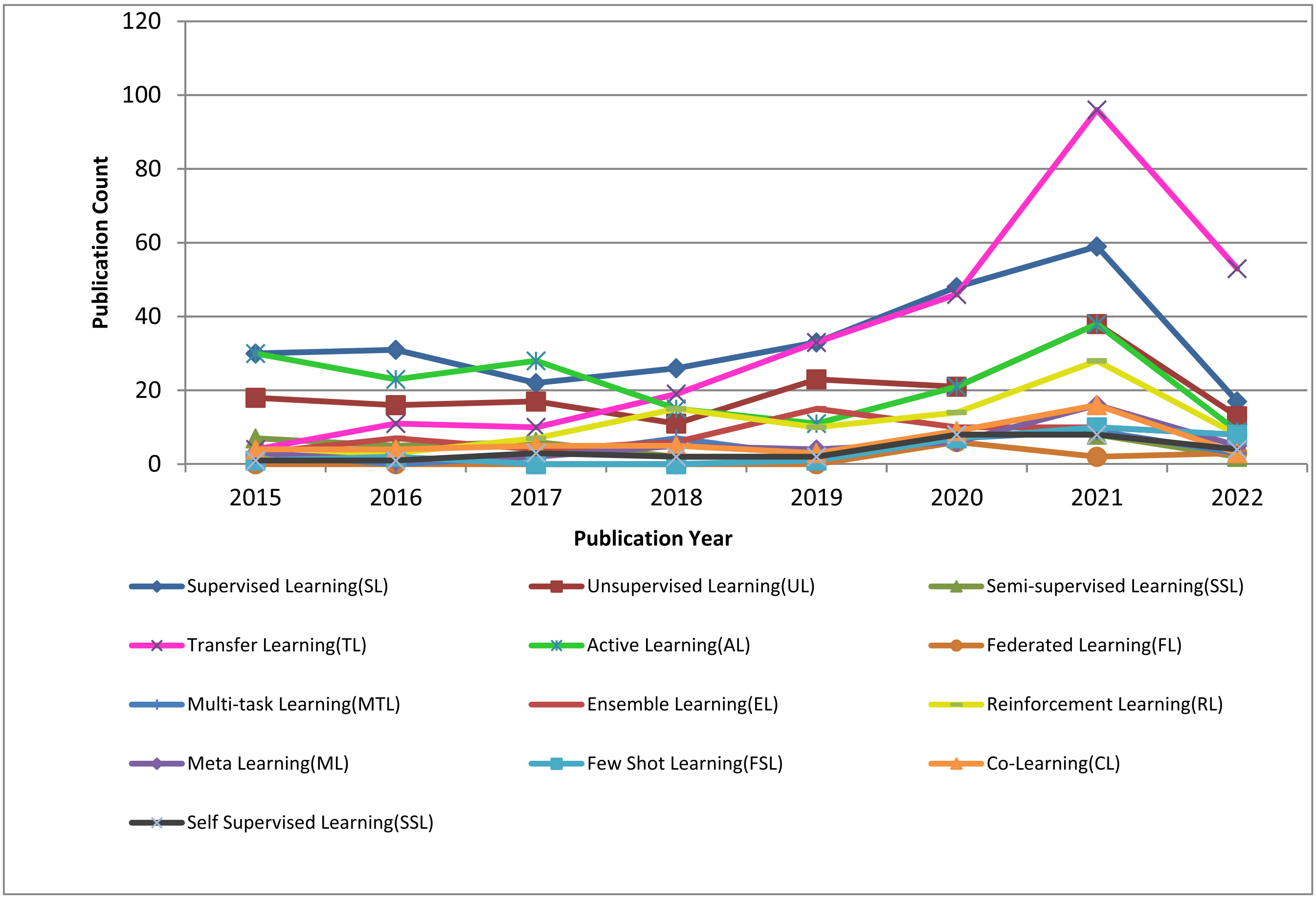

FIGURE 2. 2015-2022 year-wise publication count of learning styles.

future directions. The existing literature lacks a comprehensive survey focused on evolution of each ML style with its architecture, CV applications, research gaps and future directions.

The table 1 and figure 4 shows how few papers have dealt with datasets. The same datasets are available; however, each one only has a specific amount of data. Different machine learning styles are used in almost every application, such as Cyber security, Object detection, Spam detection, Health Care sector, Agriculture, etc. So, based on the above literature, we suggested when and where to use a particular machine learning style. The main goal of this review is to highlight current strategies, datasets that are accessible, applications, difficulties, and potential future directions of various machine learning approaches used in computer vision. In the last section this survey describes the current research gaps with the possible ML styles as solution and the future directions in the field of computer vision.

B. CONTRIBUTION OF WORK

The contributions of this comprehensive literature review are as follows:

- The authors comprehensively review the literature on ML styles in computer vision, emphasizing methodology, datasets, applications, associated problems, and potential future directions.
- The authors discuss and investigate the ML approaches and methodologies used and how they revive the computer vision field.
- The authors also give a summary of various publicly accessible datasets that are used to support this field of study.
- In addition, the authors examine distinct application domains while assessing machine learning techniques' function.
- Authors outline difficulties with various machine learning approaches, such as datasets, the accuracy of existing systems, and processing high-quality data.

C. PAPER ORGANIZATION

To summarize, this study provides the following important contributions:

1) This research aims at how different machine learning styles are used in computer vision, analyses its uses, and predicts future trends.

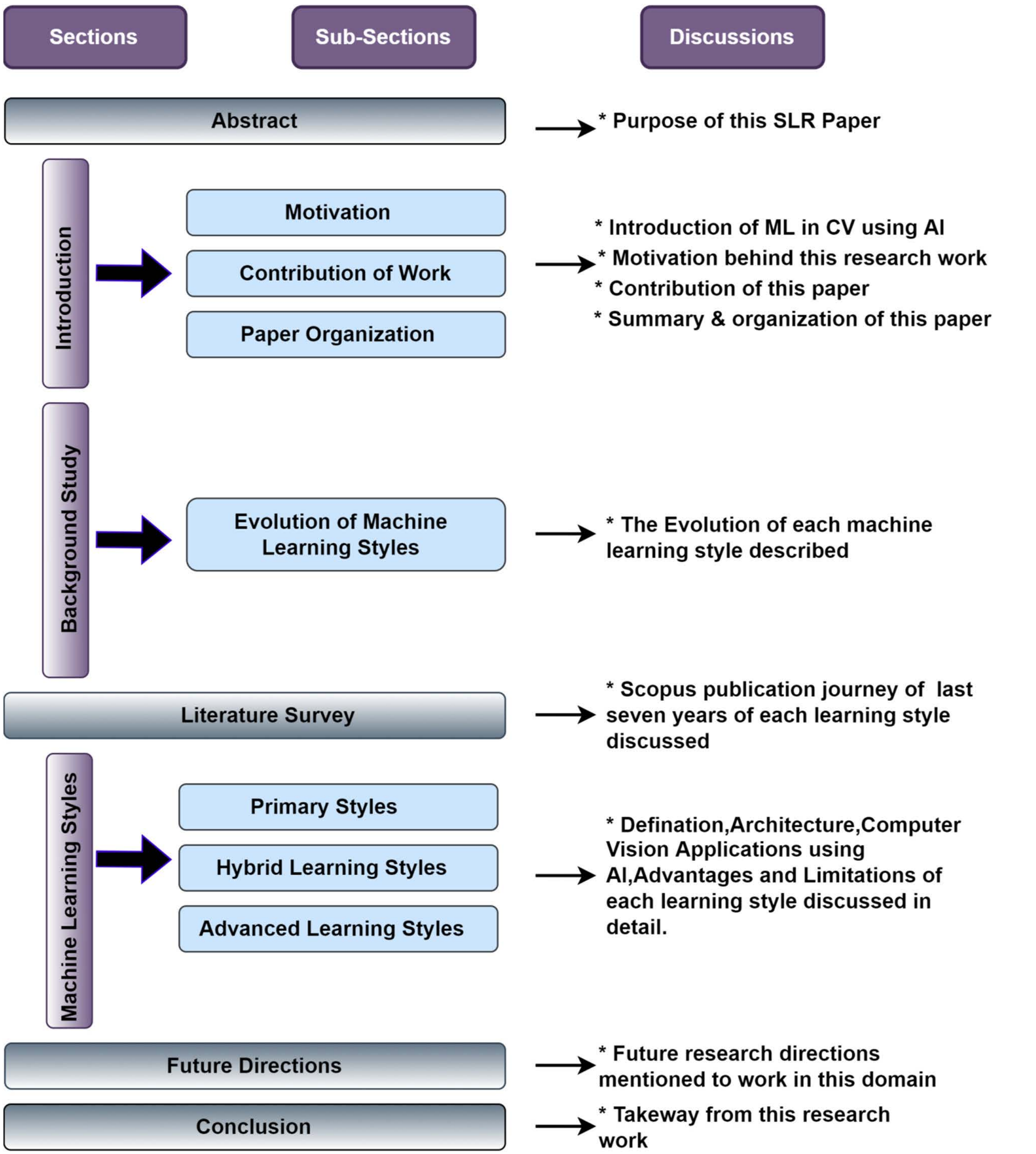

FIGURE 3. Outline of paper.

2) Evolution and literature review of different machine learning styles used in various domains of computer vision.

3) The study discovered brief overview of architecture, working, CV applications, datasets used, advantages and limitations of primary,hybrid and advanced ML styles.

4) Recent research gaps identified and highlight the future directions.

The rest of this article is organized as follows: Section 2 gives background knowledge in terms of Evolution of all ML styles. Section 3describes literature review in

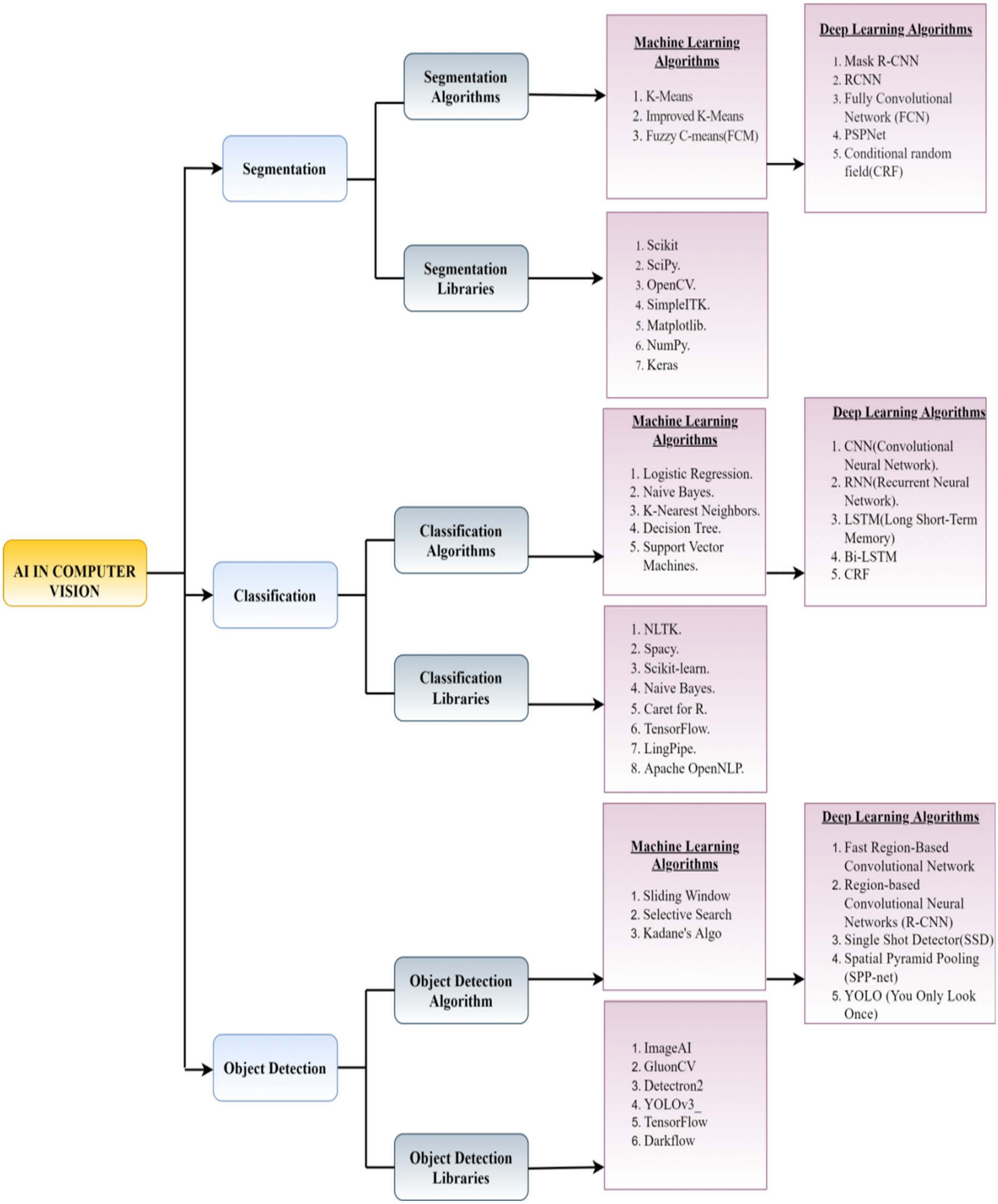

FIGURE 4. AI in computer vision.

terms of past eight year's publication count of papers in Scopus of each learning style and existing survey status of each ML style. Section 4 describes the introduction, framework, computer vision applications, datasets and techniques used with accuracy achieved of all machine learning styles. The comparative analysis of different ML styles, research gaps identified with future directions discussed in Section 5.Section 6 gives the conclusion of this study. Figure 3 shows the overall outline of the paper.

II. BACKGROUND STUDY

A. EVOLUTION OF MACHINE LEARNING STYLES

Figure 5 shows the evolution of each learning style from past to current time.In Oct 1946, Denny M.R published the first paper on reinforcement learning in which learning with 50% reinforcement & 100% reinforcement was compared in control groups [7]. Then in 1958, the study of the transfer of training & considered their implications for the study of perceptual learning recognition was explained by Vanderplas [8]. After this, in Oct 1964, the advantage of the developed system under unsupervised learning in pattern recognition problems was discussed by Pu and Chen [9]. In the same year, the first paper on Active Learning was published. By stacking multiple images onto a board of various colors, the subject of this single experiment was required to understand the link between color pictures. Everyone of any age can benefit from this type of active learning. Chen published the first work on supervised learning [10].

In July 1970, an initial investigation was done to test whether a perceptual learning process learns the visual symbols & transfer procedure was used with deaf first-grade children. This experimental study's authors found evidence for distinctive feature learning [11]. Based on Vygotsky's theories, Sir James Britton and others in England developed Collaborative Learning(Co-Learning) in the 1970s as an active learning method. According to Britton, a student's learning comes from a community of learners composed of other students [12]. Then, A theoretical rationale elaborates upon the concepts of meta goals. Meta-learning was provided in April 1975. In 1979, Seltzer Donald S. explained how robots could learn from different methods. This author explained how sensory information is used for improved Robot learning. Then, in 1980, scientists presented an adaptive model for self-supervised learning that uses a single pattern training technique to recognize vowel sounds on a computer [13].

In 1987 Littlestone Nick published his first paper on online learning. In this, online learning of various classes of boolean functions from examples is studied. Board later rediscovered semi-supervised learning in 1989. With the learning algorithm only having access to incomplete information, several unrelated concepts were learned at once. In 1990, Suddarth & Kergosien developed multi-task learning, the main concept of which is sharing what is known by various tasks while activities are trained concurrently. Then in the same year Transduction term was coined by Vladimir Vapnik. After this, in July 1994, the first paper on Co-learning on recursive functions was published. In the same year, Macoun & Richard developed a constructivist learning model helpful for ethics education.

The first study on ensemble learning, which discussed a decorrelation network training technique for enhancing the efficacy of regression learning with ensemble neural networks, was published in 1996. Then, on June 24, 1999, the authors published a study on association rule learning. They claimed that induced rules were not primarily designed for categorization and that the new measures employed for association rule learning were support and confidence. After that, the first papers on multi-view and multi-instance learning were published in 2003. Then in 2006 CELEBRATE project developed and demonstrated a federated learning object brokerage system architecture by Massart & David. Later in Dec 2013, Marcus& Ebert published the first paper on few-shot learning when datasets with few labels are available.

In 2021, the research will be directed towards all these newly introduced machine learning styles integrating computer vision applications in various domains using large pre-trained models.

III. LITERATURE REVIEW

The Scopus publication count per year is examined in exploratory data analysis. A total of eight years are considered for publishing years between 2015 and 2022. We can see the year-by-year publishing of each machine learning style by analyzing the data. In 2018, these advanced machine learning styles area has drawn the attention of numerous researchers.

As compared to traditional machine learning styles, Transfer learning and Multi-task learning styles of advanced machine learning have gradually increased yearly shown in figure 2. With 338 publications collected from Scopus in 2021, the publication count shows a strong increase.

There is tremendous scope of work in this area to work on advanced machine learning styles using AI techniques in upcoming years..

IV. MACHINE LEARNING STYLES

Many learning techniques depend on the way algorithms use many layers to extract progressively higher-level information from the raw input. Figure 1 illustrates several learning methods. In computer vision applications, including image segmentation, object detection, text recognition from an image, and association rule, these learning techniques are evolving into cutting-edge trends.

Figure 7 depicts the various machine learning styles. Primary learning styles include Supervised, Unsupervised, and Reinforcement learning. Multi-instance, Transductive, Active, Meta, and Multi-task learning are the styles of supervised learning where the input data is labeled.

Classification, Regression, and prediction are the everyday tasks performed with these styles. Unsupervised learning includes self-supervised Learning, Constructive Learning, and Association Rule types. Association rule mining, clustering analysis, data summary visualization, and time series analysis are the essential tasks performed using Unsupervised Learning. Reinforcement learning is used mainly for sentiment analysis, robotics, and gaming. It is based on the decision taken to achieve the reward.

Other machine learning styles are becoming popular in various applications using AI. Those popular styles are

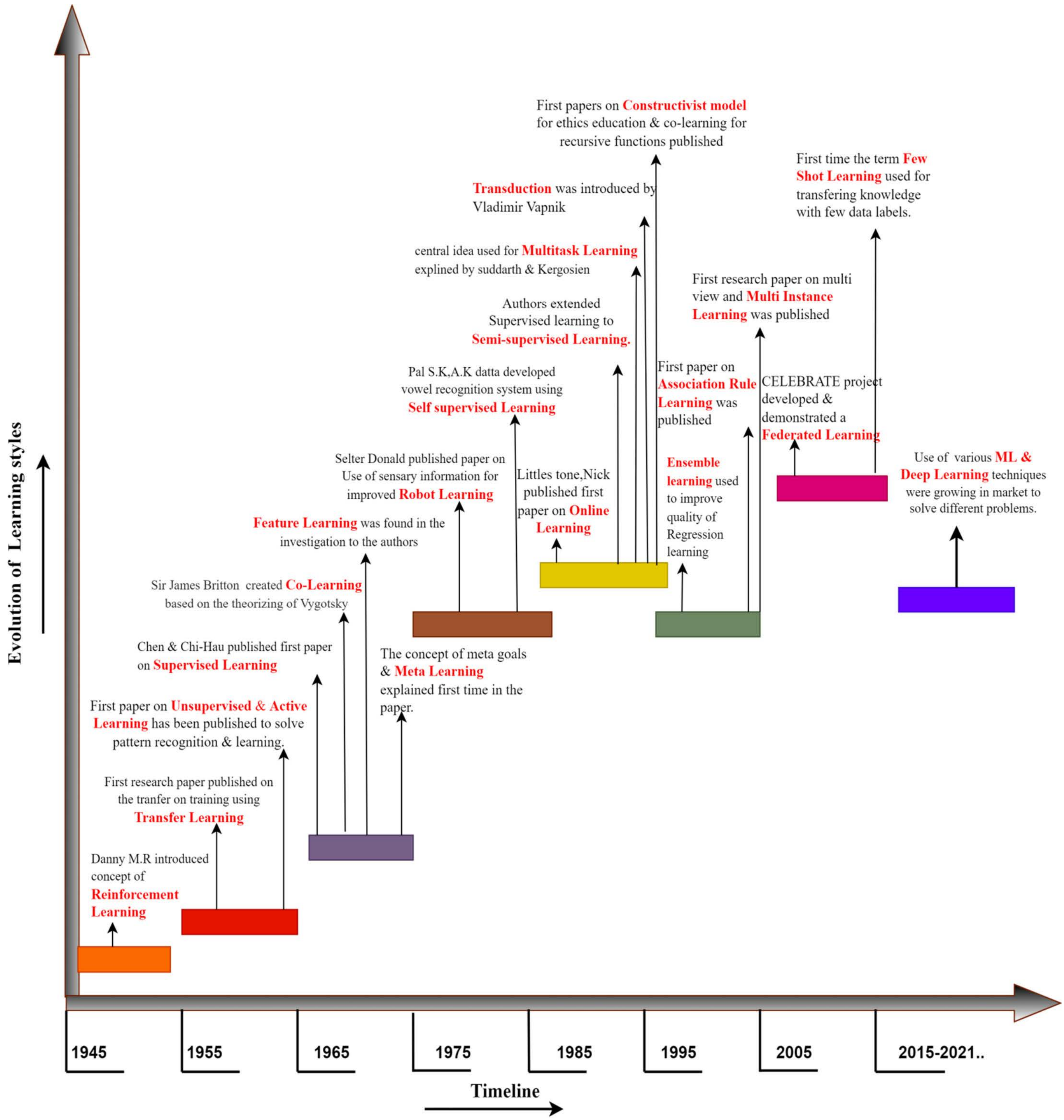

FIGURE 5. **Evolution of machine learning styles-a brief history of machine learning styles.**

Transfer learning, Federated learning, Self-taught Learning, Multi-view Learning, Online Learning, Co-learning, Few-shot learning, etc.

A. PRIMARY STYLES

Supervised, Unsupervised, Reinforcement learning, hybrid, and other learning styles are the basic categories into which machine learning styles in computer vision are divided.

1) SUPERVISED LEARNING

A machine learning task called supervised learning converts every input item to the required class label value. An object is mapped by the computer with the intended output after training. It includes a broad selection of algorithms for various supervised learning issues. Over time, applications in computer vision and machine learning have increased dramatically, with society as the only gainer. Supervised learning

TABLE 1. Summary of existing surveys related to machine learning in computer vision.

Type of Survey	Applications	Data Set used	Pre-processing Techniques	Feature Extraction (Segmentation, Classification& Object Detection)	Machine Learning Styles	Performance Metrics	Challenges discussed in the survey paper				Overview	Future Directions
							Dataset	Other learning paradigms	Technique	Future Directions		
Literature Review[14]	×	×	√	√	×	×	√	√	×	√	This survey examines the theories, overall strategies, forward-looking strategies, datasets employed, performances, and flaws in the contemporary learning styles.	Meta-learning, Lifelong Learning, and evolutionary approaches are a few more learning paradigms that get their inspiration from human thought and abstraction, and they may offer fascinating study opportunities.
Literature Review[15]	√	√	√	√	×	√	√	×	×	√	In this, image labeling for computer vision applications is surveyed. It covers key points, approaches, algorithms, datasets, and preferred deep learning models for image labeling in CV.	Other learning paradigms, such as active or reinforcement learning, may reduce the amount of manual labeling in the long term.
Literature Review[16]	√	×	√	√	×	×	√	×	×	√	It focuses on network threats and zero-day attacks using computer vision techniques. It presents learning styles, features, merits, and demerits for detecting network threats.	It's critical to create far more extensive and recent security datasets to create benchmarks and compare alternative solutions.
Literature Review[17]	×	√	×	√	×	√	×	√	×	√	When working with heterogeneous data, this review focuses on the similarities between inter-and intra-modal learning and explores the application and future perspectives.	different fusion strategy, ranging from classical to deep learning techniques, is required for each multimodal issue. For accuracy and efficiency, the appropriate fusion strategies are needed.
Literature Survey[18]	×	×	×	×	×	√	√	√	√	√	It thoroughly analyzes deep learning models for medical diagnosis, including technical contributions and shortcomings.	In this survey author suggested with more machine learning styles, the performance can be further improved
Literature Survey[19]	√	√	√	√	×	×	×	√	√	×	It examines frameworks for multimodal affect analysis and affective computing. It generally uses text, visual, and audio data while researching multimodal data fusion approaches.	-
Review[20]	×	√	√	×	×	×	×	√	√	√	It surveys deep learning techniques on some datasets and discusses future directions for Cyber security.	The author suggested that larger datasets and other learning styles can be applied to improve performance.
Literature Survey[21]	×	×	√	√	×	×	√	√	√	×	It reviews different methods employed in deep learning for computer vision and suggests the application to which we can apply the same.	-
Literature Survey[22]	√	√	√	√	×	√	×	√	√	√	It reviews different approaches for object detection and found many challenges, so they suggested the technique of how that challenges can be solved using machine learning styles.	In this survey author suggested the different machine learning styles to resolve the various challenges of object detection
Proposed Paper Survey	√	√	√	√	√	√	√	√	√	√	In this survey, we have summarized each learning style's challenges and provided solutions to those challenges.	We outlined which learning style is used to perform which CV task in brief.

is broadly divided into two categories, i.e., Classification and Regression. Objects will be categorized based on recognized class categories in classification to solve various real-world challenges. In Regression, however, the correlation between dependent and independent variables is calculated and displayed using scatter plots [6].

Figure 6 depicts the Supervised Learning process flow, where the input is labeled data from which features are extracted, and the model is trained. The trained model will be applied to the test dataset to forecast the result. The performance accuracy will be calculated by comparing the predicted and actual output.

Advantages of Supervised Learning-

- It gives more accurate results of classification than Unsupervised learning.
- It is simple to train and test the model with labeled dataset.

Disadvantages of Supervised Learning-

- Lack of training dat;
- Poor data qualit;

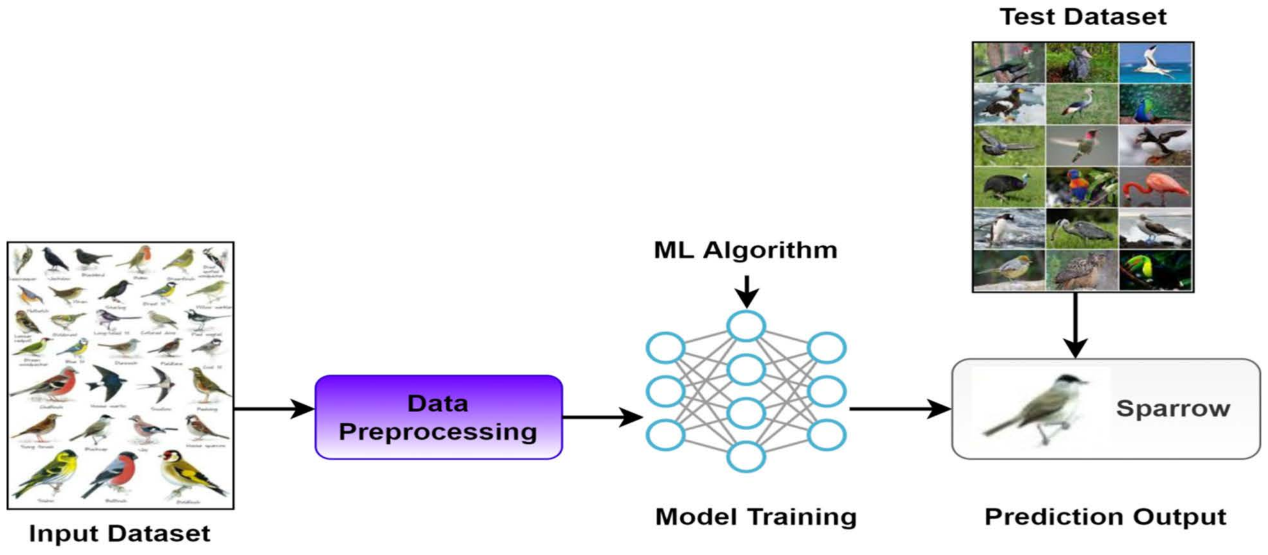

FIGURE 6. Framework of supervised learning.

- Underfitting or overfitting of training data;
- The process of machine learning is complicated;

a: MULTIPLE INSTANCE LEARNING (MIL)

MIL is a form of poorly supervised learning in which training examples are collected into bags and given labels for the whole bag rather than for the specific cases. It enables the use of poorly labeled data, which is common in many business challenges due to the high cost of labeling data [29]. Figure 8 shows the framework for MIL training Phase. In this, we give training images as input. In training bags, image segmentation and feature extraction are performed, based on the size of dictionary bag features computed and finally applied, a classifier; the model predicts the result [30].

The majority of tasks using computer vision in medicine are either:

1) Image classification for diagnosis or
2) Segmentation to detect and separate lesions.

Most contemporary MIL approaches presume that positive and negative cases from a positive and negative distribution are sampled separately. Due to the co-occurrence of several relationships, this is frequently not the case:

Similarities Within the Bag:

Similarities exist between examples from the same bag that does not exist between instances from other bags. In computer vision applications, all segments are likely to have certain commonalities in capture conditions (e.g., illumination). Overlapping patches in an extraction process is another possibility.

Co-Occurrence of Instances:

When instances share a semantic link, they co-occur in bags. Or when particular objects are frequently discovered together, or it is more likely to be kept in one place, this form of correlation occurs.

Table 3 gives experimental results of the classification problem using KNN, SVM & Bagging-APR algorithms on MUSK1 & MUSK2 datasets [31].

In table 3: MUSK1 and MUSK2 are benchmark datasets [3] consisting of 92, 102 bags, 5.17, 64.69 instances, and 47, 39 positive bags, respectively.

Advantages of Multiple Instance learning-

- Multiple instances learning deep neural networks are able to learn the features that optimally represent the given training data.
- It works with worse classification performance.

Disadvantages of Multiple Instance learning-

- Pooling functions are predefined and non trainable.
- Hyper parameter r is global, thus, it is not adaptive to new instances.

b: ACTIVE LEARNING

In order to quickly train an algorithm, Active Learning (AL) tries to simplify data collection by automatically identifying which instances an annotator should categorize. The premise of active learning is that unlabeled data is readily available but costly to label. Given this, active learning aims to extensively use unlabeled data without incurring the expense of labeling it. Active learning has been a popular study area in many machine learning applications. Active learning has recently been the subject of ongoing research and hypotheses that it can outperform typical supervised learning algorithms in some situations, such as when there is a lot of unlabeled data, and manual labeling is expensive [34], have defined the Active Learning in Object Detection application. Weak supervision techniques with active learning, such as: using other forms of inadequate supervision with active learning, expressing the problem of integrating weak and strong supervision as an optimization problem under budget

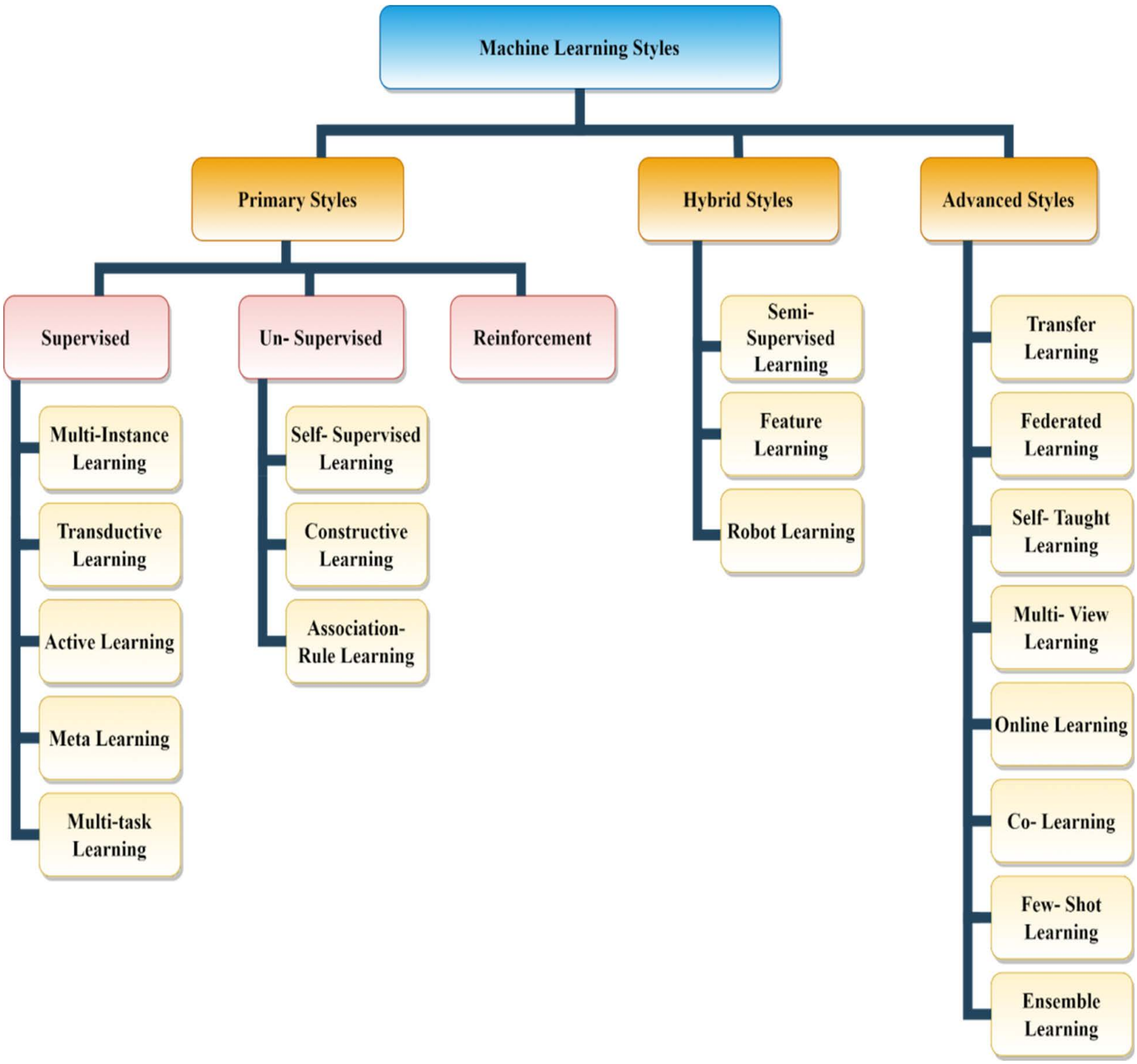

FIGURE 7. Taxonomy of machine learning styles classification.

restrictions, and merging active learning techniques with data programming-based weak supervision approaches, are just a few examples.Figure 9 shows the architecture diagram of how active learning works.

Given a set of items called "I" and a machine learning algorithm called "M," AL seeks to provide a method for gradually selecting items from "I" to obtain actual labels so that "M" can be taught with a reduced dataset for random item sampling. The essential premise is that obtaining training data is expensive, hence it is very advantageous to reduce the amount of such a dataset for certain target accuracy. [35].

Applications of Active Learning-

- Both in-person and online classes can incorporate active learning.
- When there is too much data to classify, or intelligent labeling of the data needs to take precedence over other tasks, active learning might be used.
- By adaptively choosing which samples to classify for prediction, active learning produces highly accurate predictive models at a low cost.

To apply active learning to an unlabeled data collection, follow these steps-

1. A small subset of this data must be manually labeled as the first step.
2. The model needs to be trained after a small amount of labeled data has been gathered. The model won't be perfect, of course, but it will help us choose which areas of the parameter space to label first to improve it.

TABLE 2. Supervised learning applications.

Ref.	Application Area	Dataset used	Technique used	Accuracy
[23]	Breast cancer diagnosis	1. Wisconsin Breast Cancer (WBC) 2. Wisconsin Diagnostic Breast Cancer (WDBC)	• SVM with six kernel functions is utilized	96.31% 96.67%
[24]	Spam email detection	1. Email dataset	• Naïve Bayes • decision tree	88.12%
[25]	Brain Tumor Classification	1. Magnetic resonance imaging (MRI)	• Neural Network algorithms are used.	99.25%
[26]	Hand Gesture Recognition	1. Two centroid distance datasets	• Gesture Learning Module Architecture (GeLMA)	99 %
[27]	Snake Species Identification	iNaturalist, HerpMapper	• CNN	79%
[28]	Inspecting Buildings Using Drones	Crack dataset	• CNN	93%

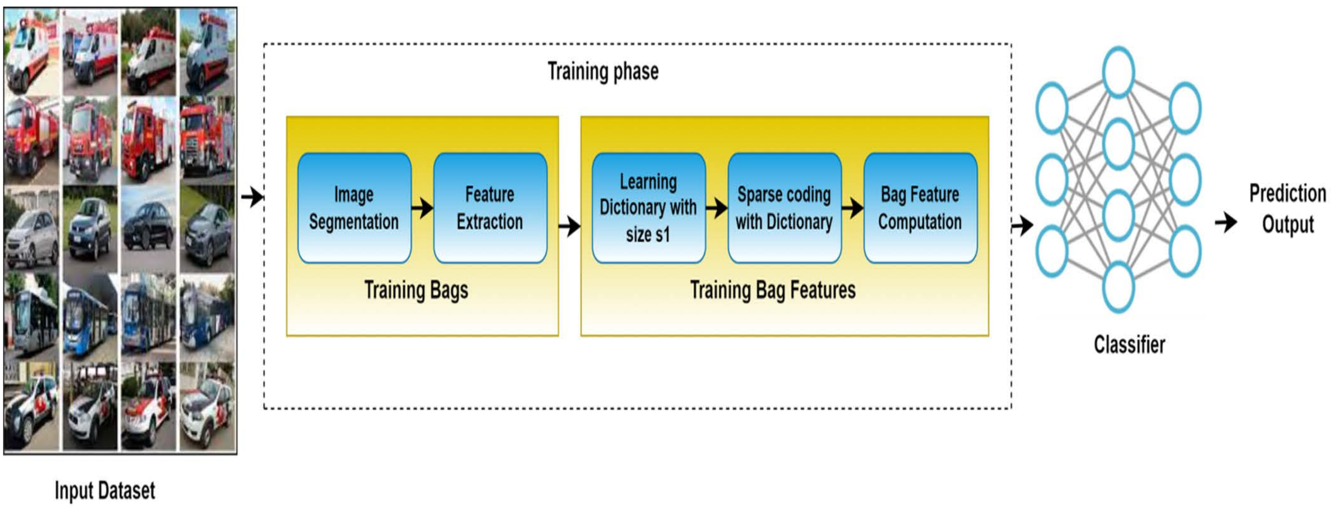

FIGURE 8. Framework of multiple instance learning.

3. Each remaining unlabeled data item's class is predicted using the model after it has been trained.
4. Depending on the prediction made by the model, a score is given to each unlabeled data point.

Once the appropriate strategy for prioritizing the labeling has been identified, a new model can be trained on a new labeled data set that has been labeled based on the priority score. After the new model has been trained on the subset of data, it can analyze the unlabeled data points to update the prioritizing scores and continue labeling [33]. In this manner, as the models advance, the labeling method may be continually improved.

Advantages of Active Learning-
- To minimize the need for labeling issues such as image annotation, recognition, object detection, segmentation, and posture estimation.
- Active learning (AL) aims to maximize the performance increase of a model.

Disadvantages of Active Learning-
- It is time consuming
- Sometimes memorization is necessary
- Not all outcomes are predictable

c: META-LEARNING

Meta-learning is the process of studying itself. The most frequent examples of meta-learning are machine learning

TABLE 3. Multiple instance learning.

Ref.	Application Area	Dataset	Technique used	Accuracy
[31]	Classification	MUSK1 MUSK2	• Bagging-APR • 10 Fold cross-validation	92.3% 93.1%
[32]	Content-based image retrieval and Classification	Benchmark	• MIL methods with 10 fold cross validation	88%
[31]	Classification	MUSK1 MUSK2	• SVM • 10 fold cross validation	87.4% 83.6%
[33]	Medical Image Analysis	Benchmark	• Semi-supervised, multiple instances, and transfer learning for diagnosis and segmentation tasks in medical imaging.	89%
[31]	Classification	MUSK1 MUSK2	• Bayesian-KNN • 10 Fold cross-validation	90.2% 82.4%
[30]	Biology and Chemistry	Benchmark and Musk	• CNN	78%

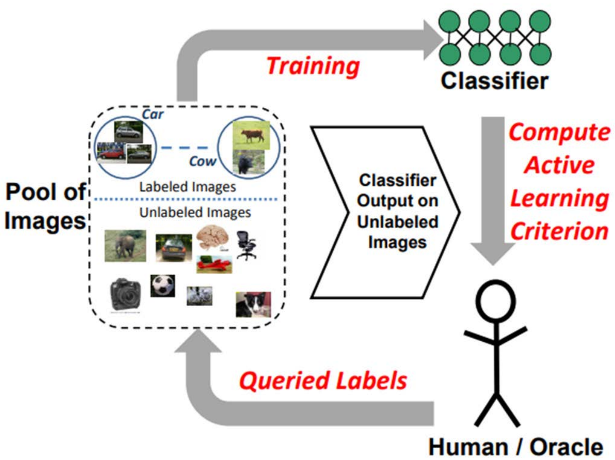

FIGURE 9. Framework of active learning.

algorithms that learn from other machine learning algorithms. Usually refers to the use of machine learning algorithms capable of combining predictions from different machine learning algorithms in the most effective way possible. Multi-task learning algorithms capable of learning across several related prediction tasks are also referred to as meta-learning. There is meta-learning within the framework of supervised learning [41]. Figure 10 shows the work flow of meta learning style.

There are two perspectives on meta-learning.

1) Mechanistic View:

- Deep neural network model that can scan a complete dataset and generate predictions for fresh data points
- This network is trained using a meta-data set, which comprises numerous datasets for a particular task.
- This viewpoint simplifies the implementation of meta-learning algorithms.

2) Probabilistic View:

- To effectively learn new tasks from a probabilistic perspective, extract prior knowledge from a set of (meta-training) tasks.
- This last and (limited) training set is used to predict the most likely posterior parameters when learning a new task.
- This viewpoint simplifies the comprehension of meta-learning algorithms.

Meta-learning techniques take notes on other data-driven machine learning algorithms' outputs. Meta-learning, then, necessitates the presence of different learning algorithms that have been trained on data [42]. For classification and regression problems, supervised meta-learning algorithms, for example, learn how to translate output instances from other learning algorithms (such as projected numbers or class labels) onto examples of target values. On the other hand, meta-learning algorithms forecast a number or a class label by using the output of current machine learning algorithms as input. Meta-learning, also known as meta-machine learning, learns how to use predictions provided by machine learning algorithms in the same way machine learning learns how to use data to produce forecasts.

Meta-Learning Algorithms

- Non-parametric techniques,
- optimization-based inference,
- and black-box adaptation
- Bayesian meta-learning

The table no. V shows a brief about meta-learning. This author used two small Datasets, i.e., AwA (Animals with Attributes) has about 30,000 photos of about 50 distinct animal classes. There are 218 occurrences, 1000 visible categories, and 360 unseen categories in CUB-200-2011 Birds (CUB). Based on the ILSVRC2012 and ILSVRC2010 datasets, ImNet-2 offers 1000 classes for visible courses and 360 classes for unseen classes. The authors use of semantic auto-encoder allows them to choose the best function for mapping semantic space and feature space so that it also functions for classes and semantic space that aren't visible.

Advantages of Meta Learning-

- It improves the speed and adaptability of AI systems to environmental changes.
- The key artifact for comprehending and learning the entire system.

TABLE 4. Application area of active learning.

Ref.	Application Area	Dataset	Technique Used	Accuracy
[36]	Image Classification	MNIST and CIFAR-10	• CNN	90%
[37]	Object Detection	PASCAL 2007 2012 datasets	• CNN	96.5% 81.9%
[38]	Gaussian Processes for Object Categorization	Caltech-4 Caltech-101	• Support vector machine(SVM) • Pyramid Match Kernel Gaussian Processes techniques are used.	85%
[39]	Facial Action Unit Detection	AM-FED	• **SVM classifiers:** using a linear kernel, RBF kernel and approximated RBF kernel	87% to 94%
[40]	Classification of Labeled & Unlabeled data, semantic segmentation	CIFAR10	• Task-aware variation adversarial Active Learning (TA-VAAL)	75%

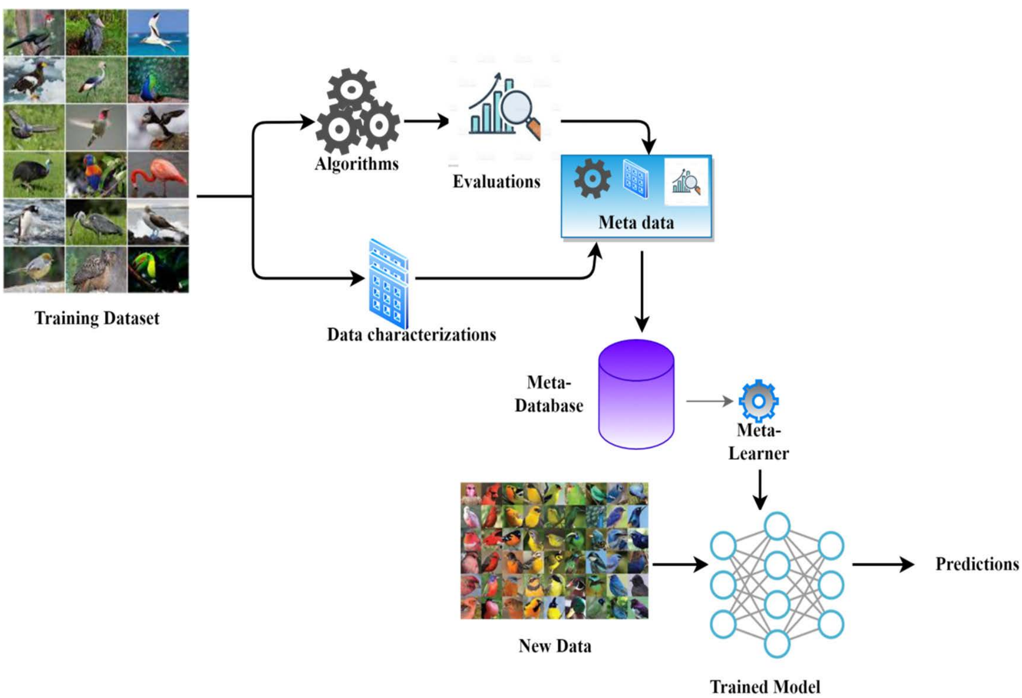

FIGURE 10. Meta-Learning framework.

- To improve their predictions, meta learning algorithms can learn to incorporate the best results from machine learning algorithms.

Disadvantage of Meta Learning-

- The natural restrictions of measuring the true performance of the dataset may make performance estimation incorrect.

d: MULTITASK LEARNING-

Multi-task learning describes a training paradigm in which a single training model learns multiple tasks concurrently. This Learning style enables the usage of beneficial connections found in related jobs. Compared to separately trained models, they increase generalization across all functions, increasing prediction accuracy for specific tasks.

TABLE 5. Applications of meta-learning.

Ref.	Application Area	Dataset	Technique used	Accuracy
[43]	Parameter Tuning	AwA CUB	• Semantic Auto-Encoder technique (SAE).	84.7% 61.4%
[44]	Clustering problem	AwA CUB	• Semantic Auto-Encoder with tuning technique. • SAE + Zero-shot Learning	84.0% 60.9%
[45]	Optimization	Mini-ImageNet	• Used LSTM technique. • Meta Learning + Few shot Learning with 95% confidence intervals	1shot - 77% 5-shot – 71%
[46]	Memory Augmented Neural Networks	Omniglot ImageNet	• One-shot learning is much easier • Matching Networks	93.8% 88.0%
[47]	For object recognition	CUB-200-2011	• Meta Learning + Few shot Learning	71%

As shown in Figure 11, MTL takes input data from text, images, or numbers. For shared layers, an encoder or auto-encoder is used to train the model, which will solve the task-specific problems simultaneously to solve multiple similar issues.

Multiple machine learning applications, including natural language processing, speech recognition, computer vision, and drug discovery, have exploited multi-task learning. Many predictions from training models, such as semantic segmentation and picture classification, can be made on a single sample.

Two MTL methods for Deep Learning-

1) Hard Parameter Sharing-It is typically implemented by preserving several distinct output layers to each task while sharing the hidden levels across all tasks.

2) Soft Parameter sharing-The second approach to MTL is soft parameter sharing, where the shared layers learn their parameters independently.

Advantages of Multi-task Learning-

- By utilizing MTL, the data model can better develop a valuable representation of the data, minimizing data overfitting and boosting generalization.
- It saves model training time as single training model learns multiple tasks concurrently.

Disadvantages of Multi-task Learning-

- Multi-task learning does not necessarily work better with fewer input data. It can be the limitation of MTL.
- The MTL technique has the potential to reduce overall performance in some circumstances. Tasks can compete with one another during the training of an MTL network to produce a more robust learning representation, meaning that one or more tasks may take control of the training process. Along with additional exercises, learning how to recognize things at the pixel level in an MTL scenario is taught. The latter task frequently dominates the learning process unless a task-balancing technique is implemented, such as segmenting a different mask for each object in an image [48].

Additionally, several increased losses may result in a more complex loss function for MTL, making optimization more challenging. In many situations, collaborating on several tasks has a negative effect, and individual networks trained on a single task may perform better.

2) UNSUPERVISED LEARNING

Unsupervised learning majorly works on unlabelled data objects. This type of learning is frequently employed for feature extraction, spotting important patterns and structures, matching together related objects, and practical purposes [51]. Anomaly detection, clustering, density estimation, feature learning, dimensionality reduction, and association rule discovery are some of the most popular unsupervised learning tasks. Figure 12 shows the workflow of an unsupervised learning process for computer vision applications.

a: SELF-SUPERVISED LEARNING

In some ways, self-supervised learning is a sort of unsupervised learning because it adheres to the condition that no labels are assigned. Self-supervised learning, on the other hand, instead of looking for high-level patterns for clustering, tries to tackle tasks typically addressed by supervised learning (e.g., image classification) without any labeling provided. Figure 13 displays the working of self-supervised learning from input data till the final output generation.

Instead of recommending new self-supervised learning techniques, this learning aims to examine how current self-supervised learning strategies might be applied to address domain adaption problems [53]. The primary task can learn a domain invariant feature representation thanks to the pretext

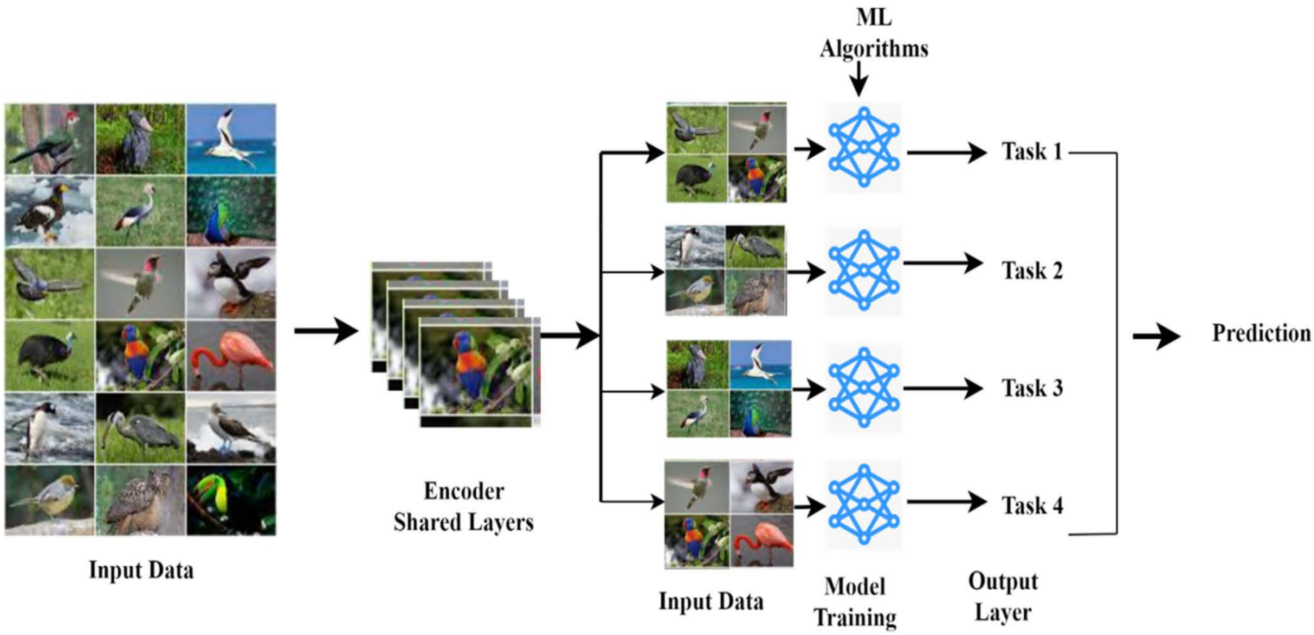

FIGURE 11. Framework of multitask learning.

TABLE 6. Multitask learning applications.

Ref.	Application Area	Dataset	Summary/Technique used	Accuracy
[49]	Image Retrieval`	UT-Zappos50K	Tasks like being overweight, having a double chin, and all three characteristics associated with beards—a goatee, a mustache, and sideburns—are combined.	88.17%
[50].	Object Detection	PASCAL	Cross-stitch units model shared representations as linear combinations and can be learned end-to-end in a ConvNet	63.0%
[51]	Learning system	CityScapes.	It has established a cohesive strategy that uses task-sharing and balance techniques for the learning system.	70% to 77 %
[112]	Networking	Taskonomy	Three techniques for MTL: 1) Optimal solution (OS), 2) Early stopping approximation (ESA), and 3) Higher-order approximation (HOA)	Around 50%
[52].	Brain Tumor Detection & Segmentation	BRATS	Faster-RCNN	Around 85.78%

job connecting the source and destination domains. In the source domain, the primary job has labels; however, in the destination domain, there is no labeling requirement. In other words, we develop unsupervised domain adaptation through self-supervised learning. The forwarded data flow is represented by solid lines in the diagram, while the optional data flow is indicated by dotted lines [53]. Through multi-task learning, the pretext and main task (such as object identification, classification, or semantic segmentation) are simultaneously learned.

Advantages of Self-Supervised learning-

- The frequency of labeling needed may be reduced with the use of self-supervised learning.
- Self-supervised learning can enhance the effectiveness of robotic surgeries by determining the dense depth of the human body.

Disadvantages of Self-Supervised learning-

- It is very difficult to duplicate samples.
- Semantic distributions of collected data

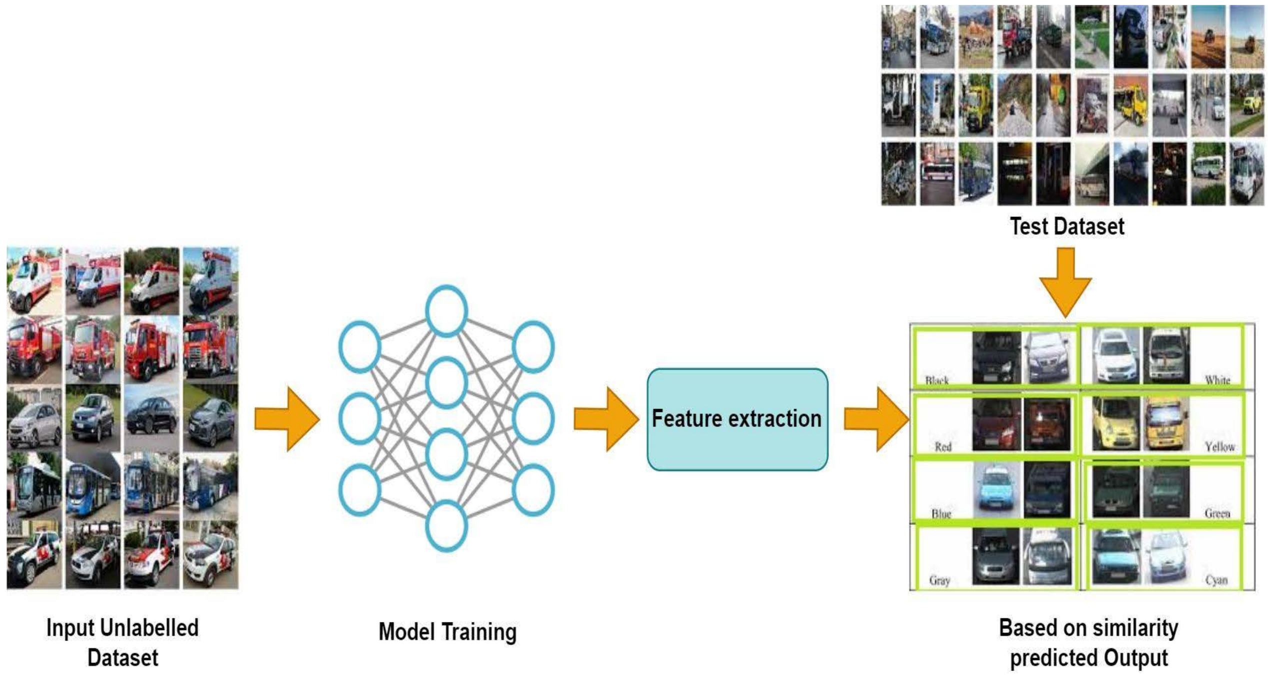

FIGURE 12. Framework of unsupervised learning.

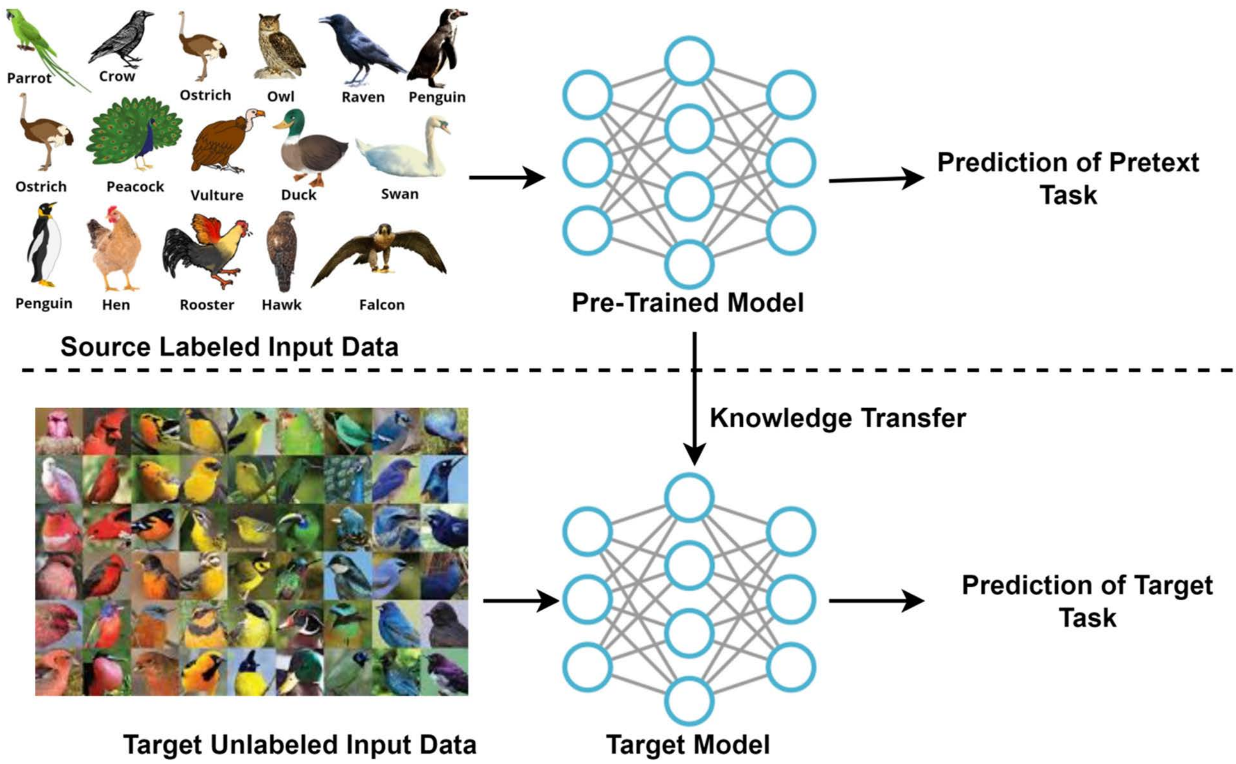

FIGURE 13. Framework of self-supervised learning.

b: CONSTRUCTIVIST LEARNING

Constructivist learning changes the network structure as it learns, resulting in a network that is automatically the appropriate fit. This method begins with a ''small'' initial network. Other hidden units and/or hidden layers are gradually added until a preset error criterion is fulfilled or no performance improvement is visible [55]. Figure 14 provides practical insight into constructive learning. The relationship between

TABLE 7. Self-supervised learning.

Ref.	Application Areas	Dataset used	Summary/ Technique used	Accuracy(%)
[53]	Object Recognition Amazon → Webcam	Office dataset ResNet-50).	Self-supervised domain adaptation disguised as a jigsaw puzzle challenge	86.9%
	Semantic Segmentation	Multi-source Domain Adaptation ResNet-18		84.93%
	Object Recognition Amazon → Webcam	Office dataset (ResNet-50).	Self-supervised domain adaptation using a pretext task for image rotation	90.1%
	Semantic Segmentation	Multi-source Domain Adaptation ResNet-18		88.67%
	Object Recognition Amazon → Webcam	Office dataset (ResNet-50).	Self-supervised domain adaptation using a challenge for rotation prediction that takes into account space.	87.3%
	Semantic Segmentation	Multi-source Domain Adaptation ResNet-18		86.57%
[54]	Medicine	Manual Dataset	CNN	95.0%
[113]	Visual Categorization	Ilsvrc-2012	S^4L: Self-Supervised Semi-Supervised Learning. S^4L -Rotation S^4L Exemplar	83.3 %

cognitive dynamics and emotion axes in the learning process Negative emotions are on the left side of the horizontal axis, while positive emotions are on the right. The vertical axis represents constructive learning, whereas the vertical axis represents destructive learning. A learner's affective and emotional state should be kept within the first two quadrants to maintain a reasonable learning rate. Suppose the tutor notices that the student's emotional state is shifting into the third or fourth quadrants. In that case, they must take immediate action to prevent the dynamic transfer, which could restart the entire learning process [55].

Advantages of Constructive learning-

- Constructive learning techniques make it easy to design the initial network architecture.
- Constructive techniques are more effective in terms of network complexity and structure as well as training time.
- Because of their incremental learning nature, constructive algorithms tend to develop tiny networks.
- Effective algorithms require the specification or selection of several problem-dependent parameters; they are not restricted to "acceptable" and "great" networks producing good performance outcomes.

Disadvantages of Constructive learning-

- Even though learners won't always actively generate meaning and construct a suitable knowledge structure, learners will appreciate this new method to learning.
- The learner may be restricted by conceptualizing learning in that, at least initially, they may not be able to create abstractions and transfer information and skills in new settings.

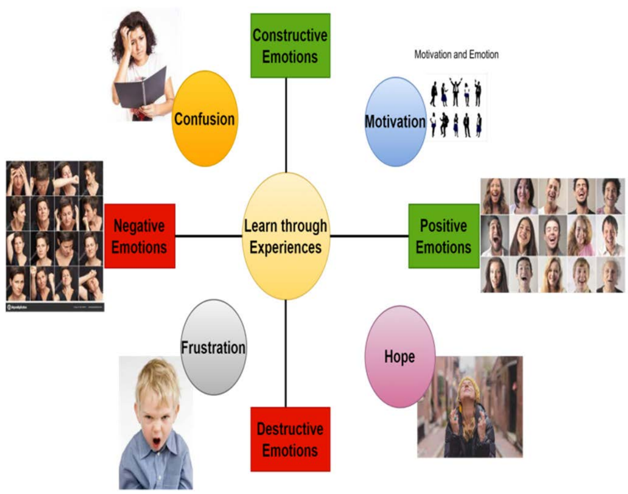

FIGURE 14. Constructive learning.

c: ASSOCIATION RULE LEARNING

An unsupervised learning method called association rule learning looks at how one data item depends on another and then maps to make it more profitable. It seeks to uncover exciting connections or interactions between the dataset's variables [58].

A technique of machine learning based on rules called association rule learning can be used to find "IF-THEN" sentences or other meaningful correlations between variables in large datasets. One example is that if a consumer purchases a computer or laptop (one thing), they will also purchase anti-virus software (another item) simultaneously. IOT services, medical diagnosis, usage behavior analytics, web usage

TABLE 8. Applications of constructive learning.

Ref.	Application Area	Working
[56]	Facial Expression Recognition	• To classify face expressions, they used a useful one-hidden-layer feed-forward neural network. • For training and generalizing images, the best recognition percentage is 100% (without rejection) and 93.75% (without rejection), respectively.
[55]	Constructive learning for Human-robot interaction, Face tracking, Facial Expression Recognition	• A human instructor lectures the students in a classroom, and the robot mimics the teacher. • To keep the students' emotional state from transferring to the next topic, he immediately attempts to make them understand the issue being studied.
[57]	Cross-model correlation algorithm for 3D model retrieval.	• They developed constructive learning for the cross-modal correlation algorithm and put it to the test on two datasets to show how effective their suggested approach is. • The outcome demonstrates that, in comparison to the most recent state-of-the-art methods, their proposed Method performed better.

mining, cutting-edge phone applications, cybersecurity applications, and bioinformatics are a few examples of contemporary uses for association rules. The order of events within or across transactions is rarely considered by association rule learning, in contrast to sequence mining. Commonly, the "support" and "confidence" metrics are employed to evaluate the value of association rules [2].

Association rule algorithms measure the frequency of complementary occurrences, or associations, across an extensive collection of things or activities. The idea is to uncover relationships that occur more frequently than a random selection of alternatives would reveal. This rule-based strategy is a quick and effective way to mine non-numeric, classified datasets. It is shown with the help of market basket analysis in the figure15.

Example: One well-known application of this methodology is the analysis of retail sales to ascertain the best way to arrange items in a store. Newborn baby diapers may be sold 10,000 times at a business with a million transactions annually, but razor blades may be sold 100,000 times. At first inspection, there is no statistically significant correlation between newborn diapers and razors. On the other hand, rule mining would go further into transaction frequency and find that 5,000 sales involve both products.

The association system introduces a new rule indicating that 50% of all buyers buying newborn diapers also purchase razor blades, which might be helpful to information for marketing efforts rather than just knowing that 1% of customers purchase diapers and 10% buy razor blades.

Moreover, when additional data is analyzed, the rule-based Method improves performance and develops new rules. With a sizable enough dataset, it enables the computer to simulate the human brain's feature extraction and abstract association abilities from unstructured input.

Application Areas of Association Rule learning-

1) **Basket data analysis -** Association mining can help you determine what your customers desire, whether you're planning product placement in a storefront, running a marketing campaign, or producing a business catalog [58].
2) **Web usage mining and intrusion detection** - A powerful prediction tool for identifying new security dangers and network performance issues that haven't been assessed by humans yet is finding these hidden correlations [59].
3) **Bioinformatics (bioinformatics)** - One of the essential methods for uncovering underutilized but potentially valuable processes across a wide range of disciplines, from biology to engineering and everything in between, is association mining [58].

3) REINFORCEMENT LEARNING

Using input from its actions and experiences, an agent is trained in an interactive environment to achieve this machine learning technique's reward and punishment mechanisms. The agent receives rewards for successful attempts and punishment for unsuccessful ones. The agent attempts to minimize inappropriate actions and maximize appropriate ones by learning from their experiences and activities [64]. When a series of decisions are required, reinforcement learning is used. The mathematical foundation of Markov decision processes is used in most reinforcement learning contexts. Reinforcement learning is utilized in computer vision applications for object detection, video analysis, gaming, and animation.

Figure 16 shows the work flow of reinforcement learning process to achieve the reward.

Advantages of Reinforcement learning-

- It is used when online computation time is important.
- It is less tractable both computationally and analytically compared to tracking or regulations problems.

Disadvantages of Reinforcement learning-

- Reproducibility is required.
- Sample inefficiency is a problem.
- Wisely choose reward structure

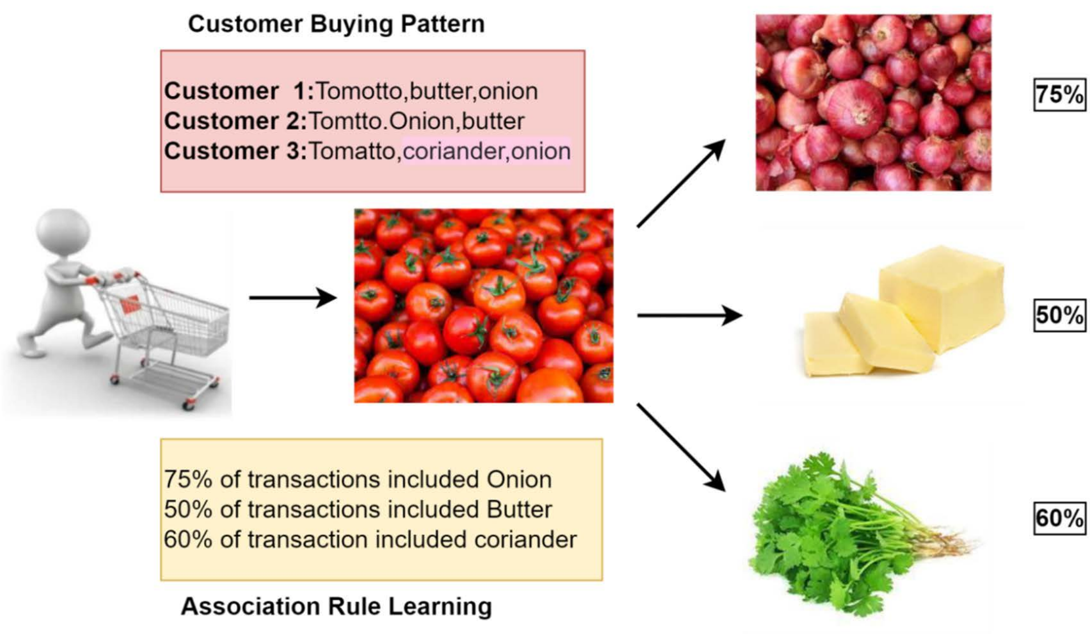

FIGURE 15. Association rule learning.

TABLE 9. Applications of association rule learning.

Ref.	Application Area	Dataset used	Technique used/ Summary	Accuracy
[60]	Image Extraction	Monk's dataset	• CNN • Faster RCNN	99%
[61]	Classification and Detection	PASCAL VOC	• Detection and alignment of 3D CAD chair models performed.	32.1–57.5%

TABLE 10. Algorithms of association rule learning.

Ref.	Algorithm	Working /Advantages	Challenges
[59]	AIS	This approach requires many passes over the entire dataset to generate the rules.	• There are too many candidate itemsets generated. • It requires more space & consumes a lot of time.
[62]	SETM	Effective performance Consistent performance across time	• An excessive number of candidate itemsets are produced, taking up more room and time.
[63]	Apriori	It is a widely used technique. Follows bottom-up technique	• To condense the search space, Apriori uses the property that "all subsets of a frequent itemset must be frequent; and if an itemset is infrequent, then all its supersets must be infrequent."
[64]	FP-Growth/ FP-tree	It uses an interactive mining environment It uses the divide and conquer approach.	• Huge amounts of data would not fit in memory using FP-Tree.

4) HYBRID LEARNING STYLES

a: SEMI-SUPERVISED LEARNING

These algorithms are trained on data that are both labeled and unlabeled. There is a lot of labeled data and a lot of unlabeled data. Figure 17 shows how semi-supervised learning works with labeled and unlabeled data.

The basic approach entails clustering similar data first. Using an unsupervised learning method and then applying it

TABLE 11. Reinforcement Learning (RL) application Areas.

Reference	Application Area	Dataset used	Summary/ Technique used	Accuracy
[65]	RL in Object Detection Active Object Localization	Pascal VOC2007, 2012 Image Dataset	• States: the observed region's feature vector and activity history. • Reward: A different IOU. • 8 actions: bigger, smaller, heavier, taller, left, right, up, down • Pre-trained CNN with 5 layers	54.2%
[66]	Active Breast Lesion Detection	T1-weighted anatomical dataset and DCE-MRI	• States: the region's current feature vector. • Localization improvement is the reward. • Nine actions: six translations, two scalings, and one trigger	TPR -0.8 FPR 3.2 (True and false positive rate)
[67].	Landmark Detection	3D CT Scan	• An axis-aligned box centered at the voxel position is the current state. • Action: move from one point to another. • Reward: distance-based feedback • 6 action: 2 per axis	20-30%
[68]	Monocular 3D Object Detection	KITTI	• State: the 2D image of an object cropped using 2D's identified bounding box and 3D bounding box parameters. Gaining accuracy after taking a certain action is the reward.	67.54%
. [69]	Efficient Object Detection in Large Images	Caltech Pedestrian dataset (CPD)	• Search at the CPNet and FPNet levels. States: the chosen area. • Reward: cost of image capture for detection recall. • Policy: REINFORCE .Binary action array	61.7%

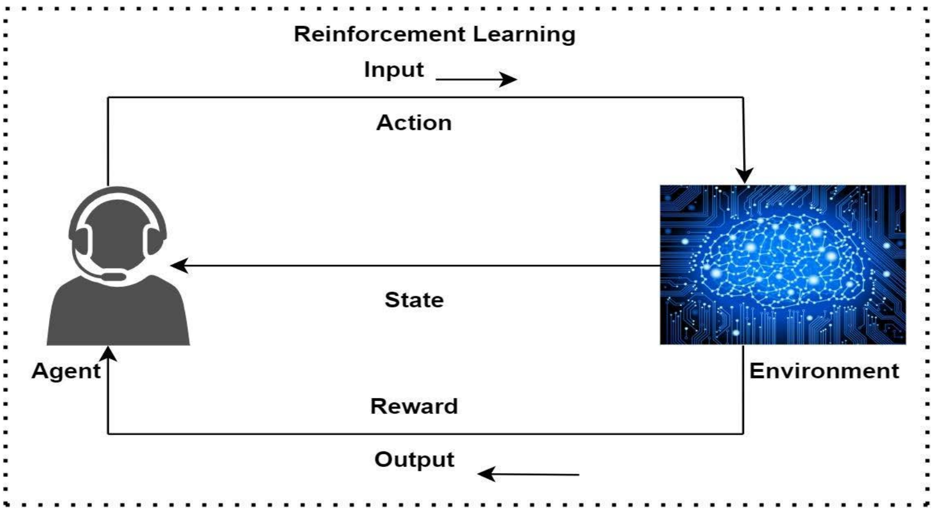

FIGURE 16. Framework of reinforcement learning.

to existing data. The rest of the unlabeled data is labeled using the labeled information [58].

Advantages of Reinforcement learning-

- In this, labeled data can contribute significantly to accurate pattern extraction.
- Semi-supervised learning can result in better convergence by having greater effects on models.

Disadvantages of Reinforcement learning-

- The issue of extending labeled data
- The difficulty of constructing the final classifier

b: FEATURE LEARNING

Current machine learning algorithms rely heavily on manually creating features, and the quality of human

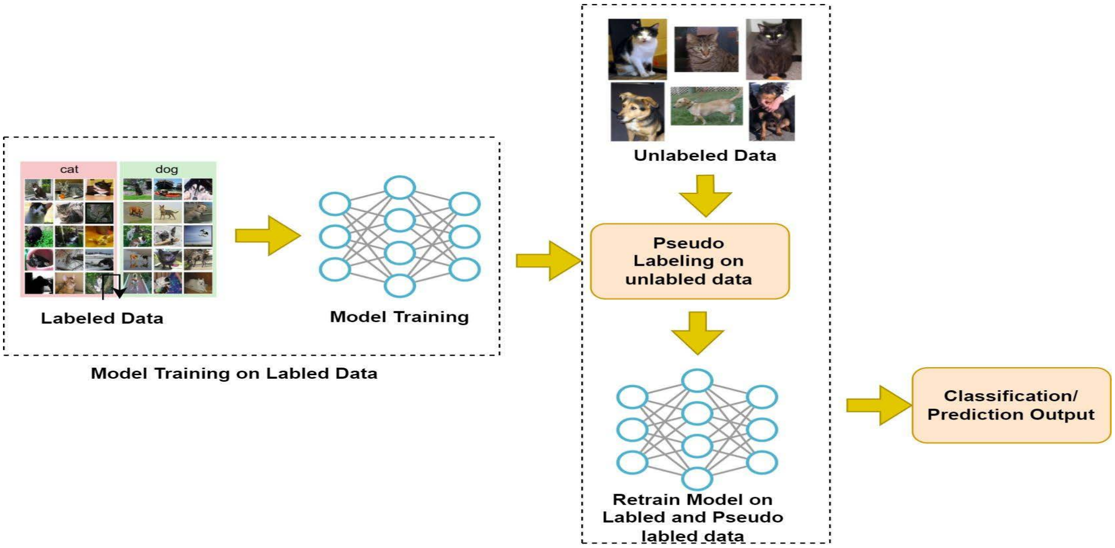

FIGURE 17. Framework of semi-supervised learning.

TABLE 12. Semi-supervised learning.

Ref.	Application Area	Dataset	Techniques Used	Accuracy
[70]	Object Segmentation	DAVIS dataset,	One-Shot Video Object Segmentation (OSVOS)	71.5%
[71]	Video Object Segmentation	DAVIS	Mask Track	74.8
[106]	Visual Categorization	Ilsvrc-2012	• S^4L: Self-Supervised Semi-Supervised Learning • S^4L -Rotation • S^4L Exemplar	83.3 %
[72]	Video Classification	Ucf-101	Semi-Supervised Learning Of Video Classifier.	32.3% to 54.3 %
[73]	Graph Data	MNIST	Graph Learning-Convolution Network (GLCN) for semi-supervised learning	93.70%
[74]	Video Sequences for Urban Scene Segmentation	Cityscapes dataset	Iterative semi-supervised learning	Average precision (AP) : 42.6% mean intersection-over-union (mIOU): 85.2%
[73]	Graph Convolution	MNIST dataset.	Graph Learning-Convolution Networks-GLCN on semi-supervised learning	93.89%

TABLE 13. Feature learning.

Reference.	Application Area	Dataset Used	Summary	Accuracy
[76]	Discovering Visual Patterns in Art Collections	Brueghel dataset	Feature learning with r one-shot cross-modal detection	Cosine similarity 75.3%
[75]	Feature Matching Problem	Synthetic and real-world feature matching dataset	Graph neural network with message passing techniques	69.9 %

representations largely determines their performance. We may never be able to construct the best and most diverse set of features that accurately characterize all variations in our data if we do it by hand. Figure 18 shows the framework of FL.

Images learn features by inducing scarcity using a pool of potential features, a belief network, convolution, or a combination of these methods. These methods have the critical elements listed below (Nithin & Siva Kumar, 2015),

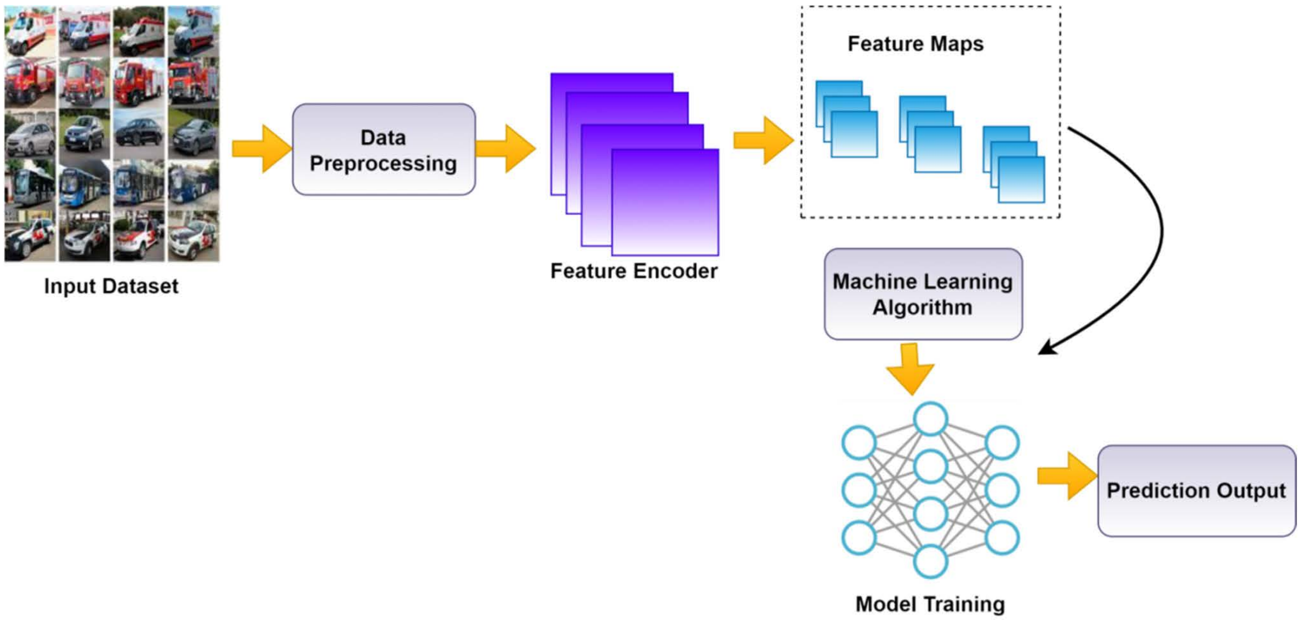

FIGURE 18. Framework of feature learning.

TABLE 14. Robot learning.

Reference.	Application Area	Dataset Used	Summary	Accuracy
[107]	Image Classification	ImageNet Household, UW RGBD and Caltech-256	Image Classification: Root network with random init	25% 46.8% 24.2%
[115]	Vision Processing	The training set consisted of 180 images, and the test set comprised 40 images.	This method deals with noisy, inconsistent, or contradictory data. The robot can cluster images into appropriate categories.	The classification of photos was accurate in 70% of cases. 10% were not classified at all, while 20% were classified incorrectly.
[114]	Robotics	ImageNet Household, UW RGBD and Caltech-256	Auto-encoder trained on all robot data	29.6% 65.7% 28.0%
[77]	Robotics and Computer-Integrated Manufacturing	OpenSign dataset	levels of human-robot interaction, various safety concerns, the anticipation of human motion intentions, teaching by example, and manipulation that falls under the category of human-robot collaboration	80 to 90 % Approx.

which is one of the reasons they can learn features while ensuring that they are generic to any task.

Key elements of Feature Learning are:

1. Hierarchical layer learning
2. Dimensionality
3. Generalization of manifold
4. Disentanglement.

A key issue in many computer vision applications, such as image registration, tracking, and motion analysis, is the feature matching problem. An essential component of effective feature matching techniques is rich local representation. However, it becomes difficult to extract rich local representations when the local characteristics are constrained to the coordinates of important places. Traditional methods solve

NP-hard assignment issues in order to match robustly using pairs or higher order handcrafted geometric features. To solve this issue, a graph neural network model that converts feature point coordinates into local features is suggested. The conventional NP-hard assignment problems are replaced with a straightforward assignment problem that may be handled quickly [75]. Table 13 gives a summary of how feature learning is used in computer vision with datasets used and accuracy achieved.

c: ROBOT LEARNING

Robot learning is a field of study that combines machine learning and robotics. It investigates learning algorithms that allow a robot to learn new skills or adapt to its surroundings. Numerous analytical systems, such as robots, are integrated with visual sensors from which they know the status of their surroundings by solving matching computer vision challenges in multiple applications. These tasks' solutions are utilized to make decisions regarding possible future actions [78].

B. ADVANCED LEARNING STYLES

1) TRANSFER LEARNING

The system's capacity to recognize and apply information and abilities acquired during previous tasks to new ones. There is a need for Transfer learning to minimize the model training time and usage of the resources to solve similar kinds of functions.

In this, if you train a simple classifier to predict whether an image contains a particular set of objects, you could use the same knowledge the model gained during its training to recognize different but related groups of new things [79].

As shown in Figure 19, transfer learning takes a pre-trained model and dataset as input. It works on data and trains the model on that data to perform the machine learning tasks. Then that trained model knowledge will be used to solve similar problems. There are two types of transfer learning: one is positive transfer learning, and another is negative transfer learning. In positive transfer learning, pre-trained models can improve the performance of new tasks and the accuracy of results generated. At the same time, the negative transfer is when the implementation of new tasks degrades due to the previously trained knowledge transfer of the model.

Transfer learning is used in various domains like Medical applications, Biometrics, transportation, recommendation systems, and urban computing applications like traffic monitoring, health care, social security, etc. Pre-training a neural network on the source domain is a way to transfer learning that is frequently employed. for instance, ImageNet, a library of over 14 million annotated pictures divided into more than 20000 categories, then fine-tune it using examples from the target domain.

Machine learning models that deal with natural language processing incorporate transfer learning. Examples include teaching a model to recognize various linguistic components or embedding pre-trained layers that comprehend certain terminology or dialects. To translate models into different languages, transfer learning is used. Models' features are developed and trained using the English language.

Table 15. Summarizes the different strategies used in transfer learning. Despite having the same source and target domains, the source and target tasks are different. The algorithms take advantage of the inductive biases of the source domain to enhance the target job. In the case of transductive transfer learning, the related domains are different even though the source and target tasks are comparable. For unsupervised transfer learning, the main focus is on unsupervised tasks in the target domain where the source and target domains are similar, but the tasks are different. The reusable aspects of a computer vision algorithm will be applied to a new model through transfer learning in computer vision for image and video data processing. Deep learning, a kind of machine learning that aims to emulate and duplicate the processes of the human brain, is reliant on artificial neural networks. Due to the intricacy of the models, neural network training consumes a large number of resources. To increase process efficiency and decrease resource demand, transfer learning is applied.

Advantages of Transfer learning-

- Removing the requirement for each new model to have a significant collection of labeled training data.
- They are increasing the effectiveness of developing and deploying machine learning for several models.
- Using several algorithms to overcome new problems is a more generalized method of machine problem-solving.
- Instead of real-world situations, simulations can be used to train models.

Disadvantages of Transfer Learning-

- One of the most significant limitations to transfer learning is the problem of negative transfer.
- Transferring knowledge from a less related source (where labeled data is less) may inversely hurt the target performance, a phenomenon known as a negative transfer

2) ENSEMBLE LEARNING

To achieve better results, ensemble learning employs strategies that expand models and combine them. Different models used as inputs for ensemble methods in this learning are referred to as base models, which provide better prediction accuracy than a single trained model. Figure 20 shows the framework of Ensemble learning.

Ensemble Learning has three methods as follows-

1. Boosting- Each training tuple is given a weight in boosting. Iterative learning is used to learn several classifiers consecutively. Each classifier's weights are modified as a result of learning. The weight of each classifier's vote is determined by its accuracy, and the final boosted classifier combines those votes.
2. Bagging- Bagging/Bootstrap Aggregating- This ensemble technique creates a model ensemble for

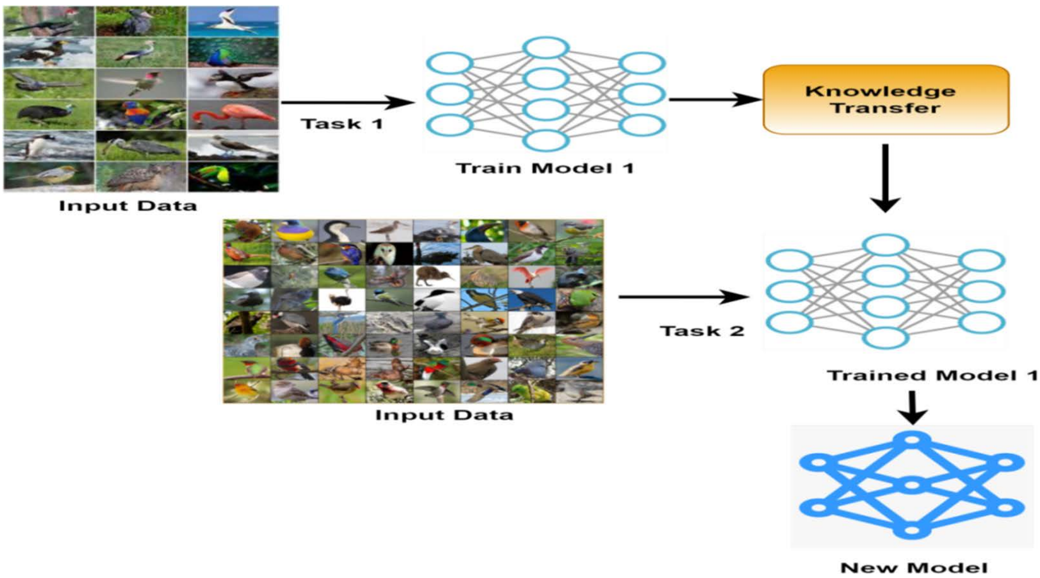

FIGURE 19. Framework of transfer learning.

TABLE 15. Transfer learning strategies.

Learning Strategy	Source and Target Domains	Source and Target Tasks	Source Domain data	Target Domain data	Task Performed
Inductive Transfer Learning	Same	Different but related	Labeled	Labeled	• Classification • Regression
Transductive Transfer Learning	Different but related	Same	Labeled	Unlabeled	• Classification • Regression •
Unsupervised Transfer Learning	Different but related	Different but related	Unlabeled	Unlabeled	• Clustering • Association Rule Mining • Dimensionality Reduction

a learning scheme where each model gives an equally weighted prediction.

3. Stacking –The input data is divided into training and testing. The training dataset is trained using different classifiers and will be taken as input to create a meta classifier. The result of the meta classifier is the final trained model. It will then be applied to the testing dataset to check the classifier's (meta) prediction accuracy [85].

Advantages of Ensemble learning-

- Compared to most other ML styles, ensemble approaches are more accurate predictors than individual models.
- When a dataset contains both linear and non-linear types of data, ensemble approaches are incredibly useful; several models can be coupled to manage this type of data.
- With ensemble approaches, bias and variance can be decreased, and the model is typically neither underfitted nor overfitted.
- A model ensemble is always more stable and less noisy.

Disadvantages of Ensemble learning-

- Ensemble learning is a difficult task to learn, and any poor decision might result in a model with lower prediction accuracy than an individual model.
- Time and storage costs of ensemble model is high.

3) FEW SHOT LEARNING

Few-Shot Learning is a type of meta-learning in which a learner works on multiple similar tasks during the meta-training phase so that it may generalize successfully to unknown (but related) tasks with only a few examples [89] shown in figure 21. This Learning is commonly used to represent many tasks and train task-specific classifiers; on top of this, representation is a practical approach to the Few-Shot Learning problem.

TABLE 16. Applications of transfer learning.

Ref.	Year	Application Area	Dataset used	Summary	Accuracy
[80]	2017	Data-Driven Pavement Distress Detection	Federal Highway Administration's (FHWA's) Long-Term Pavement Performance (LTPP)	• Random Forest Method • Support Vector Machine • Logistic Regression	86 to 88%
[81]	2019	Efficient Hardware for Mobile Computer Vision	ImageNet CIFAR100	• A fixed-weight feature extractor that generates ubiquitous CNN features	70.9 % to 81.7%
[82]	2019	Image classification, Object Detection	ResNet-50, Inception-V3	• Convolution Neural Networks (CNNs) for image classification	81.30%
[83]	2020	Medical Application	Thickness prediction-244 samples Anomaly detection-10,000(imbalanced) Cell Classification -369 sample	• A neural networks model using transfer learning is used on imagenet dataset.	87% to 95%
[84]	2020	British Sign Language Recognition	American Sign Language (ASL) data samples	• late fusion approach to multimodality in sign language recognition improves the image classification accuracy	82.55%

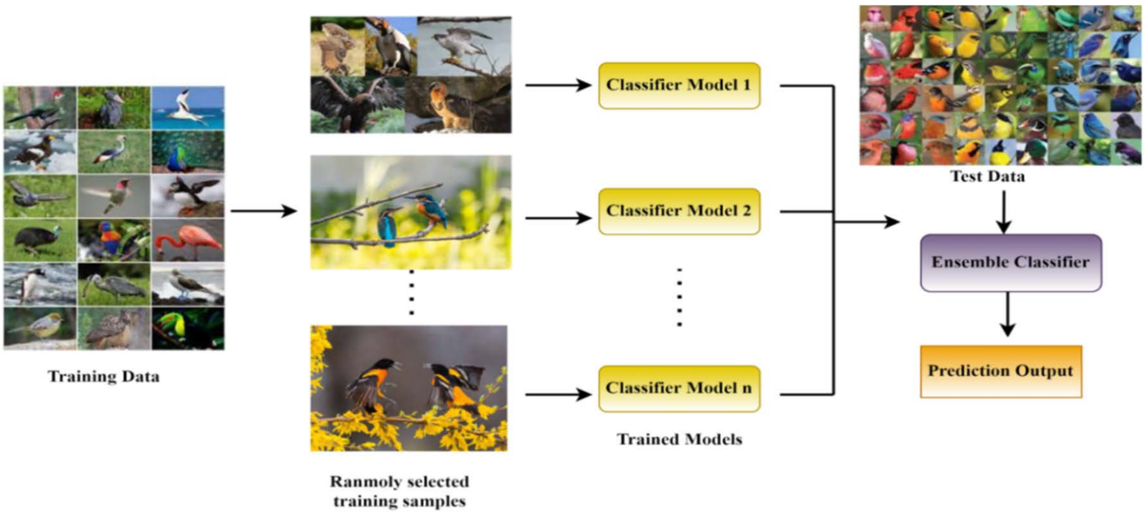

FIGURE 20. Ensemble learning framework.

Applications of Few-shot Learning in computer vision

- Few-Shot Image Recognition Human Motion And Pose Prediction
- Domain Adaptation
- Few-Shot Segmentation
- Learning To Learn from Weak Supervision
- Generating Talking Heads from Images

4) ZERO-SHOT LEARNING

The difficulty of wanting to recognize objects from classes that our model has not seen during training is known as zero-shot learning shown in figure 22. The data for zero-shot learning comprises the following:

1. Observed/Seen classes: These are classes for which we labeled images during training.
2. Unobserved/Unseen classes: These classes do not have any tagged images throughout the training phase.

Types of Zero-Shot Learning-

1) Inductive Zero-Shot: We can get tagged image data from classes that have been observed in this. The key objective is to translate semantic knowledge into visible image space so that the model can identify objects from unobserved classes during testing.

2) Transductive Zero-Shot: We also have access to unlabelled images from unobserved classes in this labeled image data from seen classes.

Methodologies Used for Zero-Shot Learning:

1) Embedding Based Method- The primary purpose of embedding-based approaches is to use a projection function trained using deep networks to map picture features and

TABLE 17. Application areas of ensemble learning.

Ref.	Application Area	Dataset	Summary/ Technique Used	Accuracy
[86]	Road damage detection	Yolov-4	Meta-learning algorithms	F1 score nearer to 62 %
.[87]	Neural Network for Non-homogeneous Dehazing	**Real-world dataset-** : • O-Haze • NTIRE2018,2019,2020, 2021 • RESIDE	Use the fusion tail to map the features from the transfer learning sub-net and the current data fitting sub-net to haze-free images.	NTIRE21-84.2 NTIRE20-70.4 NTIRE19- 58.2 NTIRE18- 78.3 RESIDE- 99.1
[88]	Identification of milling processes	589 images as a dataset	Extremely Randomized Trees classifier	96.20%

semantic attributes into a shared embedding space. The visual and semantic spaces can be used as the common embedding spaces.

2) Generative Model-Based Method- Based on generative models, this approach: The primary drawback of embedding-based techniques is that they exhibit bias and domain shifting. It suggests that because the projection function is developed using only seen classes during training, it will be biased towards predicting seen class labels as the output [84]. The learned projection function may not, at test time, accurately translate unseen class picture features to the pertinent semantic space. The deep network has only been trained to map picture data from observed classes to semantic space. It might not be able to do so successfully during testing for particular, unseen classes.

We must train our zero-shot classifier on unseen and unseen class images to overcome this limitation. It is when methods based on generative models come into play. Productive approaches use semantic properties to produce picture features for unseen classes. Typically, a conditional generative adversarial network is used, creating image features based on the semantic characteristic of a specific type.

5) ONLINE LEARNING

Online learning involves instruction using data that is made available in a stepwise order. The whole training data samples are always available in batch sampling-based learning, which is different from this method. It is useful when algorithms need to change their behavior in response to changing data patterns from all incoming input. For online learning to succeed, three essential needs must be met [90]. Figure 23 shows On the Left: The user is pointing to an object, and on the Right: Process Flow.

1. The neural system's flexibility allows for the rapid assimilation of new information without requiring an entire training cycle;
2. Near-real-time processing throughout the entire system;
3. The natural presentation of fresh object information is made possible by a subsystem for human-machine interaction.

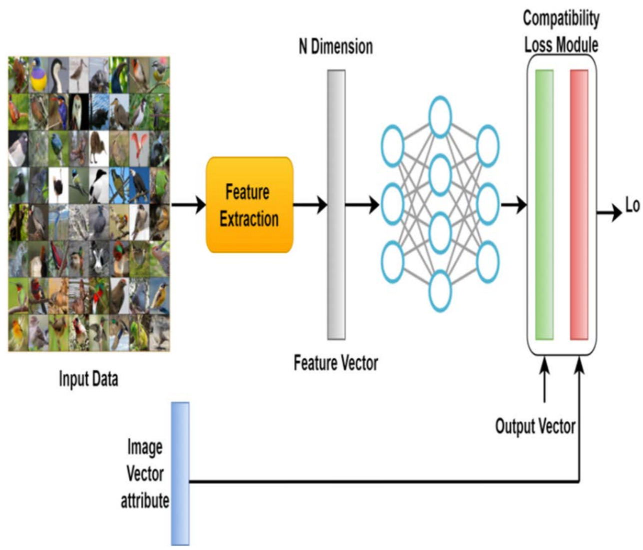

FIGURE 21. Embedding-based zero-shot method.

Advantages of Online learning-

- Low cost required
- Flexible to implement
- Covers time and mass audiences

Disadvantage of Online learning-

- Less Accuracy achieved or success rate is low.

Advantages of Zero-shot learning-

- A strong and promising learning paradigm is called zero-shot learning, in which the classes that training instances cover and the classes that we want to classify are not related.
- To improve the generalization ability of the model
- To improve scalability and robustness

Disadvantages of Zero-shot learning-

- Extensive smoothing
- Sparsely labeled

Using skin color segmentation, one instance of the VPL-classifier, and a search for pointing movements (upper branch), the scene is examined for objects that can be

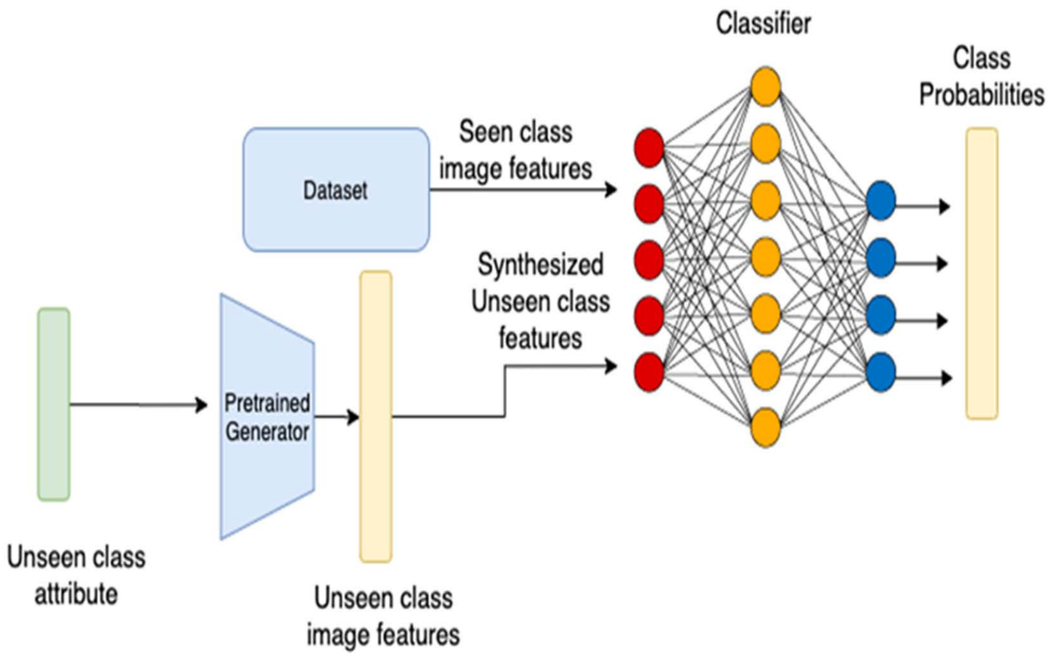

FIGURE 22. Zero-shot learning using generative model-based methods.

identified (lower branch). The "online loop" (right ellipse) starts when a story about an object is recounted. After establishing the position of the pertinent entity, pictures are taken. The database is expanded with views that have been artificially warped (scale/shear transformation, translatory offset). The candidate regions are categorized by the VPL classifier. The result is either a class number for a previously taught material or a reserved class label for "unknown." The VPL is a neural classification architecture that performs exceptionally well when being trained and retrained online with minimal data. Three processing stages called "VPL" combine pixel-level feature extraction with LLM-networks, PCA, and vector quantization are used for classification [88]. The following is how the VPL is defined: Using vector quantization, the input space is divided at the first level ("V") (VQ). For VQ, the suggested algorithm is applied. In the second level ("P") of each of the generated reference vectors, the neural Method computes the principal components (PCs) for the training data gathered in the Voronoi tessellation cells.

6) FEDERATED LEARNING

Federated Learning means fed a large number of cases. FL is a machine learning approach in which numerous participating clients who maintain their training data locally train a single shared global model.

It is a distributed learning approach that builds a universal or customized model using decentralized datasets on edge devices. Model performance in FL, however, falls well short of centralized training in the field of computer vision because there isn't any investigation in a variety of tasks with a common FL framework. FL works well in complex computer vision applications such as object recognition, picture segmentation, and image classification [90].

The FL figure 24 shows the Federated learning workflow stepwise-

1) On the Common Server, train a global model.

2) Using local datasets, deploy global models to edge devices (local models).

3) Use the local datasets from each edge device to improve the model (local models).

4) Post updates to locally trained models on a shared server.

5) Calculate the average of the update values and apply it to the overall model.

6) Repetition of steps 2 through 5

FL has been used, for instance, to train prediction models for mobile keyboards without sending confidential typing information to servers [91].

Federated Learning enables mobile devices to collectively create a shared prediction model while maintaining all of the training data on the device, separating the ability to do machine learning from the obligation to store training data on the cloud. Beyond the use of local models that make predictions on mobile devices, model training is extended to the device. According to how it works, your smart phone downloads the most recent model updates it by using data from your phone to learn from it, and then compiles the changes into a brief, targeted update. Using encrypted communication, only this particular change to the model is transferred to the cloud, where it is instantaneously averaged with updates from other users to enhance the shared model. There is no cloud storage for individual updates, and the training data is kept locally on your device.

Example: Google Keyboard: Your phone keeps the information about the current context and whether you clicked a suggested query when Gboard displays one. The history is processed on-device by Federated Learning, which offers suggestions for enhancements for the upcoming version of Gboard's query recommendation model. Federated learning functions without the necessity for cloud-based user data storage.

FedCV is a benchmarking system and federated learning library that assesses FL on the three most common computer vision tasks, including object identification, image segmentation, and classification.

A distributed machine learning (ML) framework is federated learning (FL). FL allows numerous clients to work together to solve common distributed ML issues while maintaining their local privacy. This is done under the control of a central server. FL differs from distributed ML in that the data that each participant uploads to the server is a trained sub-model rather than the original data. The FL also permits asynchronous transmission at the same time, allowing for a suitable reduction in the communication needs [94]. Each unit builds a model and transmits its input data to the server for aggregation. Data is stored on devices, and knowledge sharing with peers uses an aggregated paradigm [10]. For edge network optimization, the federated learning technique (FL) facilitates the cooperative training of deep learning and machine learning models. There is a challenge in this area even though a complex edge network with diverse devices with different restrictions can affect its performance.

Advantages of Federated learning-

- Federated learning enables several components to develop an identical, reliable machine learning model without sharing data, enabling for the resolution of crucial concerns such data privacy, security, access rights, and heterogeneous data availability.

Challenges of implementing Federated Learning-

- Resource Allocation: To manage large amounts of dispersed data without compromising privacy or health informatics, we intend to provide helpful tools for computational research on machine learning approaches.
- Data Imbalance:

1. Size imbalance: When the size of the data sample at each edge node varies widely.
2. Global imbalance: Data that is class uneven across all nodes is referred to as global imbalance.
3. Local Imbalance: Because not all nodes have the same data distribution, this is also referred to as a local imbalance, non-identical distribution, or independent distribution.

- Statistical Heterogeneity: The edges frequently gather and distribute data among the network in a non-i.i.d. fashion. Cellular phone users have access to a large range of languages for word prediction. Additionally, there may be an underlying structure that reflects the interaction between devices and the distributions connected with them, and the amount of data on various edges may vary.
- Privacy Concerns:

FL takes a step toward protecting user data by releasing only model changes (such as gradient information) rather than the complete data. Transmitting local model updates throughout the training process, however, can reveal private information to the main server or a third party. Despite current efforts to strengthen federated learning's privacy through the use of technologies like differential privacy and safe multiparty computation, these methods sometimes sacrifice system efficiency or model accuracy to ensure privacy.

V. FUTURE DIRECTIONS

This section highlights about the research gaps identified from the survey and the machine learning styles suitable to solve such challenges that could aid in this field's advancement. Table 21 gives the summary of research gaps with future directions.

The development of computer vision technology is continuing as AI becomes more pervasive in our daily lives. Due to developments in cloud computing, Auto ML pipelines, transformers, mobile-focused DL libraries, and mobile computer vision applications, as this technology scales, there will be an increased need for experts in computer vision systems.

A. IMBALANCED DATA

If a different number of images for each of the classes is existing in the input dataset. This problem is called as class imbalance. Similarly, if a set of images is not evenly distributed in the input dataset is called imbalanced data. Transfer learning, Multi-task learning and Federated learning help to overcome this unbalanced distribution of data problem. As in case of transfer learning once the model has been trained on sample dataset can be applied to solve the similar problems with the same model. In case of the Multi-task learning model can be trained with a small number of dataset. The same knowledge generated can be applied to solve all related tasks. From supervised learning Logistic regression algorithm is very useful to tackle this issue as it resample's the original training dataset to decrease the overall level of class imbalance. The authors proposed a monitoring scheme that can infer the composition of training data for each Federated Learning(FL) round, and design a new loss function - Ratio Loss to mitigate the impact of the imbalance [110].

B. SCARCITY OF DATA

Data scarcity occurs when: There is little or no labeled training data available, or there is insufficient data for certain labels in comparison to other labels are present in the dataset. Zero-shot learning, Few-shot learning and Transfer learning can be the solution for this type of dataset-related challenge. As Zero-shot or few-shot learning works properly with less or no labeled data. Transfer learning, where information from one dataset is used to inform a model on another, can be an effective solution on this challenge.

TABLE 18. Classification of zero-shot learning.

Ref.	Application Area	Datasets/ Model	Summary/ Technique used	Accuracy
[91]	Object Recognition	Animals with Attributes 2 (AWA2)	• They evaluate reports of per-class accuracy and per-image accuracy, which are often greater when the dataset is class-imbalanced.	82.00%
[92]	For object recognition,	Zero-Shot 117M All Data	• Learning from unlimited sources of unlabelled data. • Robust • Deployable • Self-supervised Approach	AGNews40.20% DBPedia - 39.60% Yahoo answers- 26.10%
		Zero-Shot 117M All Data		AGNews - 68.30% DBPedia - 52.50% Yahoo Answers 49.50%
		Zero-Shot 355M All Data		AGNews - 65.50% DBPedia - 44.80% Yahoo Answers - 52.20%
[47]	For object recognition,	CUB-200-2011	• On the CUB and mini-ImageNet test datasets, the Baseline model performs competitively with current state-of-the-art meta-learning methods when leveraging a deeper feature backbone.	71%
	For fine-grained classification	mini-ImageNet →CUB		65.57%
[93]	Low-Light Image/Video Enhancement	Dark City Scape (DCS)	• Without paired images, unpaired datasets, or segmentation annotation, the semantic-guided zero-shot low-light enhancement network (SGZ) is trained.	80%

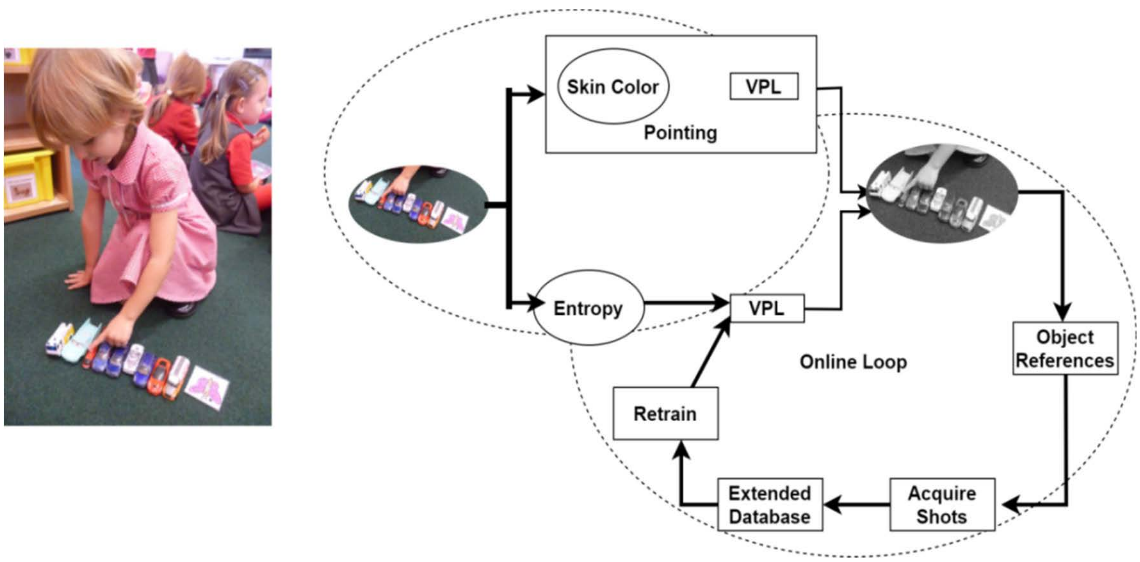

FIGURE 23. On the Left: The user is pointing to an object, and on the Right: Process Flow.

C. OVERFITTING/UNDERFITTING OF DATA

When our machine learning model is unable to recognize the data's underlying trend, underfitting occurs. Whereas overfitting is a problem when the evaluation of machine learning algorithms on training data differs from the evaluation on unknown data. Ensemble learning, Meta-Learning and Active learning are helpful to solve this problem.

D. DETECTION AND CLASSIFICATION OF BLUR IMAGES

In order to restore images, blur identification is frequently required. Authors proposed a classification technique utilizing ensemble Support Vector Machine (SVM) structure, a novel blur type classification method for digital images. Each image is considered to be prone to no more than one of the three types of blur: haze, motion, and defocus. In the

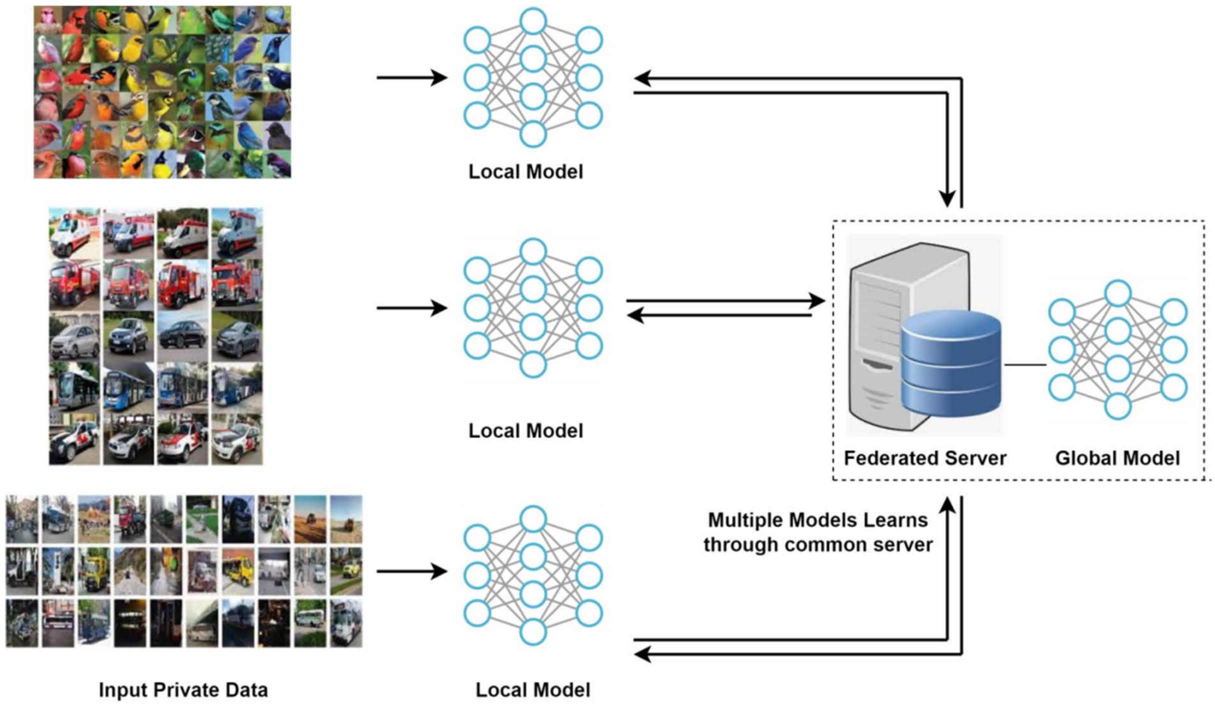

FIGURE 24. Federated learning system.

TABLE 19. Applications of federated learning.

References	Application of FL	Dataset	Techniques used	Accuracy
[95]	Image Classification	CIFAR-100	Lightweight CNNs techniques: • Efficient Net • MobileNet-V3	60.58%
[95]	Image Classification	LD-23k	• CNN	88.26%
[96]	Real-World Image Datasets for Federated Learning	Street-20 Dataset	• Faster RCNN	80%
[95]	Object Detection	YOLOV5	• FedAvg algorithm	85%
[97]	COVID-19 Chest X-ray Images	COVIDx	• MobileNet, • ResNet18, • MoblieNet and • COVID-Net	90% to 98%

suggested method, the Radial Basis Function (RBF) kernel parameters of the SVMs are additionally optimized using the SVM-Recursive Feature Elimination (SVM-RFE) method, which is used to rank the 35 blur features that were first derived from the spatial and transform domains of the picture. Additionally, the Support Vector Rate (SVR) is used to calculate the ideal number of features for classifiers to use. To categorize the various types of blurred images, the bagging random sampling method is used to build an ensemble SVM classifier based on a weighted voting mechanism [111]. In this way supervised learning SVM ensemble and SVM multiclass algorithms useful to solve blur images problem.

TABLE 20. Comparative analysis of different learning styles.

Learning type & Ref.	Definition/ Working	Application Domain	CV Task Application	Challenges	Accuracy Achieved
Supervised Learning [98][20][21][22][23][24][25]	This learning process combines input data with target variables or outputs.	• Healthcare • Industry 4.0 • Spatial data analysis • Weather prediction	• Classification • Regression • Pattern Recognition	• Poor quality of data • Training Data Underfitting • Overfitting Training Data Fit • The Process of Machine Learning is Complex • Insufficient training data.	79 to 99%
Unsupervised Learning [58] [40] [89]	It describes a class of concerns wherein relationships in data must be described or extracted using a model.	• Education • Healthcare • Industries • Market basket Analysis	• Clustering • Visualization • Projection	• Due to a large amount of training data, there is significant computational complexity. • Longer periods of training • There is a greater chance of getting erroneous results. • Validation of output variables requires human intervention. • The premise on which data was grouped was not transparent.	83 to 95%
Reinforcement Learning [99][65] [66] [67] [68] [69]	It describes a class of challenges where an agent must function in a given context while learning to do so through feedback.	• Machine Translation • Film Industry • Animated Games • Self-Driving Cars	• Clustering • Pattern Recognition • Projection	• Reproducibility is required. • Sample inefficiency is a problem. • Wisely choose reward structures.	20 to 67%
Semi-Supervised Learning [100][70][71][72][73][110]	There are many unlabeled samples in the training data instead of many labeled ones, making it supervised learning.	• Image Data • Automatic Speech Recognition (Audio Data)	• Image Classification • Segmentation	• The issue of extending labeled data • The difficulty of constructing the final classifier	32 to 93%
Self-Supervised Learning [101][53] [110]	It can leverage the learned representations for numerous downstream tasks and self-defined pseudo labels as supervision.	• Robotics • Medicine • Disease prediction	• Learning Cross-Lingual Representations • Capturing Sentence-Level Semantics	• It is very difficult to duplicate samples. • Semantic distributions of collected data	83 to 95%
Multi-Instance Learning [30][29][28][30]	Individual instances are not labeled in this supervised learning problem; bags or groups of samples are.	• Agriculture • Healthcare • Gaming	• Image Segmentation • Video Object Segmentation • Classification	• The intra-bag similarity between instances: Because the patches overlap, they are similar. • Instance Co-occurrence	78 to 93%
Inductive Learning Statistical Inference [102]	The inductive learning system employs inferred rules to classify new instances after learning—classification from training examples. A default rule is frequently applied if a choice cannot be implied from the system rule base.	• Healthcare • Industries	• Medical diagnostic • Industrial visual inspection	• Failure Detection: Not all defects can be fixed or even recovered automatically. Transient faults (network outage, memory overflow, disc space outage, and so on) • Intractability: how much time should we allow the algorithm to spend.	60 to 90 %
Transductive Learning [103]	Transductive learning is predicated on two fundamental tenets: (1) Nearby samples frequently share the same label, and (2) samples that are located on the same manifold should also share the same label.	• Video Object Segmentation • Surveillance • Self-Driving Cars • Robotics • Video Editing	• Image segmentation • Object Categorization • Image Classification	• Identification of local and global dependencies	60 to 90 %
Multi-Task Learning [47] [48] [49] [46] [50]	When there is a large amount of input data that has been labeled for one task and can be shared with another task that has considerably less labeled data, multi-task learning can be a useful way of problem-solving.	• Medical/Healthcare • Multimedia Data Processing • Multi-Modality Data Analysis	• Speech Recognition • Multimedia Data Processing • Biomedical Imaging • Socio-Biological Data Analysis • Multi-Modality Data Analysis	• Requires Model maintenance	50 to 85%
Active Learning [104][34][35][36][37][38]	To resolve uncertainty throughout the learning process, the model can ask a human user operator using this technique.	• Computational Biology. • Object Categorization • Image Classification • Image Segmentation • Scene Classification	• Computational Biology Applications • Object Categorization • Image Classification • Image Segmentation • Scene Classification	• Abundance of data	75 to 96 %
Online Learning [105] [90] [106]	It entails analyzing the data at hand and changing the model immediately before a prediction is needed or following the most recent observation.	• Object Recognition • Moving Object Detection	• Object Recognition • Moving Object Detection	• Accuracy is less.	50 to 70%

TABLE 20. *(Continued.)* **Comparative analysis of different learning styles.**

Transfer Learning [114][79][76][77][78]	It is a method of learning in which a model is initially trained on one task and then used partially or entirely as the foundation for another activity that is linked to it.	• Agriculture • Healthcare • Geoinformatics • Gaming	• Image Classification • Image Segmentation • Video Analysis •	• Dataset Scarcity problem	70 to 88%
Ensemble Learning [108][109][80][81][82]	This method involves fitting two or more modes to the same set of data and combining the results from each model.	• Cyber Security • Road damage detection • Healthcare • Handwriting pattern recognition	• Road damage detection stereo vision • Neural Networks for Pattern Recognition	• Diversity of individual model Accuracy	62 to 96%
Federated Learning **[90][92][93]**	It is a distributed learning paradigm that uses decentralized datasets on edge devices to construct a global or personalized model.	• Healthcare • Image classification	• Chest X-ray Images • Image Classification • Object detection	• Expensive communication • It is comprised of a massive number of devices.	60 to 98 %
Zero-Shot Learning [85][84][86]	Zero-shot learning refers to the problem where we want to recognize objects from classes that our model has not seen during training.	• Video Enhancement • Image Classification • Healthcare	• Image Classification • Object detection	• It involves little human intervention	26 to 80%
Meta-Learning [41][42][43][44][45]	Multi-task learning algorithms that are capable of learning across a group of connected prediction tasks are also referred to as meta-learning.	• Neural Networks • Clustering	• Image Classification • Object detection	• Discrete notation of task is required	60 to 93%

TABLE 21. **Summary of research gaps with future directions.**

Sr. No.	Challenges/Research Gaps	Suggested Learning Styles
1.	Imbalanced Data	• Transfer Learning • Multi-task learning • Federated Learning • Supervised Learning
2.	Scarcity of Data	• Transfer Learning • Zero-shot Learning • Few-Shot Learning
3.	Overfitting/Underfitting of data	• Ensemble Learning • Meta-Learning • Active Learning
4.	Detection and Classification of blur images	• Supervised Learning (Multi-class SVM)
5.	Human intervention required	• Constructive Learning • Meta-Learning
6.	Robustness of ML techniques	• Ensemble Learning • Meta-Learning • Association Rule Learning
7.	High operational cost(Model training)	• Reinforcement Learning • Meta-Learning • Transfer Learning
8.	More power consumption	• Federated Learning • Self-Supervised Learning
9.	Storage consumption	• Federated Learning
10.	Large scale of unlabeled images	• Few-shot Learning • Zero-shot Learning • Self-supervised Learning

E. HUMAN INTERVENTION REQUIRED

Meta learning and Constructive learning is the solution where human intervention is not required in the model training and testing process.

F. ROBUSTNESS OF ML TECHNIQUES

By integrating various models, ensemble learning enhances machine learning outcomes. In comparison to using a single model, this strategy enables the generation of greater prediction performance. Ensemble models are more robust than a single train model as it combines multiple models.

G. HIGH OPERATIONAL/MODEL TRAINING COST

Reinforcement learning, Transfer learning, Meta learning saves the model training operational cost as existing trained model is used to solve similar problems without the training model again.

H. MORE POWER AND STORAGE CONSUMPTION

Federated learning is a distributed learning paradigm that uses decentralized datasets on edge devices to construct a global or personalized model. So the storage consumption issue can be resolved using FL.

I. LARGE SCALE OF UNLABELED IMAGES

When input dataset consists of large scale of unlabeled images then accuracy degrades. To overcome this issue Self-supervised learning, Few-shot, and Zero-shot learning can improve CV operations performance. As these learning styles allow to train the model with few labeled samples easily.

The development of algorithms with lower training data requirements than present models is key to the future of computer vision technology. The industry has started investigating a few potentially ground-breaking research themes to address this difficulty by applying reinforcement learning, transfer learning and multi-task learning

In 2022, as augmented and virtual reality (VR) applications advance, computer vision developers will be able to expand their expertise into new fields, such as creating simple, effective ways to replicate and interact with physical things in a 3D environment. In the future, computer vision applications will likely continue to develop and have an impact.

VI. CONCLUSION

Machine learning and deep learning popularity have grown recently across several industries. The combination of machine learning and artificial intelligence methods has been used in many applications to carry out various computer vision tasks.

The literature on machine learning techniques applied in computer vision applications is reviewed in-depth in this article. The findings of a systematic literature review on machine learning styles are presented in this review. The authors intended to draw attention to the utilized learning types, adopted feature extraction techniques, methodologies, approaches, approved data sets, adopted application domains, and difficulties related to ML approaches in diverse sectors. This study planned, executed, and carried out different SLR phases on ML styles. In the literature review for computer vision applications, other artificial intelligence methods—such as those based on deep learning and machine learning—have been used. Deep learning and machine learning-based techniques are popular thanks to easily accessible datasets and automated feature extraction methods. The authors investigated publicly accessible computer vision datasets. Future possibilities for ML techniques in CV based on artificial intelligence are described, along with research obstacles in the realm of ML styles in computer vision, such as domain dependency and imbalanced dataset. Future directions are mentioned in the article which will be helpful for the researchers who are working in this domain.

REFERENCES

[1] S. Sah, "Machine learning: A review of learning types," Jul. 2020, doi: 10.20944/preprints202007.0230.v1.

[2] L. Cheng and T. Yu, "A new generation of AI: A review and perspective on machine learning technologies applied to smart energy and electric power systems," *Int. J. Energy Res.*, vol. 43, no. 6, pp. 1928–1973, 2019, doi: 10.1002/er.4333.

[3] S. Kumar, T. Kolekar, S. Patil, A. Bongale, K. Kotecha, A. Zaguia, and C. Prakash, "A low-cost multi-sensor data acquisition system for fault detection in fused deposition modelling," *Sensors*, vol. 22, no. 2, pp. 1–33, 2022, doi: 10.3390/s22020517.

[4] S. Swain, B. Bhushan, G. Dhiman, and W. Viriyasitavat, "Appositeness of optimized and reliable machine learning for healthcare: A survey," *Arch. Comput. Methods Eng.*, vol. 29, no. 6, pp. 3981–4003, Oct. 2022, doi: 10.1007/s11831-022-09733-8.

[5] H. H. Martens, "Two notes on machine 'learning,'" *Inf. Control*, vol. 2, no. 4, pp. 364–379, 1959, doi: 10.1016/S0019-9958(59)80014-0.

[6] A. A. Khan, A. A. Laghari, and S. A. Awan, "Machine learning in computer vision: A review," *EAI Endorsed Trans. Scalable Inf. Syst.*, vol. 8, no. 32, pp. 1–11, 2021, doi: 10.4108/eai.21-4-2021.169418.

[7] D. M. Dwyer, P. Gasalla, and M. López, "Partial reinforcement and conditioned taste aversion: No evidence for resistance to extinction," *Quart. J. Exp. Psychol.*, vol. 72, no. 2, pp. 274–284, 2019, doi: 10.1080/17470218.2017.1347191.

[8] J. M. Vanderplas, "Transfer of training and its relation to perceptual learning and recognition," *Psychol. Rev.*, vol. 65, no. 6, pp. 375–385, 1958, doi: 10.1037/h0040233.

[9] K. S. Pu and C. H. Chen, "A sequential decision approach to problems in pattern recognition and learning," in *Proc. 3rd Symp. Adapt. Process. (SAP)*, 1964, pp. 48–71, doi: 10.1109/SAP.1964.271134.

[10] C. H. Chen, "A note on sequential decision approach to pattern recognition and machine learning," *Inf. Control*, vol. 9, no. 6, pp. 549–562, 1966, doi: 10.1016/S0019-9958(66)80015-3.

[11] C. W. McIntyre, R. D. Odom, and M. Byassee, "The influence of congenital deafness on processes concerned with reading: An initial investigation," *J. Reading Behav.*, vol. 3, no. 1, pp. 36–41, Jul. 1970, doi: 10.1080/10862967009546922.

[12] D. W. Johnson, R. T. Johnson, and K. A. Smith, "Cooperative learning: Improving university instruction by basing practice on validated theory," *J. Excellence. Coll. Teaching*, vol. 25, no. 4, pp. 85–118, 2014. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/10180297.

[13] S. K. Pal, A. K. Datta, and D. D. Majumder, "A self-supervised vowel recognition system," *Pattern Recognit.*, vol. 12, no. 1, pp. 27–34, Jan. 1980.

[14] L. Chum, A. Subramanian, V. N. Balasubramanian, and C. V. Jawahar, "Beyond supervised learning: A computer vision perspective," *J. Indian Inst. Sci.*, vol. 99, no. 2, pp. 177–199, Jun. 2019, doi: 10.1007/s41745-019-0099-3.

[15] C. Sager, C. Janiesch, and P. Zschech, "A survey of image labelling for computer vision applications," *J. Bus. Anal.*, vol. 4, no. 2, pp. 91–110, Jul. 2021, doi: 10.1080/2573234X.2021.1908861.

[16] J. Zhao, R. Masood, and S. Seneviratne, "A review of computer vision methods in network security," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 3, pp. 1838–1878, Jun. 2021, doi: 10.1109/COMST.2021.3086475.

[17] K. Bayoudh, R. Knani, F. Hamdaoui, and A. Mtibaa, "A survey on deep multimodal learning for computer vision: Advances, trends, applications, and datasets," *Vis. Comput.*, vol. 38, no. 8, pp. 2939–2970, Aug. 2022, doi: 10.1007/s00371-021-02166-7.

[18] M. Ferdous, J. Debnath, and N. R. Chakraborty, "Machine learning algorithms in healthcare: A literature survey," in *Proc. 11th Int. Conf. Comput., Commun. Netw. Technol. (ICCCNT)*, Jul. 2020, pp. 1–6, doi: 10.1109/ICCCNT49239.2020.9225642.

[19] M. Crawford, T. M. Khoshgoftaar, J. D. Prusa, A. N. Richter, and H. Al Najada, "Survey of review spam detection using machine learning techniques," *J. Big Data*, vol. 2, no. 1, pp. 1–24, Dec. 2015, doi: 10.1186/s40537-015-0029-9.

[20] A. Halbouni, T. S. Gunawan, M. H. Habaebi, M. Halbouni, M. Kartiwi, and R. Ahmad, "Machine learning and deep learning approaches for cybersecurity: A review," *IEEE Access*, vol. 10, Mar. 2021.
[21] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopapadakis, "Deep learning for computer vision: A brief review," *Comput. Intell. Neurosci.*, vol. 2018, pp. 1–13, Feb. 2018, doi: 10.1155/2018/7068349.
[22] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *Int. J. Comput. Vis.*, vol. 128, pp. 261–318, Feb. 2020, doi: 10.1007/s11263-019-01247-4.
[23] H. Wang, B. Zheng, S. W. Yoon, and H. S. Ko, "A support vector machine-based ensemble algorithm for breast cancer diagnosis," *Eur. J. Oper. Res.*, vol. 267, no. 2, pp. 687–699, Jun. 2018, doi: 10.1016/j.ejor.2017.12.001.
[24] R. Nayak, S. A. Jiwani, and B. Rajitha, "Spam email detection using machine learning algorithm," *Mater. Today, Proc.*, pp. 1–5, Apr. 2021, doi: 10.1016/j.matpr.2021.03.147.
[25] I. A. El Kader, G. Xu, Z. Shuai, S. Saminu, I. Javaid, and I. Salim Ahmad, "Differential deep convolutional neural network model for brain tumor classification," *Brain Sci.*, vol. 11, no. 3, p. 352, Mar. 2021, doi: 10.3390/brainsci11030352.
[26] J. Tavares and N. Jorge, *Lecture Notes in Computational Vision and Biomechanics*. Springer, 2012.
[27] A. M. Durso, G. K. Moorthy, S. P. Mohanty, I. Bolon, M. Salathé, and R. R. de Castañeda, "Supervised learning computer vision benchmark for snake species identification from photographs: Implications for herpetology and global health," *Frontiers Artif. Intell.*, vol. 4, pp. 1–15, Apr. 2021, doi: 10.3389/frai.2021.582110.
[28] H. S. Munawar, F. Ullah, A. Heravi, M. J. Thaheem, and A. Maqsoom, "Inspecting buildings using drones and computer vision: A machine learning approach to detect cracks and damages," *Drones*, vol. 6, no. 1, p. 5, Dec. 2021, doi: 10.3390/drones6010005.
[29] M. Soelch, A. Akhundov, P. van der Smagt, and J. Bayer, "On deep set learning and the choice of aggregations," in *Artificial Neural Networks and Machine Learning—ICANN 2019: Theoretical Neural Computation* (Lecture Notes in Computer Science), vol. 11727. 2019, pp. 444–457, doi: 10.1007/978-3-030-30487-4_35.
[30] M.-A. Carbonneau, V. Cheplygina, E. Granger, and G. Gagnon, "Multiple instance learning: A survey of problem characteristics and applications," *Pattern Recognit.*, vol. 77, pp. 329–353, May 2018, doi: 10.1016/j.patcog.2017.10.009.
[31] Y. Chen, J. Bi, and J. Z. Wang, "MILES: Multiple-instance learning via embedded instance selection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 1931–1947, Dec. 2006, doi: 10.1109/TPAMI.2006.248.
[32] C. Yang and T. Lozano-Perez, "Image database retrieval with multiple-instance learning techniques," in *Proc. 16th Int. Conf. Data Eng.*, Feb. 2000, pp. 233–243, doi: 10.1109/ICDE.2000.839416.
[33] V. Cheplygina, M. de Bruijne, and J. P. Pluim, "Not-so-supervised: A survey of semi-supervised, multi-instance, and transfer learning in medical image analysis," *Med. Image Anal.*, vol. 54, pp. 280–296, May 2019, doi: 10.1016/j.media.2019.03.009.
[34] S. V. Desai, A. C. Lagandula, W. Guo, S. Ninomiya, and V. N. Balasubramanian, "An adaptive supervision framework for active learning in object detection," in *Proc. 30th Brit. Mach. Vis. Conf.*, 2019, pp. 1–13.
[35] B. Sayin, E. Krivosheev, J. Yang, A. Passerini, and F. Casati, "A review and experimental analysis of active learning over crowdsourced data," *Artif. Intell. Rev.*, vol. 54, no. 7, pp. 5283–5305, Oct. 2021, doi: 10.1007/s10462-021-10021-3.
[36] W. H. Beluch, T. Genewein, A. Nurnberger, and J. M. Kohler, "The power of ensembles for active learning in image classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9368–9377, doi: 10.1109/CVPR.2018.00976.
[37] C. Kao, T. L. B, P. Sen, and M. Liu, *Localization-Aware Active Learning*, vol. 2. Springer, 2019.
[38] A. Kapoor, K. Grauman, R. Urtasun, and T. Darrell, "Active learning with Gaussian processes for object categorization," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Oct. 2007, pp. 1–8, doi: 10.1109/ICCV.2007.4408844.
[39] T. Senechal, D. McDuff, and R. E. Kaliouby, "Facial action unit detection using active learning and an efficient non-linear kernel approximation," in *Proc. IEEE Int. Conf. Comput. Vis. Workshop (ICCVW)*, Dec. 2015, pp. 10–18, doi: 10.1109/ICCVW.2015.11.
[40] K. Kim, D. Park, K. I. Kim, and S. Y. Chun, "Task-aware variational adversarial active learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 8162–8171, doi: 10.1109/CVPR46437.2021.00807.
[41] S. Qiao, C. Liu, W. Shen, and A. Yuille, "Few-shot image recognition by predicting parameters from activations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7229–7238, doi: 10.1109/CVPR.2018.00755.
[42] H. Venkateswara, J. Eusebio, S. Chakraborty, and S. Panchanathan, "Deep hashing network for unsupervised domain adaptation supplementary material," in *Proc. 30th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, USA, 2017, vol. 1, no. 10.
[43] F. G. Mohammadi, M. H. Amini, and H. R. Arabnia, "On parameter tuning in meta-learning for computer vision," 2020, *arXiv:2003.00837*.
[44] E. Kodirov, T. Xiang, S. Gong, and Q. Mary, "Semantic autoencoder for zero-shot learning," in *Proc. CVPR*, Jul. 2017, pp. 3174–3183. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2017/papers/Kodirov_Semantic_Autoencoder_for_CVPR_2017_paper.pdf
[45] S. Ravi and H. Larochelle, "Optimization as a model for few-shot learning," in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1–11.
[46] O. Vinyals, C. Blundell, T. Lillicrap, K. Kavukcuoglu, and D. Wierstra, "Matching networks for one shot learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3637–3645.
[47] W. Y. Chen, Y. C. F. Wang, Y. C. Liu, Z. Kira, and J. B. Huang, "A closer look at few-shot classification," in *Proc. 7th Int. Conf. Learn. Represent. (ICLR)*, 2019, pp. 1–17.
[48] A. Kendall, Y. Gal, and R. Cipolla, "Multi-task learning using uncertainty to weigh losses for scene geometry and semantics," in *Proc. CVPR*, Jun. 2018, pp. 7482–7491.
[49] X. Zhao, H. Li, X. Shen, X. Liang, and Y. Wu, "A modulation module for multi-task learning with applications in image retrieval," in *Computer Vision* (Lecture Notes in Computer Science), vol. 11205. 2018, pp. 415–432, doi: 10.1007/978-3-030-01246-5_25.
[50] I. Misra, "Cross-stitch networks for multi-task learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2016.
[51] J. Yang, Y. Liu, M. Qian, C. Guan, and X. Yuan, "Information extraction from electronic medical records using multitask recurrent neural network with contextual word embedding," *Appl. Sci.*, vol. 9, no. 18, p. 3658, Sep. 2019, doi: 10.3390/app9183658.
[52] N. Le, K. Yamazaki, K. G. Quach, D. Truong, and M. Savvides, "A multi-task contextual atrous residual network for brain tumor detection & segmentation," in *Proc. Int. Conf. Pattern Recognit.*, Jan. 2020, pp. 5943–5950, doi: 10.1109/ICPR48806.2021.9412414.
[53] J. Xu, L. Xiao, and A. M. López, "Self-supervised domain adaptation for computer vision tasks," *IEEE Access*, vol. 7, pp. 156694–156706, 2019, doi: 10.1109/ACCESS.2019.2949697.
[54] M. Y. Lu, R. J. Chen, J. Wang, D. Dillon, and F. Mahmood, "Semi-supervised histology classification using deep multiple instance learning and contrastive predictive coding," 2019, *arXiv:1910.10825*.
[55] A. Singh, S. Karanam, and D. Kumar, "Constructive learning for human-robot interaction," *IEEE Potentials*, vol. 32, no. 4, pp. 13–19, Jul. 2013, doi: 10.1109/MPOT.2012.2189443.
[56] L. Ma and K. Khorasani, "Facial expression recognition using constructive feedforward neural networks," *IEEE Trans. Syst., Man, Cybern., B, Cybern.*, vol. 34, no. 3, pp. 1588–1595, Jun. 2004, doi: 10.1109/TSMCB.2004.825930.
[57] J. Yang, J. Zhao, and Q. Sun, "3D model retrieval using constructive-learning for cross-model correlation," *Neurocomputing*, vol. 275, pp. 1–9, Jan. 2018, doi: 10.1016/j.neucom.2017.01.030.
[58] I. H. Sarker, "Machine learning: Algorithms, real-world applications and research directions," *Social Netw. Comput. Sci.*, vol. 2, no. 3, pp. 1–21, May 2021, doi: 10.1007/s42979-021-00592-x.
[59] R. Agrawal, T. Imielinski, and A. Swami, "Mining association in large databases," in *Proc. ACM SIGMOD Int. Conf. Manag. Data (SIGMOD)*, 1993, pp. 207–216.
[60] T. T. Dhivyaprabha, P. Subashini, and M. Krishnaveni, "Computational intelligence based machine learning methods for rule-based reasoning in computer vision applications," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, 2016, pp. 1–8, doi: 10.1109/SSCI.2016.7850050.
[61] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, no. 2, pp. 303–338, Jun. 2010, doi: 10.1007/s11263-009-0275-4.

[62] M. Houtsma and A. Swami, "Set-oriented mining for association rules in relational databases," in *Proc. 11th Int. Conf. Data Eng.*, 1995, pp. 25–33, doi: 10.1109/icde.1995.380413.

[63] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proc. 20th Int. Conf. Very Large Data Bases*, 1994, pp. 487–499.

[64] A. Shafiee and M. Karimi, "On the relationship between entropy and Information," *Phys. Essays*, vol. 20, no. 3, pp. 487–493, Sep. 2007, doi: 10.4006/1.3153419.

[65] J. C. Caicedo and S. Lazebnik, "Active object localization with deep reinforcement learning," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 2488–2496, doi: 10.1109/ICCV.2015.286.

[66] G. Maicas, G. Carneiro, A. P. Bradley, J. C. Nascimento, and I. Reid, "Deep reinforcement learning for active breast lesion detection from DCE-MRI," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, in Lecture Notes in Computer Science, Quebec City, QC, Canada, vol. 10435. Heidelberg, Germany: Springer-Verlag, Sep. 2017, pp. 665–673, doi: 10.1007/978-3-319-66179-7_76.

[67] F.-C. Ghesu, B. Georgescu, Y. Zheng, S. Grbic, A. Maier, J. Hornegger, and D. Comaniciu, "Multi-scale deep reinforcement learning for real-time 3D-landmark detection in CT scans," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 1, pp. 176–189, Jan. 2019, doi: 10.1109/TPAMI.2017.2782687.

[68] A. Vlontzos, A. Alansary, K. Kamnitsas, D. Rueckert, and B. Kainz, "Multiple landmark detection using multi-agent reinforcement learning," in *Proc. 22nd Int. Conf. Med. Image Comput. Comput. Assist. Intervent. (MICCAI)*, in Lecture Notes in Computer Science, Shenzhen, China, vol. 11767, Oct. 2019, pp. 262–270, doi: 10.1007/978-3-030-32251-9_29.

[69] B. Uzkent, C. Yeh, and S. Ermon, "Efficient object detection in large images using deep reinforcement learning," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Mar. 2020, pp. 1813–1822, doi: 10.1109/WACV45572.2020.9093447.

[70] N. Hoang-Xuan, E.-R. Nguyen, T.-D. Pham-Le, and K. Hoang-Nguyen, "Efficient one-shot video object segmentation," in *Proc. 7th NAFOSTED Conf. Inf. Comput. Sci. (NICS)*, Nov. 2020, pp. 320–325, doi: 10.1109/NICS51282.2020.9335847.

[71] F. Perazzi, A. Khoreva, R. Benenson, B. Schiele, and A. Sorkine-Hornung, "Learning video object segmentation from static images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3491–3500, doi: 10.1109/CVPR.2017.372.

[72] L. Jing, T. Parag, Z. Wu, Y. Tian, and H. Wang, "VideoSSL: Semi-supervised learning for video classification," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 1109–1118, doi: 10.1109/WACV48630.2021.00115.

[73] B. Jiang, Z. Zhang, D. Lin, J. Tang, and B. Luo, "Semi-supervised learning with graph learning-convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 11313–11320.

[74] L. C. Chen, R. G. Lopes, B. Cheng, M. D. Collins, E. D. Cubuk, B. Zoph, H. Adam, and J. Shlens, "Naive-student: Leveraging semi-supervised learning in video sequences for urban scene segmentation," *Computer Vision* (Lecture Notes in Computer Science), vol. 12354. 2020, pp. 695–714, doi: 10.1007/978-3-030-58545-7_40.

[75] Z. Zhang and W. S. Lee, "Deep graphical feature learning for the feature matching problem," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 5086–5095, doi: 10.1109/ICCV.2019.00519.

[76] X. Shen, A. A. Efros, and M. Aubry, "Discovering visual patterns in art collections with spatially-consistent feature learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9270–9279, doi: 10.1109/CVPR.2019.00950.

[77] D. Mukherjee, K. Gupta, L. H. Chang, and H. Najjaran, "A survey of robot learning strategies for human-robot collaboration in industrial settings," *Robot. Computer-Integrated Manuf.*, vol. 73, Feb. 2022, Art. no. 102231, doi: 10.1016/j.rcim.2021.102231.

[78] A. Sophokleous, P. Christodoulou, L. Doitsidis, and S. A. Chatzichristofis, "Computer vision meets educational robotics," *Electron.*, vol. 10, no. 6, pp. 1–24, 2021, doi: 10.3390/electronics10060730.

[79] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proc. IEEE*, vol. 109, no. 1, pp. 43–76, Jul. 2021, doi: 10.1109/JPROC.2020.3004555.

[80] K. Gopalakrishnan, S. K. Khaitan, A. Choudhary, and A. Agrawal, "Deep convolutional neural networks with transfer learning for computer vision-based data-driven pavement distress detection," *Construction Building Mater.*, vol. 157, pp. 322–330, Dec. 2017, doi: 10.1016/j.conbuildmat.2017.09.110.

[81] P. N. Whatmough, C. Zhou, P. Hansen, S. K. Venkataramanaiah, J.-S. Seo, and M. Mattina, "FixyNN: Efficient hardware for mobile computer vision via transfer learning," 2019, *arXiv:1902.11128*.

[82] J. Xie, "Transfer learning with deep neural networks for computer vision," Tech. Rep., 2019.

[83] A. Brodzicki, M. Piekarski, D. Kucharski, J. Jaworek-Korjakowska, and M. Gorgon, "Transfer learning methods as a new approach in computer vision tasks with small datasets," *Found. Comput. Decis. Sci.*, vol. 45, no. 3, pp. 179–193, Sep. 2020, doi: 10.2478/fcds-2020-0010.

[84] J. J. Bird, A. Ekárt, and D. R. Faria, "British sign language recognition via late fusion of computer vision and leap motion with transfer learning to American sign language," *Sensors*, vol. 20, no. 18, pp. 1–19, 2020, doi: 10.3390/s20185151.

[85] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, "Ensemble learning," *Data Mining*, vol. 4, pp. 479–501, Oct. 2017, doi: 10.1016/b978-0-12-804291-5.00012-x.

[86] K. Doshi and Y. Yilmaz, "Road damage detection using deep ensemble learning," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2020, pp. 5540–5544, doi: 10.1109/BigData50022.2020.9377774.

[87] Y. Yu, H. Liu, M. Fu, J. Chen, X. Wang, and K. Wang, "A two-branch neural network for non-homogeneous dehazing via ensemble learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2021, pp. 193–202, doi: 10.1109/CVPRW53098.2021.00028.

[88] V. Riego, M. Castejón-Limas, L. Sánchez-González, L. Fernández-Robles, H. Perez, J. Diez-Gonzalez, and Á.-M. Guerrero-Higueras, "Strong classification system for wear identification on milling processes using computer vision and ensemble learning," *Neurocomputing*, vol. 456, pp. 678–684, Oct. 2021, doi: 10.1016/j.neucom.2020.07.131.

[89] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1122–1131, doi: 10.1109/CVPRW.2017.150.

[90] H. Bekel, I. Bax, G. Heidemann, and H. Ritter, "Adaptive computer vision: Online learning for object recognition," in *Proc. Joint Pattern Recognit. Symp.*, 2004, pp. 447–448.

[91] Y. Xian, C. H. Lampert, B. Schiele, and Z. Akata, "Zero-shot learning—A comprehensive evaluation of the good, the bad and the ugly," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 9, pp. 2251–2265, Sep. 2019, doi: 10.1109/TPAMI.2018.2857768.

[92] R. Puri and B. Catanzaro, "Zero-shot text classification with generative language models," 2019, *arXiv:1912.10165*.

[93] S. Zheng and G. Gupta, "Semantic-guided zero-shot learning for low-light image/video enhancement," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, Jan. 2022, pp. 581–590.

[94] K. Hu, Y. Li, M. Xia, J. Wu, M. Lu, S. Zhang, and L. Weng, "Federated learning: A distributed shared machine learning method," *Complexity*, vol. 2021, pp. 1–20, Aug. 2021, doi: 10.1155/2021/8261663.

[95] C. He, A. D. Shah, Z. Tang, D. F. A. N. Sivashunmugam, K. Bhogaraju, M. Shimpi, L. Shen, X. Chu, M. Soltanolkotabi, and S. Avestimehr, "FedCV: A federated learning framework for diverse computer vision tasks," 2021, *arXiv:2111.11066*.

[96] J. Luo, X. Wu, Y. Luo, A. Huang, Y. Huang, Y. Liu, and Q. Yang, "Real-world image datasets for federated learning," 2019, *arXiv:1910.11089*.

[97] B. Yan, J. Wang, J. Cheng, Y. Zhou, Y. Zhang, Y. Yang, L. Liu, H. Zhao, C. Wang, and B. Liu, "Experiments of federated learning for COVID-19 chest X-ray images," *Commun. Comput. Inf. Sci.*, vol. 1423, pp. 41–53, Jul. 2021, doi: 10.1007/978-3-030-78618-2_4.

[98] S. Chowdhury and M. P. Schoen, "Research paper classification using supervised machine learning techniques," in *Proc. Intermountain. Eng. Technol. Comput. (IETC)*, Jul. 2020, pp. 1–6, doi: 10.1109/IETC47856.2020.9249211.

[99] M. Soelch, A. Akhundov, P. van der Smagt, and J. Bayer, "Extra—3," *IEEE Access*, vol. 7, pp. 117227–117245, 2019.

[100] A. C. Le, A. Shimazu, and L. M. Nguyen, ‘‘Investigating problems of semi-supervised learning for word sense disambiguation,’’ in *Computer Processing of Oriental Languages. Beyond the Orient: The Research Challenges Ahead* (Lecture Notes in Computer Science), vol. 4285. Oct. 2014, pp. 482–489, doi: 10.1007/11940098_51.
[101] S. Purushwalkam, P. Morgado, and A. Gupta, ‘‘The challenges of continuous self-supervised learning,’’ 2022, *arXiv:2203.12710*.
[102] I. Birzniece, ‘‘Interactive inductive learning system,’’ *Front. Artif. Intell. Appl.*, vol. 224, pp. 380–393, Jan. 2011, doi: 10.3233/978-1-60750-688-1-380.
[103] Y. Zhang, Z. Wu, H. Peng, and S. Lin, ‘‘A transductive approach for video object segmentation,’’ in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6947–6956, doi: 10.1109/CVPR42600.2020.00698.
[104] I. Guyon, ‘‘Results of the active learning challenge,’’ in *Proc. Active Learn. Exp. Design Workshop Conjunct. AISTATS*, vol. 16, 2011, pp. 19–45, [Online]. Available: http://eprints.pascal-network.org/archive/00007803/
[105] J. Chai, H. Zeng, A. Li, and E. W. T. Ngai, ‘‘Deep learning in computer vision: A critical review of emerging techniques and application scenarios,’’ *Mach. Learn. Appl.*, vol. 6, Dec. 2021, Art. no. 100134, doi: 10.1016/j.mlwa.2021.100134.
[106] G. Heidemann and H. Ritter, ‘‘Efficient vector quantization using the WTA-rule with activity equalization,’’ *Neural Process. Lett.*, vol. 13, no. 1, pp. 17–30, 2005.
[107] X. Li, Y. Grandvalet, F. Davoine, J. Cheng, Y. Cui, H. Zhang, S. Belongie, Y.-H. Tsai, and M.-H. Yang, ‘‘Transfer learning in computer vision tasks: Remember where you come from,’’ *Image Vis. Comput.*, vol. 93, Jan. 2020, Art. no. 103853, doi: 10.1016/j.imavis.2019.103853.
[108] W. Wang, ‘‘Some fundamental issues in ensemble methods,’’ in *Proc. IEEE Int. Joint Conf. Neural Netw. (IEEE World Congr. Comput. Intelligence)*, Jun. 2008, pp. 2243–2250, doi: 10.1109/IJCNN.2008.4634108.
[109] R. Polikar, ‘‘Ensemble learning,’’ in *Ensemble Machine Learning*, C. Zhang and Y. Ma, Eds. Boston, MA, USA: Springer, Jan. 2012, doi: 10.1007/978-1-4419-9326-7_1.
[110] L. Wang, S. Xu, X. Wang, and Q. Zhu, ‘‘Addressing class imbalance in federated learning,’’ in *Proc. 35th AAAI Conf. Artif. Intell.*, vol. 11, 2021, pp. 10165–10173.
[111] R. Wang, W. Li, R. Li, and L. Zhang, ‘‘Automatic blur type classification via ensemble SVM,’’ *Signal Process., Image Commun.*, vol. 71, pp. 24–35, Feb. 2019, doi: 10.1016/j.image.2018.08.003.
[112] L. Liu, Q. Jin, D. Wang, H. Yu, G. Sun, and S. Luo, ‘‘PSNet: Reconfigurable network topology design for accelerating parameter server architecture based distributed machine learning,’’ *Future Gener. Comput. Syst.*, vol. 106, pp. 320–332, May 2020.
[113] X. Zhai, A. Oliver, A. Kolesnikov, and L. Beyer, ‘‘S4L: Self-supervised semi-supervised learning,’’ in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1476–1485.
[114] L. Pinto, D. Gandhi, Y. Han, Y. L. Park, and A. Gupta, ‘‘The curious robot: Learningvisual representations via physical interactions,’’ in *Computer Vision—ECCV 2016* (Lecture Notes in Computer Science), vol. 9906. Cham, Switzerland: Springer, 2016, pp. 3–18, doi: 10.1007/978-3-319-46475-6_1.
[115] U. Nehmzow, ‘‘Vision processing for robot learning,’’ *Ind. Robot, Int. J.*, vol. 26, no. 2, pp. 121–130, Mar. 1999.

SUPRIYA V. MAHADEVKAR received the M.Tech. degree (Hons.) in computer engineering from the Pimpri Chinchwad College of Engineering (PCCOE), SPPU University, Pune. She is currently pursuing the Ph.D. degree with the Symbiosis Institute of Technology, Pune.

She is also working as a Junior Research Fellow at the Symbiosis Institute of Technology. She has more than eight years of teaching and research experience in engineering colleges. She has expertise in applying innovative technology solutions to real-world problems. Her research interests include machine learning, deep learning, artificial intelligence, natural language processing, and data mining and analysis. She is also working in the application domains of computer vision in healthcare, data analysis, and machine simulation, via which she has also guided several UG and PG students as a domain expert. She has published more than 12 research articles in reputed international conferences and Scopus/web of science indexed journals and books.

BHARTI KHEMANI received the M.Tech. degree in computer engineering from the Thadomal Shahani Engineering College, Bandra, Mumbai. She is currently pursuing the Ph.D. degree with the Symbiosis Institute of Technology.

She is also working as an Assistant Professor with the A. P. Shah Institute of Technology, Mumbai. She has 12 years of academic experience in engineering colleges. She has expertise in applying innovative technology solutions to real-world problems. Her research interests include machine learning, deep learning, artificial intelligence, natural language processing, and graph neural networks (GNN). She is also working in the application domains of healthcare, misinformation detection, and machine simulation, via which she is also guiding several UG students as a domain expert. She has published more than ten research articles in national and international conferences and journals.

SHRUTI PATIL received the M.Tech. degree in computer science and the Ph.D. degree in the domain of data privacy from Pune University. She has been an industry professional in the past, currently associated with the Symbiosis Institute of Technology as a Professor and with SCAAI as a Research Associate, Pune, Maharashtra. She has three years of industry experience and ten years of academic experience. She has expertise in applying innovative technology solutions to real world problems. Her research interests include applied artificial intelligence, natural language processing, acoustic AI, adversarial machine learning, data privacy, digital twin applications, GANS, and multi-modal data analysis. She is also working in the application domains of healthcare, sentiment analysis, emotion detection, and machine simulation via which she is also guiding several UG, PG, and Ph.D. students as a domain expert. She has published more than 30 research articles in reputed international conferences and Scopus/web of science indexed journals and books.

KETAN KOTECHA has expertise and experience of cutting-edge research and projects in AI and deep learning for last more than 25 years. He has published widely in several excellent peer-reviewed journals on various topics ranging from education policies, teaching-learning practices, and AI for all. He is also a Team Member for the nationwide initiative on AI and deep learning Skilling and Research named Leadingindia.ai initiative sponsored by the Royal Academy of Engineering, the U.K. under Newton Bhabha Fund. He is currently the Head of the Symbiosis Centre for Applied Artificial Intelligence (SCAAI). He is considered a foremost expert in AI and aligned technologies. Additionally, with his vast and varied experience in administrative roles, he has pioneered Education Technology. He has worked as an Administrator at Parul University and Nirma University and has several achievements in these roles to his credit.

DEEPALI R. VORA received the Ph.D. degree in computer science and engineering from Amity University, Mumbai. She has worked as the Head of the Information Technology Department, Vidyalankar Institute of Technology, Mumbai. Currently, she is working as the Head of the Department and an Associate Professor in computer science and engineering with the Symbiosis Institute of Technology, Pune. She has more than 20 years of experience in total in teaching, research, and industry. She has published more than 50 research papers in reputed national, international conferences, and journals. She has coauthored three books and two book chapters and delivered various talks in data science and machine learning. She has conducted hands-on session in data science using python for students and faculties. She was appointed as a Syllabus Revision Committee Member of Mumbai University and developed the course content for B.E. degree in information technology course. She has received grants from government bodies, such as AICTE, ISTE, and industry for conducting research, organizing conference and training courses for faculties. She is acting as a Reviewer for many international conferences and journals, such as IEEE ACCESS, *Journal of Big Data* analytics (IGI Global), *Journal of Intelligent Systems*, and Inderscience. She has organized many values added courses for the benefit of the students. More than 18 students have completed and currently, three students are pursuing their postgraduate studies under her guidance from Mumbai University. In addition to that, three students are pursuing research (Ph.D.) under her guidance at Symbiosis International University, Pune. Her course developed on "Deep Learning" is currently available on Unschool platform for all and two technical blogs are available to read on KnowledgeHut.com site.

AJITH ABRAHAM (Senior Member, IEEE) received the Master of Science degree from Nanyang Technological University, Singapore, in 1998, and the Ph.D. degree in computer science from Monash University, Melbourne, Australia, in 2001. He is currently the Director of the Machine Intelligence Research Laboratories (MIR Laboratories), a Not-for-Profit Scientific Network for Innovation and Research Excellence Connecting Industry and Academia. The Network with HQ in Seattle, USA, is currently more than 1,500 scientific members from over 105 countries. As an investigator/a co-investigator, he has won research grants worth over more than 100 million U.S.$. Currently, he holds two university professorial appointments. He works as a Professor in artificial intelligence at Innopolis University, Russia, and the Yayasan Tun Ismail Mohamed Ali Professorial Chair of Artificial Intelligence at UCSI, Malaysia. He works in a multi-disciplinary environment. He has authored/coauthored more than 1,400 research publications out of which there are more than 100 books covering various aspects of computer science. One of his books was translated into Japanese and a few other articles were translated into Russian and Chinese. He has more than 46,000 academic citations (H-index of more than 102 as Per Google Scholar). He has given more than 150 plenary lectures and conference tutorials (in more than 20 countries). He was the Chair of IEEE Systems Man and Cybernetics Society Technical Committee on Soft Computing (which has over more than 200 members), from 2008 to 2021, and served as a Distinguished Lecturer of IEEE Computer Society representing Europe (2011–2013). He was the Editor-in-Chief of *Engineering Applications of Artificial Intelligence* (EAAI), from 2016 to 2021, and serves/served on the editorial board for over 15 international journals indexed by Thomson ISI.

LUBNA ABDELKAREIM GABRALLA received the B.Sc. and M.Sc. degrees in computer science from the University of Khartoum and the Ph.D. degree in computer science from the Sudan University of Science and Technology, Khartoum, Sudan. She became a Senior Fellow (SFHEA), in 2021. She is currently an Associate Professor with the Department of Computer Science and Information Technology, Princess Nourah Bint Abdulrahman University, Saudi Arabia. Her current research interests include soft computing, machine learning, and deep learning.