JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Review Paper on Machine Learning and different Algorithm

Ms. Monika Verma
Assistant Professor

Ms. Lata Verma
Assistant Professor

ABSTRACT:
Machine learning is a science that was found and developed as a subfield Artificial intelligence in the 1950s. The first step in machine learning goes back In the 1950s, however, there was no significant research and development on this science. However, in the 1990s, researches on this area have resumed, developed and reached today. This is a science that will improve further in the future. The reason behind this development is the difficulty of rapid analysis and processing growing data. Machine learning is based on the principle of finding the best model Thanks for the new data among the previous data for this growing data. Therefore, Machine learning research will run parallel to the growing amount of data. This Research covers history of machine learning, methods used in machine Learning, its application areas and research in this area. Purpose of the study is to disseminate knowledge on machine learning, which has become very popular nowadays, and its applications to researchers.

Keywords: Machine Learning, Machine Learning Algorithms, Artificial Intelligence, Big Data

INTRODUCTION

Machine learning is a field devoted to understanding and creating methods that allow machines to "learn"—that is, methods that take advantage of data to improve computer performance on some set of tasks.

Machine learning is a growing technology that enables computers to automatically learn from past data. Machine learning uses various algorithms to build mathematical models and make predictions using historical data or information. Currently, it is being used for various tasks such as image recognition, speech recognition, email filtering, Facebook auto-tagging, recommender systems and many more.

HOW MACHINE LEARNING WORK

Suppose we have a complex problem where we need to make some predictions, so instead of writing a code for the same, we just need to feed the data to generic algorithms, and with the help of these algorithms, the machine logic Builds logic accordingly. Predict data and outputs. Machine learning has changed the way we think about problems. The block diagram below explains the working of a machine learning algorithm:

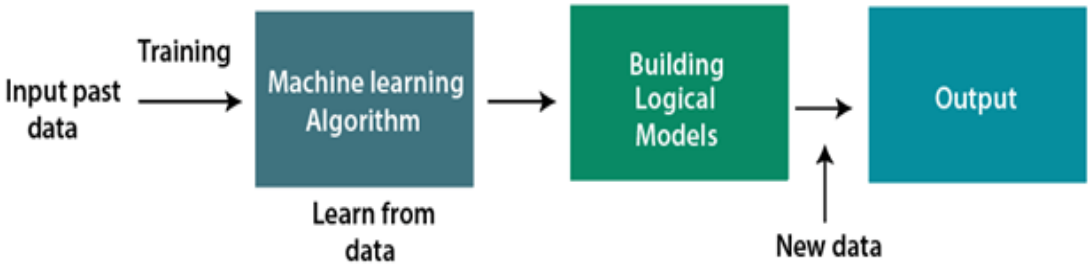


Fig. 1: WORKING OF Ml

FEATURES OF MACHINE LEARNING

Machine learning is a subset of artificial intelligence that focuses on developing algorithms and statistical models that enable computers to learn and make predictions or decisions without being explicitly programmed. The key features of machine learning include:

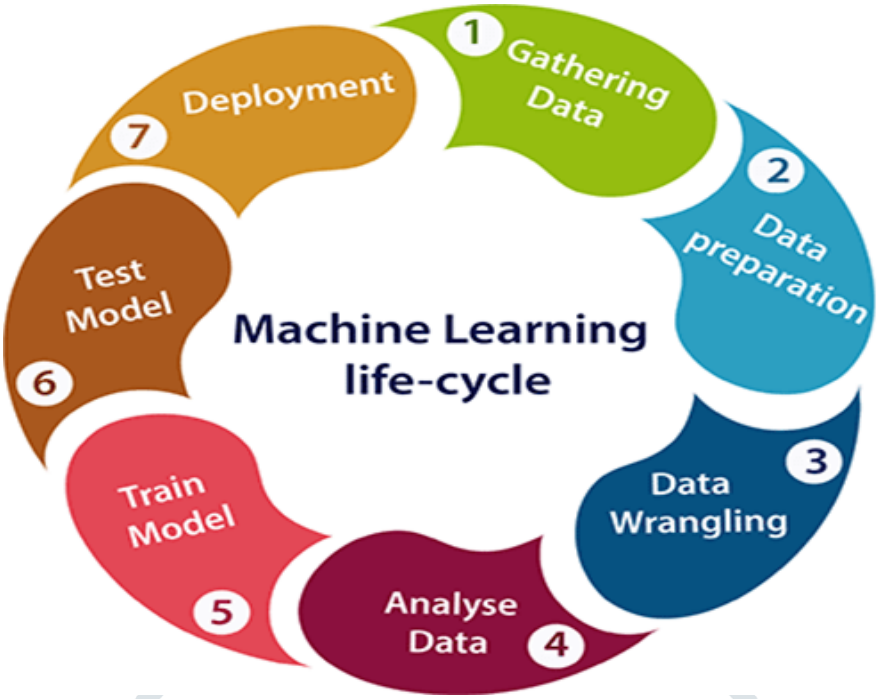
1. **Automation and Adaptation**: Machine learning algorithms automate the process of learning from data and adapt themselves to improve performance based on experience.
2. **Pattern Recognition**: Machine learning algorithms excel at recognizing patterns and extracting meaningful insights from complex and large datasets.
3. **Prediction and Generalization**: Machine learning models can make predictions or decisions based on past observations and generalize those predictions to unseen data.
4. **Data-driven Approach**: Machine learning algorithms learn directly from data, allowing them to discover patterns and relationships that may not be evident through human analysis alone.
5. **Scalability**: Machine learning techniques can handle large volumes of data and scale efficiently to process and learn from massive datasets.
6. **Iterative Improvement**: Machine learning models can be trained iteratively, continuously improving their performance as they receive more data and feedback.
7. **Unsupervised Learning**: Machine learning algorithms can perform unsupervised learning, where they identify patterns and structures in data without explicit labels or guidance.
8. **Supervised Learning**: Machine learning algorithms can also perform supervised learning, where they learn from labelled data to make predictions or classify new examples.
9. **Ensemble Methods**: Machine learning allows for the combination of multiple models or algorithms using ensemble methods to improve prediction accuracy and reduce errors.
10. **Feature Extraction**: Machine learning algorithms can automatically extract relevant features or representations from raw data, reducing the need for manual feature engineering.
11. **Real-time Decision Making**: Machine learning models can make quick and automated decisions in real-time, enabling applications such as fraud detection, recommendation systems, and autonomous vehicles.

12. **Continuous Learning**: Machine learning models can be updated with new data to adapt to changing conditions and incorporate new information.

LIFE CYCLE OF MACHINE LEARNING

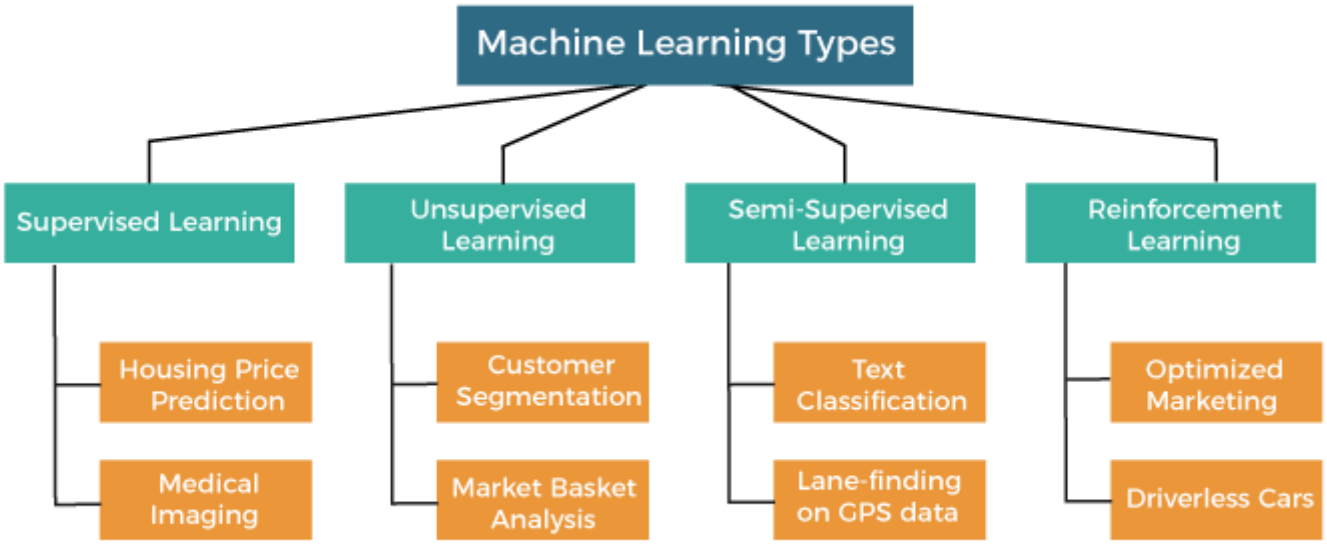
The life cycle of machine learning typically involves several stages, including:

1. **Problem Definition**: Clearly define the problem you want to solve using machine learning techniques. Understand the goals, requirements, and constraints of the problem at hand.
2. **Data Collection**: Gather relevant data that is representative of the problem domain. This may involve obtaining data from various sources, such as databases, APIs, or manual data collection.
3. **Data Pre-processing**: Clean and pre-process the collected data to remove any noise, inconsistencies, or missing values. This step may involve tasks like data cleaning, feature selection, feature transformation, normalization, and handling missing data.
4. **Data Splitting**: Divide the pre-processed data into training, validation, and testing sets. The training set is used to train the machine learning model, the validation set is used to tune hyper parameters and evaluate different models, and the testing set is used to assess the final model's performance.
5. **Model Selection**: Choose an appropriate machine learning model or algorithm that best suits your problem. Consider factors such as the type of problem (classification, regression, clustering, etc.), available data, interpretability requirements, and computational resources.
6. **Model Training**: Train the selected machine learning model using the training data. The model learns from the data and adjusts its internal parameters to minimize the error or maximize the objective function.
7. **Model Evaluation**: Evaluate the trained model's performance using the validation set. Common evaluation metrics depend on the specific problem, such as accuracy, precision, recall, F1 score, mean squared error, etc.
8. **Model Tuning**: Fine-tune the model's hyper parameters to improve its performance. This process involves adjusting parameters that are not learned directly from the training data, such as learning rate, regularization strength, or the number of layers in a neural network.
9. **Model Deployment**: Once the model meets the desired performance, deploy it to the production environment to make predictions on new, unseen data. This may involve integrating the model into an application, setting up APIs, or creating a user interface for interaction.
10. **Monitoring and Maintenance**: Continuously monitor the deployed model's performance and collect feedback from real-world usage. Update the model periodically to incorporate new data and improve its accuracy or adapt to changing conditions.



TYPE OF MACHINE LEARNING

Machine learning can be categorized into three main types based on the learning process and the availability of labeled data:



1. **Supervised Learning**: In supervised learning, the machine learning algorithm is trained on labeled data, where each input example is associated with a corresponding target or output label. The algorithm learns to map input features to the desired output based on the provided examples. The goal is to train a model that can accurately predict labels for new, unseen data. Supervised learning tasks include classification (predicting categorical labels) and regression (predicting continuous values).

The main goal of the supervised learning technique is to map the input variable (x) with the output variable (y). Some real world applications of supervised learning are risk assessment, fraud detection, spam filtering, etc.

Categories of Supervised Machine Learning

Supervised machine learning can be classified into two types of problems, which are given below:

- **Classification**
- **Regression**

a) Classification

Classification algorithms are used to solve classification problems in which the output variable is categorical, such as "yes" or no, male or female, red or blue, etc. Classification algorithms predict the categories present in the dataset. Some real world examples of classification algorithms are spam detection, email filtering, etc.

b) Regression

Regression algorithms are used to solve regression problems in which there is a linear relationship between input and output variables. These are used to predict continuous output variables, such as market trends, weather prediction, etc.

2. **Unsupervised Learning**: Unsupervised learning involves training machine learning algorithms on unlabeled data, where there are no predefined target labels. The algorithms aim to discover patterns, structures, or relationships in the data without explicit guidance. Unsupervised learning tasks include clustering (grouping similar data points), dimensionality reduction (reducing the number of features while preserving essential information), and anomaly detection (identifying unusual or abnormal instances).

Categories of Unsupervised Machine Learning

Unsupervised Learning can be further classified into two types, which are given below:

- **Clustering**
- **Association**

1) Clustering

The clustering technique is used when we want to find the inherent groups from the data. It is a way to group the objects into a cluster such that the objects with the most similarities remain in one group and have few or no similarities with the objects of other groups. An example of the clustering algorithm is grouping the customers by their purchasing behavior.

Some of the popular clustering algorithms are given below:

- **K-Means Clustering algorithm**
- **Mean-shift algorithm**
- **DBSCAN Algorithm**
- **Principal Component Analysis**
- **Independent Component Analysis**

2) Association

Association rule learning is an unsupervised learning technique that finds interesting relationships between variables within a large dataset. The main objective of this learning algorithm is to find out the dependency of one data item on another data item and map those variables accordingly so that it can generate maximum profit. This algorithm is mainly applied in market basket analysis, web usage mining, continuous production etc.

3. **Reinforcement Learning**: Reinforcement learning involves an agent that learns to interact with an environment in order to maximize a reward signal. The agent takes actions in the environment, and based on the feedback received in the form of rewards or penalties, it learns to make better decisions over time. The goal is to find an optimal policy or strategy that maximizes the long-term cumulative reward. Reinforcement learning is commonly used in scenarios such as game playing, robotics, and autonomous systems.

Categories of Reinforcement Learning

Reinforcement learning is categorized mainly into two types of methods/algorithms:

- **Positive Reinforcement Learning:** Positive reinforcement learning specifies increasing the tendency that the required behavior would occur again by adding something. It enhances the strength of the behavior of the agent and positively impacts it.
- **Negative Reinforcement Learning:** Negative reinforcement learning works exactly opposite to the positive RL. It increases the tendency that the specific behavior would occur again by avoiding the negative condition.

ADVANTAGE OF MACHINE LEARNING

Machine learning's advantages make it valuable in numerous fields, including healthcare, finance, marketing, manufacturing, cybersecurity, and many others. By harnessing the power of data and automated learning, machine learning opens up new opportunities for innovation, efficiency, and problem-solving.

Machine learning offers numerous advantages that make it a powerful and valuable tool in various domains. Some of the key advantages of machine learning include:

1. **Automation and Efficiency**: Machine learning algorithms automate complex tasks and processes, reducing the need for manual intervention and saving time and effort. They can quickly analyze large volumes of data and make predictions or decisions much faster than humans.
2. **Handling Complex and Large Datasets**: Machine learning enables the analysis and processing of large and complex datasets that would be difficult for humans to handle manually. It can uncover hidden patterns, relationships, and insights from vast amounts of data that might not be apparent through traditional analysis.
3. **Accurate Predictions and Decisions**: Machine learning models can make highly accurate predictions or decisions based on patterns and relationships learned from data. They can generalize from past examples to new, unseen data and provide reliable and consistent outcomes.
4. **Adaptability and Learning from Data**: Machine learning algorithms can adapt and improve their performance over time by continuously learning from new data. They can incorporate new knowledge and adjust their predictions or decisions as conditions change, allowing them to remain up-to-date and relevant.
5. **Handling Non-linearity and Complex Relationships**: Machine learning techniques, such as neural networks, excel at capturing non-linear and complex relationships in data. They can model intricate patterns and interactions between variables, enabling the discovery of hidden insights and predictions that traditional methods may struggle with.
6. **Handling High-dimensional Data**: Machine learning algorithms can effectively handle high-dimensional data, where the number of features or variables is large. They can automatically select relevant features, reduce dimensionality, and capture complex relationships in high-dimensional spaces.
7. **Personalization and Customization**: Machine learning enables personalized experiences and customized recommendations by analyzing individual user data and preferences. It allows businesses to deliver tailored products, services, and content to customers, enhancing customer satisfaction and engagement.
8. **Identifying Anomalies and Outliers**: Machine learning algorithms can detect anomalies or outliers in data that deviate significantly from the norm. This capability is valuable in various applications, such as fraud detection, network security, and predictive maintenance.
9. **Continuous Improvement and Optimization**: Machine learning models can be optimized and fine-tuned to improve their performance continuously. By iteratively updating the model parameters or architecture, they can enhance accuracy, reduce errors, and adapt to changing conditions.

10. **Discovering Insights and Patterns**: Machine learning can uncover hidden patterns, correlations, and insights from data that might not be apparent to human analysts. It can help identify important features, relationships, and trends, leading to better decision-making and strategic planning.

DISADVANTAGE OF MACHINE LEARNING

Understanding these disadvantages is essential to ensure responsible and ethical use of machine learning and to address the limitations while harnessing its benefits. It's crucial to be mindful of potential biases, interpretability challenges, and data quality issues to build robust and fair machine learning systems.

While machine learning has numerous advantages, it also comes with certain limitations and challenges. Some of the disadvantages of machine learning include:

1. **Need for Sufficient and Quality Data**: Machine learning algorithms require a significant amount of labeled data for training. Acquiring and preparing high-quality, labeled data can be time-consuming and expensive, especially in domains where data collection is challenging or costly.
2. **Bias and Discrimination**: Machine learning models can inherit biases present in the training data, leading to biased predictions or decisions. If the training data reflects societal biases or contains unfair disparities, the model can perpetuate or amplify these biases, resulting in discriminatory outcomes.
3. **Interpretability and Explain ability**: Some machine learning models, especially complex ones like deep neural networks, are often considered black boxes. They may provide accurate predictions, but it can be challenging to understand and explain the reasoning behind those predictions. This lack of interpretability can be a concern, particularly in critical domains like healthcare or legal systems.
4. **Overfitting and Generalization**: Machine learning models can sometimes over fit the training data, meaning they become too specific to the training examples and perform poorly on new, unseen data. Achieving a balance between capturing complex patterns and avoiding overfitting can be challenging, requiring careful model selection, regularization techniques, and validation.
5. **Computational Resources and Time**: Training complex machine learning models, especially deep learning models, can be computationally intensive and time-consuming. It may require substantial computational resources, such as high-performance GPUs or specialized hardware, and lengthy training times, which can be a practical limitation for certain applications.
6. **Data Privacy and Security**: Machine learning relies on data, and handling sensitive or private data raises concerns about privacy and security. There is a need for appropriate measures to protect data confidentiality, prevent unauthorized access, and ensure compliance with regulations such as GDPR (General Data Protection Regulation) or HIPAA (Health Insurance Portability and Accountability Act).
7. **Lack of Human Intuition and Creativity**: Machine learning models are driven by data and patterns present in the training set. They may struggle to exhibit human-level intuition, creativity, or common sense

reasoning. They might not capture contextual nuances or make decisions based on ethical or moral considerations.

8. **Dependency on External Factors**: Machine learning models are trained on historical data, and their performance can be influenced by external factors that may change over time. Shifts in data distributions, concept drift, or changes in the environment can impact the model's accuracy and effectiveness, requiring continuous monitoring and adaptation.

9. **Limited to Available Data**: Machine learning models rely heavily on the data they are trained on. If the training data does not capture the full range of scenarios or variations in the real world, the model's performance may suffer when faced with unseen or novel situations.

10. **Complexity and Expertise Requirements**: Building, training, and deploying machine learning models require expertise in various areas, including data pre-processing, feature engineering, algorithm selection, hyper parameter tuning, and model evaluation. The complexity of machine learning can make it challenging for individuals without the necessary skills and knowledge to effectively utilize and interpret the results.

References

1. I. H. Sarker, M. M. Hoque, M. D. K. Uddin, and A. Tawfeeq, "Mobile data science and intelligent apps: concepts, ai-based modeling and research directions," *Mobile Networks and Applications*, vol. 26, no. 1, pp. 1–19, 2020.
View at: Publisher Site | Google Scholar

2. M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth, "Machine learning for internet of things data analysis: a survey," *Digital Communications and Networks*, vol. 4, no. 3, pp. 161–175, 2018.
View at: Publisher Site | Google Scholar

3. I. H. Sarker, A. S. M. Kayes, S. Badsha, H. Alqahtani, P. Watters, and A. Ng, "Cybersecurity data science: an overview from machine learning perspective," *Journal of Big Data*, vol. 7, no. 1, pp. 41–29, 2020.
View at: Publisher Site | Google Scholar

4. I. H. Sarker and R. FurhadNowrozy, "Ai-driven cybersecurity: an overview, security intelligence modeling and research directions," *SN Computer Science*, vol. 2, no. 3, p. 173, 2021.
View at: Publisher Site | Google Scholar

5. B. Slusarczyk, "Industry 4.0 - are we ready?" *Polish Journal of Management Studies*, vol. 17, no. 1, pp. 232–248, 2018.
View at: Publisher Site | Google Scholar

6. Google, "Trends In," 2019, https://trends.google.com/trends/.
View at: Google Scholar

7. J. Han, J. Pei, and M. Kamber, *Data Mining: Concepts and Techniques*, Elsevier, Amsterdam, 2011.

8. I. H. Witten and E. Frank, *Data Mining: Practical Machine Learning Tools and Techniques*, Morgan Kaufmann, Burlingto,MA, USA, 2005.

9. I. H. Sarker, "Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective," *SN Computer Science*, vol. 2, no. 3, 154 pages, 2021.
View at: Publisher Site | Google Scholar

10. I. H. Sarker, P. Watters, and A. S. M. Kayes, "Effectiveness analysis of machine learning classifcation models for predicting personalized context-aware smartphone usage," *J Big Data*, vol. 6, no. 1, pp. 1–28, 2019.
View at: Publisher Site | Google Scholar
11. L. P. Kaelbling, M. L. Littman, and A. W. Moore, "Reinforcement learning: a survey," *Journal of Artificial Intelligence Research*, vol. 4, pp. 237–285, 1996.
View at: Publisher Site | Google Scholar
12. R. Jhaveri, S. Ramani, G. Srivastava, T. Reddy Gadekallu, and V. Aggarwal, "Fault-Resilience for Bandwidth Management in Industrial Software-Defined Networks," *IEEE Transactions on Network Science and Engineering*, IEEE, vol. 8, no. 4, pp. 3129–3139, Aug 2021.
View at: Publisher Site | Google Scholar
13. ICT-Fact, "ICT-Fact and Figures 2017 Accessed," 2020, https://www.itu.int/en/ITU-D/Statistics/Documents/facts/ICTFactsFigures2017.pdf.
View at: Google Scholar
14. ICT-Facts, "Telecommunication Union," 2017, https://www.itu.int/en/ITU-D/Statistics/Pages/facts/default.aspx.
View at: Google Scholar
15. D. K. Bhattacharyya and J. K. Kalita, *Network Anomaly Detection: A Machine Learning Perspective*, Chapman & Hall, London, U.K, 2013.
16. V. Ambalavanan and B. P. Shanthi, "Cyber threats detection and mitigation using machine learning," *Handbook of Research on Machine and Deep Learning Applications for Cyber Security*, IGI Global, Hershey, PA, USA, pp. 132–149, 2020.
View at: Publisher Site | Google Scholar
17. T. Thomas, A. Vijayaraghavan, and S. Emmanuel, "Machine Learning and Cybersecurity," *Machine Learning Approaches in Cyber Security Analytics*, Springer, Berlin, Germany, pp. 37–47, 2020.
View at: Publisher Site | Google Scholar
18. S. Latif, J. Qadir, S. Farooq, and M. Imran, "How 5G wireless (and concomitant technologies) will revolutionize healthcare?" *Future Internet*, vol. 9, no. 4, p. 93, 2017.
View at: Publisher Site | Google Scholar
19. Y. Zhan, Y. Zhennan, P. Zhigang et al., "Multi-instance deep learning: discover discriminative local anatomies for bodypart recognition," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1332–1343, 2016.
View at: Publisher Site | Google Scholar
20. M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou, "Lung pattern classification for interstitial lung diseases using a deep convolutional neural network," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1207–1216, 2016.
View at: Publisher Site | Google Scholar
21. W. Shen, M. Zhou, F. Yang, C. Yang, and J. Tian, "Multi-scale convolutional neural networks for lung nodule classification," *Lecture Notes in Computer Science*, vol. 24, pp. 588–599, 2015.
View at: Publisher Site | Google Scholar
22. J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for MR image reconstruction," *Lecture Notes in Computer Science*, pp. 647–658, 2017.
View at: Publisher Site | Google Scholar

23. J. Majumdar and A. Majumdar, "Rodeo: robust de-aliasing autoencoder for real-time medical image reconstruction," *Pattern Recognition*, vol. 63, pp. 499–510, 2017.
View at: Publisher Site | Google Scholar
24. A. Davy, M. Havaei, D. Warde-Farley et al., "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.
View at: Publisher Site | Google Scholar
25. H. Wang, X. Li, H. Jhaveri et al., "Sparse Bayesian learning based channel estimation in FBMC/OQAM industrial IoT networks," *Computer Communications*, Elsevier, vol. 176, pp. 40–45, 2021.
View at: Publisher Site | Google Scholar
26. L. Xing, E. A. Krupinski, and J. Cai, "Artificial intelligence will soon change the landscape of medical physics research and practice," *Medical Physics*, vol. 45, no. 5, pp. 1791–1793, 2018.
View at: Publisher Site | Google Scholar
27. S. Latif, M. Asim, M. Usman, J. Qadir, and R. Rana, "Automating motion correction in multishot MRI using generative adversarial networks," in *Proceedings of the 32nd International Conference on Neural Information Processing Syste*, Montréal Canada, December 2018.
View at: Google Scholar
28. X. W. Xiaotong and X. Lin, "Big data deep learning: challenges and perspectives," *IEEE Access*, vol. 2, pp. 514–525, 2014.
View at: Publisher Site | Google Scholar
29. R. H. Jhaveri, R. Tan, A. Easwaran, and S. V. Ramani, "Managing Industrial Communication Delays with Software-Defined Networking," in *Proceedings of the 25th IEEE International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA 2019)*, pp. 1–11, IEEE, Hangzhou, China, August 2019.
View at: Publisher Site | Google Scholar
30. C. Campolo, A. Molinaro, and R. Scopigno, "From today's VANETs to tomorrow's planning and the bets for the day after," *Vehicular Communications*, vol. 2, no. 3, pp. 158–171, 2015.
View at: Publisher Sitewqa | Google Scholar