开放世界机器学习：应用、挑战与机遇

JITENDRA PARMAR 和 SATYENDRA S. CHOUHAN，马拉维亚国家技术学院斋浦尔，印度VASKAR RAYCHOUDHURY，迈阿密大学，美国SANTOSH S. RATHORE，ABV-IIITM 瓦尔哈，印度

传统的机器学习主要是监督学习，遵循封闭世界学习的假设，即对于每个测试类别，都有一个可用的训练类别。然而，这种机器学习模型无法识别在训练期间未出现的类别。这些类别可以称为未见类别。而开放世界机器学习（OWML）则处理未见类别。本文首先介绍了开放世界机器学习的概述，强调其在现实世界中的重要性。接下来，探讨并讨论了开放世界机器学习的不同维度。开放世界机器学习领域在过去十年中才引起了研究界的关注。我们通过不同的在线数字图书馆进行了搜索，并仔细审查了过去十年所做的工作。本文对开放世界机器学习的各种技术进行了系统的回顾。它还提出了研究空白、挑战和开放世界机器学习的未来方向。本文将帮助研究人员理解开放世界机器学习的全面发展及在适当领域扩展研究的可能性。它还将有助于选择适用的方法论和数据集，以进一步探索这一领域。

CCS 概念： • **计算方法 → 学习范式**。

附加关键词和短语：开放世界学习、持续机器学习、增量学习、开放世界图像和文本分类

ACM 参考格式：
Jitendra Parmar、Satyendra S. Chouhan、Vaskar Raychoudhury 和 Santosh S. Rathore. 2021. 开放世界机器学习：应用、挑战与机遇. 1, 1 (2021年2月), 42 页。

1 引言

传统的机器学习方法在数据分析的各个领域都取得了良好的成果。在过去的几年中，它在许多数据领域发挥了重要作用，在许多领域都有效，因此它是一种不断发展的数据分析和可视化方法。然而，传统的机器学习存在一些局限性，例如 1) 它处理孤立的数据，并且在没有使用先前知识的情况下进行学习，以及 2) 训练后的机器模型只能处理用于训练目的的相似实例的输入实例 [61]。考虑在计算机视觉与图像处理 (CV-IP) 和自然语言处理 (NLP) 领域中说明 OWML 的示例（图 1）。

作者地址：Jitendra Parmar，2019rcp9044@mnit.ac.in；Satyendra S. Chouhan，sschouhan@mnit.ac.in，马拉维亚国家技术学院斋浦尔，JLN Marg，斋浦尔，拉贾斯坦邦，印度，302017；Vaskar Raychoudhury，迈阿密大学，510 E. High St.，牛津，俄亥俄州，美国，raychov@miamioh.edu；Santosh S. Rathore，ABV-IIITM 瓦尔哈，瓦尔哈，中央邦，印度。

https://doi.org/

(a) 开放世界的图像分类机器学习

(b) 用户与之间的消息交换 *abc* 银行聊天机器人

图 1. (a) 开放世界机器学习在计算机视觉和图像处理中的应用 (CV-IP) 和 (b) 开放世界机器学习在自然语言处理中的应用 (NLP)

假设一个机器学习模型被训练来识别特定的图像，比如松鼠和兔子。该模型的完整训练是通过不同的松鼠和兔子的图像完成的。如果在测试时提供松鼠或兔子，它能够识别松鼠和兔子。如果提供的图像不是松鼠或兔子，它仍然会将其分类为松鼠或兔子。

然而，在开放世界机器学习的情况下，老虎、鹿和狐狸的图像在测试时将被拒绝，因为这些图像对模型来说是未见的，即系统可以检测到不来自训练集的示例。识别未见示例或对其进行分类的能力称为开放世界学习。

在图（1b）中，一个银行聊天机器人专门设计用于处理与账户和交易相关的查询。如果用户提出一个超出范围的问题，会发生什么？由于它是在封闭世界环境中训练的（场景A），系统将给出错误的回应。在第一次询问中，用户询问关于“暂停账户服务”的问题。聊天机器人正确识别了该查询并做出了适当的回应。在第二次询问中，用户询问水的冰点，这是一个超出范围的查询。然而，聊天机器人错误地识别了这个问题，并建议“账户冻结”的程序。场景B展示了一个理想的聊天机器人系统，使用开放世界机器学习。这两个询问与场景A相同，但聊天机器人正确识别了第二个查询的意图。第二个查询的意图是超出范围的，因此聊天机器人拒绝回答该查询。

经典机器学习 [1, 8, 85, 94],，尤其是监督学习 [61],，遵循封闭世界学习 [75] 的假设。在每个测试类别中，都有一个可用的训练类别 [85, 91]。然而，在现实世界场景中，交互式和自动化应用在动态环境中工作，并且来自新类别的数据定期到达。在这种情况下，遵循封闭世界假设的模型无法应对这种情况。开放世界学习是一种普遍现象，即它并不局限于特定的机器学习 [94]。它可以一般定义为一种模型，能够在执行其预期工作时学习以前未学习的新事物。在工作中学习新事物是一种识别未知的能力。它可以填补人类与机器之间的学习风格差距。换句话说，我们可以说开放世界机器学习可以使机器像普通人一样学习。这是一种自我驱动的学习，能够在出现不同事物时进行区分。开放世界机器学习可以增强多个最近基于AI的原型在计算机视觉-图像处理中的应用，如自动驾驶汽车 [10, 152],、医疗保健和医学诊断 [19, 107],、视频监控 [50],、机器人技术 [130],、社交媒体上干扰图像的识别 [52, 58]。同样，在自然语言处理领域，开放世界机器学习也可以...

学习可以帮助改善聊天机器人系统 [27, 80],智能助手 [23, 53],电子邮件垃圾邮件检测 [142], 产品推荐 [148, 159],和网络欺凌识别 [102]。

开放世界学习是机器学习中一个相对较新的领域，尽我们所知，目前关于开放世界机器学习的综述文章非常少。现有的大多数开放世界机器学习综述文章都是任务或领域特定的，例如，在 [123], 中，作者回顾了许多可以在开放世界中发现新攻击或恶意软件的方法。在 [72, 154], 中，作者回顾了包括基于深度学习的方法在内的多种方法，这些方法用于在开放世界环境中识别人类（个体）。最近的一篇综述文章，见于 [37], ，回顾了计算机视觉和图像处理领域的大部分工作。然而，近年来，开放世界机器学习也被应用于自然语言处理，例如自动对话系统。尽我们所知，目前缺乏关于开放世界机器学习的综述文章，这些文章可以提供更广泛的开放世界学习分类。

本文对开放世界学习相关工作的系统性回顾进行了介绍。首先，我们概述了开放世界机器学习在现实世界背景下的重要性。还对计算机视觉和自然语言处理领域中使用的多种开放世界机器学习方法进行了分类。除此之外，我们还提供了现有工作的表格摘要，强调了优缺点的讨论。此外，我们讨论了一些在开放世界机器学习中用于计算机视觉和自然语言处理的基准算法。总结的讨论有助于在特定学习环境中选择适当的方法。总之，本文的贡献如下。

- 我们提出了一种基于任务的分类法，区分开放世界机器学习（OWML）的关键特征及其关系。
- 我们分析了几种技术及其在效率和其他参数方面的特征。
- 我们还讨论了在开放世界机器学习中用于计算机视觉、图像处理和自然语言处理的各种数据集及其特征，以全面理解结果。
- 我们简要介绍了各种研究空白和挑战，帮助推动开放世界机器学习（OWML）的工作。
- 此外，我们展示了一些与OWML相关的领域，通过不同的技术来确定开放世界问题。

本文的组织结构如下。第2节介绍了关于OWML的背景信息。第3节解释了本文采用的评审方法和OWML的分类法。第4节讨论了OWML在计算机视觉与图像处理 (CV-IP) 方面的相关工作。第5节讨论了OWML在自然语言处理 (NLP) 方面的相关工作。第6节回顾了多个研究者在OWML中使用的标准基准数据集。第7节讨论了OWML中使用的一些基线算法。第8节讨论了与OWML相关的领域。接下来，我们在第9节和第10节中解释了一些研究挑战和未来方向。最后，第10节总结了本文。

2 背景和正式定义

经典机器学习分为两个部分：训练和测试。对于每个测试示例，我们必须有一个训练示例来识别这些类别。因此，专家总是建议测试得分要高，但高测试得分并不能保证有意义的现实世界结果。为了在现实世界中获得良好的结果，机器需要像人类一样学习新事物。如果机器学习了新事物，特别是那些在训练期间不存在的事物，并能够识别这些事物，

testing, then the system will produce more convincing outputs. open-world machine learning can address the concerns of a dynamic environment where the input and nature of input data (size, category, frequency, etc.) are changing rapidly.

To better understand open-world learning, we have to know what *open* means. The systems are often designed for a specific task; the models are trained to identify particular objects if we consider computer vision examples. However, do similar objects come in the real world? In real-world objects are surrounded by many other things. In open-world learning, classifications are open, or models can learn incrementally. It can learn about new classes and update the existing model (without re-training). open-world machine learning also refers as cumulative learning [1] and open-world recognition [8, 28]. Before comparing classical machine learning techniques with open-world machine learning, we have defined some terms here.

1. *Seen-Seen Instances*: Instances that are labelled in the training datasets i.e., classes are know *a priori*.
2. *Seen-Unseen Instances*: Instances that are unlabelled in testing datasets but belongs to the seen classes i.e., classes are known during training time.
3. *Unseen-Unseen Instances*: Instances that are unlabelled in datasets and have not been appeared during training time.
4. *Unseen Instances*: unlabelled instance during training time.

Table 1. Different Paradigms of Machine Learning

Domain	**Techniques and Proposed Year**	**Task**	**Training Data**	**Testing Data**	**Knowledge Accumulation**	**Knowledge Retention**
ML	Supervised Learning (1988)	CL and RG	Seen-Seen	Seen-Unseen	-	-
	Unsupervised Learning (1989)	CR and AS	Unseen	Seen-Unseen	-	-
	Reinforcement Learning (1995)	CL, CR and CNT	Seen-Seen /Unseen	Seen-Unseen	-	-
	Semi-Supervised Learning (2000)	CR and CL	Seen-Seen / Unseen	Seen-Unseen	-	-
DL	Deep Neural Networks (1965)	CL, CR and RL	Seen-Seen / unseen	Seen-Unseen/ Unseen	-	-
CML	Supervised Continual Learning (1995)	CL and RG	Seen-Seen	Seen-Unseen	√	√
	Reinforcement Continual Learning (1995)	CL, CR and CNT	Seen-Seen /Unseen	Seen-unseen	√	√
	Continual Learning in Deep Neural Networks (2002)	CL, CR and Rl	Seen-Seen / Unseen	Seen-Unseen / Unseen	√	√
	Unsupervised Continual Learning (2014)	CR and AS	Unseen	Seen-Unseen and Unseen	√	√
	Semi-Supervised Continual Learning (2015)	CR and CL	Seen-Seen and Unseen	Seen-Unseen	√	√
OWML	Open-world Machine Learning (2015)	CL, CR	Seen-Seen/Unseen	Seen-Unseen and **Unseen-Unseen**/Unseen	√	√

Abbreviations: ML: Machine learning, DL: Deep Learning, CML: Continual Machine Learning, OWL: Open-world Learning, CL: classification, RG: Regression, CR: Clustering, AS: Association, CNT: Control, RL: Representation learning

Traditional machine learning has five major tasks: classification, regression, association, clustering, and control (robotics), which are done with various kinds of Machine Learning (ML), which are shown in Table 1. Supervised machine learning proposed in the early 1988s uses seen-seen data for training and testing. In contrast, unsupervised Machine learning, which is also proposed in the 1980s, uses unseen data for training and testing. Semi-supervised machine learning uses seen-unseen data for both training and testing. Reinforcement learning, recommended for classification and control, perceives and understands its context, takes actions and acquires knowledge by experiments and oversights, uses the seen data for training and seen-unseen data for testing. Deep learning is working on both classification and clustering uses seen data for both training and testing. The task of all categories of continual machine learning and training and testing data are similar to similar traditional machine learning, except traditional machine learning neither accumulates knowledge nor retains any previous knowledge in any future task. Table 1 given the broad classification of tradition and continuous machine learning based on task, required training and testing data, and knowledge accumulation and retention. Open-world machine learning, which uses seen data for training and seen, seen-unseen, and unseen data for testing, is the only method that has a rejection capability for unseen instances.

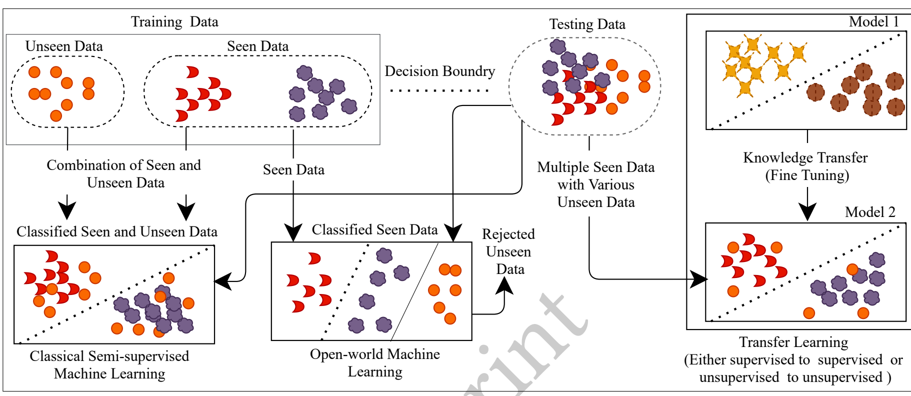

Fig. 2. Classical Semi-supervised Machine Learning *vs* Open-world Machine Learning *vs* Transfer Learning

Sometimes it might appear that, open-world machine learning is associated with semi-supervised learning or transfer learning. However, these are different methods. Figure 2 shows the comparison between semi-supervised machine learning, open-world machine learning, and Transfer Learning. Semi-supervised machine learning involves small number of labelled data and possibly large number of unlabelled data. However, it still follows the closed world assumption that the unlabelled (unseen) data belongs to seen classes. It classifies instances according to classes available in training data. In contrast, open-world machine learning trained with seen data and classified seen data and rejected unseen data. Transfer learning uses knowledge transfer and fine-tuning to classify the new data (knowledge gained from one model can be used in different models to classify instances). Subsequently it works on new data and assume the closed world assumption during testing time. For rest of the paper, we assume that seen classes means the classes which were appeared during training time, and unseen classes mean unseen-unseen classes.

Open-world machine learning problem can formally be defined as follows.

Definition 1. *Let $D = \{(x_1,\ y_1), (x_2,\ y_2), \ldots, (x_i,\ y_i), \ldots, (x_n,\ y_n)\}$, where n is the total number of instances, is the labeled training data for m seen classes. Here x_i is the i^{th} instance and $y_i \in \{s_1, \ldots, s_m\} = S$ is x_i's class label. The objective of the classifier is to classify each test example x to one of the m seen classes or identify it as unseen class.*

The process of learning in open-world machine learning can be defined in following three steps.

- Step 1: At specific time t, classification model builds by a learner that is multi-class classifier M_t based on all previous classes t of data with class labels $S^t = (s_1, s_2, \ldots, s_t)$. M_t is capable enough either classify seen classes $s_i \in S^t$ or reject them as unseen classes and put them in a rejection set R_e. The R_e may have instances of more then one new or unknown classes.

- Step 2: Now, the system can identify the hidden classes c in R_e and prepare training sets from this data to find unknown classes.
- Step 3: The model will learn from updated training dataset (previous data + new identified dataset). The model M_t is update to a new model M_{t+c}.

There are many domains in which open-world machine learning is beneficial, such as image processing, computer vision, and text processing. open-world machine learning can be a bridge for self-motivated learning systems. open-world machine learning is a relatively new field of machine learning; as we analyzed, most of the work in this area has only been done in this decade. We divided the review work into two parts: open-world machine learning in Computer Vision and Image Processing (CV-IP), and the other also known natural language processing (NLP); we have explored the literature work in these two domains.

3 Review Methodology

The methodical review summarized in this paper was done by succeeding conventional review processes that ease understanding of domains of open-world machine learning. The steps involved to write in this review article are the historical timeline, convoying the survey, describing the outcomes, discussing investigations and challenges, the reasoning of conclusions, and future direction.

3.1 Review Plan

Conveying a methodical study includes collecting initial analysis about conclusions. Typical methods of such surveys incorporate confirmation and contradiction of preceding claims, classification and examination of analysis gaps/challenges, and future direction for exiting research. There is a fundamental advantage of conveying a methodical review and beneficial for authors as it covers the information of the domain with data. The following steps are taken to complete this survey.

- Steps of Review Plan
  (1) Recognize the requirement for a methodical survey
  (2) Frame an investigation query.
  (3) Find tasks and methods around that investigation query.
- Steps of Review and Result Reporting
  (1) Explore the initial investigations
  (2) Study the initial investigations for significance and relevance of domains
  (3) Selection of methodologies of the initial investigations
  (4) Integrate and abstract the extracted studies from initial investigations
  (5) Describe and report results as it is with suitable datasets
  (6) Conclude the methodologies and investigation
  (7) Conduct analytical and tabular comparisons
  (8) Write the methodological survey

3.1.1 Investigation Queries We have formed the following generic queries to pursue the results from the readers' perspective. These are the standard parameter and findings that are required to understand any domain of research. Further, we prepare the entire draft according to the review plan and investigation queries.

(1) What is the importance of learning in the open-world?
(2) How has machine learning grown in the last decade?
(3) What are the classifications of open-world machine learning?, and
(4) How does it differ from traditional Machine Learning (ML)?

(5) Which domains are correlated with OWML and how OWML can help to improve these domains?
(6) What is the current status of research in OWML?
(7) What are the tasks of OWML?
(8) What are the methods available in OWML to handling open-world tasks?
(9) How many datasets are available to investigate or perform OWML research for explicit domains.
(10) What are the associated areas of OWML?
(11) What are the challenges in the field of OWML to learn in open environments for various domains?
(12) What are the future directions of research in OWML?

3.1.2 Sources of information and Selection Criteria There is a need for a comprehensive aspect to the boundless coverage for an immeasurable and helpful article. We have collected a piece of pertinent information and data before getting started with a comprehensive article. We have explored many articles and select profoundly associated articles only to include them in the review. To collect this data, we have used prominent electronic sources, which are listed in Table 2.

Table 2. E-source of Information

E-Sources	Content Type	Total Article
https://www.acm.org	Journal and Conference	43
https://www.springer.com/in	Journal and Conference	89
https://www.ieee.org	Journal and Conference	179
https://www.elsevier.com/en-in	Journal and Conference	94
https://www.tandfonline.com/	Journal and Conference	17
https://www.jmlr.org	Journal and Conference	26
https://www.aaai.org	Conference	35
https://www.kdd.org	Conference	23

Supplementary Sources: Other than the mainstream sources of information, we have used many repositories and other e-resources. These sources are helping us to provide additional information, technical and scientific reports, and analytical data to understand the domain. Some of the sources are listed below.

(1) https://mitpress.mit.edu (Books and Article)
(2) https://citeseerx.ist.psu.edu (Article)
(3) https://www.semanticscholar.org (Article and Technical Reports)
(4) https://www.morganclaypool.com (Book)
(5) https://www.kdd.org (Article)
(6) https://www.sciencedirect.com (Article and Technical Reports)
(7) https://www.connectedpapers.com/ (Article)
(8) https://scholar.google.co.in (Article and Technical Reports)

Article Search and Inclusion Criteria: In approximately all searches carried the keyword "open-world" in its title or abstract, we keep it in our repository. The domain is relatively new, and most of the work has been done only in the last decade, so we have to access multiple sources to collect the information. We have detailed examined these articles and kept the relevant articles only, process shown in Figure 3. Other than the keyword, we have used the most recent articles and technical reports to follow the rooted trail and find many relevant articles. To maintain the high authenticity of any claim, we used only notable and reputed sources articles in this review.

After selecting articles by applying all the criteria on obtained articles from various sources, we have comprehensively studied the selected article on open-world machine learning. Based on the study, articles are categorized in two significant domains of OWML, that is Computer Vision and Image Processing (CV-IP), and Natural Language Processing (NLP). Further, the articles are classified based on task, some of them shown in Figure 4, and the rest are discussed in the following sections.

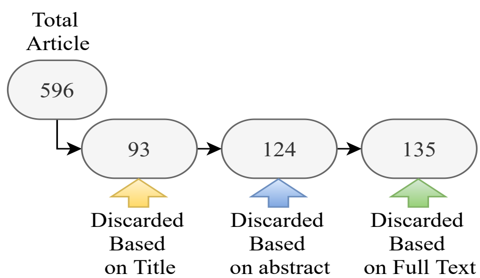

Fig. 3. Article Discard Process

The timeline for open-world machine learning for both CV-IP and NLP shows in Figure 4. The classified task shown in the figure is categorized based on a domain with proposed years. The most of researches were done between 2016 to 2021. In the timeline, we have included only essential methodologies that have enhanced existing work or introduced novel methodology for specified tasks of open-world machine learning. We can observe that there is less research available in NLP for open-world machine learning than CV-IP. The timeline clearly shows that most of the research done in open-world machine learning is generally associated with discovering unseen instances. Some of the research focuses on the detection of novel classes for both CV-IP and NLP.

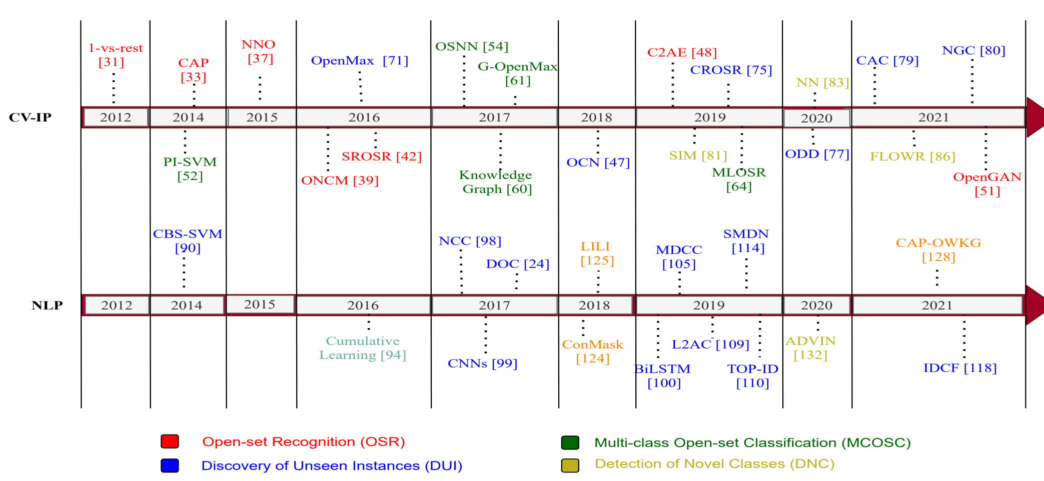

Fig. 4. Timeline of Task Done in OWML for Both CV-IP and NLP

3.2 Taxonomy of Open-world Machine Learning

To ease the understanding of readers, we have graphically summarized (Figure 5) the entire work done in open-world machine learning mentioned in this article. It involves cataloging of the domain, used or proposed methods, and dataset. There are two major fields where work has been done: computer vision & image processing (CV-IP) and natural language processing. We have further categorized the work done in CV-IP based on tasks, such as Open-set Recognition (OSR), Multi-class Open-set Classification (MCOSC), Discovery of Unseen Instances (DUI), and Detection of Novel Classes (DNC). In computer vision and image processing, Numerous approaches have been used with various datasets to evaluate methods with different evaluation parameters.

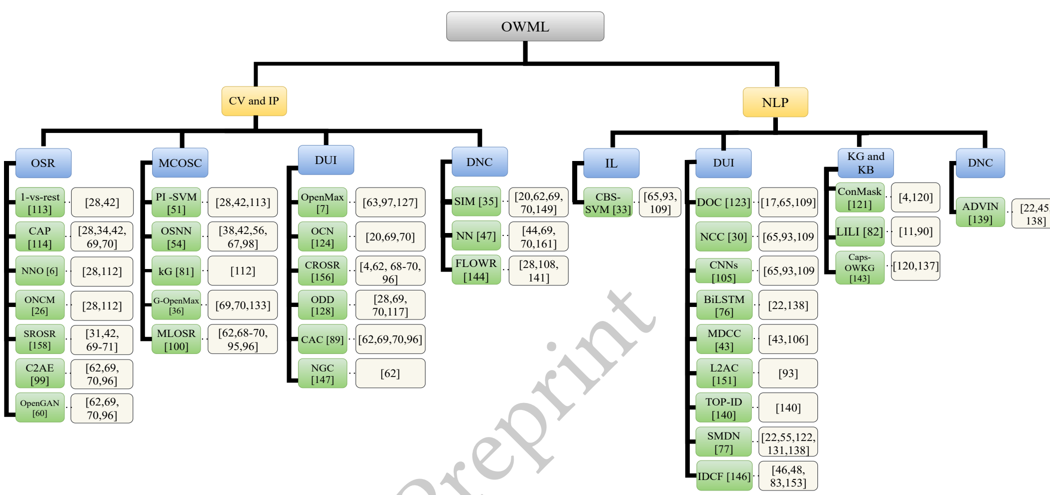

Fig. 5. Taxonomy of Open-world Machine Learning

In computer vision and image processing to achieve various task 1-vs-rest, Nearest Non-Outlier (NNO), Extreme Value Theory (EVT), Conditioned Auto-encoder (C2AE), and Deep Neural Network (DNN) has been used, and these methods are evaluated with various datasets. The baseline benchmark algorithms are improved or extended to integrate the existing changes, such as Support Vector Machines (SVM) used to minimize open-space risk. Some other methods were also used for open-space risk minimization, such as NNO. The Probability of Inclusion- Support Vector Machines (PI-SVM), Open-Set Nearest Neighbor (OSNN) model, and EVT are used in many frameworks for multiset reorganization. The unseen class identification OpenMax, Generative OpenMax (G-OpenMax), Convolutional Neural Network (CNN), Classification-Reconstruction Learning for Open-Set Recognition (CROS), and DNN has been used. CNN and Stream Classifier with Integral Similarity Metrics (SIM) have also been used to discover new classes.

We also discussed the work done in natural language processing in open-world; and further categorized it in Incremental Learning (IL), Discovery of Unseen Instances (DUI), Knowledge-graph & Knowledge-base (KG & KB), and Detection of Novel Classes (DNC).

Several researchers have been used baseline algorithms such as Center-Based Similarity Support Vector Machines (CBS-SVM) in natural language processing to reduce the open-space risk and incrementally acquire knowledge. Several methodologies mentioned here are a significant part of the framework or are used standalone for unseen class discovery. The 1-vs-rest, CBS-SVM, Nearest Centroid Class (NCC), Long Short-term Memory Networks (LSTM), Learning to Accept Classes (L2CA), SoftMax, and Deep Novelty (SMDN), and Automatic Discovery of Novel Intents

(ADVIN) were used and evaluated with different datasets. The Lifelong Interactive Learning and Inference (LILI) model is used for new class detection, and OpenMax based models are used for text classification in the open-world. Now we discuss the reviews of methodologies that were used for computer vision and image processing in detail (Section 4). Next, we discuss methodologies for natural language processing in open-world machine learning (Section 5).

4 Open-world Machine Learning in Computer Vision and Image Processing

In this segment, we discuss the literature works done in computer vision and image processing with open-world settings. The preliminary research is focused on open-set reorganization using various methods. The input images are unknown for the model, or we can say that input is novel and unseen; as the input images are not available in training data, the knowledge is incomplete for the model. The model needs to respond to these unseen (open) data. The task analysis of work done for the particular task is shown in Table 3. Further, These tasks are discussed in Section 4.1, 4.2, 4.3, and 4.4.

Table 3. : Summarized Study of Task Performed by Open-world Machine Learning in CV-IP

Author	OSR	MCOSC	DUI	DNC
W.J. Scheirer et al. [113]	√	-	-	-
W.J. Scheirer et al. [114]	√	-	-	-
L. P. Jain et al. [51]	-	√	-	-
A. Bendale and T. Boult [6]	√	-	-	-
A. Bendale and T. Boult [7]	-	-	√	
R. De Rosa et al. [26]	√	-	-	-
H. Zhang and V. M. Patel [158]	√	-	-	-
P. R. M. Junior et al. [54]	-	√	-	-
S. Demyanov et al. [36]	-	√	-	-
V. Lonij et al. [81]	-	√	-	-
L. Shu et al. [124]	-	-	√	-
P. Oza and V. M. Patel [99]	√	-	-	-
R. Yoshihashi et al. [156]	-	-	√	-
P. Oza and V. M. Patel [100]	-	√	-	-
Y. Gao et al. [35]	-	-	-	√
M. Hassen and P.K. Chan [47]	-	-	-	√
L. Song et al. [128]	-	-	√	-
D. Miller et al. [89]	-	-	√	-
J. Willes et al. [144]	-	-	-	√
S. Kong and D. Ramanan [60]	√	-	-	-
Z.-F. Wu et al. [147]	-	-	√	-

4.1 Open-set Recognition (OSR)

In a real-world environment, several external circumstances restrict the identification and distribution of tasks as inputs change frequently. It is generally challenging to accumulate training examples to employ all levels when training a classifier. A more practical situation is open-set recognition occurs, wherever inadequate system information exists during the training, and unseen classes can be provided to a system during testing. In such a situation, expect the classifiers to correctly label the seen classes and effectively deal with unseen classes. Some research approaches this obstacle and recognizes open-sets.

In [113], the authors proposed an algorithm that can accept input with incomplete knowledge. The existing algorithm cannot handle open-sets. Thus they improve algorithms with the normalization of an algorithm to handle open-sets. They introduced 1-vs-set and performed experiments on Caltech-256 [42] and ImageNet [28] sets. They perform experiments on labeled face data and compared it with their work. Also, perform experiments on different image domains and compare results with binary SVM linear kernel, binary 1-vs-set machine linear kernel, 1-class SVM linear kernel, and 1-class 1-vs-set linear kernel. Based on performance evolution, F-measure and accuracy clearly state that the binary 1-vs-set machine linear kernel performed better than the other algorithms. Object and face recognition are considered in their work to verify the experimental results. Many researchers use multi-class classification to handle open-set problems, but multi-class approaches need labels for each input class. Therefore, the entire dataset required very laborious labeling and is still not an acceptable solution to handle open-sets. In this work, openness formalizes as:

$$Openness = \sqrt{\frac{2 * |T_c|}{|T_s| + T_g}} \tag{1}$$

Where T_c= Training Class, T_s= Testing Class and T_g= Target Class. It yields openness in percentage between 0 to 100. Where 0= complete close class and 100 denote maximum openness, they conduct experiments on SVM with half-space and classify data not available in the training set. The SVM found separate classes as negative and positive, and the negative is only for known objects. The Rest of the unknown objects are left as large unclassified open-space, which could be a part of a positive set. Then they felt that this could be remedied by reducing the open-space. Instead of generalization and specialization to minimize the errors in training function, they introduce "set". Set is known class training data of 1-vs-set. It is used for open-space risk models and error minimization.

In [114], the authors proposed a 1-vs-rest machine. The 1-vs-Set algorithm handles the hazard of the unknown classes by dealing with two plane optimization, likely to result as a linear classifier. They extended open-risk classification to include non-linear classification in multi-class settings. They suggested a new model Compact Abating Probability (CAP) based on Weibull-calibrated SVM (W-SVM); it decreases the probability value of member class when points move towards open-space from known data. The CAP evaluated on publically available benchmark datasets Letter [34], MNIST [69, 70], Caltech-256 [42], and ImageNet [28]. The results show that the CAP can reduce the open-space risk for known data. In [113], it is established that by optimizing two planes, the 1-vs-set machine can manage the risk and produce a linear classifier. 1-vs-rest reduces the open-space risk by interchanging half-space but still, the open-space risk is infinite. For certain known classes, it is quite easy to find known but when handling multisets. Here three basic categories of the class are denoted, known classes (the class with abstractedly labeled, that is a positive example), known unknown classes (negative examples), and unknown classes (unseen classes) Figure. The algorithm is specially designed for unknown classes, and this algorithm deduced open-space risk from infinite to finite.

1-vs-set machine assigns class labels to examples through the testing. It used a probability decision score for multi-class. It classifies examples by multiple classifiers based on the highest probability or probability that goes beyond the threshold. Examples that are below the threshold are rejected as unknown. This work has formalized a compact abating probability to address open-set regeneration by introducing a new algorithm, W-SVM, which integrates compact abating probability model and probability estimation theory. The experimental results clearly state that the f-measure (openness) of the W-SVM is relatively high for both open-set binary object regeneration and multi-class open-set recognition. The W-SVM also performed decently on the OMNIST dataset for multi-class open-set recognition.

In [6], authors address the issuers associated with open-world recognition, such as open-space risk and practical tasks. They proposed a protocol to evaluate open-world recognition. They proposed Nearest Non-Outlier (NNO) algorithm to manage open-space risk, model efficiency, and adding object categories incrementally while detecting outliers. The proposed NNO algorithm experiment on more than 1.2 million images of ImageNet [28] dataset to validate the model. NNO is an extension of the Nearest Class Mean (NCM) [111] algorithm. They set an open-world evaluation protocol in which uses seen classes in training. However, both seen and unseen classes are used during the testing and continually add new class categories when encountering unseen classes. The training phase is further divided into two phases, the first is metric learning, and the second is the incremental learning phase.

NNO is a combination of two open-space risks for model combination and open-space risk for threshold space. NCM [111] assumes that all classes are known that are not suitable for open-world recognition; hence NNO extends its feature and considers a measurable recognition function. It is given the probability of an object being in any class. It is always greater than zero. When a measurable function is zero, and all classes reject, the NNO rejects the object as unknown. After collecting all unknown novel objects, NNO considered these objects as new classes. NNO can identify unknown objects and include these unseen classes as new classes. Hence NNO is an open-world algorithm. The experiment has been done with two different sets of categories that are 50 and 200 and applied SVM, 1-vs-set, NCM, NNO. The experiment evaluations state that NNO outperformed the rest of the algorithms and generate significantly better outcomes when testing with unknown categories of classes for both 50 and 200 sets of categories.

In [26], an author extends the work of open-world recognition [6]. They argue that to capture dynamic word recognition, needed incremental learning of underlying matrices, confidence threshold for unseen classes, and description space of class. They conduct experiments in three phases, first large-scale increment learning, second is open-world recognition, and third is an online prediction of streamed images. The ImageNet [28] and ILSVRC'10 [112] data set used in experiments, which consist of 1.2 million, 50K, and 150K images for training, validation, and testing, respectively. The large-scale increment learning method created relevant matrices and learned parameters on the initial set of 20 classes, then classes are added incrementally in the set of 10 classes. To recognize classes in an open-world environment method, learn parameters and metrics on an initial 50 classes, and images of 50 classes are added after each iteration to evaluate the performance on a test set that comprises both known and unknown classes. To predict online images, researchers used current methods Nearest Class Mean (NCM) classifier and Nearest Ball Classifier (NBC) and updated these accuracy. They predict images using ONCM and ONBC and compare results with existing NCM and NBC. Initially predicted the labels for samples with the current model, then update online accuracy based on predicted labels and genuine labels. After updating the online accuracy, updated the existing methods using accurate labels and sets. They also conduct the same experiment on Places-2 dataset [160]. The results clearly state that the ONNO, ONCM, and ONBC are performed better than existing algorithms.

In [158], the authors proposed a framework that works on sparse representation-based classification (SRC). SRC used class reconstruction error for classification. The most useful information about open-sets is available in the tail of the similar and non-similar parts of the class. To tail distribution reconstruction error SRC uses statistical Extreme Value Theory (EVT) [25]. To evaluate the result they used benchmark datasets, such as extended Yale B [71], MNIST [69, 70], UIUC Attribute dataset [31] and Caltech-256 Dataset [42]. Evaluation results show that the simple, sparse representation classification does not effort up to the mark for open-sets. Hence introduce different training modules to the trained system to handle open-sets. In training, acquired a random sample for each class then partition the class into two sets, cross-train and cross-test. These partitions are used for training and testing. Cross-train contains 80, and cross-test contains 20 percent of training samples. To evaluate the result of sparse representation-based open-set recognition (SROSR), Compare the result with existing methods that used sparsity-based rejection such as w-SVM, Sparsity Concentration Index (SCI) [145], ratio, and Naïve. The result clearly states that the SROSR is providing more accuracy and better F-Measure than existing methods.

The encoder-based model [124] to detect unseen classes are further extended and use encoder and decoder for classification and open-set detection. In [99], the authors proposed a C2AE method that used class conditioned auto-encoder to recognize open-set with novel training and testing methodologies. The proposed model worked in two parts, close-set classification and open-set identification. Encoder learned the first task for close-set classification, and decoder learned the next task for open class identifications. Training has been done using a close-set model. The close-set model consists of known classes. It trained encoder and classifier and conventionally calculated classification loss; after training encoder for close-set, it trained open-set identification module which consists of an auto-encoder network with weight and decoder for the reconstruction of the image according to label condition vectors. Testing of the model has been done for an open-set using a k-inference algorithm. To evaluate the performance of the model for open-set compare the model with W-SVM [113], SROR [158] and DOC [123], on MNIST [69, 70], SVHN [96] and CIFAR10 [62] benchmark datasets.

In [60], the authors proposed an OpenGAN to recognize the open-sets by open data generation. The OpenGAN consists of the GAN-discriminator to classify testing examples. It is a binary classifier trained for both open-set and closed-set data. There are many other techniques available for close-set classification, but each technique has limitations, but OpenGAN overcomes them by integrating them with various technical insights. In the first step, OpenGAN picked GAN Discriminator on a few actual outlier data already used in the existing research. The second step synthesizes "fake" data and adds it to the complete open training examples. The proposed method methods are evaluated with benchmark datasets (CIFAR, SVHN, and MNIST) and show promising outcomes for open-set recognition vi open data generation.

4.2 Multi-class Open-set Classification (MCOSC)

Multi-class classification in the open-world is a very challenging task. If the unknown classes remain unaddressed, the classifier either misclassifies or classifies false, known classes, and there is also a possibility to classify false unknown classes. Misclassification and false classification can be avoided if multi-class classifiers can identify unknown classes appropriately.

In [51], authors articulate the problem as one of sculpting positive training data at the decision boundary and invoke the arithmetical theory. The new algorithm termed PI-SVM recalls higher accuracy than existing. It is used for assessing the non-normalized posterior likelihood of class insertion. They convert MNIST [69, 70] data set from closed set to open-set recognition task and experiment with different sets of training and testing data. Now the following steps have been followed: (i) It used the standard supervised learning algorithm for MNIST [69, 70] classification with a 1-vs-rest SVM on Platt [104] probability estimation in which classes are seen during the training, (ii) Used only six classes from MNIST, (iii) Used all ten classes from MNIST and four unseen classes during the training, and (iv) Change the testing regime to cross-class validation. In this scenario, similar classes are held out during the training (it is just a shuffling in (ii)) but comprise in a testing (it is a shuffling in (iii)). They have performed two distinct open-set scenarios with the cross-class validation: object detection by specific classifiers and multi-class open-set recognition followed by detection of a problem and then compares it with PI-SVM. To evaluate performance measures and binary decision elements for an open-set object decision used a universe of 88 classes [113]. To train the model, used images from Caltech-256 and for testing took the images from both Caltech-256 and ImageNet [28]. The entire evaluation was done over the five fold cross dataset. The result clearly states that PI-SVM improves the F-Measure by 12 to 22 percent compared to existing methods.

In [54], the authors proposed Open-set Nearest Neighbor (OSNN) to address the issues in multi-class classifiers. It is an extension of the Nearest Neighbor (NN) [57] for open-set. OSNN used a similarity ratio instead of a similarity score and applied a threshold to find similarities between classes. They also designed a specific experimental protocol to evaluate open-set methods. Earlier proposed algorithms and frameworks are displayed as a virtuous outcome in the experiment, but in real-world applications, these algorithms and systems are straggled to perform with open-set. Hence, to overcome these issues, they also proposed a system that measures adaptation in an existing open-set classification system, that is, Normalized Accuracy (NA) and Open-set F-measures (OSFM), and evaluates the classifier's performance for both seen and unseen classes. The evaluation done on 15-Scenes [67], Auslan [56], Caltech-256 [42], ALOI [38] and Ukbench [98] datasets.

Visual recognition systems play an essential role in identifying both seen and unseen classes of images. In [81], authors proposed a knowledge graph-based approach to identify unknown visuals and recognize visuals in the open-world. Three basic methods are used to predict classes, first Standard classification settings can predict only classes that are available in training, and images that are not available in training cannot be accurately predicted by standard classification. The second, zero-shot setting can predict images that are not available in training, but some partial information is available for novel classes. The third open-world setting can predict images that are neither available in training nor any partial information about its classes. The proposed method used the knowledge graph embedding model and image embedding model. The knowledge graph model uses properties, and the image embedding model uses images for training (ILSVRC-2012 [112]). Image embedding models used properties of the knowledge graph to predict open-world images.

In [36] authors proposed Generative OpenMax (G-openMax), which calculated the decision score of unseen classes instead of seen classes. It is an extension of OpenMax, which consists of a GANs [40] network. The proposed method used visualization for both seen and unseen classes; it also used probability estimation to GANs, and previous seen class dissemination to produce reasonable and domain-adapted synthetics unseen samples. The evaluation has been done on both small and large scales datasets. A minimum of 10, and a maximum of 95 classes are utilized for the openness problem on two (HASYv2 dataset [133] and MNIST [69, 70]) handwritten datasets.

In [100], authors' proposed Multi-task Learning-based Open-Set Recognition (MLOSR). It is based on a neural network for multitasking in open-set visual recognition. The proposed method is a combination of a classification network, decoder network, and feature extractor network. It utilized a decoder network to reject an open-set, and the decoder network reconstructs the error. It also uses EVT [25] for model tail error reconstruction from seen classes. EVT improves the overall performance of the model. The feature extractor network took input and generated the latent. The classifier uses this latent decoder to predict the class labels and reconstruction of input images. The entire network has trained for both reconstructions of input images and classification. EVT modeled the trail of the reconstruction of the error distribution. The probability of reconstruction error by EVT and classification score used for open-set recognition testing. Experiment done on COIL-100 [95], MNIST [69, 70], SVHN [96], CIFAR10 [62] and Tiny-ImageNet [68] datasets, MLOSR are tested with benchmark network VGG and dance Net with SoftMax [39], OpenMax [115] and combination of ladder net, DHRNet with SoftMax, OpenMax, and CROSR, MLOSR performed better than an existing network to recognize open-set. Existing work focused on the prediction of unknown classes. The identification of newer classes from unknown classes is difficult without finding instances of unknown classes.

4.3 Discovery of Unseen Instances (DUI)

Generally, systems choose images that might not be useful or significantly meaningless. In traditional classification methods, the system has to classify the testing object in some of the classes. In comparison, an ideal system must reject the unseen classes that are meaningless and irrelevant. Some of the work presented shows how "fooling" [97] and "rubbish" [41] images appear in relevant classes as their confidence is high, whereas these are far from the class in which they appeared. Traditional deep networks have used fully connected feeds to the SoftMax layer [39] as output. SoftMax produces probability for the known labeled classes.

In [7], the authors addressed this issue by introducing a methodology that can reject the unseen classes while testing. It is an adapted deep network for open-set identification. This methodology introduced OpenMax [115] that can evaluate the likelihood of an input being for an unseen class. OpenMax rejects unrelated images, reduces the error rate, and manages open-space risk. OpenMax estimates class by measuring a distance between the model vector aimed at the limited upper classes and the activation vector for an input. OpenMax provides the likelihood of unknown classes. Here OpenMax has an extended version of SoftMax that includes probability for unknown classes. This method used meta-recognition in deep networks and found scores to estimate how far testing an object to a known class. To estimate the score activation layer has been used in deep networks. Meta-recognition and OpenMax can differentiate seen and unseen classes and avoid foolish images to classify in known classes. The proposed model was evaluated on ImageNet, a subset of the ILSVRC-2012 dataset since ILSVRC-2012 test labels are unavailable for use, experiments stated on validation set performance [63, 97, 127].

To extend neural network-based unseen class discovery and add the capability of rejection combination of networks has been used. In [124], authors proposed a framework to identify seen classes and reject unseen classes, seen classes that are available in training, and unseen classes which are available at the time of testing. It is not possible without having previous knowledge. The objective is to discover unseen classes for any given task and make a cluster by rejected examples. Open-world machine learning is quite different from knowledge transfer. In the knowledge transfer mechanism, the system sends information between supervised to supervised and unsupervised to unsupervised systems. In this work, knowledge is shared from supervised to unsupervised. To achieve the objective, they develop a model that consists Combination of two networks an Open Classification Network (OCN) and a Pairwise Classification Network (PCN). Both networks will share the same omponents for learning. OCN is Build function F(x) that can classify each seen and unseen class in S where PCN Build $g(x_p,\ x_q)$, a binary classification model. PCN will identify two test examples seen, unseen, from the same class or different classes, and Hierarchical clustering used to discover hidden classes in all rejected examples. To test compare with existing work, the proposed model, evaluated on MNIST [69, 70], and EMNIST [20] dataset.

All the methods discussed above are trained in a supervised manner and designed to classify known classes that are available at the time of training. Therefore it is tough to determine unseen or unknown classes using these methods. It also upshots the accuracy of the classification of known classes.

In [156], authors proposed Classification-reconstruction Learning for Open-set Recognition (CROSR) for robust unknown classes deprived of distressing the classification accuracy of known classes. CROSR trained networks for categorization and restoration of input data. While learning to distinguish unseen from seen and classes of seen, this technique helps to improve the implicit interpretation. To provide durable unseen recognition despite compromising the efficiency of seen-class classification, CROSR method uses implicit structures for reconstruction. CROSR is based on OpenMax formulation. It reconstructs the input data to detect uuseen classes. They use exclusionary learning algorithms in seen classes to build their classifiers. An open-set classification system based on DHRNets, CROSR combines seen classification with unseen detection. This technique outperforms existing deep open-set classifier algorithms DOC [123], SoftMax [39] and OpenMax [115], for most permutations of seen data and anomalies, according to the trials conducted on five typical picture and text datasets MNIST [69, 70], CIFAR-10 [62], SVHN [96], tiny-ImageNet [68] and DBpedia [4].

The current research scenario focuses on finding new classes in rejected data that are unseen or unknown. It will make the system more realistic and capable of working as a human being in a dynamic environment. In [128], authors focused on the impact of out-of-distribution detectors and evaluated the performance of detectors. They took six Out-of-distribution Detectors (ODD) , which are published in the best conferences in the world. They also tested detectors for corrupt images that's effect is unpredictable on the outcome; it may improve or decrease the performance. The out-of-distribution detectors ODIN, Network Agnostophobia, Mahalanobis Detector, Auto-encoder Detector, Deep-SVDD, and Outlier Exposure, are evaluated with MNIST [69, 70], VOC12, ImageNet, Internet Photos [117], Gaussian Noise and Uniform Noise with WRN-28-10 model using a different combination of in-distribution. The performance evolution states the adversarial training can improve the end-to-end strength. Adversarial training decreases discriminative influence and leads to poorer detection performance on benign out-of-distribution data.

In [89], the authors proposed a simple Deep Neural Network (DNN) based framework for open-set classification. DNN contains Open-set Layer (OS-Layer) and Closed-set Layer (CS-Layer). It splits the data of intraclass. DNN splits data into subsets and produces an atypical sample. Atypical samples are used to model then abnormal data and normal samples are used for training. Intraclass info splitting exploits the inter-class information. The closed set regularization deep neural network apprehends an extraordinary close-set precision, and it is competent to discard unseen classes. The experiment performance evaluation is done on MNIST [69, 70], SVHN [96] and CIFAR10 [62] and compares results with WSVM, OCSVM, GAN, CF, and AE-ics.

In [147], The authors present NGC, a novel graph-based noisy tag learning framework, which rectifies in-distribution noisy tags and filters out-of-distribution examples by leveraging the confidence of model predictions and geometric characteristics of the data, when it comes to testing. NGC can identify and discard out-of-distribution samples without any additional training. NGC is evaluated on CIFAR-10 and CIFAR-100 publically available benchmark datasets associated with real-world tasks. The experimental evaluation of NGC shows that shows improvement over the existing methods.

4.4 Detection of Novel Classes (DNC)

The key challenge is finding instances of newly presented data known in nature for the system. Most researches are focused on data with a low dimension dependent on coherence data and its property; therefore, detecting instances for newly known classes is hard to detect.

In [35], the authors proposed a solution to this problem. The proposed framework SIM is a semi-supervised stream classifier that performs classification and detects novel classes on high-dimensional data streams. It uses latent features space for classification, and an open-world classifier implements metric learning, stream classification, and detects novel classes in unseen data. The performance evaluation was done on both image and text data. To test image model they calculate novel misclassified instance (M_{new}) and existing instances misclassified as a novel (F_{new}) apart from slandered performance majors FASHION-MNIST [149], MNIST, EMNIST, and CIFAR-10 and for real-time text data, articles from the New York Times and Guardian have been used with ten classes of other news.

open-world machine learning has also extended its significance in security as we have new kinds of malware in every period. To recognize that type of unseen class of malware, we need a system to detect undefined classes. In [47], authors proposed a method that can detect new unseen classes of malware. In this exemplification, samples from a similar class are closed to each other while those from different classes are further apart, leading to more significant space between known classes for unknown class samples to occupy. The proposed algorithm uses three datasets to evaluate the results, MNIST [69, 70], MS challenge [44], and Android genom [161].

In [144],the authors proposed open-world classification techniques that use embedding-based few-shot learning algorithms. It comprises small context and big context few-shot open-world recognition formalization where decision-making machines must classify existing classes. Few-shot learning for open-world recognition combines Bayesian non-parametric class priors with an embedding based pre-training method. It also discovers unknown classes and then quickly adapts and generalizes classes with the limited labeled data. It adapts benchmarks approaches such as few-shot training, open-set classification, and open-world identification to this environment. The authors present a Bayesian few-shot learning technique based on Gaussian embedding. The proposed system can efficiently integrate new classes for both few-shot open-world recognition situations and Bayesian non-parametric classes. The evaluation results show that the proposed approach improves on a range of current methodologies by 12 percent in terms of H-measure. They evaluate the proposed model on Mini ImageNet [141] and TieredImageNet [108] few-shot learning datasets(Subset of ImageNet ILSVRC-12 [28]).

Table 4 shows a summarized illustration of literature on open-world machine learning in computer vision and image processing. It shows the used or recommended methodology, datasets employed for evaluation, and proposed results by the authors.

4.5 Available Software Packages and Implementations

In this section, we provided a link for some for software packages which contains the various implementation of various model of open-world machine learning in computer vision and image processing (Table 5). These are the models commonly used in various frameworks of open-world machine learning .

Available software packages can be used to improve further learning in the open-world for computer vision and image processing. The 1-vs-rest is helping to improve the rejection of unknown classes. The Nearest Non-Outlier (NNO) can normalize the open-space risk and open-set reorganization. Conditioned Auto-encoder (C2AE) is an encoder and decoder method for open-set reorganization, Multi-stage Deep Classifier Cascades (MDCC) for finding new classes. The Probability of Inclusion- Support Vector Machines (PI-SVM) and Weibull-calibrated Support Vector Machines (W-SVM) can be used for multi-class classification in open-world machine learning .

4.6 Discussion

Many algorithms and frameworks are given significant outcomes for images in real-world settings. However still, there is a need for a generic framework to deal with real-time inputs in a dynamic environment. Ideal outcomes can be achieved if models can adopt generalization or specialization and optimization of parameters. The algorithms must have the capability to handle inputs from multiple domains that may contain various classes, and these classes may have a different kind of object in nature. The multiple objects in inputs can be handled by including localization while optimizing the parameters. In the open-world applications that are working in the real world, the input rate is a significant issue because of the unpredicted flow of input in terms of size and frequency. The open-space risk minimization is a crucial challenge for every algorithm to ac hive high accuracy while learning in the open-world. The system must include prior knowledge to adapt continuity in learning that can reduce learning efforts in the future.

Image processing is one of the binding domains of computer science, and there is plenty of work has been done in this field, although there is scope to extend the research in open-world machine learning . The world is towards automation in computer vision and image processing, such as driverless cars and humanless goods delivery systems introduced by many research organizations. The real-time activities in a dynamic environment can be handled if the system is interactive and functions end-to-end to recognize the multiple objects in open-space. The interactive models

Table 4. Summarized Study of Open-world Machine Learning in CV-IP

Author(s)	Proposed /Used Methodology	Dataset	Reported Results
W.J. Scheirer et al. [113]	1-vs-rest	Caltech-256 and ImageNet	F1-score 80%, Accuracy 98%
W.J. Scheirer et al. [114]	Compact Abating Probability (CAP)	Letter, MNIST, Caltech-256, and ImageNet	F-measure 95 to 98% for 0 to 14% Openness
L. P. Jain et al. [51]	P_I-SVM	Letter, MNIST, Caltech-256, and ImageNet	F-measure 88 to 98% for 0 to 14% Openness
A. Bendale and T. Boult [6]	Nearest Non-Outlier (NNO)	ImageNet and ILSVRC'10	
A. Bendale and T. Boult [7]	OpenMax	ImageNet (ILSVRC'10)	F-measure 0.59% for Threshold values 0.20 to 0.45
R. De Rosa et al. [26]	ONCM, ONNO, and ONBC	ImageNet (ILSVRC'10)	Top-1 Accuracy 43% for known Train Classes Top-1 Accuracy 49% for Unknown Train Classes (50 Known Classes)
H. Zhang and V. M. Patel [158]	Sparse Representation-based OPen-Set Recognition (SROSR)	MNIST, Extended Yale B,UIUC attribute, and Caltech-256	F1-measure 93 to 98% for 0 to 14% Openness Accuracy 92 to 99% for for 0 to 14% Openness
P. R. M. Junior et al. [54]	Open-Set Nearest-Neighbor (OSNN)	15-Scenes, Letter, Auslan, Caltech-256, ALOI, and Ukbench	Normalized Accuracy 90%(Max. with Auslan) Micro open-set F-measure 80% (Max. with Letter) Closed Accuracy 90% (Max. with ALOI)
S. Demyanov et al. [36]	Generative Open-Max (G-OpenMax)	MNIST and HASYv2	F-measure 80 to 99% for 0 to 13% openness Accuracy 58% (Maximum with MNIST)
V. Lonij et al. [81]	knowledge-graph	ILSVRC-2012	Fracrion of Image 85% (With atleast 1 correct triple) Mean Rank 14%, and average number of true triples 19%
L. Shu et al. [124]	Open Classification Network (OCN) CNN and 1-vs-rest	MNIST and EMNIST	Mirco F1-score 91% (Max with MNIST) Accuracy 81% (Max with EMNIST)
P. Oza and V. M. Patel [99]	Class Conditioned Auto-Encoder (C2AE)	MNIST, SVHN, CIFAR10, CIFAR+10, CIFAR+50, and TinyImageNet	F-measure 82 to 94% for 0 to 100% openness.
R. Yoshihashi et al. [156]	Classification-Reconstruction learning for Open-Set Recognition (CROSR)	MNIST, CIFAR-10, SVHN, TinyImageNet, and DBpedia	F-Measure 41 to 79% for the threshold value o.1 to 0.9 (Maximum With MNIST) Micro F1-score 82.7% (Maximum with CIFAR-10)
P. Oza and V. M. Patel [100]	Multi-task Learning Based Open-Set Recognition (MLOSR)	MNIST, SVHN, CIFAR10, CIFAR+10, CIFAR+50, COIL-100, and TinyImageNet	F-measure 82 to 90% for 0 to 49%
Y. Gao et al. [35]	Stream Classifier with Integral Similarity Metrics (SIM)	Image Datasets: Fashion MNIST, MNIST, EMNIST CIFAR-10 Text Dataset: NEW YORK TIMES, GUARDIAN	Image Dataset: Accuracy = 96.94% Label Ratio = 100% Effectiveness = 96.94% Mnew = 61.3% Fnew = 47.1% Text Dataset: Accuracy = 57.95% Label Ratio =96.0% Effectiveness = 57.95% M_{new} = 62.14% F_{new} =59.0%
M. Hassen and P.K. Chan [47]	Neural-network	MNIST, MS Challenge, and Android Genom	AUC 95.88% for 100%FPR and 8.30% for 10% FPR (Maximum with MNIST)
D. Miller et al. [89]	Class Anchor Clustering (CAC)	MNIST, SVHN, CIFAR10, CIFAR+10/+50, and TinyImageNet	Area Under the ROC Curve (AUROC) 99.1% (Maximum with MNIST)
J. Willes et al. [144]	few-shot learning for open-world recognition (FLOWR).	Mini ImageNet and Tiered-ImageNet (Both are subset of ILSVRC-12)	Accuracy 51.64% , Support-accuracy 57.76% and Incremental-Accuracy 39.39% (Maximum with Mini ImageNet) H-Measure 19.06% (Maximum with TieredImageNet)
S. Kong and D. Ramanan [60]	Open Generative adversarial networks (OpenGAN)	CIFAR, SVHN, MNIST, and Cityscapes	AUC 98.0% (Maximum with CIFAR) and F1-score 58.7% (Maximum with Cityscapes)
Z.-F. Wu et al. [147]	Noisy Graph Cleaning (NGC)	CIFAR-100, TinyImageNet, and Places-365	Accuracy 94.18% (Maximum with Places-365) AUROC 94.31% (Maximum with CIFAR)

will help to scale real-time data handling capacity with multi-class objects, and they can be from different domains. Realistic results can be achieved if the system can deal with both empirical and open-space risks. The use of past knowledge to recognize unseen objects in a dynamic environment will increase the accuracy of the system and provide more realistic results. Thus the knowledge base must be updated incrementally. The following challenges we observed in OWML for CV-IP tasks.

Table 5. Available Software Packages and Implementations

Author	Model	Link
W. J. Scheirer et al. [113]	1-vs-Set	https://github.com/Vastlab/liblinear.git
A. Bendale and T. Boult [6]	NNO	http://vast.uccs.edu/OpenWorld
P. Oza and V. M. Patel [99]	C2AE	https://github.com/dhruvramani/C2AE-Multilabel-Classification
R. Yoshihashi et al. [156]	CROSR	https://nae-lab.org/~rei/research/crosr/
C.-C. Chang et al. [15]	W-SVM, PI-SVM	https://github.com/ljain2/libsvm-openset.

- Open-space and empirical risk parameters are not optimized. Therefore, many models cannot adapt generalization or specialization.
- Most of the recommended methods have used limited training sampling; hence, the real-world impacts can not be determined accurately.
- Most of the recommended methods have not been employed with localization; hence, it is insufficient to address images with multiple objects.
- There is an absence of a mechanism for the minimization of open-space risk. The learning can be improved by employing a dictionary learning-based algorithm for open-set recognition.

5 Open-world Machine Learning in Natural Language Processing (NLP)

Over the years, there has been enormous content generated on the web in the form of text. Social media is where billions of users create most of the text that can influence human beings and social sentiments in terms of thoughts, stories, expression, news, and daily life events. Social media is a crucial part of the current environment in terms of social and political perspectives. It can influence billions of people of the world positively or negatively by injecting synthetic views that can be already part of any plan. Therefore, analysis of social media content is vital to guide the world in a positive direction. Some work has been done on text data to analyze the text in different ways. open-world machine learning can help us learn about the text in a dynamic environment. Text classifications and analysis of data is the utmost imperative entity for any organization. Standard text classification includes sentiment analysis, spam filtering, movie genre reviews, and document classification. The classification of tasks and work done towards these tasks are shown in Table 6. Further, These tasks are discussed in Section 5.1, 5.2, 5.3, and 5.4.

Table 6. Summarized Study of Task Performed by Open-world Machine Learning in Neutral Language Processing

Author(s)	IL	DUI	KB&KG	DNC
G. Fei and B. Liu [32]	-	√	-	-
L. Shu et al. [123]	-	√	-	-
S. Prakhy et al. [105]	-	√	-	-
X. Guo et al. [43]	-	√	-	-
T. Doan and J. Kalita [30]	-	√	-	-
B. Shi and T. Weninge [121]	-	-	√	-
S. Mazumde et al. [82]	-	-	√	-
T.-E. Lin and H. Xu [76]	-	√	-	-
H. Xu et al. [151]	-	√	-	-
N. Vedul et al. [140]	-	√	-	-
T.-E. Lin and H. Xu [77]	-	√	-	-
G. Fei et al. [33]	√	-	-	-
N. Vedula et al. [139]	-	-	-	√
Q. Wu et al. [146]	-	√	-	-
Y. Wang et al. [143]	-	-	√	-

5.1 Incremental Learning (IL)

Incremental learning is a Machine Learning (ML) method concerns expanding artificially intelligent systems that can continue to learn new tasks from novel input while retaining previously gained knowledge. Whenever a novel task(s) appears and changes, the training method occurs. The model keeps whatever has been learned according to the novel task(s) and old knowledge. The most notable distinction of incremental learning from conventional machine learning is that it does not lose previous knowledge. However, the training samples resemble it over time.

In [32], authors proposed Center-based Similarity (CBS) method for open-world text recognition. It is a space learning method that can reduce open-space risk. The CBS is based on SVM. Center-based similarity space learning transforms each document space vector or feature vector, each feature in the center of the positive class document, and the feature vector of the document. At the same time, traditional classification directly uses training examples for trained binary text classifiers. CBS can learn multiple documents features vectors, separate for each document, and represents the center for multiple positive documents. Similarity value can be computed using multiple document similarity functions. The performance evaluations have been done on two publicly available datasets, 20-Newsgroup [65, 109], and amazon customer reviews [93].

In [33], authors extend their work and given a better system that can practice incremental learning in which the system can learn cumulatively. Whenever the system learned about new classes /unseen classes became more knowledgeable, just like humans do. They proposed a system with two specific abilities, continually detecting unknown classes and cumulatively adding the data of these new classes to the knowledge base without re-train the whole system. The proposed Center-based Similarity Space Learning SVM (CBS-SVM) was evaluated with two different datasets Amazon product reviews of 100 domains and 20-newsgroup [65, 109]. Classifying classes in the open-world uses the same unseen class rejection method based on threshold probabilities. The system used a similarity method to learn unseen/ new classes. It searched for sets of similar classes and learned to separate new classes. To learn a separate new class, it builds a binary classifier. After detecting or specifying a new class, updates the existing classifier to avoid confusion for the next unseen classes. The proposed method was evaluated by comparing the result with existing 1-vs-rest-SVM, 1-vs-set-linear, WSVM-linear, WSVM-RBF, PI-SVM-linear, PI-SVM-RBF, ExploratoryEM, CBS-SVM performed better than all existing methods with all different openness.

5.2 Discovery of Unseen Instances (DUI)

open-world machine learning has the significant importance of rejection of unseen classes; the accuracy of prediction of the known class must be justifiable. In [123], authors proposed Deep Open Classification (DOC) to identify new classes or tasks which may not belong to any training class. The ideal classifier should document both for which training class is available and the document for which training class is not available. This method is called open-world classification or open classification.

Giving the training data set $D = \{(A_1, B_1), (B_i, A_2), \ldots, (A_n, B_n)\}$ Where A_i, is i^{th} document and $B_i = \{l_1, l_2, \ldots l_m\} = B$ is A_i class label. They build classifier $f(x)$ it can classify test instances A such that, A belongs to the training class m as a seen class in B or discard it that means it is unseen class, and test instances do not belong to any of m training class or any other seen class. DOC will build a multi-class classifier with the 1-vs-rest final layer of sigmoid in place of OpenMax [115] to reduce open-space risk. DOC used the sigmoid function with Gaussian fitting to lighten the decision boundaries and reduces open-space risk. DOC used a Convolutional Neural Network (CNN) with a 1-vs-rest sigmoid layer and Gaussian fitting for classification. DOC Chose CNN because OpenMax uses CNN, and CNN performs well on the text. Doc has three layers for a different task. Layer

1: Embedded word (word vectors pre-trained from Google News that is Word2Vec) [87, 88] in x document into a dense vector. Layer 2: Perform convolution on layer 1 with the different filters with a variety of sizes. Layer3: A pooling layer selects a maximum value from the result of layer-2 and forms a K-dimension. To evaluate the performance experiment has been done on two datasets, 20 Newsgroups [109] and 50-class reviews [17]and compare results with CBS-SVM [32] and OpenMax. To extract the features, they convert the document into vectors using word2Vec [87, 88] method. They used pre-trained from Google News vector [87] that consist of three million words and 300 dimensions word for word to vectors.

In [30], the authors proposed the Nearest Centroid Class (NCC) to detect unseen classes in open-world machine learning . It is an incremental learning method, which can take sets of closest neighbors of the centroid class. There are clusters for classes, and in a cluster, each class has minimum points. These are the membership points that are associated with clusters. Each class must have a minimum membership point to join the particular cluster. The class also represents the data point, and the center of the class is the data points. New classes that have the nearest class center data point allow joining the cluster. To evaluate the performance of the algorithm experiment done on 20-newsgroups and amazon reviews datasets with different numbers of domains. The prior algorithm performed better for some of the parameters for both datasets, but NCC's overall performance is significantly better.

ChatBots can work in a dynamic open-world environment, but it is vital to recognize the user's intention. Intent classification is a technique to distinguish the perseverance or intention by estimating the text language. It refers to an intent classification or intent identification. Nowadays, many institutions use text-based chat systems to solve their customers' queries without any human interactions. ChatBots must understand the unknown intentions of the user to work as a human being.

In [105], the authors proposed another CNN-based approach. It is based on feature extraction. To extract the features, they convert the document into a vector using word2Vec [87, 88] method. To calculate the document vector, they used naïve methodology and estimated the cosine similarity among the mean of the document of the entire document vector. Deep learning models are used for open text classification with a modified Weibull layer as the final layer instead of the traditional SoftMax layer [39]. It is single-layer architecture, but the experiment has been done with different no of layers. To evaluate the performance of the proposed model, evaluation of proposed technique has been done on 20-Newsgroups [109] and Amazon product reviews dataset and compared the results with existing methods.

In [76], the authors proposed two-stage methods for detection of unknown intent in the dialog system. To extract the feature of unknown intent, it uses Bidirectional Long Short-term Memory (BiLSTM) network using a margin loss. The LSTM network minimizes the variances of intra-class and maximizes the variances of inter-class intents. Glove word embedding used to create vectors and to distinguish the unknown intent local outlier factor LOF [12] has been used. The loss layer detects the known intents from deep discriminative features, while LOF detects unknown intents. The SNIPS [22] and ATIS [138] dataset has been used to evaluate the result of the proposed method. The performance of the method is compared with Maximum Softmax Probability (MSP) [49], DOC [123], DOC SoftMax, and LOF SoftMax.

In [43] authors proposed a Deep convolutional neural network (DCNN) which is cascade architecture that can continue to learn newer classes. The framework is an end-to-end Open-world Recognition (OWR). To detect the instances from unknown classes, they proposed Multi-stage Deep Classifier Cascades (MDCC). It contains unique features for known classes and can distinguish the class as a known class at any stage of the process. Incremented leaf nodes can detect features of unknown classes and recognize newly added classes. It can learn new features of recently added classes without wounding existing features of known classes. The evolution of MDCC was done on the RF signal and Twitter dataset [43, 106]. The experimental outcomes are compared with Local Novel Detector (LOD) [9], S-Forest [92] and R-OpenMax [91].

The e-commerce industry is growing and has become a significant part of the world economy. Product classification is one of the most important aspects of any e-commerce organization. The unpredicted or unknown search about the product is critical for these industries as different categories of products appear every day. The queries which are not predefined or known for the system can affect the reliability of the entire organization. In [151], the authors proposed open-world learning (OWL) model Learning to Accept Classes (L2AC), which is based on meta-learning. L2AC maintains only dynamic known classes that allow novel classes to be added without retrained the model. In L2AC, each known class acts as a small set of the training example. The testing uses only Meta-classifier (using known and novel classes). The L2AC model has two primary mechanisms, ranker and meta-classifier. The ranker retrieves examples from known classes that are comparable or nearest to test examples. The meta-classifier is the core mechanism of L2AC, and it is a binary classifier that distinguishes the classes as known based on probability score or rejects otherwise. To evaluate the performance of L2AC outcomes compared with a different variant of DOC [123] on the Amazon data set, the L2AC shows effectiveness for some parameters.

In [140], the authors proposed a model Towards Open Intent Discovery (TOP-ID) for open intent detection. It is a two-phase mechanism that predicts the intent for the statement and then tags the intent in the input statement. The model consists of a BiLSTM [116] and Conditional Random Field (CRF) with the adversarial training method, and it increases robustness and performance through the domain. TOP-ID can detect a user's intent automatically in natural language. It does not need any prior knowledge for intent detection. The first part of TOP-ID detects existing open intent and then tags it into input words with action and objective. If there is no objective and action associated with detected intent, then it is tagged as none. To perform this task initially, convert the text into feature sequence by assembling character level representation, obtained by using a CNN with Glove word embedding [103]. To avoid combined word embedding effect on accuracy, TOP-ID used Highway Network [129]. The second module of TOP-ID is the intent discovery framework. It takes adversarial inputs (close to the original) created by adding noise in data in the form of perturbations. The overall training has been done with both original and adversarial inputs. The attention mechanism is part of the intent discovery framework. There are multiple attention functions used that attend the information of the input sequences at different positions. Finally, the CRF predicts one of the three tags for the sequence of the words. To evaluate the TOP-ID, they create a dataset by collecting 75K questions with correct answers then annotating 25K quotations data (three tags action, object, and none) with the amazon technique. The F1- Score of TOP-ID is significantly better than existing methods.

In [77], the authors proposed a Softmax and Deep Novelty (SMDN) detection model to detect unknown intents. The SMDN classifiers can be functional on any model without altering the architecture of the existing model. The model uses SoftMax that classifies by calculating the calibrated confidence score, and detects unknown intent by calculating decision boundary. The LOF [12] is used as an output layer to detect the unknown intent. To evaluate the performance outputs are compared with different variant of DOC [123] on three SNIPS [22], ATIS [138] and SwDA [55, 122, 131] benchmark datasets.

In [146], the authors propose a Inductive Collaborative Filtering (IDCF) system, which provides inductive learning for user inputs while also ensuring sufficient expressiveness and adaptability. The IDCF uses two representation models to extract user-specific embeddings, that term meta latents. It factorizes a set of essential users' data matrices, followed by an attention technique that learns concealed graphs among essential users and queries users based on their past ranking habits.

For query users, the inductive calculation of user-specific representations is enabled by the revealed associated graphs. IDCF standard version can decrease restoration loss to a similar level as vanilla matrix factorization technique under a slight circumstance. Empirically, IDCF offers actual close Root Mean Square Error (RMSE) to transductive Collaborative Filtering (CF) models. IDCF achieves improved results over the few-shot and unseen users compared to several inductive models on explicit feedback data Movielens-100K, Movielens-1M [46], and Douban [153], and implicit feedback data Amazon-Beauty, and AmazonBooks [48, 83].

5.3 Knowledge-graph & Knowledge-base (KG & KB)

The knowledge graph (KG) is one of the key methodologies for the online and offline world. The KG is helping in many important tasks such as Web search, entity linking, language processing, recommendation, and prediction. This method is also worked under the close-world assumptions as nodes are predefined. Very few research is available for open-world machine learning through the graph completion method. The relation and triples [157] are key components for knowledge graph completion methods. In [121] authors proposed a ConMask, which is a model for Knowledge Graph Completion (KGC) in the open-world. This model learns embedding of any entity by its name, description given in text fields and identifies unknown classes of entities to the knowledge graph. ConMask used relation depending content masking to extract relevant chunks and reduce the noisy text description. After extracting relevant chunks, train the model with fully connected CNN to concur chunks with entities in a knowledge graph. Knowledge graphs can be representing as (H_d, R_e, T_a) where H_d is head, R_e is relation between Head H_d and some tail entity T_a. The generalize incomplete graph $G = (e,\ r,\ t)$, where e = entity set , r= relation set and t= tuple set. This graph can be complete by finding missing tuples t'. $t' = \{\langle H_d, R_e, T_a \rangle | H_d \in e, R_e \in r, T_a \in e, \langle H_d, R_e, T_a \rangle \in t\}$ in the incomplete KG. The model consists of three modules, relationship content masking, target fusion, and target entity resolution. The first module indicates the words relevant to the task. The second module extracts target entities embedding. The last module picked the target entities based on the similarity score between target entities. The last module of the model furthermore extracts entity embedding and textual features.

As the world is going towards automation, ChatBots systems are becoming popular for customer common query solutions. Every system that takes input as a text to elucidate or respond to a User's inquiry needs to understand the query. Existing systems are working with a limited environment, which means they can answer the only query for data available in the Knowledge Base (KBs). Such systems have limitations; they cannot work in a dynamic environment because they cannot learn new knowledge. Although KBs have extensive data from appropriate sources still much information has missing from them. Many techniques have been proposed till now to complete this missing information. These methods are termed KB completion, but all the methods worked under closed-world assumptions. KBs has a limitation, and it cannot work in an open-world environment.

To address this kind of issue, in [82], the authors proposed Open-world Knowledge Base Completion (OKBC) and Lifelong Interactive Learning and Inference (LILI) technique for ChatBots to acquire knowledge in the dialogue process. This knowledge learning engine allows ChatBots to gain knowledge throughout the conversation and make it further interactive. It is based on the theory of continual learning, where ChatBots become more knowledgeable with time as they learn continually after every conversation. Lifelong interactive learning and inference analyze the query and add it to KB if it does not exist. LILI, formulate an inference strategy, learn interaction behaviors, leverage the acquired knowledge, and continuously repeated this to learn new knowledge. The evolution of LILI has been done on two benchmark datasets Freebase FB15k [11] and WordNet [90] and compare results for known, unknown, and overall classes. The result of LILI is effective for both predictive eminence and strategy formulation capability.

In [143], The authors suggest a capsule network-based approach Caps- OWKG that leverages context to describe relationships and objects in the open-world knowledge graph. The proposed Caps-OWKG consists of triplets that are the basic unit of the system. In addition, the capsule network conducts extraction of features, judgment on triplets, text synthesis, and fusion analysis. When computing triplets, the Caps-OWKG technique has the benefit of providing a stronger connection between items and relationships. These interpretations are also refined; thus, the Caps-OWKG model may be considered a dynamic embedding exploration that accurately represents the triplet. The existing known techniques such as ConMask [121], TransE-OWE [120], and DKRL [150] are used to compare the performance of Caps-OWKG on the two benchmark datasets FB15k-237 [137] and DBPedia50k [120], achieving better outcomes than existing techniques.

5.4 Detection of Novel Classes (DNC)

The existing research finds only new intent in the available domain, but the novel domain is introduced incrementally as data increases. The novel domain must be found to make the system fully automated and reduce the limitations of the system. In [139] authors proposed Automatic Discovery of Novel Intents and domains (ADVIN) to discover novel domains and intents of text from unlabeled data. ADVIN works in three stages: discovering the novel domains and intent from extensive unlabeled data, knowledge transfer, and linking related intents to corresponding novel domains. To identify the instances of novel intents ADVIN, used BERT [29] based multi-class classifiers. The DOC [123] is used for distinguishing unseen intents. In the second stage that discovers the categories of newly discovered intents, it uses a hierarchical clustering method to transfer knowledge. Finally, by linking novel intents into novel domains, ADVIN used clusters of seen classes as ideal clusters and knowledge transfer modules to represent clusters. To evaluate the performance of ADVIN results are compared with DOC, IntentCapsNet [157], LOF-LMCL [30] and different combinations of ADVIN and DOC on four benchmark datasets, SNIPS [22], ATIS [138], Facebooks' Task-oriented Semantic Parsing (FTOP) [45] and dataset from a commercial voice assistant, Internal NLU. The overall performance of ADVIN is significantly better.

Table 7 shows a summarized illustration of literature on open-world machine learning in Natural language processing. It shows the used or recommended methodology, datasets employed for evaluation, and proposed results by the authors.

Table 7. Summary of the Proposed Approaches for Natural Language Processing in open-world machine learning

Author	**Proposed or Used Methodology**	**Datastes**	**Reported Results**
G. Fei and B. Liu [32]	Center-based Similarity support vector machine (CBS-SV)	20-newsgroup and Amazon reviews	Accuracy 45 to 87.3% for 25% to 100% openness (Maximum with Amazon reviews 10 Domains)
G. Fei et al. [33]	Cumulative Learning (usine CBS-SVM)	Amazon product reviews and 20-newsgroup	Micro F1-score 66.2 to 83.5% for openness of 33% to 100% (Maximum with 20-newsgroup)
L. Shu et al. [123]	Deep Open Classification (DOC)	20 Newsgroups and Amazon reviews (50-class reviews)	Micro F1-score 82.3 to 92.6% for 25% to 100% openness (Maximum with 20 Newsgroups)
S. Prakhy et al. [105]	Convolutional Neural Networks (CNNs)	20-newsgroup and Amazon reviews	F1-score 79.7 to 82.1% for 25% to 100% openness (Maximum with Amazon reviews 10 Domains)
T. Doan and J. Kalita [30]	Nearest Centroid Class (NCC)	20 newsgroups and Amazon reviews	Accuracy 20 to 82% for 0 to 50 domains (with Amazon revies)
X. Guo et al. [43]	Multi-stage Deep Classifier Cascades (MDCC)	RF signal Datasets,Twitter dataset,	EN-Accuracy 60.45% (Maximum with RF Signal) F1-score 75% (Maximum with RF Signal)

B. Shi and T. Weninge [121]	Content Masking (ConMask)	DBPedia50k and DBPedia500k	Mean Rank 90 and Mean Reciprocal Rank 35.0 (for head) Mean Rank 16 and Mean Reciprocal Rank 61.0 (for trail) both are maximum with DBPedia50k
S. Mazumde et al. [82]	lifelong inter- active learning and inference (LILI)	Freebase (FB15k1) and WordNet	Avg. F1-score 63.43% (Maximum with FB15k1) and Avg. MCC 39.39% (Maximum with WordNet)
T.-E. Lin and H. Xu [76]	Bidirectional long short-term memory (BiLSTM)	SNIPS and ATIS	F1-score 78.8 to 79.2% for 25& to 75% openness (Maximum with SNIPS)
H. Xu et al. [151]	Learning to Accept Classes (L2AC)	Amazon Datasets	Micro F1-score 84.68 to 93.19 % for 25% to 75% openness
N. Vedul et al. [140]	Towards Open Intent Discovery (TOP-ID)	25k real-life utterances (Created dataset)	F1-score 91% (Maximum among all the versions of TOP-ID)
T.-E. Lin and H. Xu [77]	SofterMax and deep novelty detection (SMDN)	SNIPS, ATIS, and SwDA	Macro F1-score 71.1 to 79.8% for 25% to 75% openness (Maximum with SNIPS)
N. Vedula et al. [139]	Automatic Discovery of Novel Intents (ADVIN)	SNIPS, ATIS, FTOP, and Internal NLU Dataset	F1-score 92% (for discovery of unseen instances) NMI 83% Purity 92% F1-score 78.0% (for discovery of unseen classes)
Q. Wu et al. [146]	Inductive Collaborative Filtering (IDCF) model	Douban, Movielens-100K, Movielens-1M, Amazon-Books, and Amazon-Beauty	AUC 94.4% (Maximum with Amazon-Books) and Normalized discounted cumulative gain (NDGC) 95.5% (Maximum with Douban)
Y. Wang et al. [143]	capsule-network for open-world knowledge graph (Caps-OWKG)	DBPedia50k and FB15k-237-OWE	Tail prediction 64.8% (Maximum with DBPedia50k)

5.5 Available Software Packages and Implementations

In this section, we provided a link for some software packages which contains the various implementation of various model of open-world machine learning in natural language processing (Table 8). These are the models commonly used in various frameworks of open-world machine learning .

The available software packages can be used to further improve learning in the open-world for natural language processing. The Open-set Deep Networks (OSDN) can be used for open-set reorganization, Deep Open Classification (DOC) for unseen class identification, and Content Masking (ConMask) for identification of unseen entities in the knowledge graph. There are two packages word to vector (Word2Vec) and Global Vectors (Glove), for input word embedding.

Table 8. Available Software Packages and Implementation

Author	Model	Link
P. Moore and H. Van Pham [91]	OSDN	https://github.com/abhijitbendale/OSDN
L. Shu et al. [123]	DOC	https://github.com/leishu02/EMNLP2017_DOC
B. Shi and T. Weninger [121]	ConMask	https://github.com/bxshi/ConMask.
T. Mikolov et al. [88]	Word2Vec	https://code.google.com/archive/p/word2vec
J. Pennington et al. [103]	Glove	https://nlp.stanford.edu/projects/glove/
X. Guo et al. [43]	MDCC	https://github.com/xguo7/MDCC-for-open-world-recognition
Y. Kim et al. [59]	CNN Text Classification	https://github.com/dennybritz/cnn-text-classification-tf
Y. Kim et al. [59]	CNN Sentence Classification	https://github.com/alexander-rakhlin/CNN-for-Sentence-Classification-in-Keras

5.6 Discussion

There are very few works that have been done in open-world machine learning for natural language processing. The semantic similarity in the text is hard to address while learning new knowledge about the text at run time, especially when there is no training set available for such data. To achieve valuable outcomes from any framework or an algorithm, it must distinguish the semantic similarities in text. Therefore there is a need for large scale knowledge-base to learn the hierarchical structure of text words with meanings. Openness is a significant issue as we have analyzed that the accuracy has been reduced whenever the openness increased. There is a need for frameworks that can deal with dynamic values of openness and provide high accuracy with maximum openness. The automated dialog-based system that is quite popular nowadays needs a mechanism to process informal conversation in real-time.

The automated ChatBot systems and text and voice-based assistance devices are increasing rapidly in this decade, and it further improves the world of automation. To increase the accuracy of such a system, the open-space risk and distinguish the semantic similarities is one of the significant aspects of neural language processing in open-world settings. The cumulative and incremental model can help to address such issues. The system will produce more realistic outputs when it adapts scalability in input with maximum openness as the real-world inputs are unstructured. In this section, we have discussed various problems associated with NLP in open-worlds learning and proposed solutions by various authors that can help improve a text-based application working in a real-world domain and dynamic environment. The following challenges we observed in OWML for NLP tasks.

- Stable performance can be achieved to identify unseen instances only if the threshold value is within a reasonable range.
- The recommended methods show superior outcomes for sample instances. The accuracy of many of the proposed systems decreases if the number of seen classes is low.
- There is a need for improvements to use these prototypes for real-time systems, as extensive experiments with large-scale datasets are missing.
- Only a few methods are employed with cumulative learning.
- In NLP, many of the recommended models suffer when distinguishing unseen intent from seen intents where semantic meanings are similar.

6 Datasets Used in Open-world Machine Learning

Most of the researchers employed benchmark datasets to evaluate the performance of their proposed algorithms. Some of the researchers built their datasets or altered the existing dataset and evaluated the methods. In this section, we discuss some of the datasets primarily used in both the domain of open-world machine learning.

Figure 6 shows the classification of datasets for CV-IP and NLP with their proposed years. Next we discussed publicly available datasets that are used in OWML.

Caltech-256 [42]: $caltech - 256$ has set of 256 categories of object and the total 30607 images in this dataset. Each category contains minimum 80 and maximum 827 images, these categories are further labeled with three tags on the basis of image quality. the labels are *good*, *bad* and *none* (out of the category). the *good* indicates clear vision and *bad* indicates clutters or artistic example where *none* indicates the image does not belong to the particular category.

MNIST [69, 70]: Modified National Institute of Standard and Technology, wildly known as $MNIST$ is a handwriting dataset. It is a modified version of NIST. MNIST is used in optical character reorganization and is also used as a test case in pattern recognition and machine learning. We have analyzed, MNIST has become a standard for testing machine learning algorithms. There are 60000 training images; some may use for validation and 10000 images for testing purposes. All the digits are black and white and normalized in seize, the center intensity with $28 * 28$ pixels; thus, the dimension of the image is $28 * 28 = 784$, and each element is a binary. The MNIST has tested for almost all the benchmark baseline algorithms and well-known Fields of classification such as

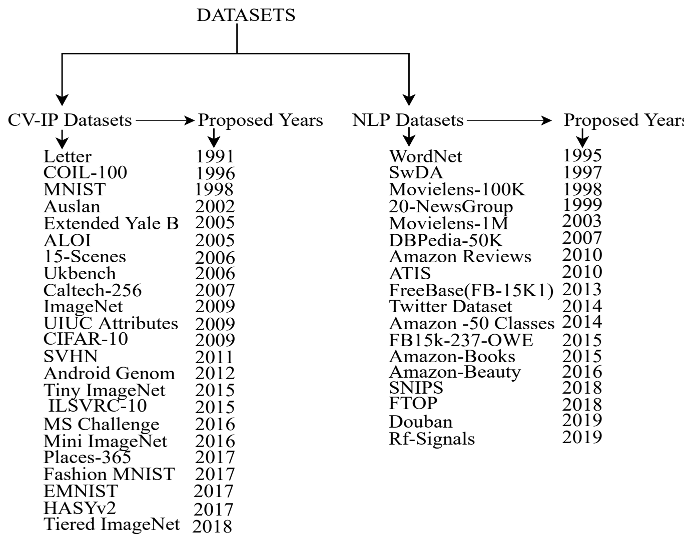

Fig. 6. Classification of Datasets Used in CV-IP and NLP with Proposed Years

linear classification, convolution neural networks, simple neural networks, K-Nearest neighbors, support vector machines (SVMs), boosted stamps, and nonlinear classification.

Fashion-MNIST [149]: The Fashion-MNIST is a dataset of Zalando's article images which containing a training set of 60,000 images, and a test set of 10,000 images. Each image is a 28x28 pixel and grayscale image related to a label from 10 different classes. Zalando aims for Fashion-MNIST to serve as a substitution for the original MNIST dataset, which comprises many handwritten digits, for benchmarking artificial intelligence and machine learning algorithms.

ImageNet [28]: $ImageNet$ is a large scale ontological dataset of visual objects. The structure of $ImageNet$ inspired from $WordNet$ dataset thus it constructs on backbone of $WordNet$. $ImageNet$ has 80000 synsets of $WordNet$ with around 1000 full resolution cleaned images and its updating continuously. The basic $ImageNet$ contains 3.2 million images with 12 sub-trees and $5247 synsets$. $ImageNet$ is hierarchical dataset like $WordNet$ which contains the synonym's of world in tree structure. The 12 sub-trees consist of the following categories: bird, reptile, vehicles, musical instruments, tools, fruits, mammals, fish, amphibians, geological formulations, furniture, and flower with 5247 synsets. The continual updating of this data aimed at 50 million images in a hierarchical structure. The evolution with various baseline methods showed that the $ImageNet$ has a 99.7 percent average precision rate.

Tiny-ImageNet [68]: $Tiny - ImageNet$ is a collection of 100000 images that are retrieved from internet. The resolution of all these images is $32 * 32$ pixels and $64 * 64$. Tiny-ImageNet has 200 categories of images, of which $100,000$ images for training, 10000 for validation, and 10000 images are reserved for testing. Images are collected by sending all search words in $WordNet$ to the image in the search engine. it is a successful dataset tested on application-specific algorithms because of a high level of noise and low resolution. $Tiny - ImageNet$ is suitable for general-purpose algorithms. It also contains synsets of high quality with an average resolution of $400 * 350$.

CIFAR-10 [62]: The CIFAR-10 data set was developed by the Canadian Institute for Advanced Research. It contains 10 categories (dog, frog, automobile, bird, horse, ship, truck, airplane, cat, and deer) of images, as it is a subset of CIFAR-100, which consists of 100 categories of images. Total

60000 color images are in CIFAR-10 with the resolutions of $32 * 32$ pixels, and every class has 600 images. The dataset is divided into training and test sets, which consist of 50000 and 10000 images, respectively. The entire dataset is divided into batches, 5 batches for training and 1 for testing. The testing batch has 1000 random images, and the rest of the images randomly contain by training batches.

SVHN [96]: The Street View House Number (SVHN) contains 600000 labeled digits that are cropped from actual street view images. The initial goal of this dataset is to identify house numbers from original street view images. There are two types of images one is whole numbers, and another is cropped digits. The whole numbers contain high-resolution full-size original images with character-level bounding boxes for the house number. The cropped digits are character-level ground truth, and all these digits are resized with the resolution of $32 * 32$ pixels. The SVHN is further divided for training and testing, and there are 73257 digits images for training and 26032 digits images for testing. The rest of the images also reserve as extra for training.

20-NewsGroup [65, 109]: It is a collection of near about 20000 new documents which are collected from different newsgroups. It is one of the most popular datasets for the application that is based on text classification in machine learning. All the newsgroups are different, but some of the new groups are related to each other. Generally, 90 percent of documents are used for training and 10 percent for testing. 20-NewsGroup is publicly available in different forms, and the original dataset is not sorted but later on is sorted by date. The headers and duplicate data are also removed in this version. The latest version of this data is available with 18828 documents with only "From" and "Subject" headers.

Amazon Product Reviews [93]: Amazon.com is one of the most successful e-commerce web site across the globe since it emerges. The amazon product review dataset contains 5.8 million reviews, written by 2.14 million for 6.7 million products from 9600 different categories when extracted from these reviews. The dataset has 8 different headers such as Product ID, Reviewer ID, Rating, Date, Review Title, Review Body, Number of Helpful Feedbacks, and Number of Feedbacks. It can be used for feature identification and construction of both reviewer and reviews, and the features can be Review Centric or Product-Centric. The amazon product reviews.

50-Class Reviews [17]: The 50-Class review dataset is a collection of reviews, and there are 50 different categories of products. The data set has two versions, one has reviews of 50 different electronic items, and the other has 50 different non-electronic items. There are 1000 reviews for each product or domain.

WordNet [90]: WordNet is a multi-language (Approx 200 languages) lexical dataset of semantic relationships among words, including meronyms, synonyms, and hyponyms. Some synsets contain synonyms in a group with short definitions and examples. The WordNet is a popular dataset for text analysis applications in artificial intelligence and machine learning. Initially, it has created for the English language only; later on, it extended for other languages, and updating is continuous to add a new language in WordNet. The WordNet contains approx 175979 words which are organized in 175979 synsets, and there is a 207016 pairs which are word-sense pair. All synsets are connected with semantic relations.

SwDa [55, 122, 131]: The Switchboard Dialog Act Corpus (SwDA) covers the SwDA-1 Telephone Speech Corpus, and some tags recapitulate semantic, syntactic, and pragmatic information about the related turn.

Some publicly available datasets with their repository link are shown in Table 9.

7 Baseline Algorithms Used in Open-world Machine Learning

Some methods and algorithms are used in open-world machine learning that are standard or base concepts to practice open-world learning.

Table 9. Publicly Available Benchmark Datasets Repositories

Dataset	Link
Caltech-256 [42]	http://www.vision.caltech.edu/Image_Datasets/Caltech256
MNIST [69, 70]	http://yann.lecun.com/exdb/mnist
Extended Yale B [71]	http://vision.ucsd.edu/ leekc/ExtYaleDatabase/ExtYaleB.html
ALOI [38]	https://aloi.science.uva.nl
UIUC Attributes [31]	https://vision.cs.uiuc.edu/attributes
Mini ImageNet [141]	https://cseweb.ucsd.edu/ weijian/static/datasets/mini-ImageNet
Fashion-MNIST [149]	https://github.com/zalandoresearch/fashion-mnist/tree/master/data
HASYv2 [133]	https://https://zenodo.org/record/259444/files/HASYv2.tar.bz2?download=1
ImageNet [28]	http://image-net.org/download
Tiny-ImageNet [68]	http://cs231n.stanford.edu/tiny-imagenet-200.zip
CIFAR-10 [62]	https://www.cs.toronto.edu/~kriz/cifar.html
RF Signal Dataset [43]	https://github.com/xguo7/MDCC-for-open-world-recognition
Twitter Dataset [106]	https://github.com/xguo7/MDCC-for-open-world-recognition
SVHN [96]	http://ufldl.stanford.edu/housenumbers
20-NewsGroup [65, 109]	http://qwone.com/~jason/20Newsgroups
Amazon Product Reviews [93]	https://jmcauley.ucsd.edu/data/amazon
WordNet [90]	https://wordnet.princeton.edu/download
SwDa [55, 122, 131]	https://web.stanford.edu/~jurafsky/ws97/
ATIS [138]	https://rasa.com/docs/rasa/nlu-training-data/#json-format
FB15k [11]	https://www.microsoft.com/en-us/download/ confirmation.aspx?id=52312
DBpedia [4]	https://wiki.dbpedia.org/datasets
EMNIST [20]	https://www.nist.gov/itl/products-and-services/emnist-dataset
Auslan [56]	https://archive.ics.uci.edu/ml/datasets/Australian +Sign+Language+signs+(High+Quality)
Ukbench [98]	https://archive.org/download/ukbench
Places-2 [160]	http://places2.csail.mit.edu/download.html

7.1 Center-Based Similarity (CBS) [32]

CBS is a classification method that classifies the data points into seen and unseen classes. It works on center-based similarity space learning technique. The CBS learns news classes incrementally and use 1-vs-rest layer to classify unseen classes [32]. The 1-vs-rest is one of key concept in open-world machine learning to discover unseen classes.
Let us assume there is new class l_{X+1}, for learning it need a model M_X. Model M_X consist set of X 1-vs-rest binary classifiers $M_X = (m_1, m_2, ...m_X)$, for the previous X classes there is training dataset $D^{p_r} = (D_1, D_2, ...D_X)$ (p_r = previous) and corresponding labels are $S^X = (l_1, l_2, ...l_X)$. Here each M_i builds a binary classifier to identify l_i, when new dataset D_{X+1} arrives for class l_{X+1}. the entire system functions for two task, to update M_X and build new M_{X+1} model to classifies all available instances in existing class $S^{X+1} = (l_1, l_2, ...l_X, l_{X+1})$ and recognize the U_s unseen classes.

Step 1:search a *set of classes* S_c that are comparable to new class l_{X+1},
Step 2: learning for Isolate the new class l_{X+1} and the previous classes in S_c.

Step 3: $M_X = (m_1, m_2, ...m_X)$ to classify instances in D_{X+1} , the similarity between old classes $(l_1, l_2, ...l_X)$ and new class l_{X+1} can be computed by using each of 1-vs-rest binary classifier m_i.

In next step, new class l_{X+1} separated and now for S_c there is two task,
Step 4: Build M_{X+1} new classifier for l_{X+1}.
Step 5: Update existing classifier as the classes which are in S_c.

7.2 Incremental Class Learning [32]

Incremental learning is encouraged by the thought of the human learning process. It learns most of the knowledge by an experience like humans do. It learns new knowledge by the time instead of finding existing knowledge.

Let us assume we have Classification model $M_X = (m_1, m_2, ...m_X)$ as input and $D^{p_r} = (D_1, D_2, ...D_X)$ is previous dataset. The new dataset is D_{X+1} and λ_s is Similarity Threshold. we need classification model $M_{X+1} = (m_1, m_2, ...m_X, m_{X+1})$ to learn incrementally using previous data. To obtain this the following steps need to be executed.
Step 1: Initialize S_c to empty set.
Step 2: Initialize the count and record total instances in D_{X+1} (positive classified by m_i).
Sept 3 : Use m_i and classify each instances in D_{X+1} and record total positive instances classified by m_i.
Step 4: check whether there are disproportionate instances in D_{X+1} as positive by m_i to reduce class l_i. as resemblance to class l_{X+1}. The λ_s is threshold which regulate how many instances in D_{X+1} should be classified l_i before considering as analogous to l_{X+1}.
Step 5: build novel classifier M_{X+1}.

7.3 Nearest Class Mean (NCM) [84, 111]

The nearest class mean (NCM) is generally used for large-scale image classification. Two methods were used in most Research for large-scale image classification, k-nearest neighbor (K-NN) and nearest class mean (NCM), the nearest class mean (NCM) is more flexible than K-NN. The NCM characterizes classes by their mean feature vectors of its components.

Let us assume we have image P, which is represented in D-dimension with the feature vector. $\vec{f} \in \mathbb{R}^D$.

Step 1: compute the class centroid c_a for each class $a \in A$.

$$c_a = \frac{1}{P_a} \sum_{i \in P_k} \vec{f}_i \tag{2}$$

where P_a is the set of images label with class a and the set of centroid (for each class) is $C = \{c_a\}$ and it has cardinality $|C| = |A|$

Step 2: The classifier of nearest class mean to classify an image P will search closest centroid in feature space.

$$a^*(P) = \overset{argmni}{a \in A} \left\| \vec{f} - c_a \right\|^2 \tag{3}$$

where $\vec{f}$ is the feature vector of P

7.4 1-vs-rest [110]

Classical machine learning uses different functions as output for multi-class classification. However, These functions can not reject unknown classes. There is a need to normalized these functions for each class across the training classes to achieve rejection capability in the output mechanism.

The *1-vs-rest* is one the method which provide rejection capability. Les us assume there is 1-vs-rest method with with s sigmoid functions, where s is known objects. We have i^{th} sigmoid function for p_i class. The 1-vs-rest distinguish the classes as positive and negative for p_i class such that $q = p_i$ is positive class and rest of all $q \neq p_i$ classes are negative. The loss can be calculated as log loss for all s sigmoid functions for the training data.

$$loss = \sum_{i=1}^{s} \sum_{j=1}^{n} -\mathbb{I}(q_j = p_i) log p_b(q_j = p_i) - \mathbb{I}(q_j \neq p_i)(1 - log p_b(q_j = p_i)) \tag{4}$$

Where $\mathbb{I}$ is the Indicator function, j= 1 to n (n = Number of instances) and probability output of s sigmoids for j^{th} input of i^{th} dimension of r is $p(y_j = x_i) = sigmoid(r_{j,i})$ reject unseen classes such that

$$\hat{q} = \begin{Bmatrix} reject,\ if sigmoid(t_i) < k_i,\ \forall\, p_i\ \in\ q_i \\ argmax_{p_i \in q}\, sigmoid(t_i),\ Otherwise \end{Bmatrix} \tag{5}$$

Where k_i is threshold which belongs to p_i , is probability p_b is less than the threshold the input will be reject. 1-vs-Rest predicts class which has highest probability.

These are the benchmark algorithms that have been used in both computer vision and natural language processing to elucidate problems in open-world settings. There are several frameworks and models suggested which are centered on these algorithms or used these algorithms. The Center-Based Similarity (CBS) does not support the artificial neural network; thus the some of the authors used its extension or modified versions [1, 32]. The Nearest Class Mean (NCM) has been used as the baseline in [7, 26]. The classical model of NCM considers examples of training a novel 1-vs-rest classifier for individual supplementary classes. Hence in the case of large-scale datasets and multiple classes, it becomes a burden for classifiers. Earlier trained classifiers will also need to be restructured to increase their performance; thus, the extension of NCM such as Nearest Non-Outlier (NNO) [6] implemented, which gives better accuracy. The Nearest Centroid Class (NCC) [30] also inspired by NCM. In [26, 30, 82, 124] the concept of incremental learning has been used to recognise objects in open-world.

Identification of unknown data has a significant impact on open-world machine learning . To adapt the capacity of rejection of unknown data, many authors use the 1-vs-rest [110] method. Some authors have used the 1-vs-rest method to identify unknown data [113, 123, 139] and some researchers used this method as part of their framework to distinguish the known and unknown data [124, 151]. As we have seen in literature, the baseline approaches are still generating auspicious outputs with current expertise such as Convolutional Neural Network (CNN) and Deep Neural Network (DNN) frameworks. We also perceive that the few modified versions of these algorithms reinforce the model in open-world machine learning .

8 Related Areas

Some related areas that are closely associated with open-world machine learning are mentioned and discussed briefly in this section.

8.1 Transfer Learning

Conventional Machine Learning (ML) techniques perform predictions on the expected data by applying analytical principles trained on previously accumulated unlabeled or labeled training examples [5, 64, 155]. Analysis on transfer learning has brought attention since 1995 in several titles: knowledge transfer, learning to learn, multitask learning, knowledge consolidation, inductive transfer, knowledge-based inductive bias, context-sensitive learning, cumulative learning [134, 136], and multitask learning framework [78]. Transfer learning involves interpreting data for a reference task to provide a productive basis for a new task. Transfer learning is often applied to specific data sets, which have some labeled value. For example, an actual demonstrative prototype of one virus would have a significant advantage to developing a distinguishing prototype for another virus, for which fewer training samples are available. While all learning involves generalization across all queries, transfer learning illustrates the transfer of information across comparable but non-consistent fields, tasks, and distributions. In distinction, the unlabeled data does not require to be obtained from a similar task in the transfer learning framework. In the prior decade, there has been substantial development in improving cross-task transfer utilizing both discriminative and generative strategies in a broad category of frames.

8.2 Active Learning

Active Learning [135] is a discipline of machine learning where the algorithm is designed for learning, can choose the data for learning, or learning strategy generated during learning (Figure 7). The active learning methodologies can play an essential role in domains that are dealing with real-time data such as speech recognition, information extraction [119], classification, and filtering. Moreover, active learning provides high accuracy with a small testing size of labeled data.

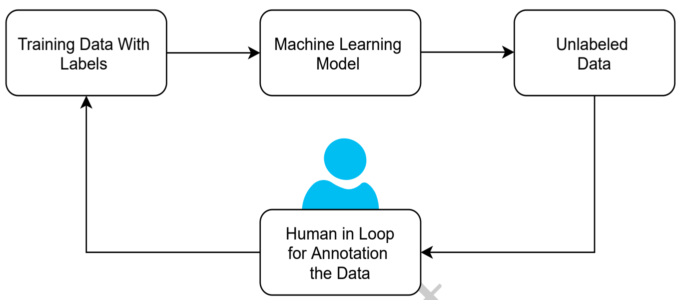

Fig. 7. Basic Framework of Active Learning [118]

There are different types of scenarios of active learning, such as membership query synthesis [2], stream-based selective sampling [21], and pool-based active learning [73]. The standard data mining methods learn models with isolated data and make a prediction based on static models [5, 64, 155]. It needs to use previous knowledge, or a learning model should transfer knowledge, and it must be used to predict future learning. It is termed as transfer learning [101]. The knowledge can be transferred in various forms such as transferring knowledge of instances [24], knowledge of feature representations [3] (for both supervised and unsupervised), knowledge of parameters [66] and relational knowledge [86].

8.3 Lifelong Learning/Continual Machine Learning

Lifelong machine learning is a system that can continuously learn from different domains, and this knowledge can be used effectively on future tasks in an efficient manner [126]. The selective knowledge is transferred when learning a novel task. Knowledge is retained from a different source and improves learning (Figure 9). The various techniques of lifelong learning as prior works in knowledge retention and improves learning a new task. The major tasks of lifelong machine learning are shown in Figure 8.

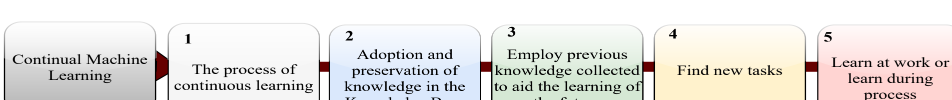

Fig. 8. Tasks of Continual Machine Learning

There are different names: constructive induction, incremental and continual learning, explanation-based learning, sequential task learning, and never-ending learning. These methods are further divided into different categories: lifelong machine learning is supervised learning, continual learning is reinforcement learning, and self-taught learning or never-ending learning is unsupervised learning.

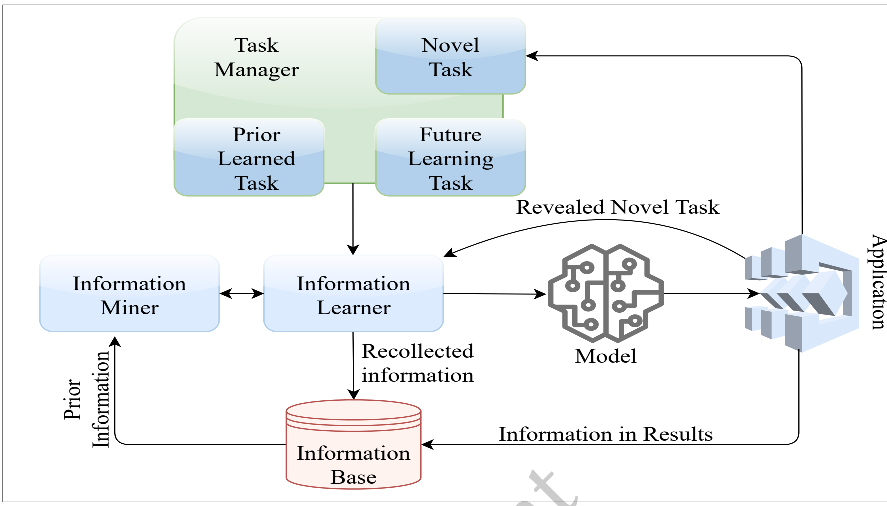

Fig. 9. Basic Framework of Lifelong Machine Learning [18]

Supervised learning lifelong learning uses Explanation-based Neural Network (EBNN) using back propagation gradient. Whenever new learning tasks occur, EBNN uses prior domain information of the task. EBNN gives more accurate results even with fewer amounts of data. In [125], authors suggested knowledge-based cascade correlation neural networks. This method uses prior trained networks and concealed units to set the new bias for a novel task. Unsupervised Lifelong learning is used to increase the system's scalability, and adaptive resonance theory had been used to map the bottom and top nodes of clusters. Set threshold to consider a new example node, a map with vigilance parameter (below threshold).

In [132], authors proposed a novel approach to ensemble clusters from the primary partition of objects; it uses labels of the cluster deprived of accessing the original features. The self-taught learning models build high-level features by using unlabeled data and test such models for various classification applications in image, web, and song genres. Lifelong learning goals can be achieved by another popular method is Never-ending Language Learner (NELL) [13]. NELL extracts data or reads information from the web and increases its knowledge, then learns how to perform a new task better than the same task done in the earlier day.

Rather than focus on conventional machine learning, the system should retain knowledge and transfer this to the system to the learning agent. The system should learn sequential tasks and increase their magnitude.

8.3.1 Challenges and Benefits of Lifelong Learning Models [126].

- **Input /output Type, Complexity and Cardinality :** The real-time environment has a variety of data from different domains; it can differ in nature. The attributes of each input may vary according to their source and required task.
- **Training Examples Vs Prior Knowledge :** In life-long learning systems, prior knowledge is a crucial part of the end-to-end system to achieve accuracy while performing a new task.

There is a need to retain valid data from the knowledge base that must have information act as a training example.

- **Effective and Efficient knowledge Retention :** The system must retain efficient information that must not be erroneous. Furthermore, it must use finite memory to store knowledge with limited computational capacity. The system must be capable of handling duplicate data and increase the accuracy of the prior knowledge.
- **Effective and Efficient knowledge Transfer :** Prior knowledge should not increase computational time and effort. Moreover, the transfer of knowledge shout not generates less accurate inputs/models for new tasks. There are three major components of lifelong learning.

(1) Retention of learned task knowledge,
(2) Selective transfer of prior while learning a new task, and
(3) The system must ensure that retention and transfer of knowledge must be efficient.

- **Scalability :** Scalability is one of the most challenging and essential aspects of almost all fields of computer science. The system must be able to adapt increments in volumes of input data. The lifelong learning systems must be able to address the space and time complexity of all these factors.
- **Heterogeneous Domain of Task :** The lifelong learning systems must handle data from different domains by establishing relations among the origin domain and targeted domain. There are so many features that are common between but diversity in transferred knowledge data also exists. The system must have the ability to map features in transferred knowledge.

8.4 Multi-task Learning

Multi-task learning (MTL) [14, 16, 74] acquires various associated tasks concurrently, beaming at delivering a more reliable representation by using the associated knowledge yielded by various jobs. The motive behind introducing inductive bias in Multi-task learning is to joint hypothesis space of every job by utilizing the task-relatedness building. It additionally inhibits over-fitting in the specific job and therefore has a more immeasurable generalization capability. Unlike transfer learning, it mainly does various jobs preferably than various areas as much of the area's existing research is based on several comparable jobs of the identical application area. Multi-task learning allows those jobs are strictly associated with each other. There are several hypotheses in terms of job-relatedness, which drive to another modeling strategy. Many researchers continue to hypothesize that all job data come from the same sources and are correlated to the standard or global models. According to this hypothesis, they created the association among jobs employing a task-coupling parameter, including regularization. In [79], authors proposed multi-task learning for the deep neural network. They classify multi-tasking tasks into two categories for deep learning. First is classification, and second is ranking; in classification, the model identifies the queried domain, whereas the ranking model finds the relevant queries.

9 Research Challenges

Learning in a dynamic environment is still a challenging task due to the unpredicted nature of the upcoming event. How we can integrate the classifier to obtain a sub-knowledge of unknown classes and reduce the open-space risk. The significant challenges in open-world machine learning are:

- **Incremental Volume of Data:** This is the era of digitization. Hence various sources are generating a large amount of data. This data is not only significant in volume but also unstructured. Managing and finding the different classes of the various domains is very difficult as continuous updates appear with additional unseen instance categories. There is a lack of a mechanism that can deal with real worlds data.

- **Identifying a Novel Classes:** Once a system identifies instances as unseen or rejects unseen classes, the system has to learn about the classes of these unseen instances. There is a need for a complete framework that can address these unseen instances and make novel classes for them. The framework must find the instances of each class to acquire new classes.
- **Updating a Knowledge-Base:** There is a lack of a mechanism that can append new knowledge to the system at run time. There are many complexities in appending a new domain and its classes in a progressive environment as input data overgrows. There is also a need to use such classes (newly recognized) for the next prediction without retraining the entire model.
- **Open-space Risk:** During the learning phase in the open environment of the open world, it is hard to manage space away from positive training examples. It is referred to as open-space. The open-space risk needs to be addressed to learn more accurately with increasing openness. There is a need to build a framework that can reduce open-space risk.
- **Open Framework:** There is a need for a generic framework to discover unseen classes in the real-world domain that can function end-to-end for learning in a progressive environment. To build the complete framework of an open-world machine learning system, one needs to execute both operations together (discovery of unseen instances and identification of novel classes). Hence it needs two or more different modules to function dynamically. These modules can use different methods; Hence modules' concatenation and synchronization are relatively complicated as different methods are involved. The model needs to precisely address both seen and unseen classes with synchronization of newly adopted classes.
- **Efficient Discovery of Novel Classes:** There is a lack of an open-world machine learning model to discover and identify the unseen classes. Various models exist in open-world machine learning to discover unseen instances, but very few can identify the novel classes out of these unseen instances. However, there is also a lack of learning models that identify the number of unseen classes efficiently.
- **Retention of Obtained Knowledge:** There is a lack of models that continuously updates the knowledge while learning a new task. To the best of our knowledge, very few learning models use a continual learning approach to learn unseen instances employing previously obtained knowledge for the subsequent prediction.

10 Future Directions

We have reviewed numerous research works of open-world machine learning in computer vision and image processing, and natural language processing. Based on the study, we have identified three significant aspects necessary to achieve learning in the open-world. open-world machine learning can be improved by enhancing these three aspects, model, rejection capability, and identification of new classes. This section discusses and analyses the limitation in brief and discusses the research directions in detail.

Open-world Models. The existing models of open-world machine learning are working in a hybrid manner and address the problem in parts. There is a lack of models available that can work in an end-to-end manner. The end-to-end model for open-world machine learning can strengthen the classification for both categories, known-known class, and known unknown class. To our best knowledge, there is no promising model available for unknown-unknown class identification, which is one of the challenging categories in open-world machine learning . The existing methods for unknown-unknown class classification are worth extending further.

Rejection of Unknown Classes. Few works are available to reject unknown-unknown classes, while automation systems entirely depend on unseen class rejection with high accuracy. Further work to increase the rejection capability of unknown classes can make the system more reliable as the real-world application faces many unknown objects while working in a dynamic environment. Existing models need more improvements to reject unseen classes with high accuracy.
Identification of Unseen Classes. Most existing models either detect known or reject unknown, but after rejecting classes as unknown, there is no promising mechanism available to further identify classes in rejected data. There is a need for models that can identify the number of hidden clauses in rejected data.

11 Conclusion

In this paper, we investigated the works in open-world machine learning proposed in the last decade. We have also discussed the prominence and many real-life applications of open-word machine learning. Many algorithms, models, and frameworks have been proposed in the literature to address numerous objectives allied to open-world settings. The domain is relatively new; thus, there are inadequate sources of information. The presented review will help in understanding the open-world scenario, working, and associated challenges. We provided a task-based classification of OWML in CV-IP and NLP. Further, we discussed various techniques and used datasets in OWML. The limitations of numerous technologies are also analyzed to facilitate promising future extensions of these methods.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Ethem Alpaydin. 2020. *Introduction to machine learning.* MIT press.
[2] Dana Angluin. 1988. Queries and concept learning. *Machine learning* 2, 4 (1988), 319–342.
[3] Andreas Argyriou, Theodoros Evgeniou, and Massimiliano Pontil. 2008. Convex multi-task feature learning. *Machine learning* 73, 3 (2008), 243–272.
[4] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. 2007. Dbpedia: A nucleus for a web of open data. In *The semantic web*. Springer, 722–735.
[5] Elena Baralis, Silvia Chiusano, and Paolo Garza. 2007. A lazy approach to associative classification. *IEEE Transactions on Knowledge and Data Engineering* 20, 2 (2007), 156–171.
[6] Abhijit Bendale and Terrance Boult. 2015. Towards open world recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 1893–1902.
[7] Abhijit Bendale and Terrance E Boult. 2016. Towards open set deep networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* 1563–1572.
[8] Christopher M Bishop. 2006. *Pattern recognition and machine learning.* springer.
[9] Paul Bodesheim, Alexander Freytag, Erik Rodner, and Joachim Denzler. 2015. Local novelty detection in multi-class recognition problems. In *2015 IEEE Winter Conference on Applications of Computer Vision.* IEEE, 813–820.
[10] Mariusz Bojarski, Davide Del Testa, Daniel Dworakowski, Bernhard Firner, Beat Flepp, Prasoon Goyal, Lawrence D Jackel, Mathew Monfort, Urs Muller, Jiakai Zhang, et al. 2016. End to end learning for self-driving cars. *arXiv preprint arXiv:1604.07316* (2016).
[11] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Durán, Jason Weston, and Oksana Yakhnenko. 2013. Translating Embeddings for Modeling Multi-Relational Data. In *Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2* (Lake Tahoe, Nevada). Curran Associates Inc., Red Hook, NY, USA, 2787–2795.
[12] Markus M Breunig, Hans-Peter Kriegel, Raymond T Ng, and Jörg Sander. 2000. LOF: identifying density-based local outliers. In *Proceedings of the 2000 ACM SIGMOD international conference on Management of data.* 93–104.
[13] Andrew Carlson, Justin Betteridge, Bryan Kisiel, Burr Settles, Estevam R Hruschka, and Tom M Mitchell. 2010. Toward an architecture for never-ending language learning. In *Twenty-Fourth AAAI conference on artificial intelligence.*

[14] Rich Caruana. 1997. Multitask learning. *Machine learning* 28, 1 (1997), 41–75.

[15] Chih-Chung Chang and Chih-Jen Lin. 2011. LIBSVM: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2, 3 (2011), 1–27.

[16] Jianhui Chen, Lei Tang, Jun Liu, and Jieping Ye. 2009. A convex formulation for learning shared structures from multiple tasks. In *Proceedings of the 26th Annual International Conference on Machine Learning*. 137–144.

[17] Zhiyuan Chen and Bing Liu. 2014. Mining topics in documents: standing on the shoulders of big data. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 1116–1125.

[18] Zhiyuan Chen and Bing Liu. 2018. Lifelong machine learning. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 12, 3 (2018), 1–207.

[19] Gouenou Coatrieux, Laurent Lecornu, Bülent Sankur, and Ch Roux. 2006. A review of image watermarking applications in healthcare. In *2006 International conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 4691–4694.

[20] Gregory Cohen, Saeed Afshar, Jonathan Tapson, and Andre Van Schaik. 2017. EMNIST: Extending MNIST to handwritten letters. In *2017 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2921–2926.

[21] David Cohn, Les Atlas, and Richard Ladner. 1994. Improving generalization with active learning. *Machine learning* 15, 2 (1994), 201–221.

[22] Alice Coucke, Alaa Saade, Adrien Ball, Théodore Bluche, Alexandre Caulier, David Leroy, Clément Doumouro, Thibault Gisselbrecht, Francesco Caltagirone, Thibaut Lavril, et al. 2018. Snips voice platform: an embedded spoken language understanding system for private-by-design voice interfaces. *arXiv preprint arXiv:1805.10190* (2018).

[23] Benjamin R Cowan, Nadia Pantidi, David Coyle, Kellie Morrissey, Peter Clarke, Sara Al-Shehri, David Earley, and Natasha Bandeira. 2017. " What can i help you with?" infrequent users' experiences of intelligent personal assistants. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–12.

[24] Wenyuan Dai, Ou Jin, Gui-Rong Xue, Qiang Yang, and Yong Yu. 2009. Eigentransfer: a unified framework for transfer learning. In *Proceedings of the 26th Annual International Conference on Machine Learning*. 193–200.

[25] Laurens De Haan, Ana Ferreira, and Ana Ferreira. 2006. *Extreme value theory: an introduction*. Vol. 21. Springer.

[26] Rocco De Rosa, Thomas Mensink, and Barbara Caputo. 2016. Online open world recognition. *arXiv preprint arXiv:1604.02275* (2016).

[27] Kerstin Denecke, Alaa Abd-Alrazaq, and Mowafa Househ. 2021. Artificial Intelligence for Chatbots in Mental Health: Opportunities and Challenges. *Multiple Perspectives on Artificial Intelligence in Healthcare* (2021), 115–128.

[28] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. 2009. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*. 248–255.

[29] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[30] Tri Doan and Jugal Kalita. 2017. Overcoming the challenge for text classification in the open world. In *2017 IEEE 7th Annual Computing and Communication Workshop and Conference (CCWC)*. IEEE, 1–7.

[31] Ali Farhadi, Ian Endres, Derek Hoiem, and David Forsyth. 2009. Describing objects by their attributes. In *2009 IEEE conference on computer vision and pattern recognition*. IEEE, 1778–1785.

[32] Geli Fei and Bing Liu. 2016. Breaking the closed world assumption in text classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 506–514.

[33] Geli Fei, Shuai Wang, and Bing Liu. 2016. Learning cumulatively to become more knowledgeable. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 1565–1574.

[34] Peter W Frey and David J Slate. 1991. Letter recognition using Holland-style adaptive classifiers. *Machine learning* 6, 2 (1991), 161–182.

[35] Yang Gao, Yi-Fan Li, Bo Dong, Yu Lin, and Latifur Khan. 2019. SIM: Open-World Multi-Task Stream Classifier with Integral Similarity Metrics. In *2019 IEEE International Conference on Big Data (Big Data)*. IEEE, 751–760.

[36] Zongyuan Ge, Sergey Demyanov, Zetao Chen, and Rahil Garnavi. 2017. Generative OpenMax for multi-class open set classification. In *British Machine Vision Conference 2017*. British Machine Vision Association and Society for Pattern Recognition.

[37] Chuanxing Geng, Sheng-jun Huang, and Songcan Chen. 2020. Recent advances in open set recognition: A survey. *IEEE transactions on pattern analysis and machine intelligence* (2020).

[38] Jan-Mark Geusebroek, Gertjan J Burghouts, and Arnold WM Smeulders. 2005. The Amsterdam library of object images. *International Journal of Computer Vision* 61, 1 (2005), 103–112.

[39] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. 2016. *Deep learning*. Vol. 1. MIT press Cambridge.

[40] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[41] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. 2014. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572* (2014).

[42] Gregory Griffin, Alex Holub, and Pietro Perona. 2007. Caltech-256 object category dataset. (2007).

[43] Xiaojie Guo, Amir Alipour-Fanid, Lingfei Wu, Hemant Purohit, Xiang Chen, Kai Zeng, and Liang Zhao. 2019. Multi-stage deep classifier cascades for open world recognition. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. 179–188.

[44] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, and Jianfeng Gao. 2016. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In *European conference on computer vision*. Springer, 87–102.

[45] Sonal Gupta, Rushin Shah, Mrinal Mohit, Anuj Kumar, and Mike Lewis. 2018. Semantic Parsing for Task Oriented Dialog using Hierarchical Representations. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 2787–2792.

[46] F Maxwell Harper and Joseph A Konstan. 2015. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)* 5, 4 (2015), 1–19.

[47] Mehadi Hassen and Philip K Chan. 2020. Learning a neural-network-based representation for open set recognition. In *Proceedings of the 2020 SIAM International Conference on Data Mining*. SIAM, 154–162.

[48] Ruining He and Julian McAuley. 2016. Ups and downs: Modeling the visual evolution of fashion trends with one-class collaborative filtering. In *proceedings of the 25th international conference on world wide web*. 507–517.

[49] Dan Hendrycks and Kevin Gimpel. 2016. A baseline for detecting misclassified and out-of-distribution examples in neural networks. *arXiv preprint arXiv:1610.02136* (2016).

[50] Andrei De Souza Inácio, Matheus Gutoski, André Eugênio Lazzaretti, and Heitor Silvério Lopes. 2021. OSVidCap: a Framework for the Simultaneous Recognition and Description of Concurrent Actions in Videos in an Open-Set Scenario. *IEEE Access* (2021).

[51] Lalit P Jain, Walter J Scheirer, and Terrance E Boult. 2014. Multi-class open set recognition using probability of inclusion. In *13th European Conference on Computer Vision*. 393–409.

[52] A Jeya Christy and K Dhanalakshmi. 2021. Content-Based Image Recognition and Tagging by Deep Learning Methods. *Wireless Personal Communications* (2021), 1–26.

[53] Jiepu Jiang, Ahmed Hassan Awadallah, Rosie Jones, Umut Ozertem, Imed Zitouni, Ranjitha Gurunath Kulkarni, and Omar Zia Khan. 2015. Automatic online evaluation of intelligent assistants. In *Proceedings of the 24th International Conference on World Wide Web*. 506–516.

[54] Pedro R Mendes Júnior, Roberto M De Souza, Rafael de O Werneck, Bernardo V Stein, Daniel V Pazinato, Waldir R de Almeida, Otávio AB Penatti, Ricardo da S Torres, and Anderson Rocha. 2017. Nearest neighbors distance ratio open-set classifier. *Machine Learning* 106, 3 (2017), 359–386.

[55] Dan Jurafsky. 1997. Switchboard SWBD-DAMSL shallow-discourse-function annotation coders manual. *Institute of Cognitive Science Technical Report* (1997).

[56] Mohammed Waleed Kadous et al. 2002. *Temporal classification: Extending the classification paradigm to multivariate time series.* Citeseer.

[57] James M Keller, Michael R Gray, and James A Givens. 1985. A fuzzy k-nearest neighbor algorithm. *IEEE transactions on systems, man, and cybernetics* 4 (1985), 580–585.

[58] Zubair Ahmed Khan and Asma Rizvi. 2021. AI BASED FACIAL RECOGNITION TECHNOLOGY AND CRIMINAL JUSTICE: ISSUES AND CHALLENGES. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12, 14 (2021), 3384–3392.

[59] Yoon Kim. 2014. Convolutional Neural Networks for Sentence Classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 1746–1751.

[60] Shu Kong and Deva Ramanan. 2021. OpenGAN: Open-Set Recognition via Open Data Generation. *arXiv preprint arXiv:2104.02939* (2021).

[61] Sotiris B Kotsiantis, I Zaharakis, P Pintelas, et al. 2007. Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering* 160, 1 (2007), 3–24.

[62] Alex Krizhevsky, Geoffrey Hinton, et al. 2009. Learning multiple layers of features from tiny images. (2009).

[63] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25 (2012), 1097–1105.

[64] Ludmila I Kuncheva and Juan J Rodriguez. 2007. Classifier ensembles with a random linear oracle. *IEEE Transactions on Knowledge and Data Engineering* 19, 4 (2007), 500–508.

[65] Ken Lang. 1995. Newsweeder: Learning to filter netnews. In *Machine Learning Proceedings 1995*. Elsevier, 331–339.

[66] Neil D Lawrence and John C Platt. 2004. Learning to learn with the informative vector machine. In *Proceedings of the twenty-first international conference on Machine learning*. 65.

[67] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. 2006. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern*

Recognition (CVPR'06), Vol. 2. IEEE, 2169–2178.
[68] Ya Le and Xuan Yang. 2015. Tiny imagenet visual recognition challenge. *CS 231N* 7, 7 (2015), 3.
[69] Yann LeCun. 1998. The MNIST database of handwritten digits. *http://yann. lecun. com/exdb/mnist/* (1998).
[70] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.
[71] Kuang-Chih Lee, Jeffrey Ho, and David J Kriegman. 2005. Acquiring linear subspaces for face recognition under variable lighting. *IEEE Transactions on pattern analysis and machine intelligence* 27, 5 (2005), 684–698.
[72] Qingming Leng, Mang Ye, and Qi Tian. 2019. A survey of open-world person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology* 30, 4 (2019), 1092–1108.
[73] David D Lewis and William A Gale. 1994. A sequential algorithm for training text classifiers. In *SIGIR'94*. Springer, 3–12.
[74] Hui Li, Xuejun Liao, and Lawrence Carin. 2009. Multi-task Reinforcement Learning in Partially Observable Stochastic Environments. *Journal of Machine Learning Research* 10, 5 (2009).
[75] Zhizhong Li and Derek Hoiem. 2017. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence* 40, 12 (2017), 2935–2947.
[76] Ting-En Lin and Hua Xu. 2019. Deep Unknown Intent Detection with Margin Loss. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 5491–5496.
[77] Ting-En Lin and Hua Xu. 2019. A post-processing method for detecting unknown intent of dialogue system via pre-trained deep neural network classifier. *Knowledge-Based Systems* 186 (2019), 104979.
[78] Qiuhua Liu, Xuejun Liao, Hui Li, Jason R Stack, and Lawrence Carin. 2008. Semisupervised multitask learning. *IEEE transactions on pattern analysis and machine intelligence* 31, 6 (2008), 1074–1086.
[79] Xiaodong Liu, Jianfeng Gao, Xiaodong He, Li Deng, Kevin Duh, and Ye-Yi Wang. 2015. Representation learning using multi-task deep neural networks for semantic classification and information retrieval. (2015).
[80] Abbas Saliimi Lokman and Mohamed Ariff Ameedeen. 2018. Modern chatbot systems: A technical review. In *Proceedings of the future technologies conference*. Springer, 1012–1023.
[81] Vincent Lonij, Ambrish Rawat, and Maria-Irina Nicolae. 2017. Open-world visual recognition using knowledge graphs. *arXiv preprint arXiv:1708.08310* (2017).
[82] Sahisnu Mazumder, Nianzu Ma, and Bing Liu. 2018. Towards a continuous knowledge learning engine for chatbots. *arXiv preprint arXiv:1802.06024* (2018).
[83] Julian McAuley, Christopher Targett, Qinfeng Shi, and Anton Van Den Hengel. 2015. Image-based recommendations on styles and substitutes. In *Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*. 43–52.
[84] Thomas Mensink, Jakob Verbeek, Florent Perronnin, and Gabriela Csurka. 2013. Distance-based image classification: Generalizing to new classes at near-zero cost. *IEEE transactions on pattern analysis and machine intelligence* 35, 11 (2013), 2624–2637.
[85] Donald Michie, David J Spiegelhalter, and Charles C Taylor. 1994. Machine learning, neural and statistical classification. (1994).
[86] Lilyana Mihalkova, Tuyen Huynh, and Raymond J Mooney. 2007. Mapping and revising markov logic networks for transfer learning. In *Aaai*, Vol. 7. 608–614.
[87] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
[88] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
[89] Dimity Miller, Niko Sünderhauf, Michael Milford, and Feras Dayoub. 2021. Class anchor clustering: a distance-based loss for training open set classifiers. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV 2021)*.
[90] George A. Miller. 1995. WordNet: A Lexical Database for English. *Commun. ACM* 38, 11 (Nov. 1995), 39–41. https://doi.org/10.1145/219717.219748
[91] Philip Moore and Hai Van Pham. 2015. On context and the open world assumption. In *2015 IEEE 29th International Conference on Advanced Information Networking and Applications Workshops*. 387–392.
[92] Xin Mu, Kai Ming Ting, and Zhi-Hua Zhou. 2017. Classification under streaming emerging new classes: A solution using completely-random trees. *IEEE Transactions on Knowledge and Data Engineering* 29, 8 (2017), 1605–1618.
[93] Susan M Mudambi and David Schuff. 2010. Research note: What makes a helpful online review? A study of customer reviews on Amazon. com. *MIS quarterly* (2010), 185–200.
[94] Kevin P Murphy. 2012. *Machine learning: a probabilistic perspective*. MIT press.
[95] Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al. 1996. Columbia object image library (coil-100). (1996).

[96] Yuval Netzer, Tao Wang, Adam Coates, Alessandro Bissacco, Bo Wu, and Andrew Y Ng. 2011. Reading digits in natural images with unsupervised feature learning. (2011).

[97] Anh Nguyen, Jason Yosinski, and Jeff Clune. 2015. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 427–436.

[98] David Nister and Henrik Stewenius. 2006. Scalable recognition with a vocabulary tree. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, Vol. 2. Ieee, 2161–2168.

[99] Poojan Oza and Vishal M Patel. 2019. C2ae: Class conditioned auto-encoder for open-set recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2307–2316.

[100] Poojan Oza and Vishal M Patel. 2019. Deep cnn-based multi-task learning for open-set recognition. *arXiv preprint arXiv:1903.03161* (2019).

[101] Sinno Jialin Pan and Qiang Yang. 2009. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* 22, 10 (2009), 1345–1359.

[102] Sayanta Paul and Sriparna Saha. 2020. CyberBERT: BERT for cyberbullying identification. *Multimedia Systems* (2020), 1–8.

[103] Jeffrey Pennington, Richard Socher, and Christopher D Manning. 2014. Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)*. 1532–1543.

[104] John Platt et al. 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers,Cambridge, MA* 10, 3 (1999), 61–74.

[105] Sridhama Prakhya, Vinodini Venkataram, and Jugal Kalita. 2017. Open set text classification using convolutional neural networks. In *International Conference on Natural Language Processing, 2017*.

[106] Hemant Purohit, Carlos Castillo, Fernando Diaz, Amit Sheth, and Patrick Meier. 2014. Emergency-relief coordination on social media: Automatically matching resource requests and offers. *First Monday* (2014).

[107] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. 2018. Deep learning for medical image processing: Overview, challenges and the future. *Classification in BioApps* (2018), 323–350.

[108] Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B Tenenbaum, Hugo Larochelle, and Richard S Zemel. 2018. Meta-Learning for Semi-Supervised Few-Shot Classification. In *International Conference on Learning Representations*.

[109] Jason Rennie and Ken Lang. 2008. The 20 Newsgroups data set. *Available in web page< URL: http://qwone. com/jason/20Newsgroups* (2008).

[110] Ryan Rifkin and Aldebaro Klautau. 2004. In defense of one-vs-all classification. *The Journal of Machine Learning Research* 5 (2004), 101–141.

[111] Marko Ristin, Matthieu Guillaumin, Juergen Gall, and Luc Van Gool. 2014. Incremental learning of ncm forests for large-scale image classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3654–3661.

[112] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. 2015. Imagenet large scale visual recognition challenge. *International journal of computer vision* 115, 3 (2015), 211–252.

[113] Walter J Scheirer, Anderson de Rezende Rocha, Archana Sapkota, and Terrance E Boult. 2012. Toward open set recognition. *IEEE transactions on pattern analysis and machine intelligence* 35, 7 (2012), 1757–1772.

[114] Walter J Scheirer, Lalit P Jain, and Terrance E Boult. 2014. Probability models for open set recognition. *IEEE transactions on pattern analysis and machine intelligence* 36, 11 (2014), 2317–2324.

[115] Walter J Scheirer, Anderson Rocha, Ross J Micheals, and Terrance E Boult. 2011. Meta-recognition: The theory and practice of recognition score analysis. *IEEE transactions on pattern analysis and machine intelligence* 33, 8 (2011), 1689–1695.

[116] M. Schuster and K. K. Paliwal. 1997. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing* 45, 11 (1997), 2673–2681. https://doi.org/10.1109/78.650093

[117] Vikash Sehwag, Arjun Nitin Bhagoji, Liwei Song, Chawin Sitawarin, Daniel Cullina, Mung Chiang, and Prateek Mittal. 2019. Analyzing the robustness of open-world machine learning. In *Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security*. 105–116.

[118] Burr Settles. 2009. Active learning literature survey. (2009).

[119] Burr Settles, Mark Craven, and Lewis Friedland. 2008. Active learning with real annotation costs. In *Proceedings of the NIPS workshop on cost-sensitive learning*. Vancouver, CA:, 1–10.

[120] Haseeb Shah, Johannes Villmow, Adrian Ulges, Ulrich Schwanecke, and Faisal Shafait. 2019. An open-world extension to knowledge graph completion models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33. 3044–3051.

[121] Baoxu Shi and Tim Weninger. 2018. Open-world knowledge graph completion. In *Thirty-Second AAAI Conference on Artificial Intelligence*.

[122] Elizabeth Shriberg, Andreas Stolcke, Daniel Jurafsky, Noah Coccaro, Marie Meteer, Rebecca Bates, Paul Taylor, Klaus Ries, Rachel Martin, and Carol Van Ess-Dykema. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and speech* 41, 3-4 (1998), 443–492.
[123] Lei Shu, Hu Xu, and Bing Liu. 2017. DOC: Deep Open Classification of Text Documents. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*. 2911–2916.
[124] Lei Shu, Hu Xu, and Bing Liu. 2018. Unseen class discovery in open-world classification. *arXiv preprint arXiv:1801.05609* (2018).
[125] Thomas R Shultz and Francois Rivest. 2001. Knowledge-based cascade-correlation: Using knowledge to speed learning. *Connection Science* 13, 1 (2001), 43–72.
[126] Daniel L Silver, Qiang Yang, and Lianghao Li. 2013. Lifelong machine learning systems: Beyond learning algorithms. In *2013 AAAI spring symposium series*.
[127] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
[128] Liwei Song, Vikash Sehwag, Arjun Nitin Bhagoji, and Prateek Mittal. 2020. A critical evaluation of open-world machine learning. *arXiv preprint arXiv:2007.04391* (2020).
[129] Rupesh Kumar Srivastava, Klaus Greff, and Jürgen Schmidhuber. 2015. Highway networks. *arXiv preprint arXiv:1505.00387* (2015).
[130] Wallace Stevens. 2021. Efficient Uncertainty Estimation for Open-Set Object Detection. *Epistemic Uncertainty Estimation for Object Detection in Open-Set Conditions* (2021), 91.
[131] Andreas Stolcke, Klaus Ries, Noah Coccaro, Elizabeth Shriberg, Rebecca Bates, Daniel Jurafsky, Paul Taylor, Rachel Martin, Carol Van Ess-Dykema, and Marie Meteer. 2000. Dialogue act modeling for automatic tagging and recognition of conversational speech. *Computational linguistics* 26, 3 (2000), 339–373.
[132] Alexander Strehl and Joydeep Ghosh. 2002. Cluster ensembles—a knowledge reuse framework for combining multiple partitions. *Journal of machine learning research* 3, Dec (2002), 583–617.
[133] Martin Thoma. 2017. The hasyv2 dataset. *arXiv preprint arXiv:1701.08380* (2017).
[134] Kristinn R Thórisson, Jordi Bieger, Xiang Li, and Pei Wang. 2019. Cumulative learning. In *International Conference on Artificial General Intelligence*. Springer, 198–208.
[135] Sebastian Thrun. 1995. Exploration in active learning. *Handbook of Brain Science and Neural Networks* (1995), 381–384.
[136] Sebastian Thrun and Lorien Pratt. 1998. Learning to learn: Introduction and overview. In *Learning to learn*. Springer, 3–17.
[137] Kristina Toutanova and Danqi Chen. 2015. Observed versus latent features for knowledge base and text inference. In *Proceedings of the 3rd workshop on continuous vector space models and their compositionality*. 57–66.
[138] Gokhan Tur, Dilek Hakkani-Tür, and Larry Heck. 2010. What is left to be understood in ATIS?. In *2010 IEEE Spoken Language Technology Workshop*. IEEE, 19–24.
[139] Nikhita Vedula, Rahul Gupta, Aman Alok, and Mukund Sridhar. 2020. Automatic discovery of novel intents & domains from text utterances. *arXiv preprint arXiv:2006.01208* (2020).
[140] Nikhita Vedula, Nedim Lipka, Pranav Maneriker, and Srinivasan Parthasarathy. 2019. Towards open intent discovery for conversational text. *arXiv preprint arXiv:1904.08524* (2019).
[141] Oriol Vinyals, Charles Blundell, Timothy Lillicrap, Daan Wierstra, et al. 2016. Matching networks for one shot learning. *Advances in neural information processing systems* 29 (2016), 3630–3638.
[142] Alex Hai Wang. 2010. Detecting spam bots in online social networking sites: a machine learning approach. In *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, 335–342.
[143] Yuhan Wang, Weidong Xiao, Zhen Tan, and Xiang Zhao. 2021. Caps-OWKG: a capsule network model for open-world knowledge graph. *International Journal of Machine Learning and Cybernetics* 12, 6 (2021), 1627–1637.
[144] John Willes, James Harrison, Ali Harakeh, Chelsea Finn, Marco Pavone, and Steven Waslander. 2021. Bayesian Embeddings for Few-Shot Open World Recognition. *arXiv preprint arXiv:2107.13682* (2021).
[145] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma. 2008. Robust face recognition via sparse representation. *IEEE transactions on pattern analysis and machine intelligence* 31, 2 (2008), 210–227.
[146] Qitian Wu, Hengrui Zhang, Xiaofeng Gao, Junchi Yan, and Hongyuan Zha. 2021. Towards Open-World Recommendation: An Inductive Model-based Collaborative Filtering Approach. In *International Conference on Machine Learning*. PMLR, 11329–11339.
[147] Zhi-Fan Wu, Tong Wei, Jianwen Jiang, Chaojie Mao, Mingqian Tang, and Yu-Feng Li. 2021. NGC: A Unified Framework for Learning with Open-World Noisy Data. (2021), 62–71.
[148] Bo Xiao and Izak Benbasat. 2007. E-commerce product recommendation agents: Use, characteristics, and impact. *MIS quarterly* (2007), 137–209.
[149] Han Xiao, Kashif Rasul, and Roland Vollgraf. 2017. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747* (2017).

[150] Ruobing Xie, Zhiyuan Liu, Jia Jia, Huanbo Luan, and Maosong Sun. 2016. Representation learning of knowledge graphs with entity descriptions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.

[151] Hu Xu, Bing Liu, Lei Shu, and P Yu. 2019. Open-world learning and application to product classification. In *The World Wide Web Conference*. 3413–3419.

[152] J Yang and Joseph F Coughlin. 2014. In-vehicle technology for self-driving cars: Advantages and challenges for aging drivers. *International Journal of Automotive Technology* 15, 2 (2014), 333–340.

[153] N Yang. 2019. Douban Movie and NetEase Music Datasets and Model Code. *The President & Fellows of Harvard College: Cambridge, MA, USA* (2019).

[154] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven CH Hoi. 2021. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021).

[155] Xiaoxin Yin, Jiawei Han, Jiong Yang, and Philip S Yu. 2006. Efficient classification across multiple database relations: A crossmine approach. *IEEE Transactions on Knowledge and Data Engineering* 18, 6 (2006), 770–783.

[156] Ryota Yoshihashi, Wen Shao, Rei Kawakami, Shaodi You, Makoto Iida, and Takeshi Naemura. 2019. Classification-reconstruction learning for open-set recognition. In *In the proceedings of the Conference on Computer Vision and Pattern Recognition*. 4016–4025.

[157] Kun Yue, Jiahui Wang, Xinbai Li, and Kuang Hu. 2020. Representation-based completion of knowledge graph with open-world data. In *2020 5th International Conference on Computer and Communication Systems (ICCCS)*. IEEE, 1–8.

[158] He Zhang and Vishal M Patel. 2016. Sparse representation-based open set recognition. *IEEE transactions on pattern analysis and machine intelligence* 39, 8 (2016), 1690–1696.

[159] Wayne Xin Zhao, Sui Li, Yulan He, Liwei Wang, Ji-Rong Wen, and Xiaoming Li. 2016. Exploring demographic information in social media for product recommendation. *Knowledge and Information Systems* 49, 1 (2016), 61–89.

[160] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. 2017. Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence* 40, 6 (2017), 1452–1464.

[161] Yajin Zhou and Xuxian Jiang. 2012. Dissecting android malware: Characterization and evolution. In *2012 IEEE symposium on security and privacy*. IEEE, 95–109.